Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Distributed Intrusion Detection for Secure Consensus Computations

Fabio Pasqualetti Antonio Bicchi Francesco Bullo

Abstract— **This paper focuses on trustworthy computation systems and proposes a novel intrusion detection scheme for linear consensus networks with misbehaving nodes. This prototypical control problem is relevant in network security applications. The objective is for each node to detect and isolate the misbehaving nodes using only the information flow adopted by standard consensus protocols. We focus on the single misbehaving node problem. Our technical approach is based on the theory of Unknown Input Observability. First, we give necessary and sufficient conditions for the misbehavior to be observable and for the identity of the faulty node to be detectable. Second, we design a distributed unknown input estimator, and we characterize its convergence rate in the "equal-neighbor" model and in the general case. Third and finally, we propose a complete detection and isolation scheme and provide some remarks on the filter convergence time. We conclude the paper with the numerical study of a consensus problem and of a robotic deployment problem.**

I. INTRODUCTION

Given a set of autonomous agents with the ability to exchange messages and perform local computations, a distributed algorithm is a procedure each agent performs to achieve a common task. Many distributed algorithms have been proposed to solve problems like average consensus [1], rendezvous [2] and agreement, where the agents try to agree on a parameter, which may be a common direction of motion, the clock speed or a decision value represented by a scalar or a vector [3]. In these algorithms, the nodes are assumed to cooperate and to follow exactly the protocol, otherwise the task is not guaranteed to be fulfilled. It is of increasing importance to design distributed control systems capable of performing trustworthy computations in the face of failures and intrusions —for instance, computer science approaches are discussed [4].

In the literature several results can be found, mostly based on reputation systems, that deal with cooperation issues among nodes of an ad-hoc network. The reputation, as defined in [5], is an index for the reliability of a node in the network, i.e., for the contribution to common network operations, and it is used to exclude uncooperative and misbehaving nodes from the network. Intentional non-cooperation is mainly caused by two types of nodes: selfish ones, that want to save power, and malicious nodes, that are not primarily concerned with power saving but that are interested in attacking the network [6]. In [7], authors develop security systems, where trust relationship and routing decision are based on routing and forwarding behavior of the nodes.

In the fault detection and isolation literature, we find several strategies to construct model-based fault detection systems [8]. One of them is the observer-based technique, whose main idea is to make the decision on possible faults in the process on the basis of the residuals generated by estimating the outputs of the process. The detection system should be immune to faults, in such a way that the differences between the outputs of the process and those of the observer give information about faults in the system. In [9], a distributed consensus algorithm is implemented by a sensors network to design a distributed fault diagnosis procedure for dynamic system, but the fault diagnosis of the consensus protocol is not considered.

In the linear consensus algorithms we consider [3], agents use the information coming from the neighbors to update their state. Since there are no forwarded messages, reputation systems are not applicable, unless we include additional communication overhead to the control mechanism. Consider for instance the prototypical problem of uniform deployment on a segment [2]. A malicious agent moving towards an extreme of the segment would push all agents on that side to the extreme point, without them being able to detect the misbehavior from the information they possess (i.e., the distance from their immediate neighbors).

The main contributions of this work are as follows. We apply a technique based on unknown-input observers to the problem of intrusion detection for a linear consensus algorithm. We give conditions for the solvability of the problem in relation to the topology of the network, and we prove that if the network is 2-connected, then a faulty node can be detected, identified and finally isolated from the network, to preserve the functionality of the algorithm. We design an embedded filter, which, only considering the information coming from the neighbors, asymptotically estimates the state of the other nodes of the network, and we analyze the properties of the estimation error, which is strictly related to the fault induced in the system. The estimation rate of the filter is also considered, and we prove that, in the "equal-neighbor" model, as the number of agents grows, the largest eigenvalue of the observer has the same upper bound of the second largest eigenvalue of the iteration matrix of the algorithm, which determines the convergence rate of the protocol. We also compute an estimation of the convergence rate of the filter in a more general case, i.e., when the associated graph is weighted and directed. The

This material is based upon work supported in part by the ARO MURI Award W911NF-05-1-0219, and in part by Contracts IST-2004-004536 (IP) "RUNES" and IST-2004-511368 (NoE) "HYCON."

Fabio Pasqualetti is with the Centro I. R. "E. Piaggio," Università di Pisa, and with the Control, Dynamical Systems and Computation, University of California at Santa Barbara

Antonio Bicchi is with the Centro I. R. "E. Piaggio," Università di Pisa

Francesco Bullo is with the Center for Control, Dynamical Systems and Computation, University of California at Santa Barbara, bullo@engineering.ucsb.edu

approach proposed to improve the security of distributed consensus is, to the best of our knowledge, original and it is shown to be effective under a reasonable set of assumptions.

The rest of this paper is organized as follows. In Section II, we introduce the consensus algorithm in the presence of a misbehaving node. In Section III, we present some properties and prove results. In Section IV, we study the convergence rate of the proposed filter. In Section V, we propose a procedure to implement an intrusion detection system for the consensus problem. Some examples and our conclusions are respectively in Section VI and VII.

II. CONSENSUS ALGORITHMS WITH MISBEHAVING NODE

We start summarizing basic notions of graph theory. A weighted directed graph G_r is defined by a triple (V, E, A), where $V = \{1, \dots, n\}$ is a set of n vertices, E is a set of ordered pair of vertices called edges, and A is a $n \times n$ weighted adjacency matrix whose entries satisfy $a_{kj} > 0$ if the pair $(j, k) \in E$. Two nodes j and k are neighbors if $(j, k) \in E$, and the set of the neighbors of the node j is denoted as

$$N_j = \{k \,|\, a_{jk} > 0\}.$$

Let $w_i(l)$ the value at time l of the node i, an agreement algorithm updates w_i as

$$w_i(l+1) = \sum_{j=1}^{n} f_{ij} w_j(l),$$

or $w(l+1) = Fw(l)$, being $w(l)$ the vector with the values of all nodes at time l, and F the matrix with entries f_{ij}. Under some conditions the agreement algorithm is guaranteed to converge [3], so that

$$\lim_{l \to +\infty} w_i(l) = c, \text{ for all } i \in \{1, \dots, n\}.$$

In order to implement a consensus algorithm, it should be ensured that the failure of one agent to perform its designated duties, or the presence of an intruder among the nodes, does not block the task completely. Let $\{e_1, \dots, e_n\}$ be the canonical base of $\mathbb{R}^n$, the failure or the misbehavior of a node can be modeled as an external input, so that the system becomes

$$w(l+1) = Fw(l) + e_i \bar{u}(l), \quad l \in \mathbb{Z}_{\geq 0}. \tag{1}$$

The index i is unknown to the good nodes of the network, and the function $\bar{u}$ represents the arbitrary behavior of the intruder. Because of the presence of an exogenous input $\bar{u}$, the network is not able to achieve agreement.

For simplicity, let

$$y_j(l) = C_j w(l), \tag{2}$$

denote the information about the status of the network available to the agent j, where $C_j = [e_{k_1} \cdots e_{k_p}]^T$, $k_1, \dots, k_p \in N_j$, $p \in \mathbb{Z}_{\geq 0}$, and define $\Sigma_{ij} = (F, e_i, C_j)$ as the system associated with a linear consensus algorithm on $\mathbb{R}^n$ as described in (1,2).

III. SOLVABILITY CONDITIONS

Given $H \subseteq \mathbb{R}^{n \times n}$, $B \subseteq \mathbb{R}^{n \times p}$, and $C \subseteq \mathbb{R}^{q \times n}$, $(n, q) \in \mathbb{Z}_{\geq 0}$, consider the linear discrete time system

$$x(l+1) = Hx(l) + Bu(l),$$
$$y(l) = Cx(l),$$

and let $\mathbf{B} = \mathrm{Im}(B)$, $\mathbf{C} = \mathrm{Ker}(C)$. Recall that an $(H, \mathbf{C})-$conditioned invariant is a subspace $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $H(\mathbf{S} \cap \mathbf{C}) \subseteq \mathbf{S}$, and denote with $\mathbf{S}^*$ the minimal $(H, \mathbf{C})-$conditioned invariant containing $\mathbf{B}$. By definition, a conditioned invariant $\mathbf{S}$ is said to be both internally and externally stabilizable, if there exists at least one real matrix G such that $(H + GC)\mathbf{S} \subseteq \mathbf{S}$ with $(H+GC)$ stable [10]. We say that a discrete system described by matrices (H, B, C) is *unknown input observable* (UIO), if it is possible to estimate the whole state in the presence of the unknown input u. From [10] we know the following result.

Theorem 3.1 (Unknown Input Observability): Assume that the pair (H, C) is detectable, the problem of asymptotically estimating the whole state, in the presence of the unknown input u, has a solution if and only if

- $\mathbf{S}^* \cap \mathbf{C} = \emptyset$;
- $\mathbf{S}^*$ is externally stabilizable.

Let $\mathbf{e}_i = \mathrm{Im}(e_i)$, $\mathbf{C}_j = \mathrm{Ker}(C_j)$, then:

Theorem 3.2: (Unknown input observability for consensus systems): Given a linear consensus algorithm over a graph G_r, the system Σ_{ij} is UIO, for all $(i, j) \in \{1, \dots, n\}$, if and only if G_r is strongly connected and $i \in N_j$.

Proof: Since $i \in N_j$, the minimal $(F, \mathbf{C}_j)-$conditioned invariant containing $\mathbf{e}_i$ coincides with $\mathbf{e}_i$. In fact $F(\mathbf{e}_i \cap \mathbf{C}_j) = F(\emptyset) = \mathbf{0}$, and hence the first condition of Theorem 3.1 is verified. The external stabilizability of $\mathbf{S}^*$ is provided by finding a matrix G such that $(F+GC_j)\mathbf{e}_i \subseteq \mathbf{e}_i$, with $(F + GC_j)$ stable. Being $i \in N_j$, we can choose G in order to nullify the i_{th} column of F, so that $(F+GC_j)\mathbf{e}_i = \mathbf{0}$. Because of the connectivity of G_r, F is irreducible, and since $F \geq | \; F + GC_j \; |$, $F \neq | \; F + GC_j \; |$, we have $\rho(F+GC_j) < \rho(F) = 1$, being ρ the spectral radius, so that the matrix $(F + GC_j)$ is stable [11], and the pair (F, C_j) is detectable [10].

Suppose now that Σ_{ij} is UIO, then $\mathbf{S}^* \cap \mathbf{C}_j = \emptyset$, and $\mathbf{e}_i \subseteq \mathbf{S}^*$. We deduce $\mathbf{e}_i \subseteq \mathrm{Im}(C_j)$, and thus $i \in N_j$. Furthermore, if the digraph associated with F is not strongly connected, then there is at least a node j that does not receive information from a partition of the network, and that node obviously can not estimate the whole state of the system, even if the unknown input is constantly zero. We conclude that the digraph associated with F is strongly connected. ■

The UIO property can be generalized to the case of several faulty nodes, as follows

Corollary 3.1: Consider the system described by

$$w(l+1) = Fw(l) + B\bar{u}(l), \quad l \in \mathbb{Z}_{\geq 0},$$

with $B \subseteq \mathbb{R}^{n \times p}$ defined as

$$B = [e_{k_1} \cdots e_{k_p}], \quad k_1, \dots, k_p \in \{1, \dots, n\}.$$

the system (F, B, C_j) is UIO, if and only if $\mathbf{B} \subseteq \mathrm{Im}(C_j)$, and the digraph associated with F is strongly connected.

Proof: Theorem 3.2. ∎

We shall consider now the problem of isolating the misbehaving node, in order to guarantee that the consensus is reached by all working agents of the network. Once a node has detected an intruder, it simply ceases keeping into consideration the information coming from that agent, and adjusts the weights of the remaining incoming messages. The resulting matrix is row stochastic, and describes a consensus algorithm that converges to an agreement configuration, if the associated digraph is connected. Note that only the neighbors of the faulty node modify the topology of the network, so that no communication among the agents is needed to detect, identify and isolate the misbehaving node. The following Theorem formalizes these considerations.

Theorem 3.3: (Convergence in 2-connected faulty networks): Let Σ_{ij} be an UIO system. If the associated digraph is 2-connected, then there exists $M \in \mathbb{R}^{n\times n}$, with entries $m_{rk} = 0$ if $k \notin N_i$, $r \in \{1, \ldots, n\}$, such that the algorithm

$$w(l+1) = (F+M)w(l) + e_i\bar{u}(l)$$

achieves agreement for all w_r, $r \neq i$, and for all possible trajectories $\bar{u}$.

Proof: Choose M such that $(F+M)_{r,i} = 0$ and $(F+M)_{r,r} = F_{r,r} + F_{r,i}$, if $r \in N_i \setminus \{i\}$. Since $(F+M)_{r,i} = 0$ for all $r \neq i$, the unknown input does not affect the variables w_r, $r \neq i$. Moreover the submatrix obtained deleting the i_{th} row and column from $(F+M)$ is, by construction, row stochastic, primitive, and its associated digraph is strongly connected, since the one associated with F is 2-connected. These conditions are sufficient for achieving agreement among the variables w_r, $r \neq i$. ∎

The filter can be designed in many ways. We could be interested in minimizing the convergence rate of the estimation process, by placing the eigenvalues of the filter as close as possible to the origin. However, given an UIO system Σ_{ij}, the dimension of the observability subspace depends on the topology of the network, so that it is not always possible to place all the eigenvalues of the filter. For this reason, we describe a design that is applicable to any topology of network, and then we investigate its convergence rate.

Theorem 3.4 (Filter design): Let Σ_{ij} be an UIO system. If a filter is designed as

$$\begin{aligned} z(l+1) &= (F+GC_j)z(l) - Gy_j(l), \\ \tilde{w}(l) &= Lz(l) + Ky_j(l), \end{aligned}$$

with

$$G = -F_{N_j}, \quad K = C_j^T, \quad L = I_n - KC_j,$$

being F_{N_j} the columns of F with indexes N_j, then

$$\tilde{w}(l) \to w(l),$$

as $l \to +\infty$, for all possible input trajectories $\bar{u}$.

Proof: Consider the equation of the estimation error,

$$\begin{aligned} r(l+1) &= z(l+1) - w(l+1) \\ &= (F+GC_j)r(l) - e_i\bar{u}(l). \end{aligned}$$

We choose $G = -F_{N_j}$, in order to nullify the N_j columns of F. Using the same procedure as in Theorem 3.2, we note that the matrix $(F+GC_j)$ is stable, and the reachable set of r is the minimum $(F+GC_j)$ invariant containing $\mathbf{e}_i$. Since $(F+GC_j)\mathbf{e}_i = \mathbf{0}$, the reachable set of the error r is $\mathbf{e}_i$. Let $K = C_j^T$, $L = I_n - KC_j$, and consider the estimated function $\tilde{w}$:

$$\begin{aligned} \tilde{w}(l) &= Lz(l) + KC_jw(l) \\ &= w(l) + Lr(l). \end{aligned}$$

Being $\mathbf{e}_i \subset \mathrm{Ker}(L)$, the term $Lr(l)$ will converge to zero, so that, as $l \to +\infty$, $\tilde{w}(l) \to w(l)$. ∎

IV. CONVERGENCE RATE OF THE UNKNOWN INPUT FILTER

Given an $n \times n$ matrix F describing an agreement algorithm, construct the $n \times n$ matrix $\tilde{F}$ such that $\tilde{F}_{r,r} = 1$, $\tilde{F}_{r,k} = 0$, and $\tilde{F} = F$ in all the other entries, being $r \in N_j$, and $k = 1, \ldots, n$. Compute now the matrix $F + GC_j$ as described in Theorem 3.4, and recall that its largest eigenvalue is positive and smaller than 1. Let $\lambda_{\max}$ be the largest eigenvalue of $F + GC_j$, and $\lambda_2 = \max \lambda(\tilde{F})$, where the maximum is taken over all eigenvalues λ of $\tilde{F}$ different than 1. With some analysis, we note that $\lambda_{\max} = \lambda_2$. Let $\tilde{F}x_2 = \lambda_2 x_2$. Since $\lambda_{\max} = \lambda_2 < 1$, the N_j components of x_2 must be equal to 0, while the other entries are positive [11]. The eigenvalue λ_2 can be computed as

$$\lambda_2 = \max(\langle \tilde{F}x, x \rangle),$$

subject to $\langle x, x \rangle = \sum_{k=1}^{n} x_k^2 = 1$, and $x_{N_j} = 0$.

We first analyze the convergence rate of the filter applied to the equal - neighbor consensus problem, and then we present a more general result valid for any kind of consensus algorithm.

A. Equal - neighbor case

We denote with equal - neighbor consensus algorithm the agreement procedure, when the graph is un-weighted and un-directed. In this case $F = D^{-1}A$, where D, A are, respectively, the vertex degree diagonal matrix and the adjacency matrix of the graph, including self loops. A trivial lower bound for $\lambda_{\max}$ is obtained from Perron Frobenius Theorem:

$$\lambda_{\max} \geq 1 - \frac{1}{\Delta},$$

where Δ is the maximum vertex degree, while an upper bound must be investigated with other techniques, because $\| F \|_\infty = \| F + GC_j \|_\infty = 1$, for all agreement algorithms of dimension $n > 3$.

Theorem 4.1 (Convergence rate): Let G_r be the digraph associated with an UIO system Σ_{ij}. Let n be the number of vertices, Δ the maximum vertex degree including self-loops, and d the diameter of the digraph, then

$$\lambda_{\max} \leq 1 - \frac{1}{nd\Delta}.$$

Proof: Construct $\tilde{F}$ as previously described, we use the methods of [12] to find an upper bound for $\lambda_2(\tilde{F})$. The problem

$$\lambda_2 = \max(\langle \tilde{F}x, x\rangle),$$

subject to $\sum_{k=1}^{n} x_k^2 = 1$ and $x_{N_j} = 0$, can be rewritten as

$$\lambda_2 = \max(\langle Ax, x\rangle), \tag{3}$$

subject to $\sum_{k=1}^{n} d_k x_k^2 = 1$ and $x_{N_j} = 0$, being d_k the degree of vertex k, and A the adjacency matrix, both including self-loops. Our next step is to rewrite $\langle Ax, x\rangle$ in a more revealing form. By expanding $\langle Ax, x\rangle$, we find

$$\langle Ax, x\rangle = \sum_{t=1}^{n} x_t \left(\sum_{k\in N_j} x_k \right) = \sum_{(t,k)\in E} x_t x_k,$$

where E denotes the set of all the edges of the graph. Moreover, we have

$$\begin{aligned}\sum_{(t,k)\in E} x_t x_k &= \frac{1}{2} \sum_{(t,k)\in E} 2x_t x_k \\ &= 1 - \frac{1}{2} \sum_{(t,k)\in E} (x_t - x_k)^2.\end{aligned}$$

On combining this with (3), we find

$$\lambda_2 = 1 - \frac{1}{2} \min \sum_{(t,k)\in E} (x_t - x_k)^2, \tag{4}$$

subject to $\sum_{k=1}^{n} d_k x_k^2 = 1$ and $x_{N_j} = 0$. Let x_m denote the component of x which is largest in magnitude, and let $\Delta = \max_k d_k$. Since $d_k \le \Delta$, we find $1 = \sum_{k=1}^{n} d_k x_k^2 \le n\Delta x_m^2$, so that $x_m \ge (n\Delta)^{-\frac{1}{2}}$. Since $x_{N_j} = 0$, we have $x_m = x_m - x_s \ge (n\Delta)^{-\frac{1}{2}}$, being $s \in N_j$. Given the connectivity of the graph, there is a sequence of vertices of length $r \le d$, which joins vertex m to s. Letting $\{x_{k_1} \dots x_{k_r}\}$ denote the set of vertices traversed by this chain, we have

$$(n\Delta)^{-\frac{1}{2}} \le (x_m - x_s) = (x_m - x_{k_1}) + \cdots + (x_{k_r} - x_s),$$

and by Cauchy–Schwarz inequality

$$(n\Delta)^{-1} \le \frac{r}{2} \sum_{(t,k)\in E} (x_t - x_k)^2 \le \frac{d}{2} \sum_{(t,k)\in E} (x_t - x_k)^2. \tag{5}$$

Combining (5) and (4) shows that $\lambda_2(\tilde{F}) \le 1 - \frac{1}{nd\Delta}$. ■

B. Upper bound for weighted digraph

The matrix $\tilde{F}$ is an absorbing Markov chain, with absorbing nodes N_j. We renumber the states, so that the transition matrix will have the following canonical form

$$\tilde{F}_c = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix},$$

where I and Q are of appropriate dimensions. A standard matrix algebra argument shows that $\tilde{F}_c^l$ is of the form

$$\tilde{F}_c^l = \begin{bmatrix} Q^l & \left(\sum_{j=0}^{l-1} Q^j\right) R \\ 0 & I \end{bmatrix}.$$

The entries of Q^l give the probabilities for being in each of the transient states after l steps, for each possible transient starting state. Clearly $Q^l \to \mathbf{0}$ as $l \to +\infty$, and we say that the process is absorbed with probability 1.

Theorem 4.2: (Convergence rate of an absorbing Markov chain): Let $\tilde{F} \in \mathbb{R}^{n\times n}$ be the transition matrix of an absorbing Markov chain, and let x_{N_j} be the absorbing states of the chain. Let G_r be the associated weighted digraph, such that the entry $\tilde{F}_{k,j}$ represents the weight of the edge from k to j. Let d be the diameter of G_r, and Δ the maximum vertex degree including self-loops, then

$$\lambda_2(\tilde{F}) < \left(1 - \frac{\bar{n}}{\Delta^d}\right)^{1/d},$$

being $\bar{n}$ the cardinality of the set $N_j \setminus \{j\}$.

Proof: Let $\tilde{F}_c$ be the canonical form of $\tilde{F}$. Recall that the largest eigenvalue of $\tilde{F}_c$ is 1. Consequently $\lambda_2(\tilde{F}_c) = \lambda_{\max}(Q)$, where Q is the transient matrix of $\tilde{F}_c$. Recall that

$$\lambda(Q) \le \rho(Q) \le \|Q^k\|^{1/k}, \quad k \in \mathbb{Z}_{>0},$$

being ρ the spectral radius of the matrix. Suppose we start a random walk on G_r from the node k more distant to one of the nodes belonging to N_j: it takes at most d steps before reaching a node belonging to N_j, with probability greater or equal to $1/\Delta^d$. This means that

$$(\tilde{F}_c^d)_{k,s} \ge \frac{1}{\Delta^d}, \quad s \in N_j \setminus \{j\}$$

that leads to

$$\rho(Q) \le \|\tilde{Q}^d\|_\infty^{1/d} \le \left(1 - \frac{\bar{n}}{\Delta^d}\right)^{1/d}.$$

■

V. INTRUSION DETECTION SYSTEM

By analyzing the property of the iteration error, we describe a methodology to detect and identify a misbehavior for an UIO system. The working conditions of the network are then restored through the exclusion of the faulty agent, i.e., by modifying the topology of the network, such that the information he provides is not considered.

A. ANALYSIS OF THE ITERATION ERROR

Given a linear discrete time UIO system (1) and the estimation filter of the node j (2), reorder the states variables, such that index set N_j comes first. The filter can be rewritten as

$$z(l+1) = F \begin{bmatrix} w_{N_j}(l) \\ z_p(l) \end{bmatrix},$$
$$\tilde{w}(l) = \begin{bmatrix} w_{N_j}(l) \\ z_p(l) \end{bmatrix},$$

with $p \notin N_j$. Consider the iteration error

$$\begin{aligned}\varepsilon(l) &= \mid \tilde{w}(l+1) - F\tilde{w}(l) \mid \\ &= \left| \begin{bmatrix} w_{N_j}(l+1) \\ z_p(l+1) \end{bmatrix} - F \begin{bmatrix} w_{N_j}(l) \\ z_p(l) \end{bmatrix} \right|.\end{aligned}$$

Recall that $L = I_n - K$, and $K = C_j^T C_j$; we can rewrite ε as

$$\varepsilon(l) = \left| LF \begin{bmatrix} w_{N_j}(l) \\ z_p(l) \end{bmatrix} + Kw(l+1) - F \begin{bmatrix} w_{N_j}(l) \\ z_p(l) \end{bmatrix} \right| = \left| \begin{array}{c} [w(l+1) - F\tilde{w}(l)]_{N_j} \\ \mathbf{0} \end{array} \right| = \left| \begin{array}{c} \bar{\varepsilon} \\ \mathbf{0} \end{array} \right|.$$

Since $\tilde{w}(l) \rightarrow w(l)$ as $l \rightarrow +\infty$, $(\tilde{w}(l+1) - F\tilde{w}(l)) \rightarrow B_i\bar{u}(l)$, so that we can detect and identify the intruder. The following three cases are possible:

1) there is no unknown input in the system. In this case $\| \bar{\varepsilon}(l) \| \rightarrow 0$, as $l \rightarrow +\infty$;
2) there is a faulty node i and $i \in N_j$. In this case the i_{th} component of $\bar{\varepsilon} \rightarrow \bar{u}$, while the other components converge to zero as fast as the convergence rate of the filter;
3) there is a faulty node i and $i \notin N_j$. In this case $\bar{\varepsilon}$ do not converge to zero in more than 1 component. In fact, the estimation error is $r(l+1) = (F + GC_j)r(l) - B_i\bar{u}(l)$, and its reachable set is not included into $\mathrm{Ker}(L)$. We deduce $\tilde{w}(l) = w(l) + Lr(l)$ does not converge to $w(l)$.

B. INTRUSION DETECTION PROCEDURE

Consider the iteration error $\bar{\varepsilon}$, we have

$$\bar{\varepsilon}(l) \leq \lambda^l_{\max}\varepsilon_0, \quad l \in \mathbb{Z}_{\geq 0},$$

where $\varepsilon_0 = \max(| \tilde{w}(1) - F\tilde{w}(0) |)$. The iteration error can be written as $\varepsilon(l) = e(l) + e_u(l)$, where e denotes here the error due to the estimation process, while e_u is due to the unknown input. If

$$e(l) + e_u(l) \geq \lambda^l_{\max},$$

then we can detect a fault in the system. Since $\lambda^l_{\max}$ vanishes as $l \rightarrow +\infty$, the recognizable misbehavior becomes smaller with l. The amount of time $\bar{l}$ we need to wait to ensure a correct estimation of the misbehavior, depends on the minimal unknown input we want to recognize and on $\lambda_{\max}$. We do not specify this parameter, and we consider that after $\bar{l}$ steps the estimation error is small enough to identify the intruder. A possible procedure is:

1) each node j builds an observer as described in Theorem 3.4;
2) each agent computes $\bar{\varepsilon} = | \tilde{w}(l+1) - F\tilde{w}(l) |, l \geq \bar{l}$; calling $\bar{\varepsilon}_k$ the k_{th} variable of the vector $\bar{\varepsilon}$, and letting $\delta = \lambda^l_{\max}$,
   a) if $\bar{\varepsilon}_k < \delta$ for all $k \in N_j$, then no fault is detected;
   b) if $\bar{\varepsilon}_k < \delta$ for all $k \in N_j \setminus \{i\}$, and $\bar{\varepsilon}_i \geq \delta$, then the i_{th} agent is an intruder;
   c) if $\bar{\varepsilon}_k \geq \delta$ for more than one $k \in N_j$, then there is a misbehavior in the network, but the identity of the faulty node is not determined.
3) while no intruder is found, step 2 is repeated. Otherwise, as soon as agent j detects that agent i is an intruder, agent j changes the topology of the network according to Theorem 3.3.

VI. EXAMPLES

Possible applications for the proposed security system are the consensus and the deployment problems. In the first case, we have n agents as node of a digraph that want to agree on one value. A possible scenario could be the one defined by n sensors of temperature that want to evaluate the mean of the temperatures they are measuring, or the one of n processors that want to synchronize their clock. In the deployment task, we have n mobile agents that want to deploy uniformly on a geometric figure like a circle or a segment.

A. THE CONSENSUS PROBLEM

Suppose we have 8 nodes disposed on a digraph as in Fig. 1 and suppose the node number 6 is the faulty node. Considering the node number 5 as observer, the system Σ_{65} is described by

$$\begin{aligned} w(l+1) &= Fw(l) + B_6\bar{u}(l) \\ y_5(l) &= C_5 w(l), \end{aligned}$$

where F is the corresponding equal-neighbor consensus matrix, $B_6 = e_6$, and $C_5 = [e_1 \ e_5 \ e_6]^T$. Note that the conditions of Theorem 3.2 are verified, thus we can build a whole state unknown input observer, whose matrices are

$$G = \begin{bmatrix} -1/3 & -1/3 & 0 \\ -1/3 & 0 & -1/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \\ 0 & 0 & -1/3 \\ 0 & 0 & 0 \\ 0 & -1/3 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let $\bar{u} = \text{constant} = 10$, we initialize the state of the nodes with random variables, and we simulate the system for 40 steps. Fig. 2 shows the iteration error related to the estimation of the variables 1 and 6. As we see, the iteration error for the variable 6 converges to $\bar{u} = 10$, while the error for the variable 1 goes to zero, so that node 5 can detect the fault, and identify the misbehaving node. The same procedure is applied by agent 2, and finally the intruder is isolated from the network (Fig. 3).

B. THE DEPLOYMENT PROBLEM

We consider 6 agents that are moving on the unit circle in order to reach an uniform distribution over it, and we refer to the algorithm described in [2]. Construct the system Σ_{43} as in (1,2). Theorem 3.2 is verified, and the related filter is

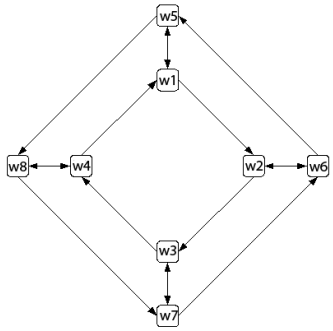

Fig. 1. Consensus network.

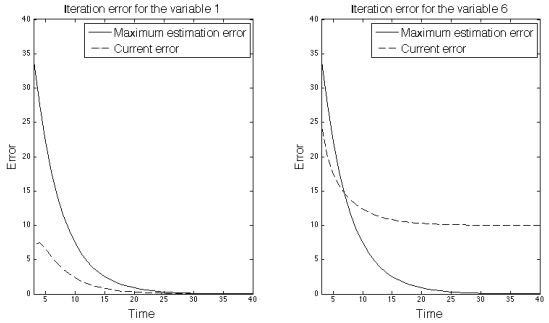

Fig. 2. Identification of the misbehaving node.

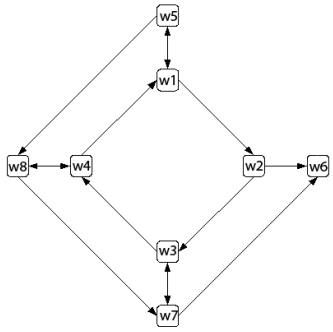

Fig. 3. Modified consensus network.

characterized by

$$G = \begin{bmatrix} -1/4 & 0 & 0 \\ -1/2 & -1/4 & 0 \\ -1/4 & -1/2 & -1/4 \\ 0 & -1/4 & -1/2 \\ 0 & 0 & -1/4 \\ 0 & 0 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Fig. 4 shows the iteration error, which leads to the identification of the malicious agent, when the unknown input is a sinusoidal signal with amplitude 0.3 and unitary frequency.

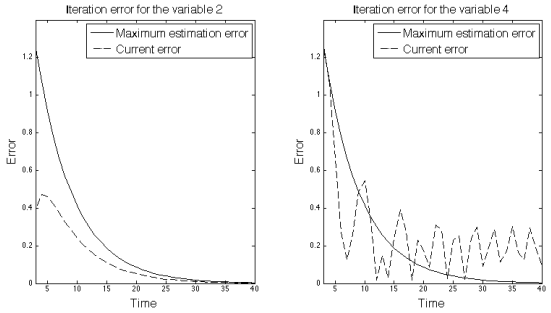

Fig. 4. Identification of the intruder.

VII. CONCLUSIONS

We considered consensus networks in the presence of a misbehaving node, and we proposed a technique based on the theory of Unknown Input Observability to detect, identify, and isolate the misbehavior from the network. We designed an embedded filter, which, only considering the information used by the control protocol, estimates the state of the nodes in the network, allowing the identification of the faulty agent. We analyzed the property of the iteration error of the unknown input filter, and we finally proposed a complete procedure to perform the intrusion detection and isolation task.

REFERENCES

[1] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, 2005.

[2] S. Martínez, F. Bullo, J. Cortés, and E. Frazzoli, "On synchronous robotic networks – Part I: Models, tasks, and complexity. Part II: Time complexity of rendezvous and deployment algorithms," *IEEE Transactions on Automatic Control*, 2007. To appear.

[3] A. Olshevsky and J. N. Tsitsiklis, "Convergence rates in distributed consensus and averaging," in *IEEE Conf. on Decision and Control*, (San Diego, CA), pp. 3387–3392, Dec. 2006.

[4] N. A. Lynch, *Distributed Algorithms*. San Mateo, CA: Morgan Kaufmann Publishers, 1997.

[5] P. Michiardi and R. Molva, "CORE: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc network," in *Communications and Multimedia Security Conference (CMS)*, (Portoroz, Slovenia), Sept. 2002.

[6] S. Buchegger and J.-Y. L. Boudec, "Cooperative routing in mobile ad-hoc networks: Current efforts against malice and selfishness," *Proceedings of Mobile Internet Workshop. Informatik*, pp. 513–517, Sept. 2002.

[7] L. Buttyan and J. P. Hubaux, "Stimulating cooperation in self-organizing mobile ad hoc networks," *ACM Journal for Mobile Networks (MONET)*, vol. 8, no. 5, pp. 579–592, 2003.

[8] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of Process Control*, vol. 7, no. 6, pp. 403–424, 1997.

[9] E. Franco, R. Olfati-Saber, T. Parisini, and M. M. Polycarpou, "Distributed fault diagnosis using sensor networks and consensus-based filters," in *IEEE Conf. on Decision and Control*, (San Diego, CA), pp. 386–391, Dec. 2006.

[10] G. Basile and G. Marro, *Controlled and Conditioned Invariants in Linear System Theory*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[12] H. J. Landau and A. M. Odlyzko, "Bounds for eigenvalues of certain stochastic matrices," *Linear Algebra and its Applications*, vol. 38, pp. 5–15, 1981.