Resilient Multi-Agent Consensus Using Wi-Fi Signals

Stephanie Gil, Cenk Baykal, and Daniela Rus

***Abstract*—Consensus is an important capability at the heart of many multi-agent systems. Unfortunately the ability to reach consensus can be easily disrupted by the presence of an adversarial agent that spawns or *spoofs* malicious nodes in the network in order to gain a disproportionate influence on the converged value of the system as a whole. In this letter, we present a light-weight approach for spoof-resiliency with provable guarantees that solely utilizes information from wireless signals. Unlike prior approaches, our method requires no additional protocol or data storage beyond signals that are already present in the network. We establish an analytical, probabilistic bound on the influence of spoofed nodes in the system on the converged consensus value. We present results of our Wi-Fi based resilient consensus algorithm and demonstrate its effectiveness for different consensus problems such as flocking and rendezvous.**

***Index Terms*—Agents-based systems, cooperative control, networked control systems, robotics, uncertain systems, fault tolerant systems, intelligent systems, sensor networks.**

I. Introduction

FROM delivery drones, to heterogeneous search and rescue teams, to autonomous vehicles, multi-robot systems are poised for impact in the real world. Consensus is at the heart of many algorithms that require coordination and the ability to reach *agreements* on a number of things from a quantity being jointly measured, to a unified heading direction formation flight [1]–[8]. While these systems find their strength in their ability to communicate and coordinate, this can also be their Achilles' heel in the case of an adversarial environment.

Unfortunately, multi-agent systems can be easy to hack. This is because optimal task performance requires shared data to be accurate and trustworthy, an assumption that is easy to break. A particularly challenging attack on this assumption is the so-called "Sybil Attack." In a Sybil attack, a malicious agent generates (or spoofs) a large number of false identities to gain a disproportionate influence on the network. These attacks are notoriously easy to implement [9] and can be detrimental to multi-agent networks, particularly those performing consensus algorithms. For instance, it was recently shown that a hacker could *spoof* non-existent airplanes in the sky [10].

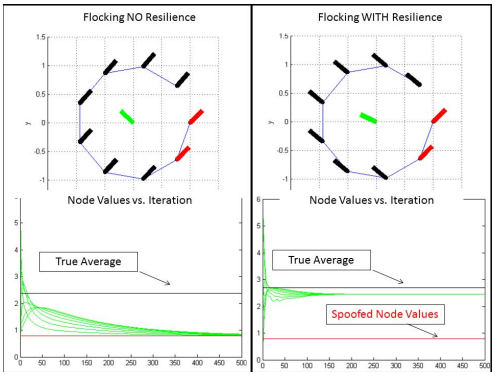

Fig. 1. Resilient consensus algorithm for the case of flocking with 7 legitimate nodes (black) and 2 spoofed nodes (red) and true average (green).

Past approaches use protocol-based and key passing methods to protect networked systems [11]–[13]. While these methods do indeed provide added security, they often require additional overhead (computation and data) and are particularly challenging to implement for mobile and distributed systems [14], [15]. The effect of an adversarial presence on the performance guarantees for multi-robot algorithms is sparsely treated in [14] and [15]. The particular needs of multi-robot networks, being often times distributed and dynamic, makes traditional solutions such as key passing difficult or impossible to implement. Some papers such as [16] and [17] start to move in this direction. A key difference between our work and previous approaches such as [11] and [16] is that the current paper aims to use information extracted from the physics of the wireless signals themselves, and not the transmitted data, in order to discard malicious node inputs. This letter also bounds the impact that adversarial activity has on the converged consensus value.

Motivated by recent developments in Wi-Fi characterization [17], [18], we present a resilient consensus approach that utilizes additional information from Wi-Fi communication signals. Unlike prior work, this approach does not necessitate the implementation of additional protocols or key-passing for providing spoof-resilient consensus with provable guarantees. This letter contributes the following:

Manuscript received March 7, 2018; revised May 22, 2018; accepted June 7, 2018. Date of publication July 6, 2018; date of current version July 23, 2018. This work was supported in part by the CyberSecurity@CSAIL and in part by NSF under Grant 1723943. Recommended by Senior Editor F. Blanchini. *(Corresponding author: Stephanie Gil.)*

S. Gil is with CIDSE, Arizona State University, Tempe, AZ 85280 USA (e-mail: sgil@asu.edu).

C. Baykal and D. Rus are with the Electrical Engineering and Computer Science Department, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139 USA (e-mail: baykalrus@mit.edu; rus@mit.edu).

Digital Object Identifier 10.1109/LCSYS.2018.2853641

1) An algorithm for spoof-resilient consensus based on existing Wi-Fi signals.
2) An analysis that probabilistically bounds and characterizes the influence of spoofed nodes in the network on the converged consensus value.
3) A simulation study that demonstrates that validates our theoretical results for different consensus problems such as flocking and rendezvous.

II. Problem

We consider the problem of consensus for multi-agent systems. The multi-robot system can be described by a weighted state-dependent graph, $\mathbb{G} = (\mathbb{V}, \mathbb{W})$, where $\mathbb{V} = \{1, \ldots, n\}$ denotes the set of node indices for n robots and $\mathbb{W} : \mathbb{V} \times \mathbb{V} \times \mathbb{R}_+ \to \mathbb{R}_+$ denotes the set of edge weights such that $w_{ij}(t) = \mathbb{W}(i, j, t)$ for $i, j \in \mathbb{V}$. The set $\mathcal{E}(t) = \{(i, j) | w_{ij}(t) > 0\}$ is called the set of undirected edges of $\mathbb{G}$. The set of *neighbors* of node i is denoted by $\mathcal{N}_i(t) = \{i \in \mathbb{V} : (i, j) \in \mathcal{E}(t)\}$ where n nodes have values $x_i(t) \in \mathbb{R}$ and indices $\mathcal{I} = \{1, \ldots, n\}$. We consider the case where a subset of nodes with indices denoted by the set $\mathcal{S}$, $\mathcal{S} \subset \mathcal{I}$, are spoofed. The set $\mathcal{S}$ is assumed to be unknown, although the cardinality $n_\mathcal{S} = |\mathcal{S}|$ is assumed to be known. Our threat model is described in detail below.

A. Threat Model

Our threat model considers one or more adversarial agents with one Wi-Fi antenna each. The adversaries can be mobile and scale power on a per-packet basis. Adversarial agents perform the "Sybil Attack" to inject packets emulating $n_\mathcal{S}$ non-existent clients according to the following definition.

Definition 1 (Sybil Attack): An adversary in the network can control the values of one or more "spoofed" nodes in the network by sending various messages over the network with unique IDs $\{j_1, j_2, \ldots\} \in \mathcal{S}$ in order to gain a disproportionate influence in the network. We assume that the graph $\mathbb{G}$ is known, but knowledge of which clients are spoofed (i.e., in $\mathcal{S}$) is unknown. If an agent j is a spoofed node such that $j \in \mathcal{S}$ then at time t its value is denoted by $x_j(t)$ and this value can be arbitrarily controlled by an adversarial agent in the network. This value is assumed to be finite for all time so that $|x_j(t)| \leq \eta$ for some $\eta > 0$ for all t.

B. Detecting Sybil Attacks Using Wireless Signals

Our previous work developed a method for measuring *directional signal profiles* (See Figure 2) using channel state information (CSI) from the wireless messages over each link (i, j) in the network [18], [19]. These profiles measure signal strength arriving from every direction in the 3D plane. Directional signal profiles (see Figure 2) display two important properties: 1) transmissions originating from the same physical agent have very similar profiles and 2) energy can be measured coming from the direct-line path between physical agents. The paper [17] quantifies these properties, providing an analysis that shows both analytically and experimentally, that a single scalar value $\alpha_{ij} \in (-0.5, 0.5)$ (shifted by -0.5 from [17]) can be computed for each signal profile that quantifies the likelihood that the transmission is coming from the same physical (spoofed) node, or a unique (legitimate) node; a property critical for thwarting Sybil Attacks. Intuitively, the α_{ij} was shown experimentally and theoretically to be close to -0.5 if one of the agents j is a spoofed node and close to 0.5 if both agents are legitimate nodes in the network [17]. This

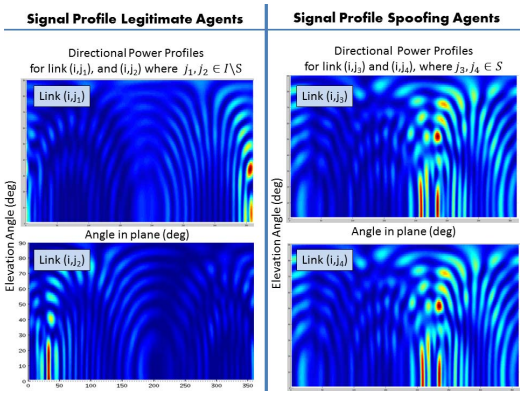

Fig. 2. Actual profiles for legitimate nodes (left col) and spoofed nodes (right col) at a single time snapshot. Red and blue indicate directions of high and low recv signal power respectively. For simultaneous transmissions from a spoofed node note that the transmission profiles are similar, to within noise, from the same physical node. This is captured quantitatively by α_{ij} [17].

is captured quantitatively by the bounds on the expectation of the α_{ij}.

Definition 2 (α_{ij} [17]): Our previous work theoretically derived and rigorously tested (in hardware experiments) the existence of weights $\alpha_{ij} \in (-1/2, 1/2)$ for the wireless channel between any two communicating agents i and j with the provable property that $\mathbb{E}[\alpha_{ij}] \leq -1/2 + \epsilon_\mathcal{S}$ if $j \in \mathcal{S}$ and $\mathbb{E}[\alpha_{ij}] \geq 1/2 - \epsilon_\mathcal{L}$ if $j \in \mathcal{I} \setminus \mathcal{S}$ where $\epsilon_\mathcal{S}$ and $\epsilon_\mathcal{L}$ are determined in closed form as a function of signal to noise ratio (SNR) of the channel, number of spoofed nodes, and channel constants. For the case that each agent reports its position $p_i(t)$ in addition to its value $x_i(t)$, as in rendezvous for example, the values of $\epsilon_\mathcal{S}$ and $\epsilon_\mathcal{L}$ can be bounded more tightly. The paper [17] presents both forms for the epsilons. In this letter, we only utilize the fact that SNR of the wireless messages is high enough (i.e., the link is of good quality) such that $\epsilon_\mathcal{S}, \epsilon_\mathcal{L} \in (0, 1/2)$.

We note that the objective of the current paper is not to prove additional properties of the signal profiles, but to develop a theoretical framework for using these signal profiles for robust consensus of multi-robot systems in the face of a Sybil Attack.

C. Threat Resilient Consensus

We consider the distributed linear consensus protocol:

$$x_i(t+1) = \mathbb{W}(i, i, t)x_i(t) + \sum_{j \in \mathcal{N}_i} \mathbb{W}(i, j, t)x_j(t). \tag{1}$$

We can write the consensus protocol from (1) in matrix form, separating out the component for spoofed and legitimate node values by defining $x(t) = [x_\mathcal{L}(t)\ x_\mathcal{S}(t)]^T$ and writing the dynamics of this system as:

$$\begin{bmatrix} x_\mathcal{L}(t+1) \\ x_\mathcal{S}(t+1) \end{bmatrix} = \begin{bmatrix} W_\mathcal{L}(t) & W_\mathcal{S}(t) \\ \Theta & \Omega \end{bmatrix} \begin{bmatrix} x_\mathcal{L}(t) \\ x_\mathcal{S}(t) \end{bmatrix}, \tag{2}$$

where $W_\mathcal{L}(t) \in \mathbb{R}^{n_\mathcal{L} \times n_\mathcal{L}}$ is the matrix multiplying the component of the state corresponding to legitimate node values, $n_\mathcal{L}$ is the number of legitimate nodes in the system, and $W_\mathcal{S}(t) \in \mathbb{R}^{n_\mathcal{L} \times n_\mathcal{S}}$ is the matrix multiplying the component of the state corresponding to spoofed node values and $\mathbb{W}(t) = [W_\mathcal{L}(t)\ W_\mathcal{S}(t)]$. The matrices Θ and Ω dictate the dynamics of the spoofed node values and are assumed to be unknown.

Intuitively, our goal is to derive weight matrices $W_{\mathcal{L}}(t)$ and $W_{\mathcal{S}}(t)$ with the property that over time (i.e., as $t \to \infty$), the influence of the legitimate nodes approaches 1 and the influence of the spoofed nodes approaches zero. Note that we focus on the problem of deriving weights for each node and not on optimizing network topology. In fact, we make the assumption that the network topology is sufficiently connected such that it would remain connected even if spoofed nodes were removed from the graph.

Assumption 1 (Sufficiently Connected Network): The graph is sufficiently connected such that removal of spoofed nodes in the system would maintain a connected network at all times. Further, a node that is spoofed remains spoofed for all time and vice versa for non-spoofed (legitimate) nodes.

Assumption 2 (Independence): For any link (i, j), the wireless channel weights $\alpha_{ij}(0), \alpha_{ij}(1), \ldots,$ are independent.

Where we also assume a uniform scattering environment. See Section V for discussion. We now formalize our problem using our consensus dynamics from Equation (2) as follows.

Problem 1: Find a weight matrix $\mathbb{W}(t) = [W_{\mathcal{L}}(t)\ W_{\mathcal{S}}(t)]$ and a tuple of problem parameters $\mathcal{P}$ such that for any $\delta \in (0, 1)$

$$\mathbb{P}\left(\lim_{t\to\infty}\left|x_{\mathcal{L}}(t) - \frac{vv^T x(0)}{n_{\mathcal{L}}}\right| = \Delta(\mathcal{P}, \delta)\right) \geq 1 - \delta,$$

for some finite $\Delta(\mathcal{P}, \delta) \leq \Delta_{\max}(\mathcal{P}, \delta)$, where

$$v_i = \begin{cases} 1, & \text{if node } i \text{ is legitimate} \\ 0 & \text{otherwise.} \end{cases}$$

While being able to handle cases of changing network topology [20] is of critical importance for multi-robot systems and an important topic for future work, optimization of the graph topology is out of scope for this letter. Instead, the focus is on the problem of solving for the *weights* applied to each node's value in a similar vein to the work in [1]. Unlike [1] however, we give explicit treatment of the adversarial case.

III. Analysis

In this section we find 1) a weight matrix $\mathbb{W}(t)$, 2) a tuple of problem parameters $\mathcal{P}$, and 3) a bound $\Delta_{\max}(\mathcal{P}, \delta)$ such that the properties described in Problem 1 hold, and analyze the properties of our proposed consensus scheme. By propagating forward the consensus dynamics from Equation (2) we obtain,

$$x_{\mathcal{L}}(t) = \prod_{k=0}^{t-1} W_{\mathcal{L}}(k) x_{\mathcal{L}}(0) + \phi_{\mathcal{S}}(0, t),$$

where $\phi_{\mathcal{S}}(0, t) = \sum_{k=0}^{t-1}(\prod_{l=k}^{t-1} W_{\mathcal{L}}(l)) W_{\mathcal{S}}(k) x_{\mathcal{S}}(k)$ is precisely the term that captures the influence of the spoofed nodes $x_{\mathcal{S}}$ on the system through time t. We wish to bound the influence of the spoofed nodes in the limit as $t \to \infty$ assuming that $x_{\mathcal{S}}(t)$ is bounded for all t. Note that $\max_{i\in\mathcal{I}\setminus\mathcal{S}} |\phi_{\mathcal{S}i}(0, t)|$ is bounded by the expression

$$\max_{i\in\mathcal{I}\setminus\mathcal{S}}\left[\sum_{k=0}^{t-1}\sum_{j\in\mathcal{S}\cap\mathcal{N}_i}\left|\left(\prod_{l=k}^{t-1} W_{\mathcal{L}}(l)\right) W_{\mathcal{S}}(k)\right|_{ij}\right]\eta,$$

where $\eta = \max_j \sup_k |x_{Sj}(k)| < \infty$ is taken to be the maximum absolute value that a spoofed node can take.

Let us choose each element in our weighting matrix $\mathbb{W}(t)$ in the following way, where $\beta_{ij}(t) = \sum_{l=0}^{t} \alpha_{ij}(l)$:

$$\mathbb{W}(i, j, t) = \begin{cases} \frac{1}{n_{\mathcal{L}}}(1 - e^{-\beta_{ij}(t)/2}), & \text{if } \beta_{ij}(t) \geq 0 \\ \frac{1}{2n_{\mathcal{L}}} e^{\beta_{ij}(t)}, & \text{if } \beta_{ij}(t) < 0 \\ 1 - \sum_l \mathbb{W}(i, l, t), & \text{if } i = j \end{cases} \tag{3}$$

This definition of the weights leads to two expressions of influence that constitute the overall spoofer influence $\phi_{\mathcal{S}_i}(0, t)$. The first term, which we denote as $\phi^F_{\mathcal{S}_i}(0, t)$, is made up of the *worst-case* sum of spoofer influences at time steps where *failures* occur; we label the tuple (i, j, τ) a failure if $\beta_{ij}(\tau) \geq 0$ for a spoofed link (i, j), $i \in \mathcal{I} \setminus \mathcal{S}$, $j \in \mathcal{S}$ at time step $\tau \in \{0, \ldots, t\}$. Intuitively, this failure event should not occur too often since $\mathbb{E}[\beta_{ij}(\tau)] = \sum_{l=0}^{\tau} \mathbb{E}[\alpha_{ij}(l)]$ and $\mathbb{E}[\alpha_{ij}(l)] < 0$ for all $l \in \{0, \ldots, \tau\}$ for a spoofed link (i, j). This intuition is formalized in Lemma 1. Mathematically,

$$\phi^F_{\mathcal{S}_i}(0, t) = \sum_{k=0}^{t-1}\sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i: \\ \beta_{ij}(k)\geq 0}}\left|\left(\prod_{l=0}^{t-1} W_{\mathcal{L}}(l)\right) W_{\mathcal{S}}(k)\right|_{ij}\eta.$$

The second term, which we denote as $\phi^{\bar{F}}_{\mathcal{S}_i}(0, t)$, relates to the *worst-case* sum of influences at time steps for which failures do not occur, i.e.,

$$\phi^{\bar{F}}_{\mathcal{S}_i}(0, t) = \sum_{k=0}^{t-1}\sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i: \\ \beta_{ij}(k)< 0}}\left|\left(\prod_{l=0}^{t-1} W_{\mathcal{L}}(l)\right) W_{\mathcal{S}}(k)\right|_{ij}\eta.$$

Thus, $\phi_{\mathcal{S}_i}(0, t)$ is upper bounded by the following composition of influences

$$\phi_{\mathcal{S}_i}(0,\ t) \leq \phi^F_{\mathcal{S}_i}(0, t) + \phi^{\bar{F}}_{\mathcal{S}_i}(0, t). \tag{4}$$

In our analysis, we will upper bound the influence of each term separately and combine the individual bounds to establish an upper bound on the overall spoofer influence $\phi_{\mathcal{S}_i}(0, t)$. Moreover, the definition of the weight matrix (3) satisfies the conditions for consensus in the limit as presented in [1].

For the proceeding analysis, we let $c = (-1/2 + \epsilon_{\mathcal{S}})^2$ be a constant. In the subsequent analysis, we will refer to the natural logarithm as simply $\log(\cdot)$. The following lemma characterizes the concentration of $\beta_{ij}(t)$ for any $t \in \mathbb{N}$ and spoofed link (i, j).

Lemma 1 (Concentration of $\beta_{ij}(t)$): For any link (i, j) with $j \in \mathcal{S}$, $t \in \mathbb{N}$, and a sequence of weights $(\alpha_{ij}(0), \ldots, \alpha_{ij}(t))$, the probability of the event that $\beta_{ij}(t)$ is non-negative decays exponentially fast with t, i.e.,

$$\mathbb{P}(\beta_{ij}(t) \geq 0) \leq \exp\left(-t(-1/2 + \epsilon_{\mathcal{S}})^2\right) \leq \exp(-ct).$$

Alternatively, if $j \in \mathcal{I} \setminus \mathcal{S}$, i.e., j is a legitimate node, then $\mathbb{P}(\beta_{ij}(t) < 0) \leq \exp\left(-t(1/2 - \epsilon_{\mathcal{L}})^2\right)$.

Proof: By linearity of expectation and our bound on the expectation of each spoofed weight, we have that

$$\mathbb{E}[\beta_{ij}(t)] = \sum_{l=0}^{t} \mathbb{E}[\alpha_{ij}(l)] \leq (t + 1)(\epsilon_{\mathcal{S}} - 1/2) < 0.$$

Since $\mathbb{E}[\beta_{ij}(t)] \leq (t + 1)(\epsilon_{\mathcal{S}} - 1/2) < 0$, we obtain the lower bound on the square of the expectation $\mathbb{E}[\beta_{ij}(t)]^2 \geq (t + 1)^2(\epsilon_{\mathcal{S}} - 1/2)^2$. We proceed by applying Hoeffding's inequality [21] by observing that for all $l \in [t] = \{0, \ldots, t\}$, $\alpha_{ij}(l) \in [-1/2, 1/2]$, i.e., $\beta_{ij}(t)$ is a sum over $t+1$ independent random variables (Assumption 2), where each summand is a bounded random variable in the interval $[-1/2, 1/2]$. Thus, we conclude by the one-sided Hoeffding's inequality that

$$\begin{aligned}\mathbb{P}\big(\beta_{ij}(t) \geq 0\big) &= \mathbb{P}\big(\beta_{ij}(t) - \mathbb{E}[\beta_{ij}(t)] \geq -\mathbb{E}[\beta_{ij}(t)]\big) \\ &\leq \exp\left(-\frac{2(-\mathbb{E}[\beta_{ij}(t)])^2}{t+1}\right) \leq \exp(-2(t+1)c).\end{aligned}$$

A symmetric argument applied to the sequence of weights $(\alpha_{ij}(\tau))_{\tau=0}^{t}$ with $j \in \mathcal{I} \setminus \mathcal{S}$ in conjunction with the lower bound on $\mathbb{E}[\alpha_{ij}(t)] \geq 1/2 - \epsilon_{\mathcal{L}}$ yields the second result of the lemma. ∎

For each $t \in \mathbb{N}$, let $A_{ij}(t)$ denote the (undesirable) event that $\beta_{ij}(t) \geq 0$ if $j \in \mathcal{S}$ and alternatively the (undesirable) event $\beta_{ij}(t) < 0$ if $j \in \mathcal{I} \setminus \mathcal{S}$. The following corollary formalizes the fact that these undesirable events are not likely to occur often.

Corollary 1: For the sequence of events $\big(A_{ij}(t)\big)_{t\in\mathbb{N}}$ defined above, the probability that infinitely many of them occurs is 0. That is, $\mathbb{P}\big(\limsup_{t\to\infty} A_{ij}(t)\big) = 0$.

Proof: We will prove the corollary for the case of $j \in \mathcal{S}$ as the argument for $j \in \mathcal{I}\setminus\mathcal{S}$ follows by the same reasoning. Note that for any event $A_{ij}(t)$ in the sequence, we have by Lemma 1 that $\mathbb{P}(A_{ij}(t)) = \mathbb{P}(\beta_{ij}(t) \geq 0) \leq \exp(-tc)$. Now consider the sum of probabilities over the events $A_{ij}(0), A_{ij}(1), \ldots$ and note that

$$\sum_{t=0}^{\infty} \mathbb{P}(A_{ij}(t)) \leq \sum_{t=0}^{\infty} \exp(-tc) = \frac{\exp(c)}{\exp(c)-1} < \infty,$$

since $\epsilon_{\mathcal{S}} \in (0, 1/2)$. Thus, by the Borel-Cantelli lemma, we have that the probability that the event $A_{ij}(t)$, i.e., $\beta_{ij}(t) \geq 0$, occurs for infinitely many values of t is 0. Thus, $\beta_{ij}(t) < 0$ occurs for infinitely many values of t with probability 1 and this establishes the claim. ∎

For a spoofed link (i, j) with $j \in \mathcal{S}$ and the infinite sequence of (random) weights $\big(\alpha_{ij}(0), \alpha_{ij}(1), \ldots\big)$, define the sequence of indicator variables $X_0, X_1, \ldots$ (we omit the subscript denoting the edge (i, j) for brevity) such that,

$$\forall l \in \mathbb{N} \quad X_l = \begin{cases} 1 & \text{if } \beta_{ij}(l) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then, note that the total number of *failures* within a given time interval $\{0, \ldots, t\}$ is given by $N_{ij}^{F}(0, t) = \sum_{l=0}^{t} X_l$, where we used the subscript ij to explicitly express the fact that the number of failures is with respect to a specific edge (i, j). The following lemma establishes a probabilistic, finite upper bound on $N_{ij}^{F}(0, t)$ for any $t \in \mathbb{N}$. Let $\mathcal{Z}$ be the set of edges such that $\mathcal{Z} = \{(i, j) : i \in \mathcal{I} \setminus \mathcal{S} \text{ and } j \in \mathcal{S} \cap \mathcal{N}_i\}$.

Lemma 2 (Probabilistic Bound on $lim_{t\to\infty} N_{ij}^{F}(0, t)$): Given any desired failure probability $\delta \in (0, 1)$, it follows that

$$\mathbb{P}\Big(\forall (i, j) \in \mathcal{Z} \ \lim_{t\to\infty} N_{ij}^{F}(0, t) \leq M(\mathcal{P}, \delta)\Big) \geq 1 - \delta,$$

where $M(\mathcal{P}, \delta)$ is a function of the problem-specific parameters $\mathcal{P} = (n_{\mathcal{S}}, n_{\mathcal{L}}, \eta, \epsilon_{\mathcal{S}})$, and δ:

$$M(\mathcal{P},\ \delta) = \xi \cdot \mu_Y, \text{ where}$$
$$\mu_Y = \frac{\exp(c)}{\exp(c)-1},\ \xi = \frac{n_{\mathcal{S}} \cdot n_{\mathcal{L}}}{\delta},\ \text{and } c = (-1/2 + \epsilon_{\mathcal{S}})^2.$$

Proof: We define Y_l with the following properties. Y_l's are independent, and by Lemma 1 for any $l \in \mathbb{N}$, $\mathbb{P}(X_l = 1) \leq \exp(-cl) = \mathbb{P}(Y_l = 1)$. Let $Y(0, t) = \sum_{l=0}^{t} Y_l$ be a sequence of random variables for each $t \in \mathbb{N}$. We define the Y_l to be Bernoulli random variables

$$\forall l \in \mathbb{N} \quad Y_l = \begin{cases} 1 & \text{with probability } \exp(-c\,l) \\ 0 & \text{otherwise,} \end{cases}$$

where $c = (-1/2 + \epsilon_{\mathcal{S}})^2$ as before. Combining applications of linearity of expectation and the fact that the expectation of an indicator random variable is equivalent to its probability, we thus have for any $t \in \mathbb{N}, t \geq 0$

$$\mathbb{E}[N_{ij}^{F}(0,\ t)] = \sum_{l=0}^{t} \mathbb{E}[X_l] \leq \sum_{l=0}^{t} \mathbb{E}[Y_l] = \mathbb{E}[Y(0,\ t)]$$
$$= \sum_{l=0}^{t} \exp(-c\,l) = \frac{\exp(c) - \exp(-tc)}{\exp(c) - 1}.$$

Now observe that the sequence $(N_{ij}^{F}(0, t))_{t\in\mathbb{N}, t\geq 0}$ is non-negative and monotonically increasing. Therefore, the monotone convergence theorem yields $\mathbb{E}[\lim_{t\to\infty} N_{ij}^{F}(0, t)] = \lim_{t\to\infty} \mathbb{E}[N_{ij}^{F}(0, t)]$, and thus we obtain

$$\mathbb{E}\Big[\lim_{t\to\infty} N_{ij}^{F}(0, t)\Big] = \lim_{t\to\infty} \mathbb{E}[N_{ij}^{F}(0, t)]$$
$$\leq \lim_{t\to\infty} \mathbb{E}[Y(0, t)] = \mu_Y$$
$$= \lim_{t\to\infty} \frac{\exp(c) - \exp(-tc)}{\exp(c) - 1} = \frac{\exp(c)}{\exp(c) - 1}$$

Applying Markov's Inequality, we obtain for $\xi \cdot \mu_Y = M(\mathcal{P}, \delta)$,

$$\mathbb{P}(\lim_{t\to\infty} N_{ij}^{F}(0, t) \geq M(\mathcal{P}, \delta)) = \mathbb{P}(\lim_{t\to\infty} N_{ij}^{F}(0, t) \geq \xi \cdot \mu_Y)$$
$$\leq \frac{\mathbb{E}\Big[\lim_{t\to\infty} N_{ij}^{F}(0, t)\Big]}{\xi \cdot \mu_Y}$$
$$\leq \frac{1}{\xi} = \frac{\delta}{n_{\mathcal{S}} \cdot n_{\mathcal{L}}}.$$

where the first inequality follows by the second by Markov's inequality and the second by $\mathbb{E}[\lim_{t\to\infty} N_{ij}^{F}(0, t)] \leq \mu_Y$. To have our bound hold for all spoofed nodes that are neighboring each legitimate node i, we apply the union bound over all $n_{\mathcal{S}}(i)$ spoofed neighbors of node i and note that $n_{\mathcal{S}}(i) \leq n_{\mathcal{S}}$ by definition. Thus,

$$\mathbb{P}(\exists (i, j) \in \mathcal{Z} : \lim_{t\to\infty} N_{ij}^{F}(t) \geq M(\mathcal{P}, \delta))$$
$$\leq \sum_{(i,j)\in\mathcal{Z}} \mathbb{P}(\lim_{t\to\infty} N_{ij}^{F}(t) \geq M(\mathcal{P}, \delta)) \leq \delta.$$

We note that the union bound above over all the links holds regardless of the correlation between different links. ∎

Lemma 3 (Probabilistic Bound on $\phi_{\mathcal{S}_i}^{F}(0, t)$): For any given failure probability $\delta \in (0, 1)$ and $t \in \mathbb{N}$, we have that

$$\mathbb{P}\left(\max_{i\in\mathcal{I}\setminus\mathcal{S}} |\phi_{\mathcal{S}_i}^{F}(0, t)| \leq M(\mathcal{P}, \delta)\frac{n_{\mathcal{S}}}{n_{\mathcal{L}}^2}\eta\right) \geq 1 - \delta,$$

where $M(\mathcal{P}, \delta)$ is defined as in Lemma 2.

Proof: From Equation (3) in the first case, the maximum weight that link (i, j) can have for the case $\beta_{ij}(t) \geq 0$ is $\frac{1}{n_{\mathcal{L}}}$. We will use this to compute the worst-case influence that a spoofed node can have in the system if it is misclassified as a legitimate node (and thus $W_{\mathcal{S}} = W_{\mathcal{L}}$) $N_{ij}^{F}(0, t)$ times ($N_{ij}^{F}(t)$ for brevity).

$$\max_{i\in\mathcal{I}\setminus\mathcal{S}} |\phi_{\mathcal{S}_i}^{F}(0, t)|$$
$$= \max_{i\in\mathcal{I}\setminus\mathcal{S}} \sum_{k=0}^{t-1} \sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i: \\ \beta_{ij}(k)\geq 0}} \left|\left(\prod_{l=k}^{t-1} W_{\mathcal{L}}(l)\right) W_{\mathcal{S}}(k)\right|_{ij} \eta$$

$$\leq \max_{i\in\mathcal{I}\setminus\mathcal{S}}\left[\sum_{k=0}^{t-1}\sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i:\\ \beta_{ij}(k)\geq 0}}\left(\prod_{l=k}^{t-1}\frac{1}{n_{\mathcal{L}}}\right)\frac{1}{n_{\mathcal{L}}}\right]\eta$$

$$\leq \max_{i\in\mathcal{I}\setminus\mathcal{S}}\max_{j\in\mathcal{S}\cap\mathcal{N}_i} N_{ij}^F(t)\frac{n_{\mathcal{S}}(i)}{n_{\mathcal{L}}^2}\eta \leq \max_{i\in\mathcal{I}\setminus\mathcal{S}}\max_{j\in\mathcal{S}\cap\mathcal{N}_i} N_{ij}^F(t)\frac{n_{\mathcal{S}}}{n_{\mathcal{L}}^2}\eta,$$

where $n_{\mathcal{S}}(i)$ denotes the number of spoofers in the neighborhood of i and $n_{\mathcal{S}}(i) \leq n_{\mathcal{S}}$ by definition. Now, by Lemma 2, we have that with probability at least $1-\delta$, $\max_{i\in\mathcal{I}\setminus\mathcal{S}}\max_{j\in\mathcal{S}\cap\mathcal{N}_i} N_{ij}^F(t) \leq M(\mathcal{P},\delta)$. Thus, conditioning on this event occurring, it follows with probability greater than $1-\delta$ that $\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^F(0,t)| \leq \eta M(\mathcal{P},\delta)n_{\mathcal{S}}/n_{\mathcal{L}}^2$. ■

The following proposition states that using a weighting matrix as defined in Equation (3) results in provably bounded influence from the spoofed nodes in the network.

Lemma 4 (Bounded $\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)$): For the consensus system with linear dynamics as defined in (2), weighting matrix as defined in (3), with $n_{\mathcal{L}} \geq 2$, and with Assumption 1 holding, the influence of the spoofed node values $x_{\mathcal{S}}(t)$ on the system is bounded for all t. Namely, $\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)$ is bounded for all $t>0$.

Proof: From [22] the term $|\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)|$ in Equation (4) is bounded so long as the interior term can be bounded by an exponential function, i.e.,

$$|\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)| = \sum_{k=0}^{t-1}\sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i:\\ \beta_{ij}(k)<0}}\left|\left(\prod_{l=k}^{t-1}W_{\mathcal{L}}(l)\right)W_{\mathcal{S}}(k)\right|_{ij}\eta < \infty$$

$$\text{if}\quad \sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i:\\ \beta_{ij}(k)<0}}\left|\left(\prod_{l=k}^{t-1}W_{\mathcal{L}}(l)\right)W_{\mathcal{S}}(k)\right|_{ij}\eta \leq c_1 e^{-c_2(t-k)},$$

where $c_1, c_2 > 0$ are positive constants. We will show that this is the case for our choice of weights $\mathbb{W}$ as defined in Equation (3).

$$\sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i:\\ \beta_{ij}(k)<0}}\left|\left(\prod_{l=k}^{t-1}W_{\mathcal{L}}(l)\right)W_{\mathcal{S}}(k)\right|_{ij}\eta$$

$$\leq \sum_{\substack{j\in\mathcal{S}\cap\mathcal{N}_i:\\ \beta_{ij}(k)<0}}\left(\prod_{l=k}^{t-1}\frac{(1-e^{-\beta_{ij}(l)/2})}{n_{\mathcal{L}}}\right)\frac{e^{\beta_{ij}(k)}}{2n_{\mathcal{L}}}\eta$$

$$\leq \frac{n_{\mathcal{S}}}{2n_{\mathcal{L}}}\eta\left(\frac{1}{n_{\mathcal{L}}}\right)^{(t-k)} \leq \frac{n_{\mathcal{S}}}{2n_{\mathcal{L}}}\eta\; e^{-(\log 2)(t-k)},$$

where the first inequality follows from an application of the definition of the weights from Equation (3), the second inequality follows from the fact that $\beta_{ij}(k) < 0$, and the last inequality follows from the condition of the lemma that $n_{\mathcal{L}} \geq 2 = \exp(\log 2)$. This satisfies the aforementioned condition with $c_1 = \frac{n_{\mathcal{S}}}{2n_{\mathcal{L}}}\eta$ and $c_2 = \log 2 > 0$. ■

We show that in addition to allowing for a finite bound on the influence of the spoofed nodes on the converged consensus value, under our weighting matrix definition from (3) we can find a closed-form solution for the bound itself, that holds with any desired probability $1-\delta \in (0,1)$.

Theorem 1 (Probabilistic Bound on Spoofer Influence): Given any desired failure probability $\delta \in (0,1)$ and $n_{\mathcal{L}} \geq 2$, the maximum amount of influence that a group of $n_{\mathcal{S}}$ spoofed nodes can have on the network is bounded by

$$\lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}(0,t)| \leq \frac{n_{\mathcal{S}}\,\eta}{n_{\mathcal{L}}}\left(\frac{M(\mathcal{P},\delta)}{n_{\mathcal{L}}}+1\right),$$

with probability at least $1-\delta$, where $M(\mathcal{P},\delta)$ is defined as in Lemma 2. That is, Problem 1 can be solved using the weights $\mathbb{W}$ defined in Equation (3), $\mathcal{P} = (n_{\mathcal{S}}, n_{\mathcal{L}}, \eta, \epsilon_{\mathcal{S}})$, and with $\Delta_{\max}(\mathcal{P},\delta) = \frac{n_{\mathcal{S}}\,\eta}{n_{\mathcal{L}}}\left(\frac{M(\mathcal{P},\delta)}{n_{\mathcal{L}}}+1\right)$.

Proof: We have by the decomposition described in Section II that

$$\lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}(0,t)| \leq \lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^F(0,t)| + |\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)|.$$

Invoking Lemmas 3 and 4, we obtain with probability greater than $1-\delta$

$$\lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^F(0,t)| + |\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)|$$

$$\leq \lim_{t\to\infty}\left(M(\mathcal{P},\delta)\frac{n_{\mathcal{S}}}{n_{\mathcal{L}}^2}\eta + \max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)|\right)$$

$$= M(\mathcal{P},\delta)\frac{n_{\mathcal{S}}}{n_{\mathcal{L}}^2}\eta + \lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)|,$$

where the last equality holds by the fact that the expression $M(\mathcal{P},\delta)$ is independent of t. Now, we know by Lemma 4 with $n_{\mathcal{L}} \geq 2$, that we can bound the second term above by a convergent geometric series

$$\lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)| \leq \lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}\sum_{k=0}^{t-1}\frac{n_{\mathcal{S}}e^{-\log(2)k}}{2n_{\mathcal{L}}}\eta,$$

$$= \lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}\frac{n_{\mathcal{S}}(2-2^{1-t})}{2n_{\mathcal{L}}}\eta,$$

and thus $\lim_{t\to\infty}\max_{i\in\mathcal{I}\setminus\mathcal{S}}|\phi_{\mathcal{S}_i}^{\bar{F}}(0,t)| \leq \frac{n_{\mathcal{S}}}{n_{\mathcal{L}}}\eta$, which concludes the proof. ■

We conclude the section by observing that the bound provided by Theorem 1 contains highly intuitive expressions. In particular, we observe that the bound becomes larger as the ratio of the spoofed nodes to legitimate nodes increases, quantifying the intuition that it is not possible to bound the spoofers' influence by a small value if the network is dominated by spoofed nodes. Moreover, the bound becomes larger as the δ term, i.e., the user-specified failure probability, decreases, capturing the intuition that in order for our bound to hold with higher probability, we require the bound to be larger.

IV. Results

In this section we implement the consensus protocol from (1) with weights $\mathbb{W}(i,j,t)$ as defined in (3) for all agents i and $j \in \mathcal{N}_i$. We show in simulation that all agent states converge to a value that is bounded near the true average value taken over the legitimate nodes only. We present results for different topologies of 7 to 20 legitimate nodes and 1 to 5 spoofed nodes and in different consensus objectives of flocking and rendezvous.

The set of figures in Fig. 3 shows the result of implementing our resilient consensus algorithm in a rendezvous context where the value of each node is its position in $\mathbb{R}^2$. The plots in the leftmost column show agent positions (top left) over

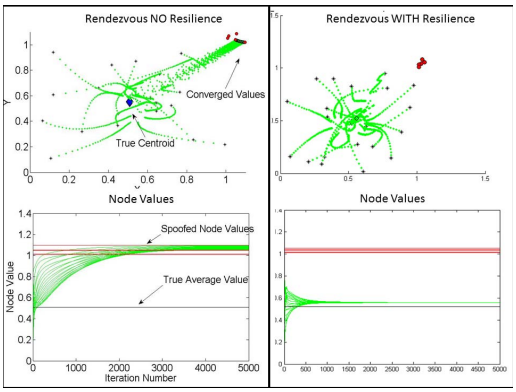

Fig. 3. Resilient consensus algorithm for rendezvous of over 20 legitimate nodes (green) and 5 spoofed nodes (red) to true average (blue diamond).

time in the case of standard consensus without using our resilient weights. The accompanying node value graph (bottom left) shows that all agents converge to the values of the spoofed nodes. In contrast, on the rightmost column the node trajectories (top right) and node values (bottom right) show convergence very close to the true average taken over only the legitimate node values.

The set of figures in Fig. 1 show the result of implementing our resilient consensus algorithm in a flocking context where agents must agree on an average heading value. The spoofed node heading is shown in red and the true average over the legitimate nodes is shown in green. The leftmost column shows the result of flocking when using a standard consensus algorithm where all nodes converge to the spoofed node input. In contrast on the rightmost column our resilient consensus algorithm is implemented and all node values converge to a heading that is within a small bounded constant of the true average heading (green).

V. Discussion

We note that a few of the assumptions made in the current paper can be relaxed in future versions of this letter. For example, allowing for changes in topology, and further, *controlling* changes in topology for mitigating the influence of spoofed nodes may be possible through the application of results like those in [20]. We now comment on independence of the α_{ij} values. Conditioned on j being legitimate or spoofed, computation of the $\alpha_{ij}(t)$ values will be subject to noise experienced over the wireless channel. This is widely assumed to be Gaussian white noise [23] that is independent with respect to time. In addition for uniform scattering environments signal paths decorrelate spatially [24]. However, it is possible that since the α_{ij} terms also have a dependence on the environment, if the environment and all robots stay perfectly static then complete independence may no longer hold. Relaxing these assumptions is deferred to future work.

VI. Conclusion

We presented a novel algorithm for resilient consensus in multi-agent networks based on Wi-Fi signals already present in the network. We analyzed the theoretical properties of our approach and characterized the influence of spoofed nodes on the converges consensus value as a function of problem-specific parameters. Our analytical results provide a closed-form expression for the probabilistic bounds that can be achieved for adversarial consensus using information from the wireless channel captured by the channel weights. Our results in simulation validate the favorable theoretical properties of our spoof-resilient consensus algorithm and demonstrate its practical effectiveness for consensus problems.

Acknowledgment

The Authors also acknowledge anonymous reviewer comments and SK for improving the ideas and quality of this letter.

References

[1] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.
[2] N. A. Lynch, *Distributed Algorithms*. San Francisco, CA, USA: Morgan Kaufmann, 1996.
[3] D. P. Bertsekas *et al.*, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA, USA: Athena Sci., 1997.
[4] H. Garcia-Molina, "Elections in a distributed computing system," *IEEE Trans. Comput.*, vol. C-31, no. 1, pp. 48–59, Jan. 1982.
[5] M. J. Fischer, N. A. Lynch, and M. S. Paterson, "Impossibility of distributed consensus with one faulty process," *J. ACM*, vol. 32, no. 2, pp. 374–382, 1985.
[6] D. P. Bertsekas and J. N. Tsitsiklis, "Some aspects of parallel and distributed iterative algorithms—A survey," *Automatica*, vol. 27, no. 1, pp. 3–21, 1991.
[7] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
[8] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.
[9] B. Sheng and Q. Li, "Verifiable privacy-preserving range query in two-tiered sensor networks," in *Proc. INFOCOM*, 2008, pp. 1–9.
[10] D. Moser *et al.*, "Investigation of multi-device location spoofing attacks on air traffic control and possible countermeasures," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2016, pp. 375–386.
[11] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 766–781, Apr. 2013.
[12] L. Lamport, "Paxos made simple," *ACM SIGACT News*, vol. 32, no. 121, pp. 51–58, 2001.
[13] L. Lamport, "Fast Paxos," *Distrib. Comput.*, vol. 19, no. 2, pp. 79–103, Oct. 2006.
[14] F. Higgins, A. Tomlinson, and K. M. Martin, "Threats to the swarm: Security considerations for swarm robotics," *Int. J. Adv. Security*, vol. 2, no. 2, pp. 288–297, 2009.
[15] I. Sargeant and A. Tomlinson, "Modelling malicious entities in a robotic swarm," in *Proc. DASC*, 2013, pp. 7B1-1–7B1-12.
[16] K. Saulnier, D. Saldaña, A. Prorok, G. J. Pappas, and V. Kumar, "Resilient flocking for mobile robot teams," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1039–1046, Apr. 2017.
[17] S. Gil, S. Kumar, M. Mazumder, D. Katabi, and D. Rus, "Guaranteeing spoof-resilient multi-robot networks," *Auton. Robots*, vol. 41, no. 6, pp. 1383–1400, 2017.
[18] S. Gil, S. Kumar, M. Mazumder, D. Katabi, and D. Rus, "Adaptive communication in multi-robot systems using directionality of signal strength," *Int. J. Robot. Res.*, vol. 34, no. 7, pp. 946–968, 2015.
[19] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proc. MobiCom*, 2014, pp. 483–494.
[20] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.
[22] H. Freeman, *Discrete-Time Systems: An Introduction to the Theory*. Huntington, NY, USA: R. E. Krieger, 1980.
[23] D. Tse and P. Vishwanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[24] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.