52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Gossip Average Consensus in a Byzantine Environment Using Stochastic Set-Valued Observers

Daniel Silvestre, Paulo Rosa, Rita Cunha, João P. Hespanha, Carlos Silvestre

Abstract— **We address the problem of a consensus system in the presence of Byzantine faults seen as an attacker injecting a perturbation in the state of the nodes. We propose the use of Set-Valued Observers to detect if the state observations are compatible with the system dynamics. The method is extended to the stochastic case by introducing a strategy to construct a set that is guaranteed to contain all possible states with, at least, a pre-specified desired probability. The proposed algorithm is stable in the sense that it requires a finite number of vertices to represent polytopic sets while also enabling the a priori computation of the largest magnitude of a disturbance that an attacker can inject without being detected.**

I. Introduction

The problem of consensus relates to a set of agents agreeing on a common value using a distributed algorithm. We are interested in randomized gossip average consensus in that nodes are allowed to send messages to a random neighbor in order to compute the average of their initial state. By nature, such algorithms are designed to cope with "crash type" faults by using redundancy and randomization. However, Byzantine faults, such as an intruder in the system, can prevent convergence or drive the system steady state to any value [1]. The applicability of Byzantine fault detection schemes ranges from consensus algorithms to information dissemination and distributed control of industry processes.

The problem of detecting Byzantine faults using unreliable fault detectors was introduced in [2], where multiple classes of theoretic detectors are presented. For the specific case of a consensus system, an algorithm is proposed that makes use of an unreliable detector to solve the problem of consensus. However, [2] differs from our work in that the consensus value is assumed to be one of the initial values, whereas we are looking at asymptotically reaching the average of the initial values by using a linear dynamic system.

The consensus problem has been widely studied when considering rather non-antagonistic failure models which include packet drops and nodes crashing but, to enable a more comprehensive model, Byzantine faults must be considered. The research interest in this issue has motivated a number of contributions including the scenario of unreliable networks in distributed systems. In particular, [1] considers the problem of detecting and correcting the state of the system in the presence of a Byzantine fault. The case of malicious agents and faulty agents is studied and the authors provide, in both cases, bounds on the number of corrupted nodes to ensure detectability of the fault. In [1], the system dynamics are described by a linear time-invariant model that constrains the communication network to be fixed in all time slots. Here, however, a randomized gossip algorithm is considered, thus dropping the assumption that the same sets of nodes are involved in message exchanges at every time instant.

The main contributions of this paper are as follows:

- the analysis of the problem of detecting an intruder in a randomized gossip consensus algorithm is recast into the framework of Linear Parameter-Varying (LPV) systems with uncertain dynamics;
- an upper bound on the magnitude of the attacker signal is derived beyond which the attacker can be detected;
- the concept of stochastic Set-Valued Observers (SVOs) is introduced by taking advantage of the use of α-confidence sets, i.e., sets where the state of the system is known to belong with a desired pre-specified $1-\alpha$ probability; which can be viewed as a generalization of confidence intervals;
- finally, it is also shown that this method inherits the main properties regarding the computational stability by having a bound on the volume and number of vertices needed to define the polytopic sets where the state is contained with a pre-specified probability; guarantees of intruder detection are also provided.

The choice for representing the set of possible states depends on a mathematical formulation that enables fast and non-conservative intersections and unions of sets, as those are major and normally time-consuming operations when implemented in a computer. One alternative is to use the concept of zonotopes, described in [3] and further developed in [4] and [5]. In this article, an alternative approach is adopted, based on the concept of Set-Valued Observers (SVOs) first introduced in [6] and [7] and further information can be found in [8] and [9] and the references therein.

The applicability of the proposed method for fault detection in a randomized gossip algorithm is not limited to the consensus problem, as several challenges in the Fault Detection and Isolation (FDI) literature - [10], [11] - share the framework described in the sequel. In [12], the authors propose the use of SVOs for fault detection as a model falsification problem. This paper extends the results in [12] to detect Byzantine faults in consensus systems, by rewriting its dynamics as an LPV. Moreover, unlike the approach in [12], the method proposed herein takes into account the probability information related to having a given communication.

The remainder of this paper is organized as follows. In Section II, we describe the problem of randomized gossip consensus in the presence of Byzantine faults. The proposed solution is given in Section III and the main properties of the obtained results are stated in Section IV and illustrated

D. Silvestre, R. Cunha, C. Silvestre are with the Dep. of Electrical and Computer Engineering, Instituto Superior Técnico, ISR, 1046-001 Lisboa, Portugal. {dsilvestre,rita,cjs}@isr.ist.utl.pt. This work was supported by the FCT project [PEst-OE/EEI/LA0009/2013]. The work from Daniel Silvestre is partially funded with grant SFRH//BD/71206/2010, from Fundação para a Ciência e a Tecnologia.

P. Rosa is with Deimos Engenharia, Lisbon, Portugal. paulo.rosa@deimos.com.pt.

João P. Hespanha is with the Dept. of Electrical and Computer Eng., University of California, Santa Barbara, CA 93106-9560, USA. J. Hespanha was supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under grants No. W911NF-09-1-0553 and W911NF-09-D-0001. hespanha@ece.ucsb.edu

through simulations in Section V. Concluding remarks and potential future work are presented in Section VI.

Notation : The transpose of a matrix A is denoted by $A^\intercal$. For vectors a_i, $(a_1, \ldots, a_n) := [a_1^\intercal \quad \ldots a_n^\intercal]^\intercal$. We let $1_n := [1 \ldots 1]^\intercal$ and $0_n := [0 \ldots 0]^\intercal$ indicate n-dimensional vector of ones and zeros, respectively, and I_n denotes the identity matrix of dimension n . Dimensions are omitted when clear from context. The vector e_i denotes the canonical vector whose components are equal to zero, except for the ith component. The notation $||.||$ refers to $\|v\| := \sup_i |v_i|$ for a vector, and $\|A\| := \bar{\sigma}(A)$.

II. Problem Statement

We consider a set of m agents with scalar state $x_i(t)$, $1 \leq i \leq m$ running a distributed iterative algorithm that guarantees convergence of the state to its initial average value, i.e.,

$$\lim_{t\to\infty} x_i(t) = x_{\text{av}} := \frac{1}{m}\sum_{i=1}^{m} x_i(0). \tag{1}$$

We refer to this problem as the *average consensus problem*.

In a gossip framework, at each transmission time, each node i chooses a random out-neighbor j according to the probability w_{ij}. Only nodes involved in the communication can change their state according to the received information. Thus, we regard transmission times as a discrete variable k that corresponds to the continuous variable t as between communications the state remains constant.

The communication topology is modeled by a graph $G = (V, E)$, where V represents the set of m agents, also denoted by nodes, and $E \subseteq V \times V$ is the set of communication links. Node i can send a message to node j, if $(i, j) \in E$. If there exists at least one $i \in V$ such that $(i, i) \in E$ we say that the graph has self-loops, which can model, for example, packet drops, since node i only has access to its own value at that transmission time. We associate to graph G a *weighted adjacency matrix* W with entries:

$$W_{ij} := \begin{cases} w_{ij}, \text{ if } (i,j) \in E \\ 0, \text{ otherwise} \end{cases}, \tag{2}$$

where the weights $w_{ij} \in [0, 1]$.

We consider here radomized gossip algorithms in a Byzantine environment of the form

$$x(k+1) = U(k)x(k) + B(k)u(k), \tag{3}$$

where the matrix $U(k)$ is selected randomly from a set $\{Q_{ij}, (i,j) \in E\}$ and $u(k)$ models Byzantine faults. The random selection of $U(k)$ models the process by which nodes select a random out-neighbor, as described above, i.e., with probability $\frac{w_{ij}}{m}$ a transmission between node i and j occur which changes the states x_i and x_j according to the *column stochastic* matrix Q_{ij} (i.e. $1^\intercal Q_{ij} = 1^\intercal$) to preserve the average. The input $u(k)$ models the behavior of nodes reporting incorrect state values or updating their state by something other than the "fault-free" averaging rule

$$x(k+1) = U(k)x(k) \tag{4}$$

In this paper, we assume symmetry in the communication and use as definition for the matrices Q_{ij} the algorithm proposed in [13], which we recall here for readability:

$$Q_{ij} = I - \frac{(e_i - e_j)(e_i - e_j)^\intercal}{2}$$

A consensus system S, as defined above, refers to the pair of equations:

$$\begin{cases} x(k+1) = U(k)x(k) + B(k)u(k) \\ y(k) = C(k)x(k) \end{cases} \tag{5}$$

From node i point of view, the output of the system $y(k)$ is the states that it can measure at time instance k. If node i transmitted to node j this will be the vector with the state x_i and x_j ($C(k) = [\mathsf{e}_i, \mathsf{e}_j]^\intercal$) and will only have its own state if the node did not communicate ($C(k) = [\mathsf{e}_i, \mathsf{e}_i]^\intercal$). With a slight abuse of notation, we use $y(k)$ to refer to the output of the system at time k and $y_k(x(0), u_k)$ to express the same output as a function of the initial state $x(0)$ and input u_k, where u_k denotes the vector of inputs up to time k.

The main goal of this paper can therefore be stated as: developing algorithms for detecting nonzero inputs $u(k)$ in (3) that do not require knowledge of the matrices $B(k)$ and signal $u(k)$ in (3) and, instead, only use the measurements y_k which stands for all the measurements up to time k.

Definition 1 (undetectable faults): Take the consensus system (5). A nonzero input sequence u_k (corresponding to a fault) is said to be *undetectable in N iterations* if:

$$\forall k < N, \exists x(0), x'(0) \in \mathbb{R}^m : y_k(x(0), u_k) = y_k(x'(0), 0).$$

The intuition behind this definition is that a fault is only detectable if there is no possible set of initial conditions such that the sequence $y(0) \cdots y(N)$ of measurable states can be generated without an attacker signal.

Assumption 2 (detectable faults): Each fault considered can be defined by means of an input sequence u_k, and is detectable in the sense of Definition 1.

The fault being detectable as in Assumption 2 relates to the observability of the system as in [14].

Assumption 3 (bounded state): $\forall k < N, \|x(k)\| < c$ for a given constant c.

Assumption 3 is sustained by the fact that a non-faulty gossip algorithm has a bounded state. Therefore, there exists a constant c that if the state is larger one could trivially detect the occurrence of the fault.

III. Proposed Solution

We start by introducing the problem from a worst-case scenario perspective and by finding suitable computational machinery to detect Byzantine faults. Progress is then made to generalize the algorithm to include the probability information and convert the detection into a stochastic scenario.

A. Worst-case scenario

In a worst-case scenario, all the realization of $U(k)$ are assumed possible and the "low probability" of specific events cannot be used to infer the likelihood of a fault.

The dynamics of the system can be cast into a Linear Parameter-Varying (LPV) model with uncertainty in the time-varying matrix $U(k)$ by rewriting them as a central matrix and a sum of uncertainties resulting in (3) being:

$$x(k+1) = \Big(A_0 + \sum_{\ell=1}^{n_\Delta} \Delta_\ell(k) A_\ell\Big)x(k) + B(k)u(k) \tag{6}$$

where n_Δ is the number of required uncertainties and each Δ_ℓ is a scalar uncertainty with $|\Delta_\ell| = 1$.

Detecting a fault in a worst-case scenario reduces to finding whether a given sequence of observations, y_k, can be

generated by the dynamics in (6) with zero terms $B(k)u(k)$ or not, for any admissible initial conditions $x(0)$.

Definition 4 (Distinguishability [15]): Let S_A and S_B be two systems of the form (5) with outputs at discrete time k, $y_A(k)$ and $y_B(k)$, and initial states $x_A(0)$ and $x_B(0)$, respectively. Then, S_A and S_B are said distinguishable in $N \geq 0$ measurements if, for any

$$(x_A(0), x_B(0), \phi(0), \cdots, \phi(N-1)) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^{n_\phi}, \cdots, \mathbb{R}^{n_\phi}$$

there exists $k \in \{0, 1, \cdots, N-1\}$ such that

$$y_A(k) \neq y_B(k).$$

where ϕ is the vector of disturbances, uncertainties and inputs.

This notion of distinguishability in Definition 4 is closely related to the notion of fault detectability in Definition 1 introduced before. We study fault detectability by considering the distinguishability between the measurements of the real system and the measurements generated by a virtual "fault-free" model characterized by zero input.

We use the SVO framework from [15] and take advantage of the distinguishability concept to derive a bound on the magnitude of the injected attacker signal, such that the attack is detected whenever this bound is exceeded. This will be one of the main contributions of this paper.

Notice that the distinguishability definition used for detecting a Byzantine fault is combinatorial by nature. Computing the set of possible state realizations is a matter of making the union of the possible state realizations for each combination of transmissions. In (6), that behavior is modeled by the vector of uncertainties $\Delta_\ell(k)$. Thus, the problem grows exponentially as the number of measurements N increases.

We adopt a similar notation as in [16] and define, at transmission time k, $X(k) := \text{Set}(M(k), m(k))$ where $\text{Set}(M, m) := \{q : Mq \leq m\}$ represents a convex polytope containing the vectors $q \in R^n$ that satisfy the constraint $Mq \leq m$, where $\leq$ is a component-wise operation between the two vectors. The aim of an SVO is to find an approximation of the smallest set containing all possible states of the system, at time k, $\tilde{X}(k)$ with the knowledge that $\forall 0 \leq i \leq N, x(k-i) \in \tilde{X}(k-i)$ and that the dynamics of the system are as in (6). Assumption 3 is needed to include the initial state into a polytope and use the SVO technique to compute a convex over-approximation. In other words, at each time k, $\tilde{X}(k)$ is an approximation of the set containing all possible states, $X(k)$, such that $X(k) \subseteq \tilde{X}(k)$.

More precisely, the initial state $x(0) \in X(0)$ where $X(0) := \text{Set}(M_0, m_0)$ and we can select M_0 and m_0 such that the corresponding polytope is guaranteed to contain the initial state. For a given uncertainty instantiation $\Delta^\star$, the set $X(k+1) := \text{Set}(M_{\Delta^*}(k+1), m_{\Delta^*}(k+1))$, which contains all the possible states of the system at time $k+1$, can be described by the set of points, $\mathbf{x}$, satisfying

$$\underbrace{\begin{bmatrix} M(k)(A_0 + A_{\Delta^*})^{-1} \\ C(k+1) \\ -C(k+1) \end{bmatrix}}_{M(k+1)} \mathbf{x} \leq \underbrace{\begin{bmatrix} m(k) \\ y(k+1) \\ -y(k+1) \end{bmatrix}}_{m(k+1)} \tag{7}$$

where

$$A_{\Delta^*} = \sum_{\ell=1}^{n_\Delta} \Delta_\ell^* A_\ell$$

and Δ_ℓ^* is the realization of the uncertainty for the current transmission time. This procedure assumes an invertible transmission matrix. When this is not the case, we can adopt the strategy in [17] and solve the inequality

$$\begin{bmatrix} I & -A_0 - A_{\Delta^*} \\ -I & A_0 + A_{\Delta^*} \\ C(k+1) & 0 \\ -C(k+1) & 0 \\ 0 & M(k) \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^- \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ y(k+1) \\ -y(k+1) \\ m(k) \end{bmatrix} \tag{8}$$

by applying the Fourier-Motzkin elimination method [18] to remove the dependence on $\mathbf{x}^-$ and obtain the set described by $M(k+1)\mathbf{x} \leq m(k+1)$.

Let the coordinates of each vertex of the hypercube $H := \{\delta \in \mathbb{R}^{n_\Delta} : |\delta| \leq 1\}$ be denoted by $\theta_i, i = 1, \cdots, 2^{n_\Delta}$. Using (7) (or (8)) we compute $X_{\theta_i}(k)$ with $\Delta^* = \theta_i$. Thus, the set of all possible states at time $k+1$ can be obtained by

$$X(k+1) = \bigcup_{\theta_i \in H} \text{Set}(M_{\theta_i}(k+1), m_{\theta_i}(k+1))$$

where we make the union for all the vertices θ_i that appear in the graph, and where M_{θ_i} and m_{θ_i} are obtained using (7) or (8). It should be noticed that not all the vertices of the hypercube H are of interest, as some communications never take place due to the limited connectivity of the network. The convex hull, $\tilde{X}(k+1)$, of set $X(k+1)$ is then obtained by using the methods described in [12], since, in general, the set $X(k+1)$ is non-convex even if $X(k)$ was convex. We recall Propositions 6.1 and 6.2 in [16] for completeness.

Proposition 5 (Membership of $X(k)$ in $\tilde{X}(k)$): Consider a system described by (6) and assume that $X(0) \subseteq \tilde{X}(0)$. Then, $X(k) \subseteq \tilde{X}(k)$ for all $k \in \{0, 1, 2, \cdots\}$.

Proposition 6 (Growth of $\tilde{X}(k)$): Suppose that a system described by (6) with $x(0) \in X(0)$ and $u(k) = 0, \forall k$, satisfies, for a sufficiently large N^*,

$$\max_{\substack{\Delta(k), \cdots, \Delta(k+N) \\ |\Delta(m)| \leq 1, \forall m \\ k \geq 0}} \left|\left| \prod_{j=k}^{k+N} \mathcal{A}(j) \right|\right| < 1,$$

for all $N \geq N^*$, and where

$$\mathcal{A}(j) := \Big[A_0 + \sum_{i=1}^{n_\Delta} \Delta_i(j) A_i(j)\Big].$$

Then, $\tilde{X}(k)$ cannot grow unbounded.

In summary, from Propositions 5 and 6, and a bound on the perturbation, one concludes that $X(k) \subseteq \tilde{X}(k)$, $\forall k$ if $X(0) \subseteq \tilde{X}(0)$. Moreover, the set $\tilde{X}(k)$ cannot grow without bound, in the sense that its volume is bounded, and it exists a hyper-parallelepiped $\tilde{X}(k)$ that, at each time, contains the set $X(k)$, and has a bounded distance between any two vertices.

Proposition 6 does not apply directly to our case since, by definition, the norm of any product of matrices Q_{ij} is equal to 1 since the matrices are contraction matrices except for the eigenvector 1. However, since there are no disturbances, we can relax the assumption and require the norm of the product to be smaller or equal to one and it follows the same reasoning as the proof of the proposition.

Notice that using the method provided before to compute $M(k)$ and $m(k)$ for the "fault-free" model gives a set where

our measurements can take values (Proposition 5 is essential to establish it). By doing the intersection with the vector of measurements $y(k)$ for the real system, if it results in an empty set, the real system is distinguishable from the virtual "fault-free" model and a fault is detectable.

An important and relevant issue regarding the construction of the SVO is its decentralized construction, which is fundamental when analyzing distributed systems. We point out that a node just requires the following: the information of the signal y that it measures when communicating with others; the matrix $C(k)$ and thus needs to be able to determine which node it is transmitting to; and its own previous estimate set which corresponds to know $M(k)$ and $m(k)$. All the matrices $A_{\Delta^\star}$ for the necessary combinations of the uncertainties Δ_ℓ can be determine if the node knows the global network structure or the node can use all possible links when it is not known. However, in a practical scenario, in order to optimize the convergence rate, the nodes will compute the matrix W in a distributed fashion and the global network structure can be inferred as the support graph of the matrix W.

B. Stochastic Set-Valued Observers

The worst-case scenario considered in the previous section completely ignores the probabilistic structure behind the selection of the matrices $U(k)$.

To understand how this information can help in detecting Byzantine faults, consider a complete 5-node network ($m = 5$) and time horizon to detect the fault $N = 20$. Each node i takes a measurement $x_i(0)$ of a quantity of interest and then a consensus procedure starts, in order to calculate the average of the initial values of the nodes $x_{\text{av}} = \sum_{i=1}^{5} x_i(0)/5$. Let us assume that the packet drop probability $p_{\text{drop}} = 0.01$, which means that the matrix of probabilities guaranteing the fastest convergence, calculated using a semi-definite program as in [13], is given by

$$W = \frac{(1 - p_{\text{drop}})}{m-1} 11^\intercal + \frac{m p_{\text{drop}} - 1}{m-1} I$$

where a packet drop is represented as a transmission from node i to itself, using the transmission matrix $Q_{ii} = I$.

If a node did not communicate, it is only able to determine its own state. Suppose that the states of the agents start dissimilar from each other but that during the first N time steps, all agents are faulty and keep their states unchanged, i.e., $x(k) = x(0), \forall k \leq N$. This fault is undetectable according to Definition 1 since the sequence of N failed transmissions due to the physical medium mimics the same behavior. Consequently, using the algorithm in the previous section, $x(0)$ must remain in the set $\tilde{X}(k), \forall k$ and therefore a fault is not detected. However, the probability of obtaining the sequence $x(k) = x(0), \forall k \leq N$ is exceedingly small:

$$\text{Prob}[\{x(k) = x(0), \forall 0 \leq k \leq N\}] = 10^{-40}$$

and is more likely to be a Byzantine fault. The inability of the SVO to incorporate the probability associated with each event is, therefore, a main drawback. Such an example motivates the introduction of Stochastic Set-Valued Observers (SSVOs) where the polytope containing the possible state is associated with a probability. The objective of this section concerns with extending the SVO concept to cope with the probability of getting a given sequence of measurements.

Consider the algorithm described in the previous subsection to generate the sets $\tilde{X}(k)$ and recall that we rewrote each Q_{ij} as in (6), therefore, associating with each hypercube vertex θ_{ij} a transmission matrix Q_{ij} with correspondent probability w_{ij}. Take the map $\psi : H \mapsto E$ which gives the correspondence between the vertices of the hypercube H and the edges in set E and let us collect the minimum number of vertices θ_{ij} in Θ such that $\sum_{\theta_{ij}} w_{\psi(\theta_{ij})} \geq 1 - \alpha$. The set for the SSVO $\bar{X}(k)$ is then an α-confidence set defined as:

$$\bar{X}(k) := \bigcup_{\theta_{ij} \in \Theta} \text{Set}(M_{\theta_{ij}}(k), m_{\theta_{ij}}(k)) \tag{9}$$

Computationally, it requires to sort the vertices θ_{ij} according to probabilities $w_{\psi(\theta_{ij})}$ as to construct Θ and then determining $M_{\theta_{ij}}(k)$ and $m_{\theta_{ij}}(k)$ as before. θ_{ij} depends on the selected edges and there can be multiple sets Θ generating an α-confidence set and any of them will suffice.

IV. Main Properties

In this section, we start by providing a result showing that the set generated by the SSVO is a α-confidence set.

Proposition 7: Take the definition of $\bar{X}(k)$ as in (9). Then, $\forall k, \bar{X}(k)$ is a α-confidence set.

Proof: The result is straightforward from the fact

$$\begin{aligned}\text{Prob}\Big[x(k) \in \bigcup_{\theta_i \in \Theta} \text{Set}(M_{\theta_i}(k), m_{\theta_i}(k))\Big] &\geq \sum_{\theta_i \in \Theta} w_{\psi(\theta_i)} \\ &\geq 1 - \alpha\end{aligned}$$

In view of Proposition 7, if we have a detectable fault as in Definition 1, and declare the fault to be detected when the set $\bar{X}(k)$ becomes empty, then, the probability of having a false positive is less or equal to α. In addition, by the construction of the set $\bar{X}(k)$, we observe that $X(k)$ is α-confidence set with $\alpha = 0$ and we have $\bar{X}(k) \subseteq X(k)$.

Borrowing the definitions in [16] for a horizon of N:

$$(A_N, b_N) = LFM\left(\begin{bmatrix} M_N \\ -M_N \\ \tilde{M}_0 \\ \tilde{M}_W \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \tilde{m}_0 \\ \tilde{m}_W \end{bmatrix}, 2m\right) \tag{10}$$

where the LFM stands for the left Fourier-Motzkin elimination method and:

$$\tilde{M}_0 = [diag(M_0, M_0) \quad 0 \quad 0 \quad 0], \tilde{m}_0 = \begin{bmatrix} m_0 \\ m_0 \end{bmatrix},$$

$$\tilde{M}_W = [0 \quad diag(M_d, \cdots, M_d)], \tilde{m}_W = [m_d^\intercal \quad \cdots \quad m_d^\intercal],$$

$$M_N = \left[\begin{array}{cc|c} C_A & -C_B & \\ C_A A_A & -C_B A_B & \bar{R} \\ \vdots & \vdots & \\ C_A A_A^N & -C_B A_B^N & \end{array}\right],$$

$$\bar{R} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ R_1^1 & 0 & \cdots & 0 \\ R_1^2 & R_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ R_1^N & R_2^N & \cdots & R_N^N \end{bmatrix},$$

$$R_i^k = \begin{bmatrix} C_A A_A^{k-i} B_A & -C_B A_B^{k-i} B_B \end{bmatrix}.$$

where the M_d and m_d define the set for the signal u, i.e., M_d and m_d are defined such that $u(k) \in \text{Set}(M_d, m_d)$; and

A_i, B_i and C_i with $i \in \{A, B\}$ are the matrices from the two systems. With a slight abuse of notation, we are writing the product of N matrices $A(k)$ as A^N for shorter notation.

The next proposition provides a theoretical bound on the attacker signal magnitude to guarantee detection.

Proposition 8 (Attacker signal bound [16]): Consider a "fault-free" system S_A and a faulty system S_B as in (5):

$$S_A = \begin{cases} x_A(k+1) = A(k)x_A(k) \\ y_A(k) = C(k)x_A(k) \end{cases}$$

$$S_B = \begin{cases} x_B(k+1) = A(k)x_B(k) + B(k)u(k) \\ y_B(k) = C(k)x_B(k) \end{cases}$$

where $u \in \mathbb{R}^{n_u}$, $x_i \in \mathbb{R}^m$, $y_i \in \mathbb{R}^2$, initialized with the same initial conditions and let the $X(k) = \text{Set}(A_N, b_N)$, where the current state is contained, be defined as in (10).

Further define:

$$P = \frac{1}{N} diag(0_{n_u}, I_{n_u}, \cdots, 0_{n_u}, I_{n_u})$$

and let $\gamma_{min} \geq 0$ be defined such that

$$\gamma_{min} \geq \max_{A_N \xi \leq b_N} \xi^\intercal P \xi$$

where the vector ξ stacks all the measurements, initial states and perturbation from the attacker. Then, system S_A and S_B are distinguishable in N measurements if

$$\frac{1}{N} \sum_{k=0}^{N} ||u(k)||^2 > \gamma_{min} \tag{11}$$

Notice that in Proposition 8, the parameter γ_{min} is the smallest "disturbance" that an attacker can inject in the system before system S_A and S_B are distinguishable in the sense that the measured output of the faulty system cannot be generated by the dynamics of the non-faulty one. Using the same line-of-thought to derive the following result.

Corollary 9 (Attacker signal bound for SSVO): Consider a non-faulty system S_A and a faulty system S_B as in Proposition 8. Then, the Byzantine fault is detectable in N measurements with a false positive probability lower or equal than α if (11) is verified.

V. Simulation Results

In this section, we present simulation results for some meaningful scenarios illustrating specific features of the proposed fault detection scheme. We consider the scenario of detecting an unresponsive node and also an attacker employing the best available strategy. A third type of fault is detected by the SSVO to motivate the use of the stochastic information, where a worst-case detection is not suitable.

We consider a 5-node network with nodes labelled $i, i \in \{1, 2, 3, 4, 5\}$ and initial state $x_i(0) = i - 1$ and a nominal bound for the state magnitude of $|x_i| \leq 5$. In order to reduce complexity and to study the properties of the algorithms in a disadvantageous setting, we considered $N = 1$, meaning that we only use the information from the previous iteration for the estimates. This is a worst-case scenario, as the algorithm only takes into account the dynamics of the system with one time step from the last estimate and discards prior observations. A missed detection is considered if the algorithm is not able to detect the fault within 300 time steps. Each result presented corresponds to 1000 Monte-Carlo runs. For convenience, node 1 is the node that performs the detection and node 2 is the failing node, and no faults occur in the first 10 transmissions. Note that if a node introduces Byzantine faults from the start of the algorithm, it can do so without being detected since the network has no information about the initial state of the Byzantine node. The following probability matrix is used:

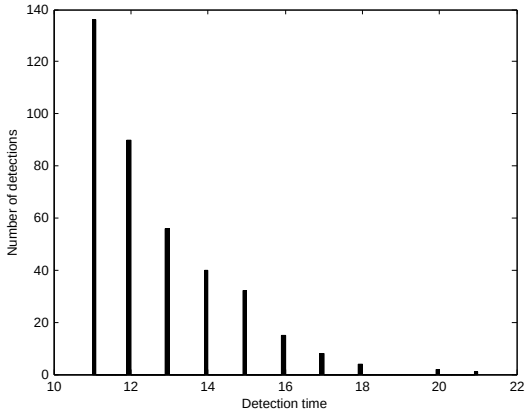

Fig. 1: Detection times for the unresponsive fault.

$$W = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.25 & 0 & 0.25 \\ 0.5 & 0.25 & 0 & 0.125 & 0.125 \\ 0 & 0 & 0.125 & 0.25 & 0.625 \\ 0 & 0.25 & 0.125 & 0.625 & 0 \end{bmatrix}$$

To show the properties of the SVO detection, we devise a scenario where a node becomes unresponsive due to CPU load or software crash, does not perform the consensus update and, therefore, replies always with the same value.

Figure 1 depicts the detection time for the unresponsive fault with a detection rate of 38.4%. We observe that the fault is more likely to be detected as time progresses. The $\gamma_{min} = 76.56$ and $\|u\|^2 = 2.997$ which indicates that the injected signal is detected even though its energy is less than the theoretical bound. Since the detection rate was so small, we also simulated for the case where each node runs an SVO with its own local observations. The detection times follow a similar pattern but the detection rate increased to 100%.

A more interesting simulation is when the attacker emulates the strategy of the fault detection scheme and uses that information to attack the system by reproducing the worst possible fault without being detected. We assume that the attacker has full knowledge of the network structure, initial states and sequence of communicating pairs for all k.

Figure 2 shows a representative case of the injected fault signal magnitude for each of the 300 time slots. The profile of the plot changes depending on the sequence of transmissions but presents the exponential decay as shown in Figure 2. We also computed the minimum, maximum and mean deviation of the steady state to be 0.2268, 2.8003 and 1.8463, respectively. We point out that since the true steady state is 2, the attacker could only deviate it to 4.8003 which is less than the trivial bound of 5.

To illustrate the advantage of the SSVO when detecting faults, we consider a scenario where a node takes advantage of the network and initiates communication with a neighbour regardless of the probability matrix W but does not change any of the nodes state. Notice that using an SVO, such faults

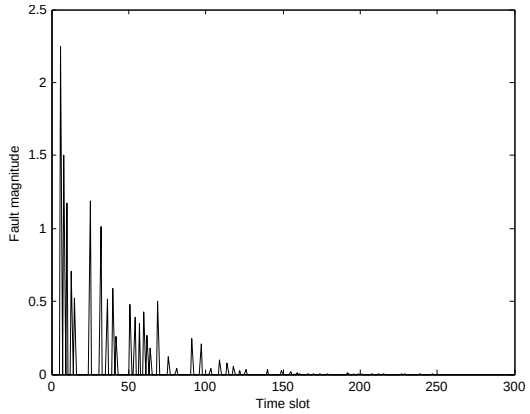

Fig. 2: Representative example of an undetected signal.

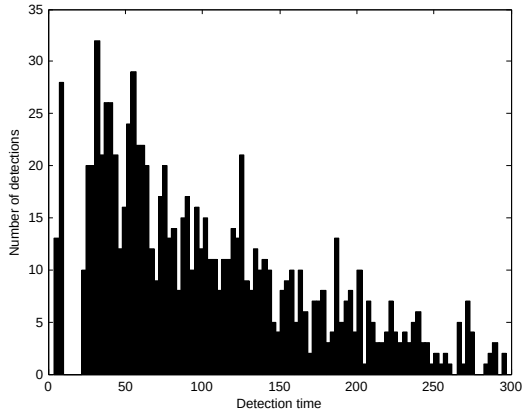

Fig. 3: Detection times for the SSVO.

would not be detected as any communication pattern that is possible is considered regardless of its probability. Between transmission time $10 < k < 20$ the communication takes place between node 3 and 4 and $\alpha = 0.1$.

Figure 3 depicts the detection times for the SSVO case with a detection rate of 92.8%. Even though the behaviour is similar, in this case there is a probability of at most α of the detection being a false positive.

VI. Conclusions and Future Work

The paper focuses on gossip algorithm with stochastic transmissions in a Byzantine environment, which considers a larger group of faults than "crash type" failures, namely: intermittent failures, state corruption, nodes executing erroneous algorithm steps or corrupted messages that pass the medium checksums and other detection mechanisms, amongst others. We are particularly interested in faults that change the final consensus value and by modeling transmissions as uncertainties in the system dynamics enabled us to provide proofs of detection if the fault is *detectable* in a given sense and bounds on the magnitude of the attacker signal.

The proposed approach adopts the concept of set-valued observers to generate, at each transmission time, a convex set containing all possible state realizations and converting the problem to that of detecting whether the intersection between this set and the set of observations is void or not (i.e. rewriting the problem as a distinguishability problem rather than a detectability one). We also provide an algorithm to construct confidence sets that take into account the probabilistic nature of the stochastic gossip algorithm, which reduces the conservativeness of the worst-case approach, at the expense of introducing a probability of false positives.

Our main contribution is two-folded: we introduce a novel type of observer for discrete-time stochastic systems, showing that the estimated set containing the state, at each time instant, is bounded; and a bound is derived on the attacker signal magnitude above which detection is guaranteed.

We envisage some directions of future work by considering asymmetric communications and a broader class of distributed algorithms. Some additional work is necessary in reducing the computational complexity of calculating the set where the state of the system is. Finally, different types of faults may be analyzed in order to pursue less conservative conditions that can still guarantee the detectability of faults.

References

[1] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *Automatic Control, IEEE Transactions on*, vol. 57, no. 1, pp. 90 –104, jan. 2012.

[2] K. P. Kihlstrom, L. E. Moser, and P. M. Melliar-Smith, "Solving consensus in a byzantine environment using an unreliable fault detector," in *Proceedings of the International Conference on Principles of Distributed Systems (OPODIS)*, 1997, pp. 61–75.

[3] D. Bertsekas and I. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *Automatic Control, IEEE Transactions on*, vol. 16, no. 2, pp. 117 – 128, apr 1971.

[4] C. Combastel, "A state bounding observer for uncertain non-linear continuous-time systems based on zonotopes," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*, dec. 2005, pp. 7228 – 7234.

[5] T. Alamo, J. Bravo, and E. Camacho, "Guaranteed state estimation by zonotopes," *Automatica*, vol. 41, no. 6, pp. 1035 – 1043, 2005.

[6] H. Witsenhausen, "Sets of possible states of linear systems given perturbed observations," *Automatic Control, IEEE Transactions on*, vol. 13, no. 5, pp. 556 – 558, oct 1968.

[7] F. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *Automatic Control, IEEE Transactions on*, vol. 13, no. 1, pp. 22 – 28, feb 1968.

[8] F. Schweppe., *Uncertain Dynamic Systems*. Prentice-Hall, 1973.

[9] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica*, vol. 27, no. 6, pp. 997 – 1009, 1991.

[10] R. J. Patton, "Fault-tolerant control systems: The 1997 situation," in *IFAC symposium on fault detection supervision and safety for technical processes*, vol. 3, 1997.

[11] J. Bokor and Z. Szabó, "Fault detection and isolation in nonlinear systems," in *Annual Reviews in Control 33.2*, 2009, pp. 113–123.

[12] P. Rosa, C. Silvestre, J. Shamma, and M. Athans, "Fault detection and isolation of LTV systems using set-valued observers," in *Proceedings of the 49th IEEE Conference on Decision and Control*, December 2010, pp. 768–773.

[13] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2508 – 2530, Jun. 2006.

[14] M. Grewal and K. Glover, "Identifiability of linear and nonlinear dynamical systems," *IEEE Trans. on Automatic Control*, vol. 21, no. 6, pp. 833–837, 1976.

[15] P. Rosa and C. Silvestre, "On the distinguishability of discrete linear time-invariant dynamic systems," in *Proceedings of the 50th IEEE Conference on Decision and Control*, December 2011.

[16] P. Rosa, "Multiple-model adaptive control Multiple-Model Adaptive Control of Uncertain LPV Systems," Ph.D. dissertation, Technical University of Lisbon, 2011.

[17] J. Shamma and K.-Y. Tu, "Set-valued observers and optimal disturbance rejection," *Automatic Control, IEEE Transactions on*, vol. 44, no. 2, pp. 253 –264, feb 1999.

[18] S. Keerthi and E. Gilbert, "Computation of minimum-time feedback control laws for discrete-time systems with state-control constraints," *Automatic Control, IEEE Transactions on*, vol. 32, no. 5, pp. 432 – 435, may 1987.