Dynamic Crowd Vetting: Collaborative Detection of Malicious Robots in Dynamic Communication Networks

Matthew Cavorsi*, Frederik Mallmann-Trenn*, David Saldaña, and Stephanie Gil

***Abstract*— Coordination in a large number of networked robots is a challenging task, especially when robots are continuously moving around the environment and there are malicious attacks within the network. Various approaches in the literature exist for detecting malicious robots, such as message sampling or suspicious behavior analysis. However, these approaches require every robot to sample or observe every other robot in the network, leading to a slow detection process that degrades team performance. This paper introduces a method that significantly decreases the detection time for legitimate robots to identify malicious robots in a scenario where legitimate robots are randomly moving around the environment. Our method leverages a concept that we refer to as "Dynamic Crowd Vetting," whereby, by utilizing observations from random encounters in combination with trusted neighboring robots' opinions, legitimate robots can quickly improve the accuracy of detecting malicious robots. The key intuition is that as long as each legitimate robot accurately estimates the legitimacy of at least some fixed subset of the team, the second-hand information they receive from trusted neighbors is enough to correct any misclassifications and provide accurate trust estimations of the rest of the team. We show that the size of this fixed subset can be characterized as a function of fundamental graph and random walk properties. Furthermore, we formally show that the detection time remains constant with respect to team size for a fixed ratio of legitimate to malicious robots. We develop a closed form expression for the critical number of time-steps required for our algorithm to successfully identify the true legitimacy of each robot to within a specified failure probability. Our theoretical results are validated through simulations demonstrating significant reductions in detection time when compared to previous works that do not leverage trusted neighbor information.**

I. Introduction

Multi-robot teams can cooperate to solve a plethora of tasks that a singular robot could not achieve alone [1], [2] such as coverage or persistent surveillance [3], [4], efficient exploration of a large area [5], and flocking [6], among others. However, whenever a task requires the coordination of multiple robots for successful task completion, there exists potential for *malicious*, or non-cooperating robots, to hinder the team's performance. Recent works have leveraged the concept of "observing" robots, and gathering information in order to identify robots that are potentially untrustworthy [7], [8], [9], [10]. This information, hereafter referred to as a *trust observation*, centers on the principle that gathering more information, or multiple trust observations, can improve the accuracy of this inter-robot trust [11], [12].

Previous works have taken a controls perspective to gathering trust observations by developing strategies that favor frequent encounters between robots. For example, the work in [13] designs specific routes for the team to follow as they patrol an environment that strategically increases the inter-robot interaction opportunities. However, this requires that robots cooperate and follow their pre-defined routes. The papers [14], [15], [16] consider environments that are discretized into regions, called *sites*, where robots provide persistent surveillance by patrolling while maintaining a desired distribution of robots over sites. In our previous work [14], we present a strategy that reduces the detection time by encouraging frequent encounters between robots using a stochastic site transition rule akin to a random walk. However, this strategy requires that every robot encounters every other robot many times in order to develop an accurate trust estimation, which can be a significantly long process, especially as the number of robots or sites increases.

A recent algorithm called *Crowd Vetting* [11] exploits opinion dynamics [17], where *the opinions of trusted neighbors can be used to fortify a robot's opinion of another*. In the Crowd Vetting algorithm, legitimate robots share opinions, which is shown to improve their trust estimation while requiring fewer observations than if robots rely solely on their own direct observations. However, the existing Crowd Vetting algorithm is limited to static networks where each robot observes the same subset of robots over time. While dynamic networks increase the diversity of

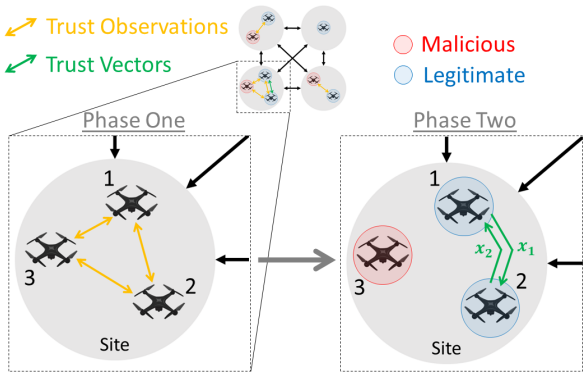

Fig. 1. Depiction of the Dynamic Crowd Vetting algorithm. In Phase 1 robots transition across sites while collecting trust observations of other robots they encounter which are stored in a local trust vector. In Phase 2 robots share their trust vector with trusted neighbors in order to improve their probability of correctly classifying the trustworthiness of others.

The authors gratefully acknowledge partial support through the Air Force Office of Scientific Research [grant number: FA9550-22-1-0223], the Office of Naval Research (ONR) Young Investigator Program (YIP) [grant number: N00014-21-1-2714], and by the EPSRC [grant number: EP/W005573/1] .

(*Co-primary authors). Matthew Cavorsi and Stephanie Gil are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA (e-mail: mcavorsi@g.harvard.edu; sgil@seas.harvard.edu)

Frederik Mallmann-Trenn is with the Department of Informatics, King's College, London, UK (e-mail: frederik.mallmann-trenn@kcl.ac.uk)

David Saldaña is with the Autonomous and Intelligent Robotics Laboratory –AIRLab– at Lehigh University, Bethlehem, PA, USA (e-mail: saldana@lehigh.edu)

robot's encounters, thus alluding to a greater benefit from using the opinions of trusted neighbors, it becomes increasingly difficult to regulate the number of trust observations every pair of robots has of each other. This introduces new challenges, since every robot can have a different number of observations of every other, and thus the information shared between robots comes with different levels of accuracy, making it difficult to arrive at any analytical performance guarantees regarding the trust estimation. Furthermore, a naive usage of indirect information from untrustworthy sources gives the potential for errors to propagate through the team.

The main contribution of this paper is the development of an algorithm, called *Dynamic Crowd Vetting* (DCV), that significantly reduces the detection time by allowing legitimate robots to leverage second-hand (indirect) opinions of trusted neighbors in dynamic networks. Our DCV algorithm computes the trust estimation in two phases. In Phase 1, robots transition between sites to estimate the trustworthiness of the team, which they store in a vector called a *trust vector*. The goal in Phase 1 is to accurately classify *at least some fixed subset* of the team correctly. Then, in Phase 2, robots continue transitioning between sites, while this time sharing their opinions about other robots, i.e., their trust vector. Finally, we show that a relatively simple majority rule algorithm for deciding trust from shared information is enough to correct misclassifications and stem the propagation of wide-scale misinformation as long as each legitimate robot classifies a sufficient proportion of the team correctly in Phase 1. Furthermore, we show that the sufficient proportions can be characterized as a function of fundamental graph and random walk properties. Additionally, we formally show that as the number of robots in the team increases, the detection time remains constant if the proportion of legitimate and malicious robots remains constant. This is in contrast to the logarithmic growth in the number of time-steps seen by an alternative *Direct Protocol*, where robots do not leverage neighboring opinions.

II. Problem Formulation

Consider a team of $N_{\mathcal{R}}$ robots, $\mathcal{R}=\{1,2,...,N_{\mathcal{R}}\}$, that move through a discrete environment composed of regions, also called sites. The environment is a topological map modeled as an undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E})$, where the vertices $\mathcal{V}=\{1,2,...,N_{\mathcal{V}}\}$ represent the sites, and the edges $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ represent paths between sites, where the operator $\times$ represents the Cartesian product of two sets. A robot can move from site ι to site ω if there is an edge $(\iota,\omega)\in\mathcal{E}$. Furthermore, robots can always remain at a current site, i.e., $(\iota,\iota)\in\mathcal{E}$ for all $\iota\in\mathcal{V}$. We assume the graph $\mathcal{G}$ is connected, so that there always exists a path between any pair of sites. Robots can communicate or observe each other if they are at the same site. The neighborhood of a robot i, denoted by $\mathcal{N}_i(t)$, consists of all robots $j\in\mathcal{R}$ that robot i can observe at time-step t. For the sake of analysis, we include each robot i in its own neighborhood $\mathcal{N}_i(t)$. A time-step is defined as an opportunity for a robot to make a transition between adjacent sites and observe the robots at that site, i.e., any robot can transition to a new site and gather new observations any time-step.

A. Background

1) Gathering Trust Observations in Dynamic Networks: In this paper, we are interested in the class of problems where an unknown subset of the team may be *malicious*, denoted by $\mathcal{M}\subset\mathcal{R}$, and *legitimate* robots $\mathcal{L}=\mathcal{R}\backslash\mathcal{M}$, can validate information and the legitimacy of neighboring robots by utilizing observations of one another, which we call *trust observations*. A trust observation of robot j by robot i, denoted by $\alpha_{i,j}(t)\in[0,1]$, represents a noisy, imperfect measurement of the legitimacy of robot j. [1] We assume that trust observations are independent for any pair (i,j) and at any time t, and that robots can only gather observations of one another when they are neighbors, i.e., $j\in\mathcal{N}_i(t)$. Furthermore, while we do not make any assumptions on the distributions of the trust observations, we impose particular assumptions on their expectation for analytical purposes. Similar to the works in [7], [11], [12], we assume the trust observations satisfy

$$\begin{aligned}\mathbb{E}[\alpha_{i,j}(t)\mid j\in\mathcal{L}]&\geq 1/2+\varepsilon_\alpha,\\ \mathbb{E}[\alpha_{i,j}(t)\mid j\in\mathcal{M}]&\leq 1/2-\varepsilon_\alpha,\end{aligned}\tag{1}$$

where the value $\varepsilon_\alpha\in(0,1/2]$ represents the quality of the observation. A low value $\varepsilon_\alpha\approx 0$ means the observation is completely ambiguous, while an observation with $\varepsilon_\alpha\approx 1/2$ gives almost certain information about the legitimacy of the transmitting robot. The quality of the observation ε_α can be found experimentally, as was done in [14] and [18], which we will use later in Section V. In [14], each robot keeps a *trust vector*, denoted by $\tilde{\mathbf{x}}_i(t)$. The goal of the trust vector is to store the correct legitimacy of every other robot in the team, where a 1 in the j^{th} entry of vector $\tilde{\mathbf{x}}_i(t)$, denoted by $[\tilde{\mathbf{x}}_i]_j(t)$, represents that robot i believes robot j to be trustworthy, and $[\tilde{\mathbf{x}}_i]_j(t)=0$ otherwise. Since the trust observations $\alpha_{i,j}(t)$ are assumed to be noisy, each robot requires multiple observations of their neighbors in order to arrive at some confidence in the validity of their trust vector. In [14] the specific number of observations required by every robot of every other is denoted by n_α, and is assumed to be an arbitrary, but given quantity. In this paper we will analytically determine the proper number of observations n_α that yields a desired success probability $1-\delta/N_{\mathcal{R}}$ when our proposed algorithm is used, for some user-defined failure probability δ. We seek to minimize the number of trust observations needed, as well as the time window T required to gather them, using the *Crowd Vetting Algorithm*.

2) The Crowd Vetting Algorithm: The Crowd Vetting algorithm [11] utilizes opinion dynamics and offers a way for robots to share their trust vectors with their trusted neighbors in order to not only reach an agreement between all legitimate neighbors on their trust vectors, but also improve the probability that the agreed upon trust vector is correct. However, the Crowd Vetting algorithm in its current form is limited to the case where all robots communicate with the same set of neighbors each time-step (static communication network). The goal is for all legitimate robots to reach an agreement on a final trust vector, $\mathbf{x}_i$, such that for every robot $j\in\mathcal{R}$,

$$[SG:[\mathbf{x}_i^*]_j]=\begin{cases}1, & \text{if } j\in\mathcal{L},\\ 0, & \text{if } j\in\mathcal{M}\end{cases}\tag{2}$$

[SG: is achieved.] To do so, each robot $i\in\mathcal{L}$ will gather n_α trust observations of every other robot and form an interim trust vector

[1] One example of such observations comes from the works in [7], [11], [12]. In these works, the trust observations are stochastic and are determined from physical properties of wireless transmissions.

from those observations. Then, each robot shares its interim trust vector with its trusted neighbors, and uses majority rule between its own and its trusted neighbors' opinions to determine whether or not to trust the other robots. In this paper, we extend the Crowd Vetting algorithm to scenarios where the robots move, and thus potentially encounter different robots each time-step.

3) Random Walks on Graphs: In this paper, our results partially depend on the topology of the site graph, and the random walk done by the legitimate robots as they transition between sites. Robots performing random walks over the environment, such as what is done in this paper, can be applicable to persistent surveillance-type tasks where robots need to constantly visit many different areas of the environment. Additionally, there exists methods to control the motion of robots, i.e., no longer use random walks, once all malicious robots are detected, such as the work in [14]. We define a *trajectory* of a legitimate robot i by a set of states, denoted $\chi_i(1),\chi_i(2),...,\chi_i(t_f)$ corresponding to the site that the robot occupied at each time-step from some arbitrary starting time $t=1$ to some arbitrary finishing time $t=t_f$. The trajectory of a robot depends on the random walk that it performs over the site graph. We assume that the random walks performed by the robots are irreducible and aperiodic, leading them to have a unique stationary distribution π. We represent the time required for robots to gather trust observations of each other as a function of the meeting time of the graph, denoted by T_{meet}, the hitting time of the graph, denoted by T_{hit}, and the mixing time of the graph, denoted by T_{mix}. See [19] for an intuition about these quantities. The meeting time is defined as $T_{\text{meet}}=\max_{\iota,\omega}T_{\text{meet}}(\iota,\omega)$, where $T_{\text{meet}}(\iota,\omega)$ is the expected time it takes for random walks starting on nodes ι and ω to meet. We say two random walks done by robots i and j meet if $\chi_i(\kappa)=\chi_j(\kappa)$ for some time κ. The hitting time is defined as $T_{\text{hit}}=\max_{\iota,\omega}T_{\text{hit}}(\iota,\omega)$, where $T_{\text{hit}}(\iota,\omega)$ is the expected time it takes for a random walk starting on node ι to reach node ω. The meeting time and hitting time are well-studied Markov Chain quantities and the interested reader can find bounds for common graphs in [20, page 169]. Additionally we compute the hitting time and meeting time for the graphs used in our simulations using [21, Theorem 3.1] and [22, Theorem 1], respectively. Finally, the mixing time T_{mix} is the time required for the distribution of the sites each robot occupies over time to approximately converge to the stationary distribution π.

B. Problem Statement

In this paper we extend the Crowd Vetting algorithm to support dynamic scenarios where a robot's set of neighbors may change with time. When the graph modeling the site transitions, $\mathcal{G}$, is connected, it was shown in [23] that all robots $i\in\mathcal{L}$ will eventually visit every site in the graph $\mathcal{G}$. This implies that any robot will encounter all the other robots given a long enough time window, $t=\{t_0,t_0+T\}$, characterized by the length of time T from some arbitrary initial time t_0, since each of them will visit every site.

Problem 1. *Given a desired failure probability δ and trust observations $\alpha_{i,j}(t)$ satisfying* (1)*, design an algorithm that reduces the length T of the time window $t\in\{t_0,t_0+T\}$ required for all robots $i\in\mathcal{L}$ to return the correct final trust vector $\mathbf{x}_i$ with probability at least $1-\delta/N_{\mathcal{R}}$.*

III. Algorithms

In order to extend the Crowd Vetting algorithm to support dynamic scenarios we first introduce the concept of *time-window neighborhoods* that capture the history of encounters between robots over a time window T.

Definition 1 (Time-window neighborhood)**.** *A* time-window neighborhood *of a robot i is defined as the union of its set of neighbors over a time-window, T, i.e., $\mathcal{N}_i^T(t)=\bigcup_{\kappa=t-T}^{t}\mathcal{N}_i(\kappa)$, for any $t>T$.*

In a time-window neighborhood, since the neighbors of each robot may change each time-step, it is difficult to ensure that a robot gathers trust observations of all others a sufficient number of times. Next, we describe the process for estimating trust vectors solely using each robot's individual (direct) observations of other robots, which we call the *Direct Protocol*.

A. Direct Protocol

In our previous work [14], when the robots need to estimate the legitimacy of their neighbors, they gather trust observations by transitioning frequently between sites, which is called the *fast transition state*. In this work, we focus our analysis on this state, and denote the transition matrix used by a robot i, by $\mathbf{P}_i$. We define the fast transition state as a lazy random walk

$$[\mathbf{P}_i]_{\iota,\omega}:=\begin{cases}\frac{1}{2}, & \iota=\omega,\\ \frac{1}{2|\{\omega'|(\iota,\omega')\in\mathcal{E}\}|}, & \iota\neq\omega,(\iota,\omega)\in\mathcal{E},\\ 0, & \text{otherwise,}\end{cases}\tag{3}$$

where $[\mathbf{P}_i]_{\iota,\omega}$ represents the (ι,ω) entry of matrix $\mathbf{P}_i$. We note that other choices of transition matrices are also valid as long as the Markov Chain is positive recurrent and aperiodic. Additionally, we note that the transition matrix $\mathbf{P}_i$ is the same for all robots $i\in\mathcal{L}$ since they are all performing random walks on the same site graph, but we include the index i to clarify that it is the transition matrix used by robot i since our later analysis is often done from the perspective of a particular robot $i\in\mathcal{L}$. Furthermore, any robot $j\in\mathcal{M}$ does not necessarily use the transition matrix designed in (3).

As the robots move throughout the environment, they gather trust observations of their neighbors. Let the vector $\mathbf{o}_{i,j}$ be a $\eta_{i,j}(t)\times 1$ vector that consists of every trust observation gathered by robot i of robot j over the time window T, where $\eta_{i,j}(t)\leq T$ represents the total number of observations gathered for the pair up to time t. Then, robot i determines a value $\beta_{i,j}(t)$, known as the *trust function*, from the vector $\mathbf{o}_{i,j}$ as follows:

$$\beta_{i,j}[t]=\sum_{\kappa=1}^{\eta_{i,j}(t)}\left([\mathbf{o}_{i,j}]_\kappa-\frac{1}{2}\right),\tag{4}$$

where $[\mathbf{o}_{i,j}]_\kappa\in[0,1]$ is the κ^{th} entry in vector $\mathbf{o}_{i,j}\in[0,1]^{\eta_{i,j}(t)}$. From (1), we know that $\alpha_{i,j}(t)<\frac{1}{2}$ in expectation if $j\in\mathcal{M}$, and so over the summation in (4) we have by the linearity of expectation [12], that $\beta_{i,j}(t)<0$ in expectation. Similarly, $\beta_{i,j}(t)>0$ in expectation if $j\in\mathcal{L}$. Therefore, each robot $i\in\mathcal{L}$ develops their interim trust vector, $\tilde{\mathbf{x}}_i(t)$, using the trust function $\beta_{i,j}(t)$ where

$$[\tilde{\mathbf{x}}_i]_j(t)=\begin{cases}1, & \text{if } \beta_{i,j}(t)\geq 0,\\ 0, & \text{if } \beta_{i,j}(t)<0.\end{cases}\tag{5}$$

The full process for estimating the legitimacy of each robot using individual trust observations is described in Algorithm 1. The algorithm requires robots to transition between sites for $\tau_{\text{d}} = \frac{4\log(2N_{\mathcal{R}}^3/\delta)}{\varepsilon_\alpha^2} T_{\text{meet}}$ time-steps, using the transition matrix $\mathbf{P}_i$ in (3), with the goal of gathering at least n_α observations for every other robot. Robots that do not gather at least n_α observations choose to not trust each other by default. We derive this time duration τ_{d} and the number of observations n_α, and show that it leads to a success probability of $1-\delta/N_{\mathcal{R}}$ in Section IV.

Algorithm 1 Direct Protocol for a robot i (**Direct**)
Input: time window τ_{d} (Theorem 1), transition matrix $\mathbf{P}_i$ in (3), number of observations n_α (Theorem 1)
Output: trust vector $\tilde{\mathbf{x}}_i(t)$

1: Using the fast transition matrix $\mathbf{P}_i$ in (3), gather trust observations of neighboring robots for τ_{d} time-steps. Keep track of the number of total observations $\eta_{i,j}(t)$, gathered for each robot $j \in \mathcal{R}$ over that time.
2: Compute the *trust vector* $\tilde{\mathbf{x}}_i(t) \in \{0,1\}^{N_{\mathcal{R}}}$: For every $j \in \mathcal{R}$ compute the entry $[\tilde{\mathbf{x}}_i]_j(t)$ using (5) if the number of observations gathered of robot j is at least n_α, otherwise $[\tilde{\mathbf{x}}_i]_j(t) = 0$. Set $[\tilde{\mathbf{x}}_i]_i(t) = 1$.

B. Dynamic Crowd Vetting Algorithm

The DCV algorithm seeks to utilize trusted neighboring opinions in order to reduce the time τ_{d} required to achieve a success probability of at least $1-\delta/N_{\mathcal{R}}$ by requiring the robots to only transition long enough to gather n_α observations of a subset of the network, rather than the entire network. Define the trusted neighborhood of a robot i at time t as the set

$$\Theta_i^T(t) := \{j \in \mathcal{N}_i^T(t) \mid [\tilde{\mathbf{x}}_i]_j(t) = 1\}. \tag{6}$$

The process for running the DCV algorithm is described in Algorithm 2. Similarly to the Direct Protocol, the algorithm requires that every legitimate robot use transition matrix $\mathbf{P}_i$ in (3). This time, the robots transition between sites for $\tau = \min\left\{ f(N_{\mathcal{R}}, |\mathcal{L}|, \delta) T_{\text{hit}}, \frac{4\log(4N_{\mathcal{R}}^3/\delta)}{\varepsilon_\alpha^2} T_{\text{meet}} \right\}$ time-steps in two phases, with the goal of gathering at least n_α trust observations of a large subset of the overall team, where $f(N_{\mathcal{R}}, |\mathcal{L}|, \delta) = \frac{26}{(1-1/e)^2} n_\alpha$, $n_\alpha = \frac{8}{0.1\varepsilon_\alpha^2} \log\left(\frac{e^{2e} N_{\mathcal{R}}}{0.1\delta|\mathcal{L}|}\right)$, and e is the Euler constant. We derive the time duration τ and number of observations needed n_α, and show that it leads to a success probability of $1-\delta/N_{\mathcal{R}}$ in Section IV.

Algorithm 2 has the robots arrive at the final trust vector faster by running the Direct Protocol for a shorter length of time in Phase 1, and then utilizing trusted neighboring opinions to fortify their own in Phase 2. In this way, robots do not need to sufficiently observe the trustworthiness of **every** other robot since they can rely on trusted neighbors to give them information about robots they have not encountered enough.

IV. Analysis

We organize this section similarly to Section III. First, we provide our theoretical analysis regarding the time required for robots to estimate the true legitimacy of all other robots using the Direct Protocol (Algorithm 1). Then, we provide analysis regarding the time required for robots to estimate the true legitimacy of all others using our proposed DCV algorithm (Algorithm 2) and show the reduction in the time required compared to the Direct Protocol (based on previous work [14]).

A. Direct Protocol

We start by deriving the time required for the Direct Protocol to return the correct final trust vector $\tilde{\mathbf{x}}_i(t)$ for all $i \in \mathcal{L}$.

Theorem 1. *Given a user-specified failure probability $\delta > 0$, site topology $\mathcal{G}$ with meeting time T_{meet}, and trust observations $\alpha_{i,j}(t)$ satisfying* (1)*. If all legitimate robots $i \in \mathcal{L}$ use the Direct Protocol (Algorithm 1) with the transition matrix $\mathbf{P}_i$ in* (3)*, $\tau_d = \frac{4\log(2N_{\mathcal{R}}^3/\delta)}{\varepsilon_\alpha^2} T_{meet}$ time-steps, and $n_\alpha = \log(2N_{\mathcal{R}}^3/\delta)/(2\varepsilon_\alpha^2)$ observations of every other robot, then the trust vector $\tilde{\mathbf{x}}_i(t)$ will be correct, i.e., satisfy* (2)*, for all $i \in \mathcal{L}$ with probability at least $1 - \frac{\delta}{N_{\mathcal{R}}}$.*

Proof: Consider any legitimate robot $i \in \mathcal{L}$. By Lemma 2 in Appendix A, robot i will correctly classify another robot j with probability at least $1-\delta/(2N_{\mathcal{R}}^3)$ if it gathers $n_\alpha = \log(2N_{\mathcal{R}}^3/\delta)/(2\varepsilon_\alpha^2)$ trust observations of robot j.

If $j \in \mathcal{M}$, then there are two cases: 1) if robot i meets robot j at least n_α times, then the probability of correctly classifying robot j is at least $1-\delta/(2N_{\mathcal{R}}^3)$. 2) if robot i meets robot j fewer times, then by default, robot i will correctly decide to not trust robot j. Taking the Union bound [24] over all pairs of robots in $\mathcal{L} \times (\mathcal{L} \cup \mathcal{M})$ gives a probability of failure of at most $|\mathcal{L} \times (\mathcal{L} \cup \mathcal{M})| \cdot \frac{\delta}{2N_{\mathcal{R}}^3}$.

It remains to argue that robot i will meet any robot $j \in \mathcal{L}$ at least n_α times in $\frac{4\log(2N_{\mathcal{R}}^3/\delta)}{\varepsilon_\alpha^2} T_{\text{meet}}$ time-steps. The probability of robot i and robot j meeting after $2T_{\text{meet}}$ time-steps is at least $1/2$ by Markov's inequality [25, Chapter 3.1], regardless of the sites that they start on. The expected number of meetings μ after $\frac{4\log(2N_{\mathcal{R}}^3/\delta)}{\varepsilon_\alpha^2} T_{\text{meet}} = 8n_\alpha T_{\text{meet}}$ time-steps is at least $\mu \geq 2n_\alpha$. Thus, by the Chernoff bound (Proposition 2 in Appendix A) setting

Algorithm 2 Dynamic Crowd Vetting (DCV) for a robot i
Input: time window τ (Proposition 1), transition matrix $\mathbf{P}_i$ in (3), number of observations n_α (Proposition 1)
Output: final trust vector $\mathbf{x}_i$

Phase 1:
1: Compute the *interim trust vector* $\tilde{\mathbf{x}}_i(t) \in \{0,1\}^{N_{\mathcal{R}}}$ using **Direct**$(\tau, \mathbf{P}_i, n_\alpha)$

Phase 2:
2: Transition for another τ time-steps while gathering the interim trust vector $\tilde{\mathbf{x}}_j(t)$ from all trusted neighbors $j \in \{\mathcal{N}_i(t) | [\tilde{\mathbf{x}}_i]_j(t) = 1\}$.
3: Compute the *final trust vector* $\mathbf{x}_i \in \{0,1\}^{N_{\mathcal{R}}}$: Assign each entry $[\mathbf{x}_i]_k$ by majority rule, i.e., $[\mathbf{x}_i]_k = 1$ if $\left(\sum_{j \in \Theta_i^\tau(t)} [\tilde{\mathbf{x}}_j]_k(t)\right) \geq \frac{|\Theta_i^\tau(t)|}{2}$, and $[\mathbf{x}_i]_k = 0$ otherwise.

$\gamma = 1/2$, we get that for the number of meetings, X,

$$\mathbb{P}[X \leq (1-\gamma)\mu] = \mathbb{P}[X \leq n_\alpha] \leq e^{-\frac{\gamma^2 \mu}{2}} \leq e^{-\frac{n_\alpha}{4}} \leq \frac{\delta}{2N_{\mathcal{R}}^3}. \quad (7)$$

Taking the Union bound over all pairs of legitimate robots gives a failure probability of at most $|\mathcal{L}\times\mathcal{L}|\cdot\frac{\delta}{2N_{\mathcal{R}}^3}$.

Summing the failure probabilities corresponding to misclassifying a robot and gathering an insufficient number of trust observations gives $|\mathcal{L}\times(\mathcal{L}\cup\mathcal{M})|\cdot\frac{\delta}{2N_{\mathcal{R}}^3} + |\mathcal{L}\times\mathcal{L}|\cdot\frac{\delta}{2N_{\mathcal{R}}^3} \leq \frac{\delta}{N_{\mathcal{R}}}$. □

We note that the bound of Theorem 1 is tight in the following sense: there exists an infinite class of graphs such that the required time for all pairs of robots to meet $\Omega(\log N_{\mathcal{R}})$ times is at least $\Omega(T_{\text{meet}} \log N_{\mathcal{R}})$. An example of this is the line graph over different numbers of sites.

B. Dynamic Crowd Vetting algorithm

Next we present the main result of this paper, which is the time required for the DCV algorithm to return the correct final trust vector $\mathbf{x}_i$ for all $i \in \mathcal{L}$. First, let

$$q = \frac{\delta\rho_1}{e^{2e}}\frac{|\mathcal{L}|}{N_{\mathcal{R}}}, \qquad n_\alpha = \frac{8}{\rho_1\varepsilon_\alpha^2}\log\left(\frac{1}{q}\right), \quad (8)$$

$$f(N_{\mathcal{R}},|\mathcal{L}|,\delta) = \frac{26}{(1-1/e)^2}n_\alpha. \quad (9)$$

Theorem 2. *Given a user-specified failure probability $\delta > 0$, site topology $\mathcal{G}$ with hitting time T_{hit} and meeting time T_{meet}, and trust observations $\alpha_{i,j}(t)$ satisfying (1). If all legitimate robots $i \in \mathcal{L}$ use DCV (Algorithm 2) with the transition matrix $\mathbf{P}_i$ in (3) and n_α observations given in (8), then the final trust vector $\mathbf{x}_i$ will be correct, i.e., satisfy (2), for all $i \in \mathcal{L}$ in time $O(\min\{T_{hit}\log(\frac{N_{\mathcal{R}}}{\delta|\mathcal{L}|}), T_{meet}\log(\frac{N_{\mathcal{R}}}{\delta})\})$ with probability at least $1-\frac{\delta}{N_{\mathcal{R}}}$.*

To prove Theorem 2, we start by defining an event E which, when it holds, implies that all final trust vectors are correct deterministically. The event E consists of four conditions where a certain proportion, denoted by $\rho_1,\rho_2,\rho_3,\rho_4 \in (0,1]$, of the legitimate robots satisfies the condition. Let E be the event that for every legitimate robot $i \in \mathcal{L}$, all of the following properties hold:

1) Robot i meets at least $(1-\rho_1)|\mathcal{L}|$ legitimate robots at least n_α times in Phase 1.
2) Robot i misclassifies at most $\rho_2|\mathcal{L}|$ legitimate robots in Phase 1.
3) Robot i misclassifies at most $\rho_3|\mathcal{L}|$ malicious robots in Phase 1.
4) Robot i meets at least $(1-\rho_4)|\mathcal{L}|$ legitimate robots at least once in Phase 2.

The four properties of event E are visually depicted in Fig. 2. We prove Theorem 2 by proving that when event E holds, all final trust vectors are correct, which we prove in Lemma 1, and the probability that event E holds is at least $1-\delta/N_{\mathcal{R}}$, which we prove in Proposition 1.

Lemma 1. *Assume that event E holds and that $1 > 3\rho_2+\rho_3+\rho_4$, where $\rho_2, \rho_3, \rho_4 \geq 0$. If the DCV algorithm is used with parameter $\tau = \min\left\{f(N_{\mathcal{R}},|\mathcal{L}|,\delta)T_{hit}, \frac{4\log(4N_{\mathcal{R}}^3/\delta)}{\varepsilon_\alpha^2}T_{meet}\right\}$ where $f(N_{\mathcal{R}},|\mathcal{L}|,\delta)$ is given in (9), then any legitimate robot i classifies any other robot j correctly.*

Proof: We analyze the process that robot $i \in \mathcal{L}$ uses to determine the trustworthiness of another robot j by taking information from each trusted neighbor $k \in \Theta_i^\tau(t)\backslash\{j\}$. We distinguish between two cases for any robot j.

For the first case, assume robot j is legitimate. Let F_i^{L+} be the number of legitimate robots that robot $i \in \mathcal{L}$ trusts, that also trust robot j, and that robot i met in Phase 2. This represents the number of legitimate robots that advocate for robot i's correct classification of robot j after Phase 2. For a legitimate robot k not to be counted in F_i^{L+}, one of three things must have happened: 1) robot i misclassified robot k, 2) robot k misclassified robot j, or 3) robot k did not meet robot i in Phase 2. We have, by Union bound, $F_i^{L+} \geq |\mathcal{L}| - 2\rho_2|\mathcal{L}| - \rho_4|\mathcal{L}| = (1-2\rho_2-\rho_4)|\mathcal{L}|$.

Let F_i^{L-} be the number of legitimate robots that robot i trusts, that do not trust robot j, and that robot i met in Phase 2. This represents the number of legitimate robots that advocate for robot i's *incorrect* classification of robot j after Phase 2. We have, by Union bound, $F_i^{L-} \leq \rho_2|\mathcal{L}|$.

Finally, let F_i^{M-} be the number of malicious robots that robot i trusts that claim that robot j is malicious. This represents the number of malicious robots that advocate for robot i's *incorrect* classification of robot j after Phase 2. We can assume that no malicious robot communicates that it trusts robot j. We have $F_i^{M-} \leq \rho_3|\mathcal{L}|$.

The sufficient condition for robot i to classify robot j correctly would be for the number of robots giving robot i the correct information to be greater than the number of robots giving robot i the wrong information, i.e., $F_i^{L+} > F_i^{L-} + F_i^{M-}$. It follows that we have $F_i^{L+} > F_i^{L-} + F_i^{M-}$ as long as $1 > 3\rho_2+\rho_3+\rho_4$. Hence, robot i classifies robot j correctly for any valid choice of ρ_2, ρ_3, and ρ_4. The process of trusted robots $k \in \Theta_i^\tau(t)\backslash\{j\}$ sharing information with robot i to help robot i make a decision about robot j is depicted in Fig. 3.

For the second case, assume that robot j is malicious. Let F_i^{L+} be the number of legitimate robots that robot i trusts, that classify robot j as malicious, and that robot i met in Phase 2. We have, $F_i^{L+} \geq |\mathcal{L}| - 2\rho_2|\mathcal{L}| - \rho_4|\mathcal{L}| = (1-2\rho_2-\rho_4)|\mathcal{L}|$.

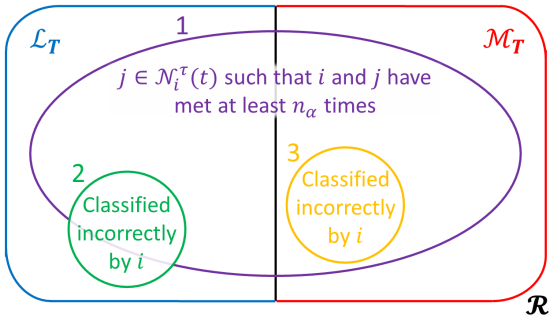

Fig. 2. Depiction of the regions specified by event E. Of all legitimate and malicious robots, the proportion that robot i meets at least n_α times during Phase 1 are represented by Property 1 in purple. Robot i will misclassify some of the legitimate robots after Phase 1, represented by Property 2 in green. This can happen if robot i misclassifies another robot having met it at least n_α times, or by classifying it as malicious by default having not met it at least n_α times. Additionally, robot i will misclassify some of the malicious robots after Phase 1, represented by Property 3 in orange. Property 4 corresponds to the robots that robot i meets in Phase 2, which is similar to the region represented by Property 1 in purple, with the distinction that robots only need to meet once in Phase 2 to be included in that region.

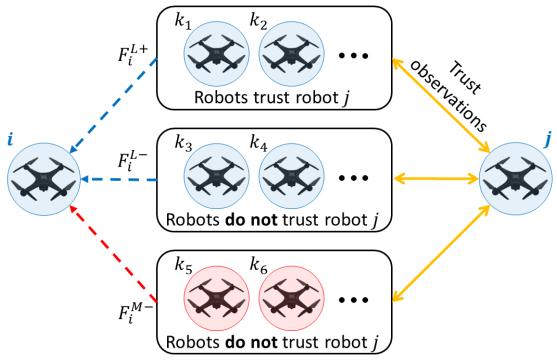

Fig. 3. The robots k share information with robot i about their opinion of robot j to help robot i determine the trustworthiness of robot j. Among the robots k, some number of them (F_i^{L+}) will be legitimate and give robot i the correct information, some number (F_i^{L-}) will be legitimate but misclassify robot j, and therefore give the wrong information, and some number (F_i^{M-}) will be malicious and purposely share the wrong information. Classification is done correctly if $F_i^{L+} > F_i^{L-} + F_i^{M-}$.

Let F_i^{L-} be the number of legitimate robots that robot i trusts, that classify robot j as legitimate, and that robot i met in Phase 2. We have, $F_i^{L-} \leq \rho_2 |\mathcal{L}|$.

Finally, let F_i^{M-} be the number of malicious robots that robot i trusts that claim that robot j is legitimate. We have $F_i^{M-} \leq \rho_3 |\mathcal{L}|$.

Clearly, we have $F_i^{L+} > F_i^{L-} + F_i^{M-}$ as long as $1 > 3\rho_2 + \rho_3 + \rho_4$. Hence, robot i classifies robot j correctly for any valid choice of ρ_2, ρ_3, and ρ_4. □

Proposition 1. *Let $|\mathcal{L}| \geq 1$, and choose $\rho_1, \rho_2, \rho_3, \rho_4 \geq 0$ such that $1 > 3\rho_2 + \rho_3 + \rho_4$ and $\rho_2, \rho_3 \geq 2\rho_1$. Given trust observations $\alpha_{i,j}(t)$ satisfying* (1)*, if all legitimate robots $i \in \mathcal{L}$ use the transition matrix $\mathbf{P}_i$ in* (3)*, then, running DCV (Algorithm 2) with parameters $\tau = \min\left\{ f(N_\mathcal{R}, |\mathcal{L}|, \delta) T_{hit}, \frac{4\log(4N_\mathcal{R}^3/\delta)}{\varepsilon_\alpha^2} T_{meet} \right\}$, and n_α, where $f(N_\mathcal{R}, |\mathcal{L}|, \delta)$ is given in* (9) *and n_α is given in* (8)*, ensures that event E holds with probability at least $1 - \frac{\delta}{N_\mathcal{R}}$.*

Remark: *Before we prove the proposition, note that if $|\mathcal{M}| = O(|\mathcal{L}|)$, then the first term of τ is $O(T_{hit})$ since $N_\mathcal{R}/|\mathcal{L}| = O(1)$, whereas the second term, regardless of the fraction of legitimate robots, is $O(T_{meet} \log(N_\mathcal{R}))$.*

Proof: Without loss of generality, we will assume that the first term of τ (the one that is a function of the hitting time) is the minimum. Otherwise, the proof trivially follows from Theorem 1 with success probability $1 - \delta/(2N_\mathcal{R})$. Consider the trajectory of any legitimate robot i given by $\chi_i(1), \chi_i(2), ..., \chi_i(t_f)$ from some arbitrary starting time $t = 1$ to some arbitrary finishing time $t = t_f$. We show that each property of event E holds with probability at least $1 - \delta/(4N_\mathcal{R})$. By doing so, a Union bound over all 4 properties yields a total success probability of at least $1 - \delta/N_\mathcal{R}$. We start with Property 1.

a) Property 1: We start by claiming that any other legitimate robot j meets robot i after $4T_{\text{mix}} + T_{\text{hit}}$ time-steps with probability at least $(1 - 1/e)^2$, where e is the Euler constant. To show this, note that, by [26, Lemma A.5] after $4T_{\text{mix}}$ time-steps, the random walk done by robot j follows the stationary distribution π with probability at least $1 - 1/e$. Then, by Lemma 3 after T_{hit} time-steps robot j does not meet robot i with probability at most $(1 - 1/T_{\text{hit}})^{T_{\text{hit}}} \leq 1/e$, where we used that $(1 + x/n)^n \leq e^x$ for $n \geq 1, |x| \leq n$. This proves the claim.

Now, we can divide time into periods of length $4T_{\text{mix}} + T_{\text{hit}}$ and repeat this trial noting that for each trial we have independence. By, Eq. (10.35) in [27], we have $4T_{\text{mix}} + T_{\text{hit}} \leq 13T_{\text{hit}}$. Let X be the number of meetings between two legitimate robots. After $f(N_\mathcal{R}, |\mathcal{L}|, \delta) T_{\text{hit}}$ time-steps, the expected number of meetings between two legitimate robots is $\mu \geq (1 - 1/e)^2 f(N_\mathcal{R}, |\mathcal{L}|, \delta)/13 = 2n_\alpha$. By the Chernoff bound (Proposition 2 in Appendix A) setting $\gamma = 1/2$, we get that the probability that there are fewer than n_α meetings between legitimate robots is at most

$$\begin{aligned} \mathbb{P}[X \leq n_\alpha] &\leq \mathbb{P}\left[X \leq \frac{\mu}{2}\right] \leq e^{-\gamma^2 \mu/2} \\ &\leq \exp\left(-\frac{2n_\alpha}{8}\right) \leq q^{2/\rho_1}. \end{aligned} \tag{10}$$

Let Y_j be the indicator variable that is 1 if legitimate robot j meets legitimate robot i less than n_α times. It is important to realize that for any $j, k \neq i$ and $j \neq k$ that Y_j and Y_k are independent since we have fixed the trajectory of robot i ahead of time. Using this crucial independence, we can bound $Y = \sum_{j \in \mathcal{L} \setminus \{i\}} Y_j$, i.e., the number of legitimate robots that meet robot i fewer than n_α times. To do so, we apply Proposition 3 (c.f. Appendix A) with $p = q^{2/\rho_1}$. Note that we can apply Proposition 3 since $p \leq \rho_1^2 / \exp(2e(1 - \rho_1))$. We have

$$\mathbb{P}[Y \geq \rho_1 |\mathcal{L}|] \leq p^{\rho_1 |\mathcal{L}|/2} = q^{|\mathcal{L}|} \leq \frac{\delta}{4|\mathcal{L}| \cdot N_\mathcal{R}}, \tag{11}$$

by Lemma 4 (c.f. Appendix A). Taking the Union bound over all legitimate robots $|\mathcal{L}|$ proves that Property 1 of event E holds with probability at least $1 - \delta/(4N_\mathcal{R})$.

b) Property 2: Let $\mathcal{L}_1$ be the set of legitimate robots that met robot i at least n_α times. By Lemma 2, since each robot $j \in \mathcal{L}_1$ met robot i at least n_α times, each robot $j \in \mathcal{L}_1$ will classify robot i correctly with probability at least $1 - q^{16/\rho_1}$. Now let $p = q^{16/\rho_1}$. In order for $\rho_2 |\mathcal{L}|$ legitimate robots to be misclassified it must hold that at least $\rho_1 |\mathcal{L}|$ robots among the $|\mathcal{L}_1|$ are misclassified.

The probability that more than $\rho_1 |\mathcal{L}|$ are misclassified, by Proposition 3 applied with $\rho = \frac{\rho_1 |\mathcal{L}|}{|\mathcal{L}_1|}$ and $n = |\mathcal{L}_1|$, with $p \leq \rho_1^2 / \exp(2e(1 - \rho_1))$, is, by Lemma 4, at most

$$p^{\rho_1 |\mathcal{L}|/2} = q^{8|\mathcal{L}|} \leq \frac{\delta}{4|\mathcal{L}| \cdot N_\mathcal{R}}. \tag{12}$$

Therefore, of all the successful meetings of Property 1: only $\rho_1 |\mathcal{L}|$ of them are misclassified with probability at least $1 - \delta/(4|\mathcal{L}| \cdot N_\mathcal{R})$. Thus, the total number of robots that misclassified robot i are the ones that met robot i fewer than n_α times, and the ones that met robot i at least n_α times but misclassified it. This gives us $\rho_1 |\mathcal{L}| + \rho_1 |\mathcal{L}| = \rho_2 |\mathcal{L}|$. Taking the Union bound over all $|\mathcal{L}|$ robots yields Property 2 with probability at least $1 - \delta/(4N_\mathcal{R})$.

c) Property 3: To maximize the number of malicious robots that are classified as legitimate, we can assume without loss of generality that each robot $j \in \mathcal{M}$ meets robot i at least n_α times. By Lemma 2, robot i will misclassify each malicious robot $j \in \mathcal{M}$ with probability at most $p = q^{16/\rho_1}$. Without loss of generality,

assume $|\mathcal{M}| \geq 1$. We can apply Proposition 3 with $\rho = \rho_3 |\mathcal{L}|/|\mathcal{M}|$ and $n = |\mathcal{M}|$, with $p \leq \rho_1^2 / \exp(2e(1-\rho_1))$. Therefore, we get that the probability that more than $\rho_3 |\mathcal{L}| = \rho_3 \frac{|\mathcal{L}|}{|\mathcal{M}|} |\mathcal{M}|$ malicious robots are misclassified is at most $p^{\rho_3 |\mathcal{L}|/2} = q^{16|\mathcal{L}|} \leq \delta/(4|\mathcal{L}| \cdot N_{\mathcal{R}})$, by Lemma 4. Taking the Union bound over all $|\mathcal{L}|$ legitimate robots completes the proof of Property 3.

d) Property 4: The proof of Property 4 is the same as Property 1 since $n_\alpha \geq 1$. Taking a Union bound over all 4 properties yields a total success probability of at least $1 - \delta/N_{\mathcal{R}}$. □

From Lemma 1, we have that event E leads to all legitimate robots returning the final trust vector correctly. Furthermore, from Proposition 1, we have that event E holds with probability at least $1 - \delta/N_{\mathcal{R}}$, thus proving Theorem 2. Note that the bound of Theorem 2 is tight in the following sense: there exists an infinite class of graphs for which the required time matches the required time of Theorem 2 up to constants. In some graphs, e.g., a star of sufficient size $N_{\mathcal{V}}$, it requires $\Omega(T_{\text{meet}} \log N_{\mathcal{V}}) = \Omega(\log N_{\mathcal{V}})$ rounds to meet. On the other side, on an $N_{\mathcal{V}} \times N_{\mathcal{V}}$ grid for example, it requires $\Omega(T_{\text{hit}}) = \Omega(N_{\mathcal{V}}^2)$ rounds.

It is also worth noting that after $O(T_{\text{hit}})$ time steps, the probability of failure is exponentially small in $|\mathcal{L}|$. Since we take the minimum of T_{hit} and T_{meet} multiplied with $\log N_{\mathcal{R}}$, we cannot hope to always get an exponentially small failure probability.

V. Simulations

To evaluate our proposed algorithm, we include simulation studies that investigate the time saved for determining the correct trust vectors by utilizing trusted neighboring opinions in our proposed method compared to the Direct Protocol. We adapt the model for trust observations shown in (1), and use $\varepsilon_\alpha = 0.35$ which was found experimentally in [14]. In Fig. 4, we varied the number of robots from 4 to 128 and checked the average number of time-steps required for legitimate robots using the Direct Protocol (grey) and our proposed DCV protocol (blue) to determine the correct trust vectors. We set $|\mathcal{L}| = |\mathcal{M}| = N_{\mathcal{R}}/2$ for each simulation, and ran the simulation 100 times for each value of $N_{\mathcal{R}}$. We also tested with different topologies, shown in the top left of each plot in Fig. 4 using 9 sites for each topology. The top left plot used a grid site topology, and the top right used a line topology. The bottom left plot considered a random graph generated using the Barabási-Albert model where $k < N_{\mathcal{V}}$ sites begin connected in a line, and the remaining sites are added one at a time with edges connecting them to up to k of the previous sites, chosen at random with $k = 3$. The bottom right plot considered a random graph generated using the Erdös-Rényi model where an edge is assigned between each pair of sites with probability 0.2.

Regardless of the site topology, our proposed DCV algorithm takes significantly fewer time-steps to achieve success compared to the Direct Protocol. It can also be seen that the difference between the number of time-steps required for each protocol increases as the number of robots increases, showing that the DCV algorithm performs better compared to the Direct Protocol as the team size is scaled up. We note that in the left-most plots (for the grid and Barabási-Albert topologies) the number of time-steps required in simulation using the DCV algorithm actually decreases slightly as the number of robots increases. This is due to the fact that we terminate simulations when the correct final trust vector is found. In Theorem 2 we show that it takes constant time to find the correct trust vectors using DCV as the number of robots increases, but that the probability of finding the correct trust vectors increases as the number of robots increases. This phenomenon can cause the decreases evident in the two left-most plots in Fig. 4. Additionally, we include lines that show the time-steps required that is predicted by our theory (purple), i.e., from Theorem 2 using $\delta = 0.1$. The hitting time for different site topologies was computed using [21, Theorem 3.1], and the meeting time was computed using [22, Theorem 1]. The hitting and meeting times for each of the 100 topologies generated using the random graph generation models (Barabási-Albert model and Erdös-Rényi model) were averaged in order to compute the time required predicted by our theory for those cases. From Fig. 4 it can be seen that the time-steps required that is predicted by theory closely matches (up to constants) the actual number found in simulation.

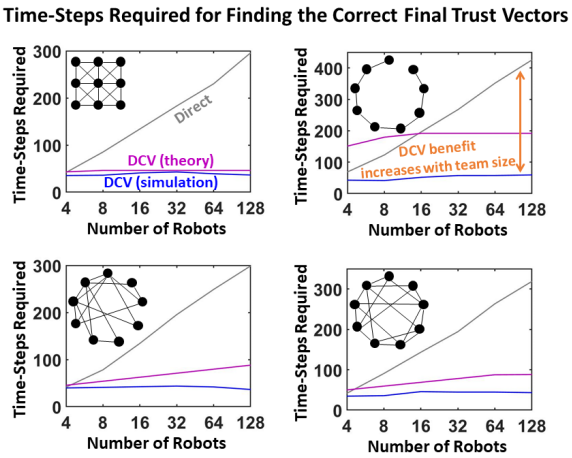

Fig. 4. Number of time-steps required for robots to find the correct final trust vectors using our proposed DCV protocol in simulation (blue) compared to the Direct Protocol (grey) and what we predict by theory for the DCV algorithm (purple). The number of robots is varied along 4 different site topologies each consisting of 9 sites: grid (top left), line (top right), Barabási-Albert (bottom left), and Erdös-Rényi (bottom right). As the number of robots increases, the ratio of legitimate to malicious robots remains constant.

In Fig. 5, we varied the number of sites in a grid topology (left), and the number of legitimate versus malicious robots (right). The left plot used $N_{\mathcal{R}} = 32$ robots, with $|\mathcal{L}| = |\mathcal{M}| = N_{\mathcal{R}}/2$. The right plot used a grid site topology with $N_{\mathcal{V}} = 9$ sites, and a constant $N_{\mathcal{R}} = 32$ robots, but varied the number of legitimate robots from $|\mathcal{L}| = 2$ to $|\mathcal{L}| = 30$ with $|\mathcal{M}| = N_{\mathcal{R}} - |\mathcal{L}|$. Both plots show that the benefits of the DCV algorithm over the Direct Protocol increase as the number of sites increases (left) and the ratio of legitimate to malicious robots increases (right).

VI. Conclusion

In this paper, we presented an algorithm for utilizing the opinions of trusted neighbors to quickly and effectively determine the true legitimacy of neighboring robots using trust observations. We show that not only does our algorithm help legitimate robots reach an agreement on their trust vectors, but it also reduces the time required to determine the trust vectors correctly by reducing the total number of time-steps that each robot has to individually gather observations for.

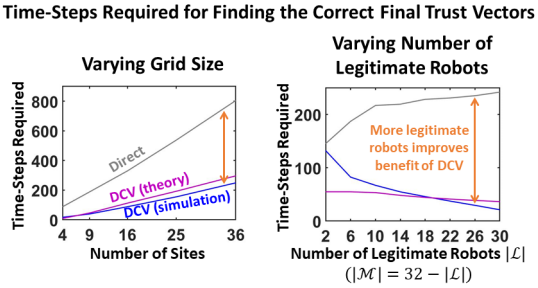

Fig. 5. Number of time-steps required for robots to find the correct final trust vectors using our proposed DCV protocol in simulation (blue) compared to the Direct Protocol (grey) and what we predict by theory for the DCV algorithm (purple). The number of sites is varied along a grid site topology (left), and the number of legitimate and malicious robots is varied using a fixed grid with $N_{\mathcal{V}} = 9$ (right).

REFERENCES

[1] Z. Yan, N. Jouandeau, and A. A. Cherif, "A survey and analysis of multi-robot coordination," *International Journal of Advanced Robotic Systems*, vol. 10, no. 12, p. 399, 2013.

[2] Y. Rizk, M. Awad, and E. W. Tunstel, "Cooperative heterogeneous multi-robot systems: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–31, 2019.

[3] A. Davydov and Y. Diaz-Mercado, "Sparsity structure and optimality of multi-robot coverage control," *IEEE Control Systems Letters*, vol. 4, no. 1, pp. 13–18, 2019.

[4] M. Boldrer, F. Pasqualetti, L. Palopoli, and D. Fontanelli, "Multi-agent persistent monitoring via time-inverted kuramoto dynamics," *IEEE Control Systems Letters*, 2022.

[5] A. Ivanov and M. Campbell, "Joint exploration and tracking: Jet," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 43–48, 2017.

[6] M. Cavorsi, B. Capelli, L. Sabattini, and S. Gil, "Multi-robot adversarial resilience using control barrier functions," *2022 IEEE Robotics: Science and Systems (RSS)*, 2022.

[7] S. Gil, S. Kumar, M. Mazumder, D. Katabi, and D. Rus, "Guaranteeing spoof-resilient multi-robot networks," *Autonomous Robots*, vol. 41, no. 6, pp. 1383–1400, 2017.

[8] C. Pippin and H. Christensen, "Trust modeling in multi-robot patrolling," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 59–66.

[9] A. Pierson and M. Schwager, "Adaptive inter-robot trust for robust multi-robot sensor coverage," in *In International Symposium on Robotics Research*, 2013.

[10] H. Parwana and D. Panagou, "Trust-based rate-tunable control barrier functions for non-cooperative multi-agent systems," *2022 IEEE International Conference on Decisions and Control*, 2022.

[11] F. Mallmann-Trenn, M. Cavorsi, and S. Gil, "Crowd vetting: Rejecting adversaries via collaboration with application to multirobot flocking," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 5–24, 2021.

[12] M. Yemini, A. Nedić, A. J. Goldsmith, and S. Gil, "Characterizing trust and resilience in distributed consensus for cyberphysical systems," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 71–91, 2021.

[13] X. Yu, D. Shishika, D. Saldana, and M. A. Hsieh, "Modular robot formation and routing for resilient consensus," *2020 American Control Conference (ACC)*, pp. 2464–2471, 2020.

[14] M. Cavorsi, N. Jadhav, D. Saldaña, and S. Gil, "Adaptive malicious robot detection in dynamic topologies," *2022 IEEE International Conference on Decisions and Control (CDC)*, 2022.

[15] S. Berman, A. Halász, M. A. Hsieh, and V. Kumar, "Optimized stochastic policies for task allocation in swarms of robots," *IEEE transactions on robotics*, vol. 25, no. 4, pp. 927–937, 2009.

[16] S. Bandyopadhyay, S.-J. Chung, and F. Y. Hadaegh, "Probabilistic and distributed control of a large-scale swarm of autonomous agents," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1103–1123, 2017.

[17] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, vol. 1, no. 1, pp. 3–49, March 2011.

[18] M. Cavorsi, O. E. Akgün, M. Yemini, A. J. Goldsmith, and S. Gil, "Exploiting trust for resilient hypothesis testing with malicious robots," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 7663–7669.

[19] L. Lovász, "Random walks on graphs," *Combinatorics, Paul erdos is eighty*, vol. 2, no. 1-46, p. 4, 1993.

[20] D. Aldous and J. Fill, "Reversible markov chains and random walks on graphs," 2002, unpublished. http://www.stat.berkeley.edu/ aldous/RWG/book.html.

[21] S. K. Rao, "Finding hitting times in various graphs," *Statistics & Probability Letters*, vol. 83, no. 9, pp. 2067–2072, 2013.

[22] M. George, R. Patel, and F. Bullo, "The meeting time of multiple random walks," *arXiv preprint arXiv:1806.08843*, 2018.

[23] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[24] G. Boole, *The mathematical analysis of logic*. Philosophical Library, 1847.

[25] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

[26] V. Kanade, F. Mallmann-Trenn, and T. Sauerwald, "On coalescence time in graphs-when is coalescing as fast as meeting?" *CoRR*, vol. abs/1611.02460, 2016. [Online]. Available: http://arxiv.org/abs/1611.02460

[27] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Society, 2006.

[28] R. I. Oliveira and Y. Peres, "Random walks on graphs: new bounds on hitting, meeting, coalescing and returning," in *2019 Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*. SIAM, 2019, pp. 119–126.

[29] T. Hagerup and C. Rüb, "A guided tour of chernoff bounds," *Information processing letters*, vol. 33, no. 6, pp. 305–308, 1990.

[30] M. Cavorsi, F. Mallmann-Trenn, D. Saldaña, and S. Gil, "Dynamic crowd vetting: Collaborative detection of malicious robots in dynamic communication networks," *ArXiv*, 2023.

APPENDIX

A. Auxiliary Claims

Lemma 2 (Upper bound [11]). *If a robot $i \in \mathcal{L}$ receives $n_\alpha = \frac{\log(1/\delta)}{2\varepsilon_\alpha^2}$ trust observations from another robot j, it will know with probability at least $(1-\delta)$ whether robot j is legitimate or malicious by simply relying on the majority of the observations:*

$$\begin{aligned} \mathbb{P}\left[\sum_{\kappa=1}^{n_\alpha}\left([\mathbf{o}_{i,j}]_\kappa - \frac{1}{2}\right) > 0 \;\middle|\; j \in \mathcal{L}\right] \geq 1-\delta, \\ \mathbb{P}\left[\sum_{\kappa=1}^{n_\alpha}\left([\mathbf{o}_{i,j}]_\kappa - \frac{1}{2}\right) < 0 \;\middle|\; j \in \mathcal{M}\right] \geq 1-\delta. \end{aligned} \tag{13}$$

Proposition 2 ([25]). *Let $X_1,\dots,X_n$ be independent Bernoulli random variables with $X = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[X]$. Then, for any $0<\gamma<1$, $\mathbb{P}[X \leq (1-\gamma)\mu] \leq e^{-\gamma^2\mu/2}$.*

Lemma 3 ([28, Theorem 1.3]). *For any $t \in \mathbb{N}\setminus\{0\}$ and any sequence $\chi_i(1),\chi_i(2),\dots,\chi_i(t) \in \mathcal{V}$ we have that for any lazy random walk $\mathbf{P}_j(\kappa)$ done by a robot j for $\kappa \in \{1,\dots,t\}$ starting from the stationary distribution π, that*

$$\mathbb{P}[\forall \kappa, \chi_j(\kappa) \neq \chi_i(\kappa)] \leq (1 - 1/T_{hit})^\kappa.$$

The following is consequence of [11, Theorem 4] and Equation 10 of [29].

Proposition 3. *Let $Y = \sum_{i=1}^n Y_i$ be the sum of n independent and identically distributed random variables with $\mathbb{P}[Y_i = 1] = p$ and $\mathbb{P}[Y_i = 0] = 1-p$ with $p \leq \rho^2/\exp(2e(1-\rho))$. We have for any $\rho \in (0, 0.8]$ that*

$$\mathbb{P}[Y \geq \rho n] \leq p^{\rho n/2}.$$

Lemma 4. *Consider the notation of Lemma 1 and Proposition 1. We have, $q^{|\mathcal{L}|} \leq \frac{\delta}{4N_{\mathcal{R}}|\mathcal{L}|}$.*

Proof: The proof can be found in our extended version [30].