Information Fusion 40 (2018) 101–111

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

# A message passing approach for decision fusion in adversarial multi-sensor networks

Andrea Abrardo, Mauro Barni, Kassem Kallas*, Benedetta Tondi

*Department of Information Engineering and Mathematics, Via Roma 56, Siena, Italy*

ARTICLE INFO

*Article history:*
Received 7 November 2016
Revised 25 April 2017
Accepted 19 June 2017
Available online 22 June 2017

*Keywords:*
Adversarial signal processing
Decision fusion in adversarial setting
Decision fusion in the presence of Byzantines
Message passing algorithm
Factor graph

ABSTRACT

We consider a simple, yet widely studied, set-up in which a Fusion Center (FC) is asked to make a binary decision about a sequence of system states by relying on the possibly corrupted decisions provided by byzantine nodes, i.e. nodes which deliberately alter the result of the local decision to induce an error at the fusion center. When independent states are considered, the optimum fusion rule over a batch of observations has already been derived, however its complexity prevents its use in conjunction with large observation windows.

In this paper, we propose a near-optimal algorithm based on message passing that greatly reduces the computational burden of the optimum fusion rule. In addition, the proposed algorithm retains very good performance also in the case of dependent system states. By first focusing on the case of small observation windows, we use numerical simulations to show that the proposed scheme introduces a negligible increase of the decision error probability compared to the optimum fusion rule. We then analyse the performance of the new scheme when the FC makes its decision by relying on long observation windows. We do so by considering both the case of independent and Markovian system states and show that the obtained performance are superior to those obtained with prior suboptimal schemes. As an additional result, we confirm the previous finding that, in some cases, it is preferable for the byzantine nodes to minimise the mutual information between the sequence system states and the reports submitted to the FC, rather than always flipping the local decision.



## 1. Introduction

Decision fusion for distributed detection has received an increasing attention for its importance in several applications, including wireless networks, cognitive radio, multimedia forensics and many others. One of the most common scenarios is the parallel distributed fusion model. According to this model, the $n$ nodes of a multi-sensor network gather information about a system and make a local decision about the system state. Then the nodes send the local decisions to a Fusion Center (FC), which is in charge of making a final decision about the state of the system [1].

Here, we focus on an adversarial version of the above problem, in which a number of malicious nodes, often referred to as Byzantines [1], aim at inducing a decision error at the FC [2]. This is a recurrent problem in many situations wherein the nodes may make a profit from a decision error. As an example, consider a cognitive radio system [3–6] in which secondary users cooperate in sensing the frequency spectrum to decide about its occupancy and the possibility to use the available spectrum to transmit their own data. While cooperation among secondary users allows to make a better decision, it is possible that one or more users deliberately alter their measurements to let the system think that the spectrum is busy, when in fact it is not, in order to gain an exclusive opportunity to use the spectrum. Online reputation systems offer another example [7]. In this scenario, a fusion center must make a final decision about the reputation of an item like a good or a service by relying on users' feedback. Even in this case, it is possible that malevolent users provide a fake feedback to alter the reputation of the item under inspection. Similar examples are found in many other applications, including wireless sensor networks [2,3], distributed detection [8,9], multimedia forensics [10] and adversarial signal processing [11].

In this paper we focus on a binary version of the fusion problem, wherein the system can assume only two states. Specifically, the nodes observe the system over $m$ time instants and make a local decision about the sequence of system states. Local decisions are not error-free and hence they may be wrong with a certain error probability. Honest nodes send their decision to the fusion center, while byzantine nodes try to induce a decision error and hence flip the local decision with probability $P_{mal}$ before sending

* Corresponding author.
*E-mail address:* k_kallas@hotmail.com (K. Kallas).

http://dx.doi.org/10.1016/j.inffus.2017.06.006
1566-2535/

it to the FC. The fusion center knows that some of the nodes are Byzantines with a certain probability distribution, but it does not know their position.

## 2. Prior work and contribution

In a simplified version of the problem, the FC makes its decision on the state of the system at instant $j$ by relying only on the corresponding reports, and ignoring the node reports relative to different instants. In this case, and in the absence of Byzantines, the Bayesian optimal fusion rule has been derived in [12,13] and it is known as Chair–Varshney rule. If local error probabilities are symmetric and equal across the network, Chair–Varshney rule boils down to simple majority-based decision. In the presence of Byzantines, Chair–Varshney rule requires the knowledge of Byzantines' positions along with the flipping probability $P_{mal}$. Since this information is rarely available, the FC may resort to a suboptimal fusion strategy.

In [8], by adopting a Neyman–Pearson setup and assuming that the byzantine nodes know the true state of the system, the asymptotic performance obtainable by the FC are analysed as a function of the percentage of Byzantines in the network. By formalising the attack problem as the minimisation of the Kullback–Leibler distance between the reports received by the FC under the two hypotheses, the blinding percentage, that is, the percentage of Byzantines irremediably compromising the possibility of making a correct decision, is determined.

In order to improve the estimation of the sequence of system states, the FC can gather a number of reports provided by the nodes before making a global decision (multiple observation fusion). In cooperative spectrum sensing, for instance, this corresponds to collectively decide about the white holes over a time window, or, more realistically, at different frequency slots. The advantage of deciding over a sequence of states rather than on each single state separately, is that in such a way it is possible for the FC to understand which are the byzantine nodes and discard the corresponding observations (such an operation is usually referred to as Byzantine isolation). Such a scenario has also been studied in [8], showing that - at least asymptotically - the blinding percentage is always equal to 50%. In [14], the analysis of [8] is extended to a situation in which the Byzantines do not know the true state of the system. Byzantine isolation is achieved by counting the mismatches between the reports received from each node and the global decision made by the FC. The performance of the proposed scheme are evaluated in a cognitive-radio scenario for finite values of $n$. In order to cope with the lack of knowledge about the strategy adopted by the attacker, the decision fusion problem is casted into a game-theoretic formulation, where each party makes the best choice without knowing the strategy adopted by the other party.

A slightly different approach is adopted in [15]. By assuming that the FC is able to derive the statistics of the reports submitted by honest nodes, Byzantine isolation is carried out whenever the reports received from a node deviate from the expected statistics. In this way, a correct decision can be made also when the percentage of Byzantines exceeds 50%. The limit of the approach proposed in [15], is that it does not work when the reports sent by the Byzantines have the same statistics of those transmitted by the honest nodes. This is the case, for instance, in a perfectly symmetric setup with equiprobable system states, symmetric local error probabilities, and an attack strategy consisting of simple decision flipping. Another approach to separate nodes with diverse behaviours into different clusters is proposed in [19]. In this work, the authors propose a K-Means fault tolerant clustering algorithm *(Epidemic K-Means)* which does not require a central FC and can approximate the performance of centralized solution in separating the nodes into clusters which, in this case, can be applied to separate Byzantines from honest nodes. A soft isolation scheme is proposed in [16], where the reports from suspect byzantine nodes are given a lower importance rather being immediately discarded. Even in [16], the lack of knowledge at the FC about the strategy adopted by the attacker (and viceversa) is coped with by adopting a game-theoretic formulation. A rather different approach is adopted in [17], where a tolerant scheme that mitigates the impact of Byzantines on the global decision is used rather that removing the reports submitted by suspect nodes from the fusion procedure.

When the value of $P_{mal}$ and the probability that a node is Byzantine are known, the optimum fusion rule under multiple observation can be derived [18]. Since $P_{mal}$ is usually not known to the FC, in [18] the value of $P_{mal}$ used to define the optimum fusion rule and the value actually used by the Byzantines are strategically chosen in a game-theoretic context. Different priors about the distribution of Byzantines in the network are considered ranging from an extreme case in which the exact number of Byzantines in the network is known to a maximum entropy case. One of the main results in [18] is that the best option for the Byzantines is not to always flip the local decision (corresponding to $P_{mal} = 1$), since this would ease the isolation of malicious nodes. In fact, for certain combinations of the distribution of Byzantines within the network and the length of the observation window, it is better for the Byzantines to minimise the mutual information between the reports submitted to the FC and the system states.

### 2.1. Contribution

The main problem of the optimum decision fusion scheme proposed in [18] is its computational complexity, which grows exponentially with the length of the observation window. Such a complexity prevents the adoption of the optimum decision fusion rule in many practical situations. Also the results regarding the optimum strategies of the Byzantines and the FC derived in [18] refer only to the case of small observation windows.

In the attempt to diminish the computational complexity while minimising the loss of performance with respect to the optimum fusion rule, we propose a new, nearly-optimum, fusion scheme based on message passing and factor graphs. Message passing algorithms, based on the so called Generalised Distributive Law (GDL, [20,21]), have been widely applied to solve a large range of optimisation problems, including decoding of Low Density Parity Check (LDPC) codes [22] and BCJR codes [20], dynamic programming [23], solution of probabilistic inference problems on Bayesian networks [24] (in this case message passing algorithms are known as *belief propagation*). Here we use message passing to introduce a near-optimal solution of the decision fusion problem with multiple observation whose complexity grows only linearly with the size of the observation window, thus marking a dramatic improvement with respect to the exponential complexity of the optimal scheme proposed in [18].

Using numerical simulations and by first focusing on the case of small observation windows, for which the optimum solution can still be applied, we prove that the new scheme gives near-optimal performance at a much lower complexity than the optimum scheme. We then use numerical simulations to evaluate the performance of the proposed method for long observation windows. As a result, we show that, even in this case, the proposed solution maintains the performance improvement over the simple majority rule, the hard isolation scheme in [14] and the soft isolation scheme in [16].

As opposed to previous works, we do not limit our analysis to the case of independent system states, but we extend it to a more realistic scenario where the sequence of states obey a Markovian distribution [25] as depicted in Fig. 2. The Markovian

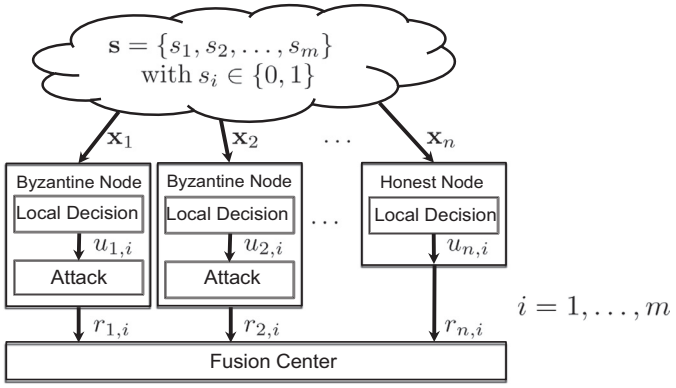


**Fig. 1.** Sketch of the adversarial decision fusion scheme.

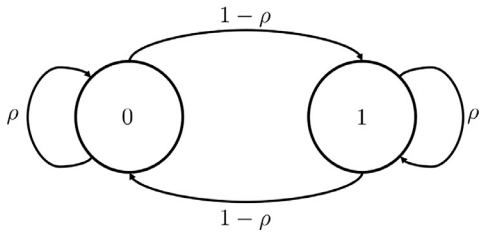


**Fig. 2.** Markovian model for system states. When $\rho = 0.5$ subsequent states are independent.

model is rather common in the case of cognitive radio networks [26–28] where the primary user occupancy of the spectrum is often modelled as a Hidden Markov Model (HMM). The Markovian case is found to be more favourable for the FC with respect to the case of independent states, due the additional a-priori information available to the FC in this case.

Last but not the least, we confirm that the dual optimum behaviour of the Byzantines observed in [18] is also present in the case of large observation windows, even if in the Markovian case, the Byzantines may continue using the maximum attack power ($P_{mal} = 1$) for larger observation windows.

The rest of this paper is organised as follows. In Section 3, we introduce the notation used in the paper and give a precise formulation of the addressed problem. In Section 4, we describe the new message passing decision rule based on factor graph. In Section 5, we first discuss the complexity of the proposed solution compared to the optimal solution. Then, by considering both independent and Markovian system states, we compare the performance of the message passing algorithm to the majority rule, the hard isolation scheme [14], the soft isolation scheme described in [16] and the optimal fusion rule. In addition, we discuss the impact that the length of the observation window has on the optimal behaviour of the Byzantines. We conclude the paper in Section 6 with some final remarks.

## 3. Notation and problem formulation

A schematic representation of the problem faced with in this paper is given in Fig. 1. We let $\mathbf{s} = \{s_1, s_2, \ldots, s_m\}$ with $s_i \in \{0, 1\}$ indicate the sequence of system states over an observation window of length $m$. The nodes collect information about the system through the vectors $\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_n$, with $\mathbf{x}_j$ indicating the observations available at node $j$. Based on such observations, a node $j$ makes a local decision $u_{i,j}$ about system state $s_i$. We assume that the local error probability, hereafter indicated as $\varepsilon$, does not depend on either $i$ or $j$. The state of the nodes in the network is given by the vector $\mathbf{h} = \{h_1, h_2, \ldots, h_n\}$ with $h_j = 1/0$ indicating that node $j$ is honest or Byzantine, respectively. Finally, the matrix $\mathbf{R} = \{r_{i,j}\}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ contains all the reports received by the FC. Specifically, $r_{i,j}$ is the report sent by node $j$ relative to $s_i$. As stated before, for honest nodes we have $u_{i,j} = r_{i,j}$ while, for Byzantines we have $p(u_{i,j} \neq r_{i,j}) = P_{mal}$. The Byzantines corrupt the local decisions independently of each other. One may argue that allowing cooperation among nodes would result in a further advantage for the Byzantines. In fact, one could envisage a scenario where the Byzantines coordinate their attacks, thus generating a more sophisticated and harmful attacking strategy. Such a coordination, however, complicates the attack that in this case would require some form of cooperation and hence communication among the byzantine nodes. This is a very interesting research direction that we are going to pursue in a subsequent work.

By assuming that the transmission between nodes and fusion center takes place over error-free channels, the report is equal to the local decision with probability 1 for honest nodes and with probability $1 - P_{mal}$ for Byzantines. Such an assumption does not diminish the generality of our analysis. In fact, the errors induced by local decisions and those introduced by the Byzantines can be modelled as the cascade of two binary symmetric channels (BSC). If we model the errors introduced by the transmission of the local decisions to the FC as the cascading of an additional BSC, the effect on our analysis would be only an increase of the overall cross-over probability of the resulting channel. For this reason, in the following, we will assume error free transmission without affecting the generality of our results.

According to the above setup, the probabilities of the reports sent by the honest nodes is given by:

$$p\big(r_{i,j}|s_i, h_j = 1\big) = (1-\varepsilon)\delta(r_{i,j} - s_i) + \varepsilon(1 - \delta(r_{i,j} - s_i)), \tag{1}$$

where $\delta(a)$ is defined as:

$$\delta(a) = \begin{cases} 1, & \text{if } a = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

On the other hand, by letting $\eta = \varepsilon(1 - P_{mal}) + (1-\varepsilon)P_{mal}$ be the probability that the fusion center receives a wrong report from a byzantine node, we have:

$$p\big(r_{i,j}|s_i, h_j = 0\big) = (1-\eta)\delta(r_{i,j} - s_i) + \eta(1 - \delta(r_{i,j} - s_i)). \tag{3}$$

As for the number of Byzantines, we consider a situation in which the states of the nodes are independent of each other and the state of each node is described by a Bernoulli random variable with parameter $\alpha$, that is $p(h_j = 0) = \alpha, \forall j$. In this way, the number of byzantine nodes in the network is a random variable following a binomial distribution, corresponding to the maximum entropy case [18] with $p(\mathbf{h}) = \prod_j p(h_j)$, where $p(h_j) = \alpha(1 - h_j) + (1-\alpha)h_j$.

Regarding the sequence of states $\mathbf{s}$, we assume a Markov model as shown in Fig. 2 , i.e., $p(\mathbf{s}) = \prod_i p(s_i|s_{i-1})$. The transition probabilities are given by $p(s_i|s_{i-1}) = 1 - \rho$ if $s_i = s_{i-1}$ and $p(s_i|s_{i-1}) = \rho$ when $s_i \neq s_{i-1}$, whereas for $i = 1$ we have $p(s_1|s_0) = p(s_1) = 0.5$.

In this paper we look for the *bitwise* Maximum A Posteriori Probability (MAP) estimation of the system states $\{s_i\}$, which reads as follows:

$$\begin{aligned}
\hat{s}_i &= \arg\max_{s_i \in \{0,1\}} p(s_i|\mathbf{R}) \\
&= \arg\max_{s_i \in \{0,1\}} \sum_{\{\mathbf{s},\mathbf{h}\}\setminus s_i} p(\mathbf{s}, \mathbf{h}|\mathbf{R}) \quad \text{(law of total probability)} \\
&= \arg\max_{s_i \in \{0,1\}} \sum_{\{\mathbf{s},\mathbf{h}\}\setminus s_i} p(\mathbf{R}|\mathbf{s}, \mathbf{h}) p(\mathbf{s}) p(\mathbf{h}) \qquad \text{(Bayes)} \\
&= \arg\max_{s_i \in \{0,1\}} \sum_{\{\mathbf{s},\mathbf{h}\}\setminus s_i} \prod_{i,j} p\big(r_{i,j}|s_i, h_j\big) \prod_i p(s_i|s_{i-1}) \prod_j p(h_j)
\end{aligned} \tag{4}$$

where the notation $\sum_{\setminus}$ denotes a summation over all the possible combinations of values that the variables contained in the expression within the summation may assume by keeping the parameter listed after the operator $\setminus$ fixed. For a given $\mathbf{h}$, the matrix of the observations $\mathbf{R}$ at the FC follows a HMM [29]. The optimisation problem in (4) has been solved in [18] for the case of independent system states. Even in such a simple case, however, the complexity of the optimum decision rule is exceedingly large, thus limiting the use of the optimum decision only in the case of small observation windows (typically $m$ not larger than 10). In the next section we introduce a sub-optimum solution of (4) based on message passing, which greatly reduces the computational complexity at the price of a negligible loss of accuracy.

## 4. A decision fusion algorithm based on message passing

### 4.1. Introduction to sum-product message passing

In this section, we provide a brief introduction to the message passing (MP) algorithm for marginalization of sum-product problems. Let us start by considering $N$ binary variables $\mathbf{z} = \{z_1, z_2, \ldots, z_N\}$, $z_i \in \{0, 1\}$. Then, consider the function $f(\mathbf{z})$ with factorization:

$$f(\mathbf{z}) = \prod_k f_k(\mathcal{Z}_k) \tag{5}$$

where $f_k$, $k = 1, \ldots, M$ are functions of a subset $\mathcal{Z}_k$ of the whole set of variables. We are interested in computing the marginal of $f$ with respect to a general variable $z_i$, defined as the sum of $f$ over all possible values of $\mathbf{z}$, i.e.:

$$\mu(z_i) = \sum_{\mathbf{z}\setminus z_i} \prod_k f_k(\mathcal{Z}_k) \tag{6}$$

where notation $\sum_{\mathbf{z}\setminus z_i}$ denotes a sum over all possible combinations of values of the variables in $\mathbf{z}$ by keeping $z_i$ fixed. Marginalization problems occur when we want to compute any arbitrary probability from joint probabilities by summing out variables that we are not interested in. In this general setting, determining the marginals by exhaustive search requires $2^N$ operations. However, in many situations it is possible to exploit the distributive law of multiplication to get a substantial reduction in complexity.

To elaborate, let associate with problem (6) a bipartite *factor graph*, in which for each variable we draw a variable node (circle) and for each function we draw a factor node (square). A variable node is connected to a factor node $k$ by an edge if and only if the corresponding variable belongs to $\mathcal{Z}_k$. This means that the set of vertices is partitioned into two groups (the set of nodes corresponding to variables and the set of nodes corresponding to factors) and that an edge always connects a variable node to a factor node.

When the factor graph is a single tree, i.e., a graph in which any two nodes are connected by exactly one path, it is straightforward to derive an algorithm which allows to solve the marginalization problem with reduced complexity. The algorithm is the MP algorithm, which has been broadly used in the last years in channel coding applications [30,31].

To describe how the MP algorithm works, let us assume that the graph is a tree. The main idea behind MP is to consider the variable whose marginal evaluation we are interested in as the root of the tree. Hence, the algorithm starts form leaf nodes and propagate the computation up to the root. At each intermediate step, a node evaluate a partial marginalization and passes the value to the parent node in the form of a message. On the other hand, each variable node in the graph can be seen as the root of the tree, and, hence, the algorithm can be parallelized with the aim of computing the marginal of all variables at the same time. In essence, when a node receives messages from all the connected nodes expect one, this one is considered the parent node and a message is delivered to it. Stated in another way, the message delivered towards a node is evaluated from the messages received from all the other nodes. To be more specific, let us define messages as 2-dimensional vectors, denoted by $\mathbf{m} = \{m(0), m(1)\}$. Such messages are exchanged between variable nodes and function nodes and viceversa, according to the following rules. Let us first consider variable-to-function messages ($\mathbf{m}_{vf}$), and take the portion of factor graph depicted in Fig. 3 as an illustrative example. In this graph, the variable node $z_i$ is connected to $L$ factor nodes, namely $f_1, f_2, \ldots, f_L$. For the MP algorithm to work properly, node $z_i$ must deliver the messages $\mathbf{m}_{vf}^{(l)}$, $l = 1, \ldots, L$ to all its adjacent nodes. Without loss of generality, let us focus on message $\mathbf{m}_{vf}^{(1)}$. Such a message can be evaluated and delivered upon receiving messages $\mathbf{m}_{fv}^{(l)}$, $l = 2, \ldots, L$, i.e., upon receiving messages from all function nodes except $f_1$. In particular, $\mathbf{m}_{vf}^{(1)}$ may be straightforwardly evaluated by calculating the element-wise product of the incoming messages, i.e.:

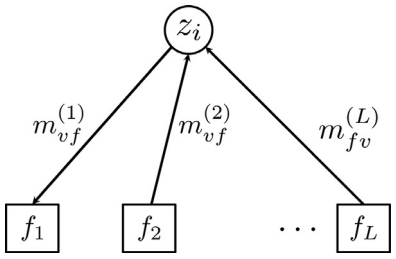


**Fig. 3.** Node-to-factor message passing.

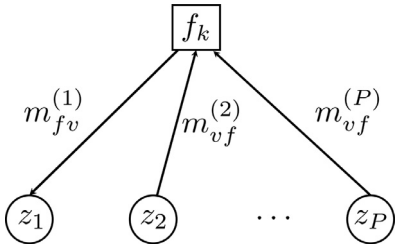


**Fig. 4.** Factor-to-node message passing.

$$m_{vf}^{(1)}(q) = \prod_{j=2}^{L} m_{fv}^{(j)}(q) \tag{7}$$

for $q = 0, 1$.

Let us now consider factor-to-variable messages, and refer to the factor graph of Fig. 4 where $P$ variable nodes are connected to the factor node $f_k$, i.e., according to the previous notation, $\mathcal{Z}_k = \{z_1, \ldots, z_P\}$. In this case, the node $f_k$ must deliver the messages $\mathbf{m}_{fv}^{(l)}$, $l = 1, \ldots, P$ to all its adjacent nodes. Let us consider again $\mathbf{m}_{fv}^{(1)}$: upon receiving the messages $\mathbf{m}_{vf}^{(l)}$, $l = 2, \ldots, P$, $f_k$ may evaluate the message $\mathbf{m}_{fv}^{(1)}$ as:

$$m_{fv}^{(1)}(q) = \sum_{z_2, \ldots, z_P} \left[ f_k(q, z_2, \ldots, z_P) \prod_{p=2}^{P} m_{vf}^{(p)}(z_p) \right] \tag{8}$$

for $q = 0, 1$.

Given the message passing rules at each node, it is now possible to derive the MP algorithm which allows to compute the marginals in (6). The process starts at the leaf nodes, i.e., those nodes which have only one connecting edge. In particular, each

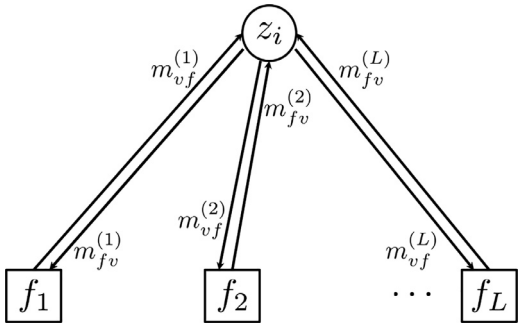


**Fig. 5.** End of message passing for node $z_i$.

variable leaf node passes an all-ones message to its adjacent factor node, whilst each factor leaf node, say $f_k(z_i)$ passes the message $m_{fv}^{(k)}(q) = f_k(z_i = q)$ to its adjacent node $z_i$. After initialization at leaf nodes, for every edge we can compute the outgoing message as soon as all incoming messages from all other edges connected to the same node are received (according to the message passing rules (7) and (8)). When a message has been sent in both directions along every edge the algorithm stops. This situation is depicted in Fig. 5: upon receiving messages from all its adjacent factor nodes, node $z_i$ can evaluate the exact marginal as:

$$\mu(z_i) = \prod_{k=1,\ldots,L} m_{fv}^{(k)}(z_i). \tag{9}$$

In the following we illustrate the machinery of MP through a toy example. Consider $N = 5$ and $f(\mathbf{z}) = f_1(z_1, z_2, z_5) f_2(z_3, z_4, z_5)$. Suppose we are interested in the marginal of $f$ with respect to $z_5$, i.e.: $\mu(z_5) = \sum_{z_1,z_2,z_3,z_4} f_1(z_1, z_2, z_5) f_2(z_3, z_4, z_5)$. To this aim, we can consider the variables $z_1$, $z_2$, $z_3$, and $z_4$ as the leaf nodes of a tree, i.e., they start the process by sending all ones messages to the parent factor nodes $f_1$ and $f_2$. Upon receiving such messages, the two factor nodes evaluate the messages to be delivered towards variable node $z_5$ according to (8), i.e.:

$$\begin{aligned} m_{f_1 v}^{(5)}(0) &= \sum_{z_1,z_2} f_1(z_1, z_2, 0) \times 1 \\ m_{f_1 v}^{(5)}(1) &= \sum_{z_1,z_2} f_1(z_1, z_2, 1) \times 1 \end{aligned} \tag{10}$$

$$\begin{aligned} m_{f_2 v}^{(5)}(0) &= \sum_{z_3,z_4} f_1(z_3, z_4, 0) \times 1 \\ m_{f_2 v}^{(5)}(1) &= \sum_{z_3,z_4} f_1(z_3, z_4, 1) \times 1. \end{aligned} \tag{11}$$

The variable node $z_5$ is now able to evaluate the required marginals as:

$$\begin{aligned} \mu_5(0) &= m_{f_1 v}^{(5)}(0) m_{f_2 v}^{(5)}(0) \\ \mu_5(1) &= m_{f_1 v}^{(5)}(1) m_{f_2 v}^{(5)}(1). \end{aligned} \tag{12}$$

With regards to complexity, the MP algorithm requires 18 operations instead of 32 as it would be required by a brute-force exhaustive approach. In general, factors to variables message passing can be accomplished with $2^P$ operations, $P$ being the number of variables in $f_k$. On the other hand, variables to nodes message passing's complexity can be neglected, and, hence, the MP algorithm allows to noticeably reduce the complexity of the problem provided that the numerosity of $\mathcal{Z}_k$ is much lower than $N$. With regard to the optimization, Eq. (9) evaluates the marginal for both $z_i = 0$ and $z_i = 1$, which represent the approximated computation of the sum-product for both hypotheses. Hence, the optimization is obtained by choosing the value of $z_i$ which maximizes it.

When the graph is not a tree, i.e., it contains cycles, the MP algorithm does not provide an exact calculation. However, although it was originally designed for acyclic graphical models, it was found that the MP algorithm can be used for general graphs, e.g., in channel decoding problems [32]. In general, when the marginalization problem is associated to a loopy graph, the implementation of MP requires to establish a scheduling policy to initiate the procedure, so that variable nodes may receive messages from all the connected factors, thus evaluating the marginals. In this case, a single run of the MP algorithm may not be sufficient to achieve a good approximation of the exact marginals, and progressive refinements must be obtained through successive iterations. However, in the presence of loopy graphs, there is no guarantee of either convergence or optimality of the final solution. In many cases, the performance of the message-passing algorithms is closely related to the structure of the graph, in general, and its cycles, in particular. Many previous works in the field of channel coding, e.g., see [33], reached the conclusion that, for good performance, the factor graph should not contain short cycles.

### 4.2. Nearly-optimal data fusion by means of message passing

The objective function of the optimal fusion rule expressed in (4) can be seen as a marginalization of a sum product of functions of binary variables, and, as such, it falls within the MP framework described in the previous Section. More specifically, in our problem, the variables are the system states $s_i$ and the status of the nodes $h_j$, while the functions are the probabilities of the reports shown in Eqs. (1) and (3), the conditional probabilities $p(s_i|s_{i-1})$, and the a-priori probabilities $p(h_j)$. The resulting bipartite graph is shown in Fig. 6.

It is worth noting that the graph is a loopy graph, i.e., it contains cycles, and as such it is not a tree. In our case, it is possible to see from Fig. 6 that the shortest cycles have order 6, i.e., a message before returning to the sender must cross at least six different nodes. We speculate that such a minimum cycles length is sufficient to provide good performance for the problem at hand. We will prove through simulations that this is indeed the case.

To elaborate further, based on the graph of Fig. 6 and on the general MP rules reported in the previous Section, we are now capable of deriving the messages for the scenario at hand. In Fig. 7, we display all the exchanged messages for the graph in Fig. 6 that are exchanged to estimate in parallel each of the states $s_i$, $i \in \{0, 1\}$ in the vector $\mathbf{s} = \{s_1, s_2, \ldots, s_m\}$. Specifically, we have:

$$\begin{aligned}
\tau_i^{(l)}(s_i) &= \varphi_i^{(l)}(s_i) \prod_{j=1}^{n} \nu_{i,j}^{(u)}(s_i) \quad i = 1, \ldots, m \\
\tau_i^{(r)}(s_i) &= \varphi_i^{(r)}(s_i) \prod_{j=1}^{n} \nu_{i,j}^{(u)}(s_i) \quad i = 1, \ldots, m \\
\varphi_i^{(l)}(s_i) &= \sum_{s_{i+1}=0,1} p(s_{i+1}|s_i) \tau_{i+1}^{(l)}(s_{i+1}) \quad i = 1, \ldots, m-1 \\
\varphi_i^{(r)}(s_i) &= \sum_{s_{i-1}=0,1} p(s_i|s_{i-1}) \tau_{i-1}^{(r)}(s_{i-1}) \quad i = 2, \ldots, m \\
\varphi_1^{(r)}(s_1) &= p(s_1) \\
\nu_{i,j}^{(u)}(s_i) &= \sum_{h_j=0,1} p\big(r_{i,j}\big|s_i, h_j\big) \lambda_{j,i}^{(u)}(h_j) \quad i = 1, \ldots, m, \quad j = 1, \ldots, n \\
\nu_{i,j}^{(d)}(s_i) &= \varphi_i^{(r)}(s_i) \varphi_i^{(l)}(s_i) \prod_{\substack{k=1 \\ k \neq j}}^{n} \nu_{i,k}^{(u)}(s_i) \quad i = 1, \ldots, m-1, \quad j = 1, \ldots, n
\end{aligned}$$

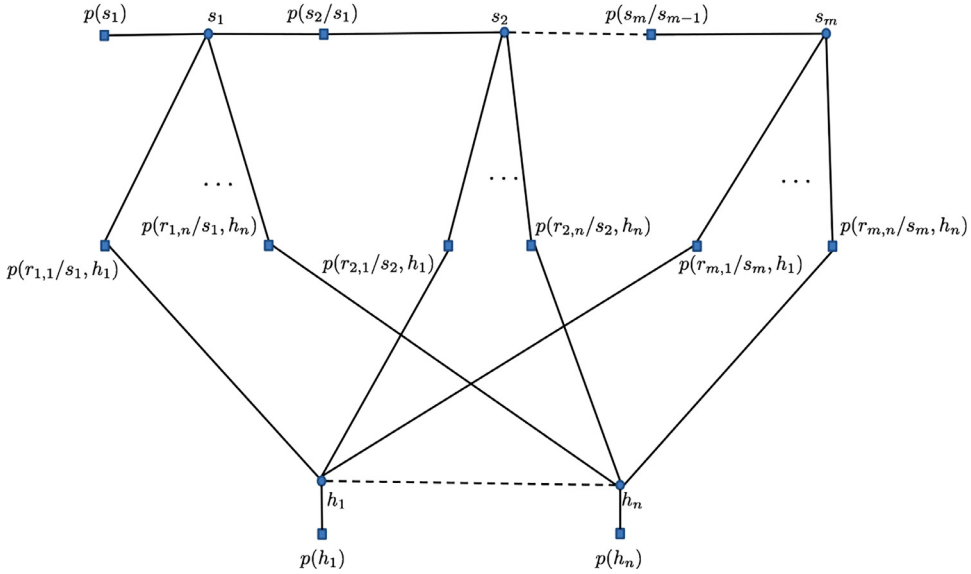


**Fig. 6.** Factor graph for the problem at hand.

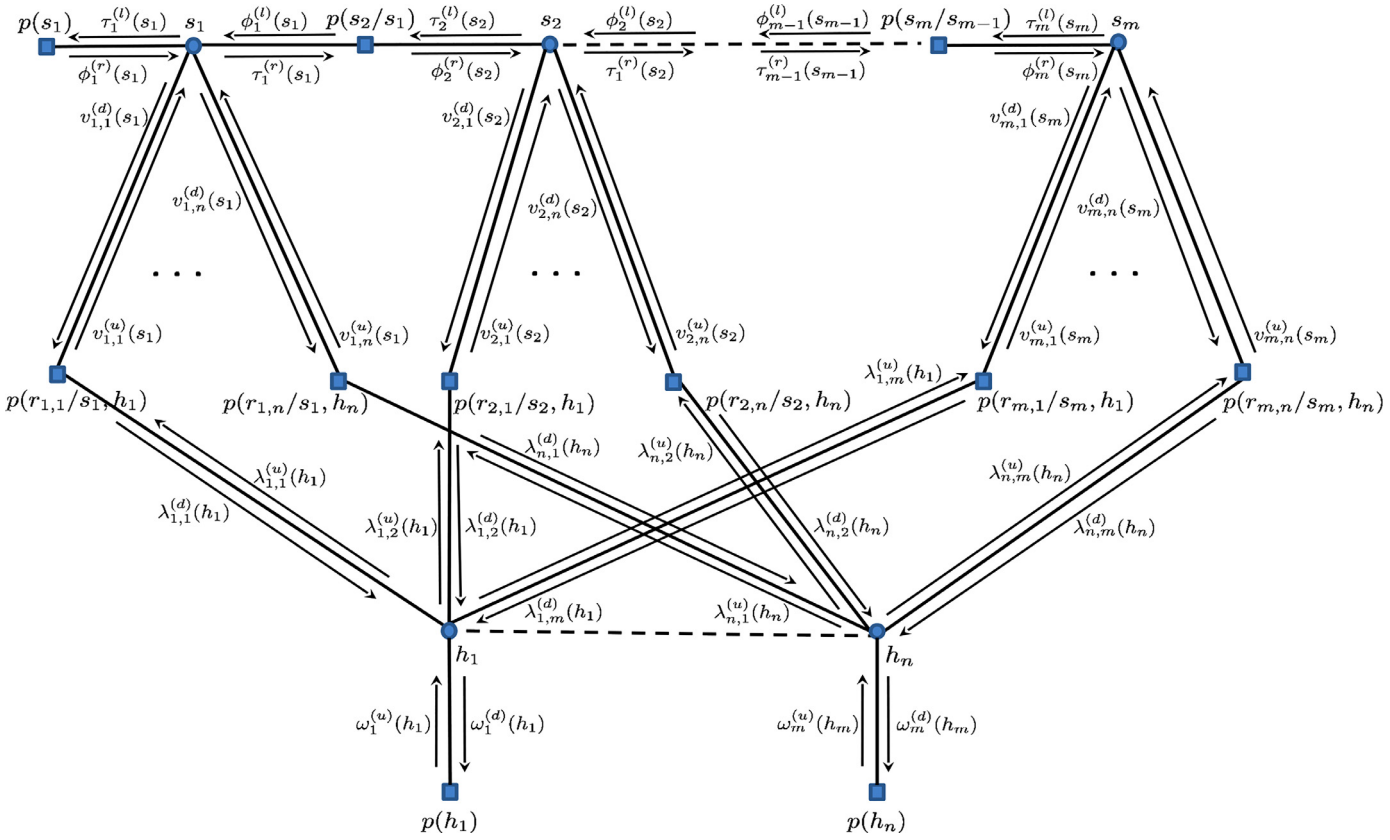


**Fig. 7.** Factor graph for the problem at hand with the illustration of all the exchanged messages.

$$\nu_{m,j}^{(d)}(s_m) = \varphi_i^{(r)}(s_m) \prod_{\substack{k=1 \\ k\neq j}}^{n} \nu_{m,k}^{(u)}(s_m) \quad j = 1,\ldots,n$$

$$\lambda_{j,i}^{(d)}(h_j) = \sum_{s_i=0,1} p\left(r_{i,j}\middle|s_i,h_j\right) \nu_{i,j}^{(d)}(s_i) \quad i = 1,\ldots,m, \quad j = 1,\ldots,n$$

$$\lambda_{j,i}^{(u)}(h_j) = \omega_j^{(u)}(h_j) \prod_{\substack{q=1 \\ q\neq i}}^{m} \lambda_{j,q}^{(d)}(h_j) \quad i = 1,\ldots,m, \quad j = 1,\ldots,n$$

$$\omega_j^{(d)}(h_j) = \prod_{i=1}^{m} \lambda_{j,i}^{(d)}(h_j) \quad j = 1,\ldots,n$$

$$\omega_j^{(u)}(h_j) = p(h_j) \quad j = 1,\ldots,n \tag{13}$$

It is worth noting that the above messages derive directly from the general MP rule shown in (7) and (8) and from the factor graph of the problem at hand depicted in Fig. 7. As an example, messages $\tau_i^{(l)}(s_i)$ in (13) take the general variable-to-factor message expression shown in (7), while messages $\varphi_i^{(l)}(s_i)$ take the general

factor-to-variable message expression shown in (8). Similar considerations can be drawn for all the messages shown in (13).

As for the scheduling policy, we initiate the MP procedure by sending the messages $\lambda_{j,i}^{(u)}(h_j) = \omega_j^{(u)}(h_j)$ to all $p(r_{i,j}|s_i, h_j.)$ factor nodes, and by sending the message $p(s_1)$ to the variable node $s_1$. Hence, the MP proceeds according to the general message passing rules, until all variable nodes are able to compute the respective marginals. When this happens, the first iteration is concluded. Then, successive iterations are carried out by starting from leaf nodes and by taking into account the messages received at the previous iteration for the evaluation of new messages. Hence, the algorithm is stopped upon achieving convergence of messages, or after a maximum number of iterations.

The MP scheme described above can be simplified by observing that messages can be normalized without affecting the normalized marginals. Henceforward, let us consider as normalization factors the sum of the elements of the messages, i.e., if we consider for example $\tau_i^{(l)}(s_i)$, the normalization factor is $\tau_i^{(l)}(0) + \tau_i^{(l)}(1)$. In this case, the normalized messages, say $\bar{\tau}_i^{(l)}(s_i)$ can be conveniently represented as scalar terms in the interval (0, 1), e.g., we can consider $\bar{\tau}_i^{(l)}(0)$ only since $\bar{\tau}_i^{(l)}(1) = 1 - \bar{\tau}_i^{(l)}(0)$. Accordingly, the normalized messages can be evaluated as:

$$
\begin{aligned}
\bar{\tau}_i^{(l)} &= \frac{\bar{\varphi}_i^{(l)} \prod_{j=1}^{n} \bar{\nu}_{i,j}^{(u)}}{\bar{\varphi}_i^{(l)} \prod_{j=1}^{n} \bar{\nu}_{i,j}^{(u)} + (1-\bar{\varphi}_i^{(l)}) \prod_{j=1}^{n} (1-\bar{\nu}_{i,j}^{(u)})} \\
&\quad i = 1, \ldots, m \\
\bar{\tau}_i^{(r)} &= \frac{\bar{\varphi}_i^{(r)} \prod_{j=1}^{n} \bar{\nu}_{i,j}^{(u)}}{\bar{\varphi}_i^{(r)} \prod_{j=1}^{n} \bar{\nu}_{i,j}^{(u)} + (1-\bar{\varphi}_i^{(r)}) \prod_{j=1}^{n} (1-\bar{\nu}_{i,j}^{(u)})} \\
&\quad i = 1, \ldots, m \\
\bar{\varphi}_i^{(l)} &= \rho \bar{\tau}_{i+1}^{(l)} + (1-\rho)(1-\bar{\tau}_{i+1}^{(l)}) \\
&\quad i = 1, \ldots, m-1 \\
\bar{\varphi}_i^{(r)} &= \rho \bar{\tau}_{i-1}^{(r)} + (1-\rho)(1-\bar{\tau}_{i-1}^{(r)}) \\
&\quad i = 2, \ldots, m \\
\bar{\varphi}_1^{(r)} &= p(s_1 = 0) \\
\bar{\nu}_{i,j}^{(u)} &= \frac{p\big(r_{i,j}|0,0\big)\bar{\lambda}_{j,i}^{(u)} + p\big(r_{i,j}|0,1\big)(1-\bar{\lambda}_{j,i}^{(u)})}{\kappa_1 + \kappa_2} \\
&\quad \text{where,} \quad \kappa_1 = p\big(r_{i,j}|0,0\big)\bar{\lambda}_{j,i}^{(u)} + p\big(r_{i,j}|0,1\big)(1-\bar{\lambda}_{j,i}^{(u)}) \\
&\quad \text{and} \quad \kappa_2 = p\big(r_{i,j}|1,0\big)\bar{\lambda}_{j,i}^{(u)} + p\big(r_{i,j}|1,1\big)(1-\bar{\lambda}_{j,i}^{(u)}) \\
&\quad j = 1, \ldots, m, \quad i = 1, \ldots, n \\
\bar{\nu}_{i,j}^{(d)} &= \frac{\bar{\varphi}_i^{(r)}\bar{\varphi}_i^{(l)} \prod_{\substack{k=1 \\ k\neq j}}^{n} \bar{\nu}_{i,k}^{(u)}}{\bar{\varphi}_i^{(r)}\bar{\varphi}_i^{(l)} \prod_{\substack{k=1 \\ k\neq j}}^{n} \bar{\nu}_{i,k}^{(u)} + (1-\bar{\varphi}_i^{(r)})(1-\bar{\varphi}_i^{(l)}) \prod_{\substack{k=1 \\ k\neq j}}^{n} (1-\bar{\nu}_{i,k}^{(u)})} \\
&\quad i = 1, \ldots, m-1, \quad j = 1, \ldots, n \\
\bar{\nu}_{m,j}^{(d)} &= \frac{\bar{\varphi}_m^{(r)} \prod_{\substack{k=1 \\ k\neq j}}^{n} \bar{\nu}_{m,k}^{(u)}}{\bar{\varphi}_m^{(r)} \prod_{\substack{k=1 \\ k\neq j}}^{n} \bar{\nu}_{m,k}^{(u)} + (1-\bar{\varphi}_m^{(r)}) \prod_{\substack{k=1 \\ k\neq j}}^{n} (1-\bar{\nu}_{m,k}^{(u)})} \\
&\quad j = 1, \ldots, n \\
\bar{\lambda}_{j,i}^{(d)} &= \frac{p\big(r_{i,j}|0,0\big)\bar{\nu}_{i,j}^{(d)} + p\big(r_{i,j}|1,0\big)(1-\bar{\nu}_{i,j}^{(d)})}{\tau_1 + \tau_2} \\
&\quad \text{where,} \quad \tau_1 = p\big(r_{i,j}|0,0\big)\bar{\nu}_{i,j}^{(d)} + p\big(r_{i,j}|1,0\big)(1-\bar{\nu}_{i,j}^{(d)}) \\
&\quad \text{and} \quad \tau_2 = p\big(r_{i,j}|0,1\big)\bar{\nu}_{i,j}^{(d)} + p\big(r_{i,j}|1,1\big)(1-\bar{\nu}_{i,j}^{(d)}) \\
&\quad j = 1, \ldots, m, \quad i = 1, \ldots, n \\
\bar{\lambda}_{j,i}^{(u)} &= \frac{\bar{\omega}_j^{(u)} \prod_{\substack{q=1 \\ q\neq i}}^{m} \bar{\lambda}_{j,q}^{(d)}}{\bar{\omega}_j^{(u)} \prod_{\substack{q=1 \\ q\neq i}}^{m} \bar{\lambda}_{j,q}^{(d)} + (1-\bar{\omega}_j^{(u)}) \prod_{\substack{q=1 \\ q\neq i}}^{m} (1-\bar{\lambda}_{j,q}^{(d)})} \\
&\quad i = 1, \ldots, m, \quad j = 1, \ldots, n \\
\bar{\omega}_j^{(d)} &= \frac{\prod_{i=1}^{m} \bar{\lambda}_{j,i}^{(d)}}{\prod_{i=1}^{m} \bar{\lambda}_{j,i}^{(d)} + \prod_{i=1}^{m} (1-\bar{\lambda}_{j,i}^{(d)})} \\
&\quad j = 1, \ldots, n \\
\bar{\omega}_j^{(u)} &= p(h_j = 0) \\
&\quad j = 1, \ldots, n
\end{aligned}
\tag{14}
$$

## 5. Simulation results and discussions

In this section, we analyze the performance of the MP decision fusion algorithm. We first consider the computational complexity, then we pass to evaluate the performance in terms of error probability. In particular, we compare the performance of the MP-based scheme to those of the optimum fusion rule [18] (whenever possible), the soft isolation scheme presented in [16], the hard isolation scheme described in [14] and the simple majority rule. In our comparison, we consider both independent and Markovian system states, for both small and large observation window $m$.

### 5.1. Complexity discussion

In order to evaluate the complexity of the message passing algorithm and compare it to that of the optimum fusion scheme, we consider both the number of operations and the running time. By number of operations we mean the number of additions, substractions, multiplications and divisions performed by the algorithm to estimate the vector of system states **s**.

By looking at expressions in Eq. (14), we see that running the message passing algorithm requires the following number of operations:

- $3n + 5$ operations for each of $\bar{\tau}_i^{(l)}$ and $\bar{\tau}_i^{(r)}$.
- 3 operations for each of $\bar{\varphi}_i^{(l)}$ and $\bar{\varphi}_i^{(r)}$.
- 11 operations for $\bar{\nu}_{i,j}^{(u)}$.
- $3n + 5$ operations for $\bar{\nu}_{i,j}^{(d)}$.
- $3n + 2$ operations for $\bar{\nu}_{m,j}^{(d)}$.
- 11 operations for $\bar{\lambda}_{j,i}^{(d)}$.
- $3m + 2$ operations for each of $\bar{\lambda}_{j,i}^{(u)}$ and $\bar{\omega}_j^{(d)}$.

summing up to $12n + 6m + 49$ operations for each iteration over the factor graph. On the other hand, in the case of independent node states, the optimal scheme in [18] requires $2^m(m + n)$ operations. Therefore, the MP algorithm is much less computationally expensive since it passes from an exponential to a linear complexity in $m$. An example of the difference in computational complexity between the optimum and the MP algorithms is depicted in Fig. 8.

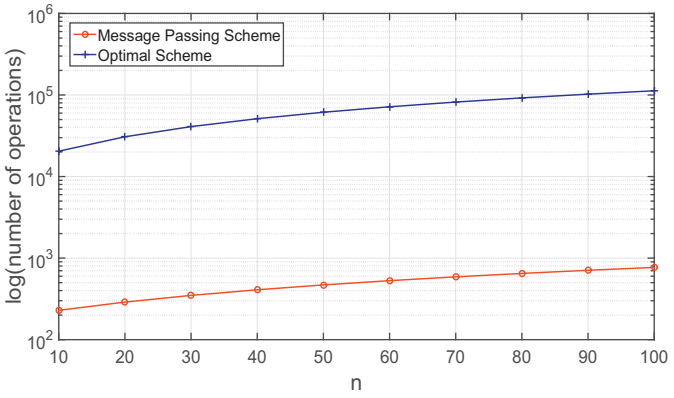


**Fig. 8.** Number of operations required for different $n$, $m = 10$ and 5 message passing local iterations for message passing and optimal schemes.

**Table 1**
Running time (in seconds) for the optimal and the message passing algorithms for: $m = 10$, $\varepsilon = 0.15$, Number of trials $= 10^5$ and Message passing iterations $= 5$.

| Setting/Scheme | Message Passing | Optimal |
|---|---|---|
| $n = 20, \alpha = 0.45$ | 943.807114 | 1.6561e + 04 |
| $n = 100, \alpha = 0.49$ | 4888.821497 | 2.0817e + 04 |

With regard to time complexity, Table 1 reports the running time of the MP and the optimal schemes. For $n = 20$, the optimal scheme running time is 17.547 times larger than that of the message passing algorithm. On the other hand, for the case of $n = 100$, the optimal scheme needs around 4.258 times more than the message passing scheme. The tests have been conducted using Matlab 2014b running on a machine with 64-bit windows 7 OS with 16,0GB of installed RAM and Intel Core i7-2600 CPU @ 3.40 GHz.

### 5.2. Performance evaluation

In this section, we use numerical simulations to evaluate the performance of the message passing algorithm and compare them to the state of the art schemes. The results are divided into four parts. The first two parts consider, respectively, simulations performed with small and large observation windows $m$. Then, in the third part, we investigate the optimum behaviour of the Byzantines over a range of observation windows size. Finally, in the last part, we compare the case of independent and Markovian system states.

The simulations were carried out according to the following setup. We considered a network with $n = 20, 100$ nodes, $\varepsilon = 0.15$, $\rho = \{0.95, 0.5\}$ corresponding to Markovian and independent sequence of system states, respectively. The probability $\alpha$ that a node is Byzantine is in the range [0, 0.45] corresponding to a number of Byzantines between 0 and 9. As to $P_{mal}$ we set it to either 0.5 or 1.[1] The number of message passing iterations is 5. For each setting, we estimated the error probability over $10^5$ trials.

#### 5.2.1. Small m

To start with, we considered a small observation window, namely $m = 10$. With such a small value of $m$, in fact, it is possible to compare the performance of the message passing algorithm to that of the optimum decision fusion rule. The results we obtained are reported in Fig. 9. Upon inspection of the figure, the superior performance of the message passing algorithm over the Majority, Soft and Hard isolation schemes is confirmed. More interestingly,

[1] It is know from [18] that for the Byzantines the optimum choice of $P_{mal}$ is either 0.5 or 1 depending on the considered setup.

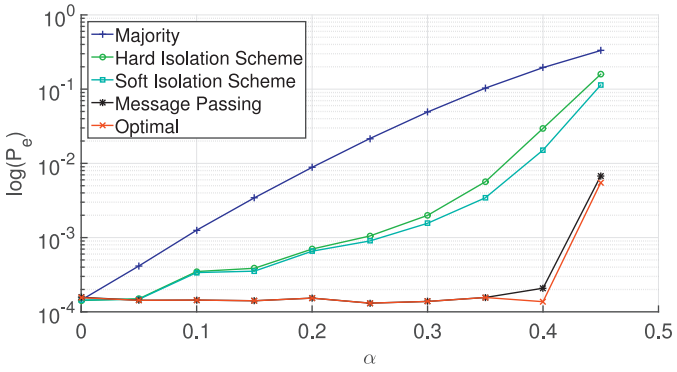


**Fig. 9.** Error probability as a function of $\alpha$ for the following setting: $n = 20$, independent Sequence of States $\rho = 0.5$, $\varepsilon = 0.15$, $m = 10$ and $P_{mal} = 1.0$.

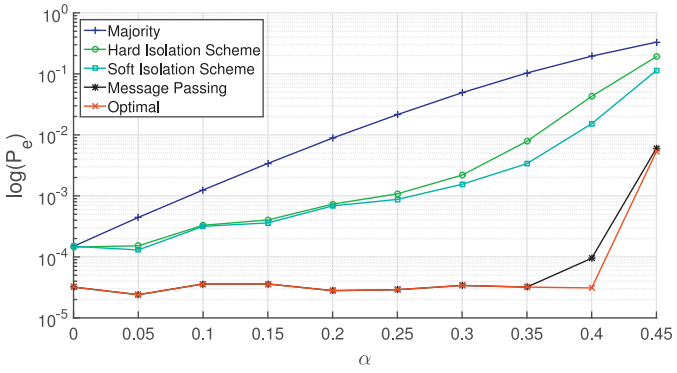


**Fig. 10.** Error probability as a function of $\alpha$ for the following setting: $n = 20$, Markovian Sequence of States $\rho = 0.95$, $\varepsilon = 0.15$, $m = 10$ and $P_{mal} = 1.0$.

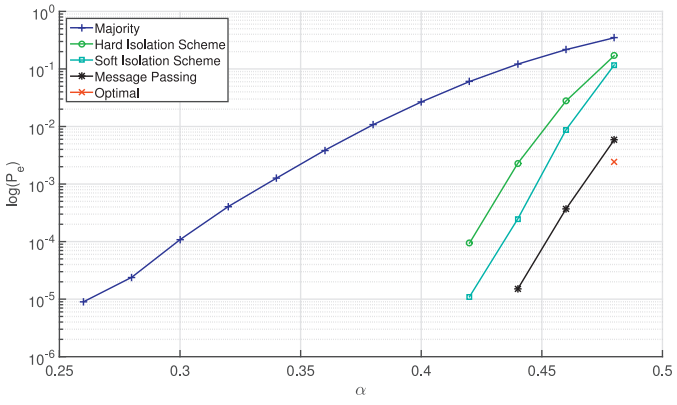


**Fig. 11.** Error probability as a function of $\alpha$ for the following setting: $n = 100$, independent Sequence of States $\rho = 0.5$, $\varepsilon = 0.15$, $m = 10$ and $P_{mal} = 1.0$. Note that, only one value is reported for the optimal scheme since it achieves $P_e = 0$ for other values of $\alpha$.

the message passing algorithm gives nearly optimal performance, with only a negligible performance loss with respect to the optimum scheme.

Fig. 10 confirms the results shown in Fig. 9 for Markovian system states ($\rho = 0.95$). These results are confirmed in Figs. 11 and 12 for a larger network size ($n = 100$), with a significant performance improvement since a larger network conveys more information at the FC.

#### 5.2.2. Large m

Having shown the near optimality of the message passing scheme for small values of $m$; we now leverage on the small computational complexity of such a scheme to evaluate its perfor-

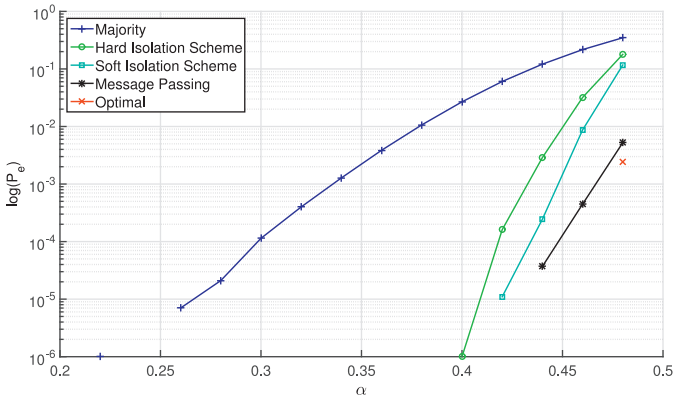


**Fig. 12.** Error probability as a function of $\alpha$ for the following setting: $n = 100$, Markovian Sequence of States $\rho = 0.95$, $\varepsilon = 0.15$, $m = 10$ and $P_{mal} = 1.0$. Note that, only one value is reported for the optimal scheme since it achieves $P_e = 0$ for other values of $\alpha$.

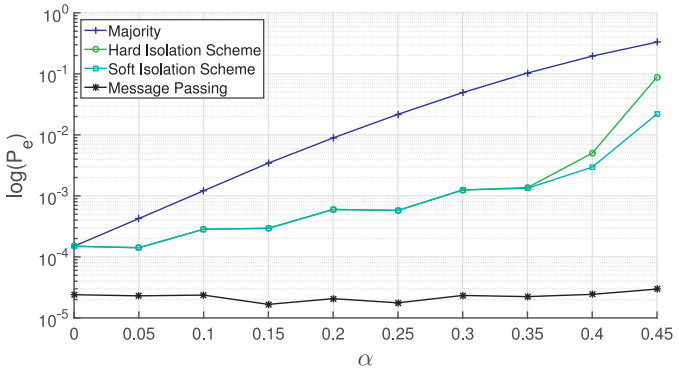


**Fig. 13.** Error probability as a function of $\alpha$ for the following setting: $n = 20$, Markovian Sequence of States $\rho = 0.95$, $\varepsilon = 0.15$, $m = 30$ and $P_{mal} = 1.0$.

mance for large values of $m$ ($m = 30$). As shown in Fig. 13, by increasing the observation window all the schemes give better performance, with the message passing algorithm always providing the best performance. Interestingly, in this case, when the attacker uses $P_{mal} = 1.0$, the message passing algorithm permits to almost nullify the attack of the Byzantines for all the values of $\alpha$. Concerning the residual error probability, it is due to the fact that, even when there are no Byzantines in the network ($\alpha = 0$), there is still an error floor caused by the local errors at the nodes $\varepsilon$. For the case of independent states, such an error floor is around $10^{-4}$. In Figs. 13 and 14, this error floor decreases to about $10^{-5}$ because of the additional a-priori information available in the Markovian case. Similar results, with lower error probabilities, are obtained for the case of $n = 100$, as it can be seen in Fig. 15.

*5.2.3. Optimal choice of* $P_{mal}$ *for the Byzantines*

One of the main results proven in [18], is that setting $P_{mal} = 1$ is not necessarily the optimal choice for the Byzantines. In fact, when the FC manages to identify which are the malicious nodes, it can exploit the fact the malicious nodes always flip the result of the local decision to get useful information about the system state. In such cases, it is preferable for the Byzantines to use $P_{mal} = 0.5$ since in this way the reports sent to the FC does not convey any information about the status of the system. However, in [18], it was not possible to derive exactly the limits determining the two different behaviours for the Byzantines due to the impossibility of applying the optimum algorithm in conjunction with large observation windows. By exploiting the low complexity of the message passing scheme, we are now able to overcome the limits of the analysis carried out in [18].

Specifically, we carried out an additional set of experiments by fixing $\alpha = 0.45$ and varying the observation window in the interval [5,20]. The results we obtained confirm the general behaviour observed in [18]. For instance, in Fig. 16, $P_{mal} = 1.0$ remains the Byzantines' optimal choice up to $m = 13$, while for $m > 13$, it is preferable for them to use $P_{mal} = 0.5$. Similar results are obtained for independent system states as shown in Fig. 17.

*5.2.4. Comparison between independent and Markovian system states*

In this subsection, we provide a comparison between the cases of Markovian and independent system states.

By looking at Figs. 16 and 17, we see that the Byzantines switch their strategy from $P_{mal} = 1$ to $P_{mal} = 0.5$ for a smaller observation window ($m = 10$) in the case of independent states (the switching value for the Markovian case is $m = 13$). We can explain this behaviour by observing that in the case of Markovian states, using $P_{mal} = 0.5$ results in a strong deviation from the Markovianity assumption of the reports sent to the FC thus making it easier the isolation of byzantine nodes. This is not the case with $P_{mal} = 1$, since, due to the symmetry of the adopted Markov model, such a value does not alter the expected statistics of the reports.

As a last result, in Fig. 18, we compare the error probability for the case of independent and Markov sources. Since we are interested in comparing the achievable performance for the two cases, we consider only the performance obtained by the optimum and the message passing algorithms. Upon inspection of the figure, it turns out that the case of independent states is more favourable to the Byzantines than the Markov case. The reason is that the FC may exploit the additional a-priori information available in the Markov case to identify the Byzantines and hence make a better decision. Such effect disappears when $\alpha$ approaches 0.5, since in this case the Byzantines tend to dominate the network. In that case, the Byzantines' reports prevail the pool of reports at the FC and hence, the FC becomes nearly *blind* so that even the additional a-priori information about the Markov model does not offer a great help.

## 6. Conclusions

In this paper, we proposed a near-optimal message passing algorithm based on factor graph for decision fusion in multi-sensor networks in the presence of Byzantines. The effectiveness of the proposed scheme is evaluated by means of extensive numerical simulations both for the case of independent and Markov sequence of states. Experiments showed that, when compared to the optimum fusion scheme, the proposed scheme permits to achieve near-optimal performance at a much lower computational cost: specifically, by adopting the new algorithm based on message passing we were able to reduce the complexity from exponential to linear. Such reduction of the complexity permits to deal with large observation windows, thus further improving the performance of the decision. Results on large observation windows confirmed the dual behavior in the attacking strategy of the Byzantines, looking for a trade-off between pushing the FC to make a wrong decision on one hand and reducing the mutual information between the reports and the system state on the other hand. In addition, the experiments showed that the case of independent states is more favorable to Byzantines than the Markovian case, due to the additional a-priori information available at the FC in the Markovian case.

As future work, we plan to focus on a scenario more favorable to the Byzantines, by giving them the possibility to access the observation vectors. In this way, they can focus their attack on the most profitable cases and avoid to flip the local decision when it is very likely that their action will have no effect on the FC decision. Considering the case where the nodes can send to the FC

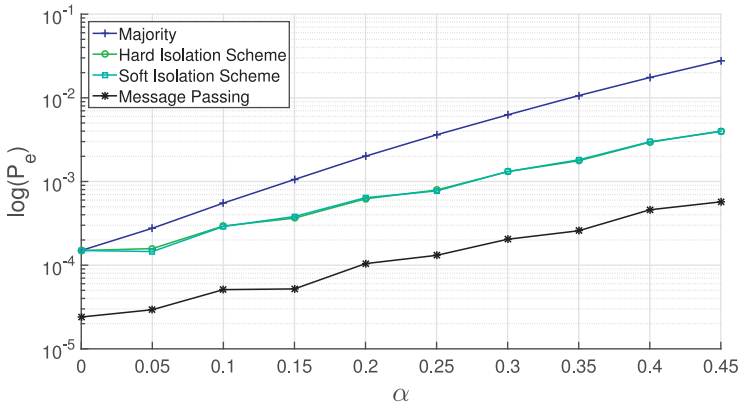


**Fig. 14.** Error probability as a function of $\alpha$ for the following setting: $n = 20$, Markovian Sequence of States $\rho = 0.95$, $\varepsilon = 0.15$, $m = 30$ and $P_{mal} = 0.5$.

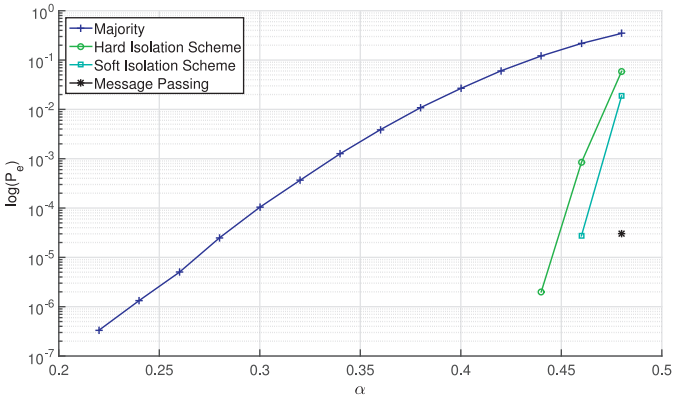


**Fig. 15.** Error probability as a function of $\alpha$ for the following setting: $n = 100$, Markovian Sequence of States $\rho = 0.95$, $\varepsilon = 0.15$, $m = 30$ and $P_{mal} = 1.0$. Note that, only one value is reported for the message passing scheme since it achieves $P_e = 0$ for other values of $\alpha$.

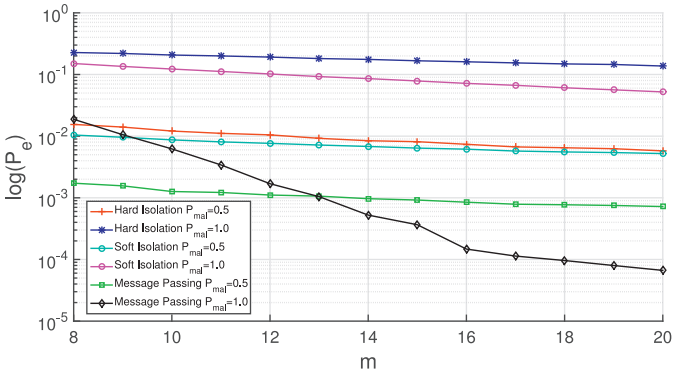


**Fig. 16.** Error probability as a function of $m$ for the following settings: $n = 20$, Markovian Sequence of States $\rho = 0.95$, $\varepsilon = 0.15$ and $\alpha = 0.45$.

more extensive reports (multi-bit case) [34] is another interesting extension. As we have already mentioned, another interesting extension, is obtained by allowing the Byzantines to coordinate their attacks. For instance, we could consider a synchronized attack, in which the Byzantines use the same pseudo-random generator with a common seed, to decide whether to flip the result of the local decision or not. Implementing the proposed message passing algorithm on real devices, i.e. test-bed, is also an interesting direction that will be the subject of our future work.

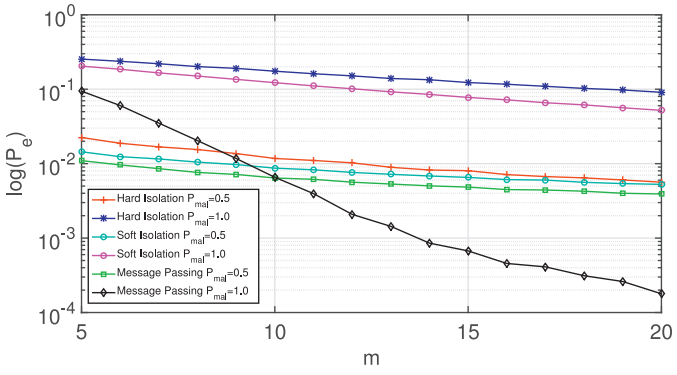


**Fig. 17.** Error probability as a function of $m$ for the following settings: $n = 20$, independent Sequence of States $\rho = 0.5$, $\varepsilon = 0.15$ and $\alpha = 0.45$.

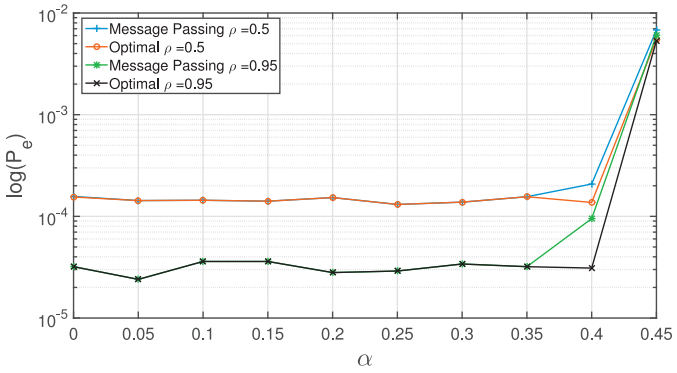


**Fig. 18.** Comparison between the case of independent and Markovian system states ($n = 20$, $\rho = \{0.5, 0.95\}$, $\varepsilon = 0.15$, $m = 10$, $P_{mal} = 1.0$).

## References

[1] A. Vempaty, T. Lang, P. Varshney, Distributed inference with Byzantine data: state-of-the-art review on data falsification attacks, IEEE Signal Process. Mag. 30 (5) (2013) 65–75.

[2] M. Abdelhakim, L. Lightfoot, J. Ren, T. Li, Distributed detection in mobile access wireless sensor networks under Byzantine attacks, IEEE Trans. Parallel Distrib. Syst. 25 (4) (2014) 950–959, doi:10.1109/TPDS.2013.74.

[3] W. Wang, H. Li, Y. Sun, Z. Han, Securing collaborative spectrum sensing against untrustworthy secondary users in cognitive radio networks, EURASIP J. Adv. Signal Process. 2010 (2010) 4.

[4] A.S. Rawat, P. Anand, H. Chen, P.K. Varshney, Collaborative spectrum sensing in the presence of Byzantine attacks in cognitive radio networks, IEEE Trans. Signal Process. 59 (2) (2011) 774–786.

[5] R. Zhang, J. Zhang, Y. Zhang, C. Zhang, Secure crowdsourcing-based cooperative spectrum sensing, in: Proceedings of INFOCOM 2013, IEEE Conference on Computer Communications, IEEE, 2013, pp. 2526–2534.

[6] W. Wang, L. Chen, K.G. Shin, L. Duan, Secure cooperative spectrum sensing and access against intelligent malicious behaviors, in: Proceedings of INFOCOM 2014, IEEE Conference on Computer Communications, IEEE, 2014, pp. 1267–1275.
[7] Y. Sun, Y. Liu, Security of online reputation systems: the evolution of attacks and defenses, IEEE Signal Process. Mag. 29 (2) (2012) 87–97.
[8] S. Marano, V. Matta, L. Tong, Distributed detection in the presence of Byzantine attacks, IEEE Trans. Signal Process. 57 (1) (2009) 16–29.
[9] B. Kailkhura, S. Brahma, P. Varshney, Optimal Byzantine attacks on distributed detection in tree-based topologies, in: IEEE International Conference on Computing, Networking and Communications (ICNC), 2013, pp. 227–231, doi:10.1109/ICCNC.2013.6504085.
[10] M. Barni, B. Tondi, Multiple-observation hypothesis testing under adversarial conditions, in: Proceedings of WIFS'13, IEEE International Workshop on Information Forensics and Security, Guangzhou, China, 2013, pp. 91–96.
[11] M. Barni, F. Pérez-González, Coping with the enemy: advances in adversary-aware signal processing, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 8682–8686, doi:10.1109/ICASSP.2013.6639361.
[12] Z. Chair, P. Varshney, Optimal data fusion in multiple sensor detection systems, IEEE Trans. Aerosp. Electron. Syst. AES-22 (1) (1986) 98–101, doi:10.1109/TAES.1986.310699.
[13] P.K. Varshney, Distributed Detection and Data Fusion, Springer-Verlag, 1997.
[14] A.S. Rawat, P. Anand, H. Chen, P.K. Varshney, Collaborative spectrum sensing in the presence of Byzantine attacks in cognitive radio networks, IEEE Trans. Signal Process. 59 (2) (2011) 774–786, doi:10.1109/TSP.2010.2091277.
[15] A. Vempaty, K. Agrawal, P. Varshney, H. Chen, Adaptive learning of Byzantines' behavior in cooperative spectrum sensing, in: Proceedings of WCNC'11, IEEE Conference on Wireless Communications and Networking, 2011, pp. 1310–1315, doi:10.1109/WCNC.2011.5779320.
[16] A. Abrardo, M. Barni, K. Kallas, B. Tondi, Decision fusion with corrupted reports in multi-sensor networks: a game-theoretic approach, in: 53rd IEEE Conference on Decision and Control, 2014, pp. 505–510, doi:10.1109/CDC.2014.7039431.
[17] R. Chen, J.M. Park, K. Bian, Robust distributed spectrum sensing in cognitive radio networks, in: Proceedings of INFOCOM 2008, 27th IEEE Conference on Computer Communications, 2008, doi:10.1109/INFOCOM.2008.251.
[18] A. Abrardo, M. Barni, K. Kallas, B. Tondi, A game-theoretic framework for optimum decision fusion in the presence of Byzantines, IEEE Trans. Inf. Forensics Secur. 11 (6) (2016) 1333–1345, doi:10.1109/TIFS.2016.2526963.
[19] G. Di Fatta, F. Blasa, S. Cafiero, G. Fortino, Fault tolerant decentralised k-means clustering for asynchronous large-scale networks, J. Parallel Distrib. Comput. 73 (3) (2013) 317–329.
[20] S.M. Aji, R.J. McEliece, The generalized distributive law, IEEE Trans. Inf. Theory 46 (2) (2000) 325–343, doi:10.1109/18.825794.
[21] P. Pakzad, V. Anantharam, A new look at the generalized distributive law, IEEE Trans. Inf. Theory 50 (6) (2004) 1132–1155, doi:10.1109/TIT.2004.828058.
[22] R. Gallager, Low-density parity-check codes, IRE Trans. Inf. Theory 8 (1) (1962) 21–28, doi:10.1109/TIT.1962.1057683.
[23] S. Verdu, H.V. Poor, Abstract dynamic programming models under commutativity conditions, SIAM J. Control Optim. 25 (4) (1987) 990–1006.
[24] J. Pearl, M. Kaufmann, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, CA, Cal., 1988.
[25] L. Rabiner, B. Juang, An introduction to hidden Markov models, IEEE ASSP Mag. 3 (1) (1986) 4–16, doi:10.1109/MASSP.1986.1165342.
[26] K.W. Choi, E. Hossain, Estimation of primary user parameters in cognitive radio systems via hidden Markov model, IEEE Trans. Signal Process. 61 (3) (2013) 782–795, doi:10.1109/TSP.2012.2229998.
[27] I.A. Akbar, W.H. Tranter, Dynamic spectrum allocation in cognitive radio using hidden Markov models: Poisson distributed case, in: IEEE Proceedings of SoutheastCon, 2007, pp. 196–201, doi:10.1109/SECON.2007.342884.
[28] T. Jiang, H. Wang, A.V. Vasilakos, Qoe-driven channel allocation schemes for multimedia transmission of priority-based secondary users over cognitive radio networks, IEEE J. Sel. Areas Commun. 30 (7) (2012) 1215–1224, doi:10.1109/JSAC.2012.120807.
[29] Y. Ephraim, N. Merhav, Hidden Markov processes, IEEE Trans. Inf. Theory 48 (6) (2002) 1518–1569.
[30] D.J. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge university press, 2003.
[31] F.R. Kschischang, B.J. Frey, H.-A. Loeliger, Factor graphs and the sum-product algorithm, IEEE Trans. Inf. Theory 47 (2) (2001) 498–519.
[32] T. Richardson, R. Urbanke, Modern Coding Theory, Cambridge University Press, 2008.
[33] Y. Mao, A.H. Banihashemi, A heuristic search for good low-density parity-check codes at short block lengths, in: Communications, 2001. ICC 2001. IEEE International Conference on, 1, IEEE, 2001, pp. 41–44.
[34] B. Kailkhura, S. Brahma, P. Varshney, On the performance analysis of data fusion schemes with Byzantines, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 7411–7415, doi:10.1109/ICASSP.2014.6855040.