Towards Differentially Private Aggregation of Heterogeneous Robots

Amanda Prorok and Vijay Kumar

Abstract We are interested in securing the operation of robot swarms composed of heterogeneous agents that collaborate by exploiting aggregation mechanisms. Since any given robot type plays a role that may be critical in guaranteeing continuous and failure-free operation of the system, it is beneficial to conceal individual robot types and, thus, their roles. In our work, we assume that an adversary gains access to a description of the dynamic state of the swarm in its non-transient, nominal regime. We propose a method that quantifies how easy it is for the adversary to identify the type of any of the robots, based on this observation. We draw from the theory of *differential privacy* to propose a closed-form expression of the *leakage* of the system at steady-state. Our results show how this model enables an analysis of the leakage as system parameters vary; they also indicate design rules for increasing privacy in aggregation mechanisms.

1 Introduction

To date, the issues of privacy and security remain poorly addressed within robotics at large. These issues are particularly important in heterogeneous multi-robot systems: by introducing heterogeneity by design, we impose a certain degree of uniqueness and specialization. Indeed, it is generally acknowledged that exploiting *heterogeneity* by design leads to more versatile systems [5, 20]. However, as a consequence, any given robot type may be critical to securing the system's ability to operate without failure. Hence, we must find ways of protecting the system to avoid threats that arise when the roles within the swarm can be determined by adversaries.

Amanda Prorok
University of Pennsylvania, PA, USA e-mail: prorok@seas.upenn.edu

Vijay Kumar
University of Pennsylvania, PA, USA e-mail: kumar@seas.upenn.edu

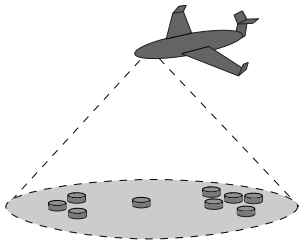

Fig. 1 Example scenario: an adversarial spy-plane can take a snapshot of a heterogeneous robot swarm. Even when the behavioral state is observable, the goal is to ensure that individual robot types/roles remain hidden.

In order to collaborate and reap the benefits of their complementarity, robots create coalitions. By doing so, the robots assemble into virtual or physical formations of higher complexity and functionality. Indeed, aggregation is a mechanism that is exploited by nature to enable interactions and information exchange between biological individuals (e.g., for predator protection and collective decision-making [10, 12, 18]). Inspired by the potential of such systems, swarm roboticists have tackled the problem of engineering and analyzing aggregation behaviors [3, 4, 11, 16, 17]. The work in this paper goes beyond previous approaches by including *privacy* as a novel component, and by posing a complementary problem formulation. Let us consider the example application shown in Fig. 1. The scenario considers a heterogeneous robot swarm that is dynamically aggregating, in order to solve an underlying problem. The swarm is observable from an outside vantage point, and is visually homogeneous. By taking a snapshot of the swarm's behavioral state, an adversary gains access to observable system-level information that may allow him to infer the role of a particular robot. By exploiting this knowledge, the adversary can then scheme attacks on the system (e.g., by targeting a specific robot type that he knows is critical to the system's operation). As a consequence, we aim at answering the following question: *How easy is it for the adversary to guess the role/type of any robot in the system?* To answer this question, we develop a model that enables the analysis of privacy as a function of the parameters that define the swarm's aggregation behavior. In other words, we measure the loss of privacy caused by the observable state of the swarm. The following sections develop our models, and elaborate the interplay between privacy and swarm behavior.

2 Model of Robot System

We define a system of robots, where each robot is classified according to its capabilities, and belongs to a *species* [20]. The system is composed of N_S species $\mathscr{S} = \{1, \ldots, N_S\}$, with a total number of robots N, and $N^{(s)}$ robots per species such that $\sum_{s \in \mathscr{S}} N^{(s)} = N$. At a high level of abstraction, we model the robots' actions (and interactions) as discrete stochastic events. We denote the states that compose the system as aggregates $a^{\mathscr{I}}$, where the superscript $\mathscr{I}$ is the list of species that compose the aggregate. For example, $a^{\{1,2\}}$ is an aggregate of species 1 and 2. Aggregation mechanisms can be expressed very elegantly using the formalism of Chemical

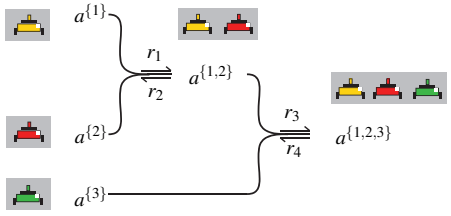

Fig. 2 Example of aggregation with three heterogeneous species that complement each other within their aggregates. Species 1 and 2 are topologically equivalent.

Reaction Networks (CRN) [15, 13, 17]. Indeed, CRNs are a powerful means of representing complex systems, and allow us to represent species interactions through linear as well as non-linear functions — though not a new field of research, many recent research findings that simplify the calculations are accelerating the adoption of CRNs into domains other than biology and chemistry [9]. Fig. 2 shows an example of three species that aggregate.

We define our CRN as a triplet $\mathscr{N} = (\mathscr{A}, \mathscr{C}, \mathscr{R})$, where $\mathscr{A}$ is the set of aggregate states, $\mathscr{C}$ is the set of complexes, and $\mathscr{R}$ is the set of reactions.

State set $\mathscr{A}$: The state set encompasses all states that arise in the system, with $\mathscr{A} = \{A_1, \ldots, A_{N_A}\}$ where N_A is the number of states. States relating to a specific species s are denoted by $\mathscr{A}^{(s)}$. The set of all states is denoted

$$\mathscr{A} = \bigcup_{s=1}^{N_S} \mathscr{A}^{(s)} \quad \text{and} \quad \mathscr{A}^{(s)} = \bigcup_{s \in \mathscr{I}} a^{\mathscr{I}} \tag{1}$$

We can identify the interactive (interspecific) states of an arbitrary subset of species $\tilde{\mathscr{S}} \subset \mathscr{S}$ by considering the intersection of sets $\cap_{i \in \tilde{\mathscr{S}}} \mathscr{A}^{(i)}$. Trivially, if $\cap_{i \in \tilde{\mathscr{S}}} \mathscr{A}^{(i)} = \emptyset$, then the species in $\tilde{\mathscr{S}}$ do not interact. The CRN is a population model, and allows us to keep track of the number of robots in each of the states (aggregates) in $\mathscr{A}$. Hence, we define a population vector $\mathbf{x} = [x_1, \ldots, x_{N_A}] \in \mathbb{N}_{\geq 0}^{N_A}$, where x_i corresponds to the population present in state A_i. We refer to the population vector $\mathbf{x}$ as the system-level state. In order to simplify the formulation of our case study later on, we will also use the notation $\mathbf{x}^{\mathscr{I}}$ to refer explicitly to the population in state $a^{\mathscr{I}}$.

Complex set $\mathscr{C}$: The complex set is defined as $\mathscr{C} = \{C_1, \ldots, C_{N_C}\}$, with N_C the number of complexes, and where $C_j = \sum_{i=1}^{N_A} \rho_{ij} A_i$ for $j = 1, \ldots, N_C$, with vector $\boldsymbol{\rho}_j = [\rho_{1j}, \ldots, \rho_{N_A j}]^\top \in \mathbb{N}_{\geq 0}^{N_A}$. A complex is a linear combination of states, and denotes the net input or output of a reaction. In other words, a complex denotes either *(i)* the states that are required for a certain reaction to take place, or *(ii)* the states that occur as an outcome of a certain reaction that took place. The non-negative integer terms ρ_{ij} are coefficients that represent the multiplicity of the states in the complexes.

Reaction set $\mathscr{R}$: We use complexes to formulate reactions $R_l : C_j \xrightarrow{r_l} C_k$. The reaction set is defined as $\mathscr{R} = \{R_1, \ldots, R_{N_R}\}$, with N_R the number of reactions, such that $R_l \in \{(C_j, C_k) | \exists C_j, C_k \text{ with } C_j \to C_k\}$ for $j, k = 1, \ldots, N_C$, and where r_l is the rate function $r_l(\mathbf{x}; \kappa_l) : \mathbb{N}_{\geq 0}^{N_A} \mapsto \mathbb{R}_{\geq 0}$ parameterized by rate constant κ_l. In this work, we use non-linear mass-action rate functions, and $r_l(\mathbf{x}; \kappa_l) = \kappa_l \prod_{i=1}^{N_A} x_i^{\rho_{ij}}$ for

all $R_l = (C_j, \cdot)$. A set of complexes that is connected by reactions is termed a linkage class. The net loss and gain of each reaction is summarized in a $N_A \times N_R$ stoichiometry matrix Γ, the columns of which encode the change of population per reaction. In particular, the i-th column of Γ corresponds to the i-th reaction $R_i = (C_j, C_k)$ and thus, the column is equal to $\boldsymbol{\rho}_k - \boldsymbol{\rho}_j$. The elements Γ_{ji} are the so-called stoichiometric coefficients of the j-th state in the i-th reaction. Positive and negative coefficients denote products and reactants of the reaction, respectively.

A simple stochastic model for CRNs treats the system as a continuous time Markov chain with state $\mathbf{x} \in \mathbb{N}_{\geq 0}^{N_A}$ (i.e., the population vector), and with each reaction modeled as a possible transition for the state. Hence, the number of transitions between two neighboring states is Poisson distributed (equivalently, the time between two transitions is exponentially distributed). In order to calibrate rate constants κ_l on hand of a real system, we proceed by measuring the effective transition rates (by observing the number of transitions, assuming the number of robots is known), and using the mass-action rate functions to solve for the parameter values.

Finally, we describe the dynamics of our system with help of two functions: an execution function $f_{\mathscr{N}}$, and a query function q:

$$\begin{aligned} f_{\mathscr{N}}(\mathbf{x_0}, t): &\ \mathbb{N}_{\geq 0}^{N_A} \times \mathbb{R}_{\geq 0} \mapsto \mathbb{N}_{\geq 0}^{N_A} \\ q(\mathbf{x}): &\ \mathbb{N}_{\geq 0}^{N_A} \mapsto \mathbb{N}^{N_O}, N_O \in \mathbb{N}_{>0} \end{aligned} \tag{2}$$

The execution function $f_{\mathscr{N}}$ samples a trajectory that describes the system's evolution over time, as defined by the states and reactions defined in $\mathscr{N}$, and returns the population vector $\mathbf{x}(t)$, evaluated at a fixed time t. The query function q allows us to formalize the notion of an *observable system-level* state. It takes the population vector $\mathbf{x}$ as input, and returns a vector of observable values $\mathbf{y}$. In its most basic form, the query function is the identity function, meaning that an observer is able to capture the exact (true) system-level state, and $\mathbf{x} = \mathbf{y}$. In this work, we assume that the observed values take the form of simple summations over the population vector. This assumption is well motivated when individual robots are not distinguishable from an outside vantage point, and thus, only aggregated values can be observed. I.e., $y_i = \sum_{j \in \Omega_i} x_j$ with $\Omega_i \subset \{1, \ldots, N_A\}$, $\cup_{i \in \{1,\ldots,N_O\}} \Omega_i = \{1, \ldots, N_A\}$, and all Ω_i disjoint. In our aggregation case-studies, we use $y_i = \sum_{\mathscr{I} \text{ s.t.} |\mathscr{I}|=i} x^{\mathscr{I}}$ for all $a^{\mathscr{I}} \in \mathscr{A}$, which counts the number of aggregates of a given size.

3 Definition of Differentially Private Swarm

In the following, we put forward a model that measures the privacy of a dynamic swarm of robots. Various measures of privacy have been proposed in the database literature so far. The early work in [1] proposes a quantification of privacy in terms of the amount of noise added to a true value, or, in other words, how closely the original value of a modified attribute can be estimated. This measure, however, omits the notion of side-information, i.e., any additional information about the underlying dis-

tribution that the adversary might own. The work in [8] extends the notion of privacy to include such prior knowledge. The proposed measure suggests a quantification of the largest difference between an adversary's a-priori to a-posterior beliefs (which corresponds to the worst-case scenario). It turns out that this model is significantly stronger, since it accounts for infrequent, but noticeable privacy breaches. In 2006, Dwork et al. introduced the notion of ε-indistinguishability, a generalization of the measure in [8], and later coined the term of *differential privacy* [7]. Today, differentially private mechanisms are enjoying tremendous success, due to their ability of dealing with arbitrary side-information (a *future-proof* quality) and worst-case scenarios [6]. For these reasons, we build our formalism on the theory of differential privacy.

In a nutshell, differential privacy is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any of its individuals. It considers two key components: a *database* that holds sensitive information pertaining to individuals, and a *query* that releases information obtained from the database via a mechanism. If an observer, who can request data from the database (through a query), cannot significantly reduce the uncertainty of her prior knowledge (i.e., side information) using the requested data, the query is considered private [1]. In particular, if an individual's presence or absence in the database does not alter the distribution of the output of the query by a significant amount, regardless of the side information, then the privacy of that individual's information is assured. Our analogy applies the concepts of database and query to the context of heterogeneous swarms. First, we consider a database that represents the composition of our robot swarm, and that records the species of each of the robots. Second, we consider an adversary who 'queries the system' by taking a snapshot of its observable state. The adversary may also own arbitrary side-information [2]. Then, our analogous definition of privacy is the notion that the adversary cannot infer to which species individual robots belong. The composition of our robot swarm is recorded in a database $\mathscr{D} \in \mathscr{S}^N$ that consists of N entries, where each entry $\mathscr{D}_i$ denotes the species of robot i. We define an *adjacency* set $\mathrm{Adj}(\mathscr{D})$ that encompasses all databases $\mathscr{D}$' adjacent to $\mathscr{D}$. Two databases $\mathscr{D}$ and $\mathscr{D}$' are adjacent if they differ by one single entry. In other words, two robot swarms (represented by $\mathscr{D}$ and $\mathscr{D}$') are adjacent if they differ by one robot i, meaning that robot i belongs to s_i in $\mathscr{D}$ (i.e., $\mathscr{D}_i = s_i$), and to a different species $s_i' \neq s_i$ in $\mathscr{D}$' (i.e., $\mathscr{D}_i \neq s_i$). As previously described, the behavior of the robot swarm can be described by tracking the system-level states. If we let the system run, it produces a trajectory that can be evaluated at a given time τ, resulting in a snapshot of the population vector $\mathbf{x}$ [3]. Our query/response model consists of a user (adversary) who is able to query this system-level state (at time τ). Hence, the query

[1] Side information can be understood as a prior probability distribution over the database [14]

[2] In our context of a robotic swarm, an example of side information could be the number of manufacturing parts ordered to build the swarm. If different robot species are made of different parts, such information can be used to construct an initial guess about the number of robots per species. Thus, one would be able to derive the probability of a robot belonging to a given species.

[3] We assume a snapshot adversary that gains system-level information at a specific time. This system-level information is a design variable, called the *observable* system-level state.

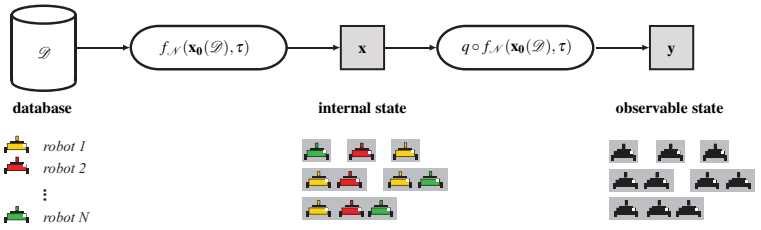

Fig. 3 The composition of the robot swarm is recorded in a database $\mathscr{D}$. The function $f_{\mathscr{N}}$ is a stochastic process that executes the swarm and returns a system-level state output $\mathbf{x}$. Query function q reads the (internal) system-level state, and returns the observable output $\mathbf{y}$. Parameter τ denotes the time at which the system is observed.

$q \circ f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}),\tau)$ depends on the swarm composition $\mathscr{D}$, and the time at which the system is observed τ. The function $\mathbf{x_0}(\mathscr{D}) : \mathscr{S}^N \mapsto \mathbb{N}_{\geq 0}^{N_A}$ distributes the robots in $\mathscr{D}$ to their initial states. A schema of this system is shown in Fig. 3.

Our aim is to analyze the differential privacy of the observed system output. To this end, we propose a definition of differential privacy that applies to dynamic swarms [19]:

Definition 1 (ε-indistinguishable heterogeneous swarm). A heterogeneous swarm with dynamics defined by a system $\mathscr{N}$ is ε-indistinguishable (and gives ε-differential privacy) if for all possible swarm compositions recorded in databases $\mathscr{D}$, we have

$$\mathscr{L}(\mathscr{D}) = \max_{\mathscr{D}' \in \mathrm{Adj}(\mathscr{D})} \left| \ln \frac{\mathbb{P}[q \circ f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}),\tau)]}{\mathbb{P}[q \circ f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}'),\tau)]} \right| \leq \varepsilon. \tag{3}$$

where $\mathbb{P}[\mathbf{y}]$ denotes the probability of the output $\mathbf{y}$, obtained through query q of the system-level state given by $f_{\mathscr{N}}$.

The value ε is referred to as the leakage. Intuitively, this definition states that if two swarm systems are close, in order to preserve privacy they should correspond to close distributions on their observable outputs. The above definition is stringent: for a pair of distributions whose statistical difference is arbitrarily small, the ratio may still result in an infinite value (leakage), when a point in one distribution assigns probability zero and the other non-zero. Later, in our evaluations, we use a smooth version of the leakage formula above, by adding an arbitrary, negligibly small value ν, uniformly over the support of the probability distributions. This allows us to differentiate between large and small mismatches of the output when one point of a probability distribution returns zero. Due to this addition, we are able to show continuous privacy trends as a function of the underlying parameters.

4 Complex-Balanced Swarms

The formula in Eq. (3) provides strong privacy guarantees. Yet, it requires that we have a way of specifying the probability distribution over the swarm's observable output. The choice of method for computing this probability distribution depends on the time at which the adversary takes the snapshot of the swarm. In [19], we presented a method that computes the probability distribution at a specific time (which can be during any regime, transient or stationary). Here, we present a method that computes the probability distribution at steady-state, in the swarm's operational mode. We show how this can be done very efficiently for a class of CRNs whose stationary distribution can be formulated analytically: *complex-balanced* CRNs. Later, in Section 5, we prove that aggregation mechanisms are complex-balanced.

4.1 Preliminaries

If a CRN is complex-balanced, it admits a single equilibrium point $\bar{\mathsf{x}} \in \mathbb{R}^{N_A}$ [21]. When modeled deterministically, the average population (rather than an exact robot count) in system states changes according to an ODE, described as follows [4]:

$$\dot{\mathsf{x}} = \mathbf{M}\mathbf{A}\boldsymbol{\psi}(\mathsf{x}), \tag{4}$$

where $\boldsymbol{\psi}(\mathsf{x})$ returns a vector in $\mathbb{R}^{N_C}$ in which each entry ψ_j is the product of states in complex j (i.e., $\psi_j = \prod_{i=1}^{N_A} x_i^{\rho_{ij}}$), where $\mathbf{M} \in \mathbb{R}^{N_A \times N_C}$ is a matrix in which each entry $\mathbf{M}_{ij}$ is the coefficient of state j in complex i, and where matrix $\mathbf{A} \in \mathbb{R}^{N_C \times N_C}$ is defined as

$$\mathbf{A}_{ij} = \begin{cases} \kappa_{ji}, & \text{if } i \neq j, (C_i, C_j) \in \mathscr{R} \\ 0, & \text{if } i \neq j, (C_i, C_j) \notin \mathscr{R} \\ -\sum\limits_{(C_i, C_k) \in \mathscr{R}} \kappa_{ki}, & \text{if } i = j \end{cases}$$

If this system admits $\mathbf{A}\boldsymbol{\psi}(\bar{\mathsf{x}}) = \mathbf{0}$, then the system is complex balanced, with equilibrium point $\bar{\mathsf{x}} \in \mathbb{R}^{N_A}$. Following Theorem (4.2) of [2], we can use the equilibrium point $\bar{\mathsf{x}}$ to define the stationary distribution $\bar{\boldsymbol{\pi}}_{\mathscr{D}}(\mathbf{x}) = \lim_{t\to\infty} \mathbb{P}[f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}), t)]$ of the stochastically modeled system. If the system is irreducible, this stationary distribution consists of a product of Poisson distributions and is given by

$$\bar{\boldsymbol{\pi}}_{\mathscr{D}}(\mathbf{x}) = \prod_{i=1}^{N_A} \frac{\bar{\mathsf{x}}_i^{x_i}}{x_i!} e^{-\bar{\mathsf{x}}_i}, \ \mathbf{x} \in \mathscr{X}_{\mathscr{N}}(\mathscr{D}) \tag{5}$$

where $\mathscr{X}_{\mathscr{N}}(\mathscr{D})$ is the set of all possible population vectors $\mathbf{x}$ that can arise from the CRN $\mathscr{N}$ and the robot species specified by $\mathscr{D}$. We note that when the system is reducible, a similar equation exists, see [2].

[4] The symbol $\mathbf{x}$ denotes the discrete state, whereas x denotes the average population.

4.2 Privacy

If the swarm's CRN model is complex-balanced, we are able to derive a stationary probability density function describing the steady-state of the system. We use this description to formulate a closed-form measure of the loss of privacy.

Proposition 1. *If a swarm's CRN is complex-balanced and irreducible, then the leakage at the steady-state of the corresponding swarm system defined by $\mathscr{D}$, and observed through the identity query $q_{\mathscr{N}}(\mathbf{x}) = \mathbf{x}$ is*

$$\mathscr{L}(\mathscr{D}) = \max_{\substack{\mathscr{D}' \in \text{Adj}(\mathscr{D}) \\ \mathbf{x} \in \mathscr{X}_{\mathscr{N}}(\mathscr{D}) \cup \mathscr{X}_{\mathscr{N}}(\mathscr{D}')}} \left| \sum_{i=1}^{N_A} x_i \ln \frac{\bar{\mathrm{x}}_i}{\bar{\mathrm{x}}'_i} - \bar{\mathrm{x}}_i + \bar{\mathrm{x}}'_i \right| \tag{6}$$

where $\bar{\mathrm{x}}$ and $\bar{\mathrm{x}}'$ are the steady-states resulting from $\mathscr{D}$ and $\mathscr{D}$'.

Proof. Starting with Eq. (3), and using query $q_{\mathscr{N}}(\mathbf{x}) = \mathbf{x}$, we have

$$\mathscr{L}(\mathscr{D}) = \max_{\mathscr{D}' \in \text{Adj}(\mathscr{D})} \left| \ln \frac{\mathbb{P}[f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}), \tau)]}{\mathbb{P}[f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}'), \tau)]} \right|. \tag{7}$$

At steady-state we have $\lim_{\tau \to \infty} \mathbb{P}[f_{\mathscr{N}}(\mathbf{x_0}(\mathscr{D}), \tau)] = \bar{\pi}_{\mathscr{D}}(\mathbf{x})$, hence

$$\mathscr{L}(\mathscr{D}) = \max_{\substack{\mathscr{D}' \in \text{Adj}(\mathscr{D}) \\ \mathbf{x} \in \mathscr{X}_{\mathscr{N}}(\mathscr{D}) \cup \mathscr{X}_{\mathscr{N}}(\mathscr{D}')}} \left| \ln \frac{\bar{\pi}_{\mathscr{D}}(\mathbf{x})}{\bar{\pi}_{\mathscr{D}'}(\mathbf{x})} \right|. \tag{8}$$

Continuing with Eq. (5) we get

$$\mathscr{L}(\mathscr{D}) = \max_{\substack{\mathscr{D}' \in \text{Adj}(\mathscr{D}) \\ \mathbf{x} \in \mathscr{X}_{\mathscr{N}}(\mathscr{D}) \cup \mathscr{X}_{\mathscr{N}}(\mathscr{D}')}} \left| \ln \left(\prod_{i=1}^{N_A} \frac{\bar{\mathrm{x}}_i^{x_i}}{x_i!} e^{-\bar{\mathrm{x}}_i} \right) - \ln \left(\prod_{i=1}^{N_A} \frac{\bar{\mathrm{x}}_i'^{x_i}}{x_i!} e^{-\bar{\mathrm{x}}'_i} \right) \right|, \tag{9}$$

which yields Eq. (6). □

Corollary 1. *If a swarm's CRN is complex-balanced and irreducible, then the leakage at steady-state of the corresponding swarm system defined by $\mathscr{D}$, and observed through the query $q_{\mathscr{N}}(\mathbf{x}) = \mathbf{y}$, with $\mathbf{y} \in \mathbb{N}_{\geq 0}^{N_O}$ and each y_i of the form $\sum_{j \in \Omega_i} x_j$, with $\Omega_i \subset \{1, \ldots, N_A\}$, $\cup_{i \in \{1,\ldots,N_O\}} \Omega_i = \{1, \ldots, N_A\}$, and all Ω_i disjoint is*

$$\mathscr{L}(\mathscr{D}) = \max_{\substack{\mathscr{D}' \in \text{Adj}(\mathscr{D}) \\ \mathbf{y}}} \left| \ln \frac{\sum_{\{\mathbf{x} | \mathbf{y} = q_{\mathscr{N}}(\mathbf{x}) \wedge \mathbf{x} \in \mathscr{X}_{\mathscr{N}}(\mathscr{D})\}} \bar{\pi}_{\mathscr{D}}(\mathbf{x})}{\sum_{\{\mathbf{x} | \mathbf{y} = q_{\mathscr{N}}(\mathbf{x}) \wedge \mathbf{x} \in \mathscr{X}_{\mathscr{N}}(\mathscr{D}')\}} \bar{\pi}_{\mathscr{D}'}(\mathbf{x})} \right|. \tag{10}$$

This formulation, even though less compact than Eq. (6) above, still provides a fast means of computing the leakage for complex-balanced swarms — in particular, the alternative to using this formulation is to compute the PMF via the Chemical Master

Equation [19], which, in our experience, is at least one order of magnitude slower. Moreover, we note that Eq. (6) is linear in $\mathbf{x}$, and can, thus, be solved by integer linear programming (ILP) methods. In summary, the analytical formulation for the privacy of complex-balanced swarms allows us to compute the leakage efficiently. In the remainder of this work, we demonstrate the benefit of this formulation with case studies on aggregation.

5 Aggregation

To apply Corollary 1, we must first show that aggregation is a complex-balanced mechanism. In order to develop our proof, we represent the topology of the reaction networks through directed acyclic graphs (DAG) (see the example in Fig. 4).

Definition 2 (Aggregation-DAG). An aggregation-DAG is a topological representation of a CRN that defines an aggregation mechanism. It is a directed acyclic graph, such that each node forms a complex C_k, the sum of its in-neighbors form complex C_j, and its incoming edges form the reactions $C_j \underset{r_n}{\overset{r_m}{\rightleftharpoons}} C_k$, with r_m, $r_n > 0$. Furthermore, a node has either 0 or > 1 in-neighbors.

Proposition 2. *The aggregation of a heterogeneous swarm of robots described by a CRN is a complex-balanced mechanism if and only if the underlying CRN can be represented by an aggregation-DAG.*

Proof. According to Theorem 4.1 of [21], a CRN is complex-balanced if *(i)* it is weakly reversible and *(ii)* it has deficiency zero. Condition *(i)* requires all complexes to be connected via some reaction pathway (cf. Def. 2.2 in [2]). If all aggregates can be decomposed as well as composed, this is trivially satisfied. The deficiency of a reaction network is $\delta = N_C - L - \text{rank}(\Gamma)$, which is the number of complexes minus the number of linkage classes, each of which is a set of complexes connected by reactions, minus the network rank, which is the rank of the stoichiometry matrix Γ. Hence, we will show that $N_C = L + \text{rank}(\Gamma)$. From Def. 2 it follows that the number of linkage classes L is equal to the number of parent nodes, and the number of complexes N_C is equal to twice the number of parent nodes. Thus, $N_C = 2L$, and it remains to be shown that $\text{rank}(\Gamma) = L$. Matrix Γ is of size $N_A \times N_R$, with $N_R = 2L$ (the network is weakly reversible). Since each new linkage class includes a new parent node (i.e., aggregate state), there are L linearly independent columns in Γ, and, hence, $\text{rank}(\Gamma) = L$. □

5.1 Example

We consider the example shown in Fig. 2. The system is composed of three species, $\mathscr{S} = \{1,2,3\}$. Aggregates are formed with one robot per species, and with species 1

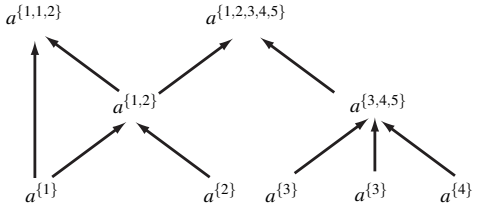

Fig. 4 Example of an aggregation mechanism that is represented as a directed acyclic graph. There are $L = 4$ linkage classes, $N_C = 8$ complexes, and $\text{rank}(\Gamma) = 4$.

and 2 aggregating prior to species 3. This behavior is formalized with the following reactions:

$$a^{\{1\}} + a^{\{2\}} \underset{r_2}{\overset{r_1}{\rightleftharpoons}} a^{\{1,2\}}, \qquad a^{\{1,2\}} + a^{\{3\}} \underset{r_4}{\overset{r_3}{\rightleftharpoons}} a^{\{1,2,3\}} \tag{11}$$

The states of this system are $\mathscr{A} = \{a^{\{1\}}, a^{\{2\}}, a^{\{3\}}, a^{\{1,2\}}, a^{\{1,2,3\}}\}$. Our population vector keeps track of the number of robots per state, and is written

$$\mathbf{x} = [x^{\{1\}}, x^{\{2\}}, x^{\{3\}}, x^{\{1,2\}}, x^{\{1,2,3\}}] \tag{12}$$

We consider an adversary who is able to observe the number of non-aggregated robots, the numer of 2-aggregates, and the number of 3-aggregates. Hence, we formulate the observable data as $\mathbf{y} = [y_1, y_2, y_3]$ with

$$y_1 = x^{\{1\}} + x^{\{2\}} + x^{\{3\}},\ y_2 = x^{\{1,2\}},\ y_3 = x^{\{1,2,3\}}. \tag{13}$$

5.1.1 Analysis

We compute the steady-state $\bar{\mathrm{x}}$ by solving the deterministic system $\mathbf{M}\mathbf{A}_\kappa \boldsymbol{\psi}(\bar{\mathrm{x}}) = \mathbf{0}$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -\kappa_1 & 0 & \kappa_2 & 0 \\ 0 & -\kappa_3 & 0 & \kappa_4 \\ \kappa_1 & 0 & -\kappa_2 & 0 \\ 0 & \kappa_3 & 0 & -\kappa_4 \end{bmatrix} \cdot \begin{bmatrix} \bar{\mathrm{x}}^{\{1\}}\ \bar{\mathrm{x}}^{\{2\}} \\ \bar{\mathrm{x}}^{\{1,2\}}\ \bar{\mathrm{x}}^{\{3\}} \\ \bar{\mathrm{x}}^{\{1,2\}} \\ \bar{\mathrm{x}}^{\{1,2,3\}} \end{bmatrix} = \mathbf{0}. \tag{14}$$

Since the number of robots per species is constant, we have

$$\begin{aligned} \bar{\mathrm{x}}^{\{1\}} + \bar{\mathrm{x}}^{\{1,2\}} + \bar{\mathrm{x}}^{\{1,2,3\}} &= N^{(1)} \\ \bar{\mathrm{x}}^{\{2\}} + \bar{\mathrm{x}}^{\{1,2\}} + \bar{\mathrm{x}}^{\{1,2,3\}} &= N^{(2)} \\ \bar{\mathrm{x}}^{\{3\}} + \bar{\mathrm{x}}^{\{1,2,3\}} &= N^{(3)}. \end{aligned} \tag{15}$$

The equations above can be plugged into Eq. (14) to give a quartic equation:

$$0 = \kappa_1 \left(N^{(1)} - N^{(3)} - \bar{\mathrm{x}}^{\{1,2\}} + \frac{\kappa_4 N^{(3)}}{\kappa_3 \bar{\mathrm{x}}^{\{1,2\}} + \kappa_4} \right) \cdot \left(N^{(2)} - N^{(3)} - \bar{\mathrm{x}}^{\{1,2\}} + \frac{\kappa_4 N^{(3)}}{\kappa_3 \bar{\mathrm{x}}^{\{1,2\}} + \kappa_4} \right) - \kappa_2 \bar{\mathrm{x}}^{\{1,2\}} \tag{16}$$

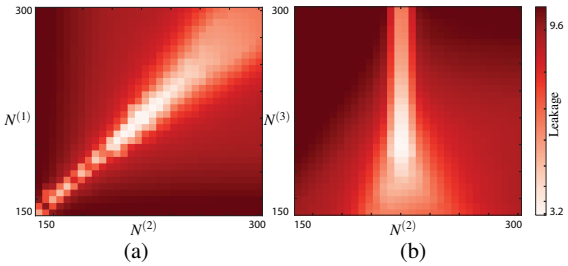

Fig. 5 Leakage for varying populations in the range [150, 300], while keeping the third species fixed, and with fixed reaction rates $\boldsymbol{\kappa} = 1$. In (a) $N^{(1)}$ and $N^{(2)}$ vary, with $N^{(3)}$ =200, and in (b) $N^{(2)}$ and $N^{(3)}$ vary, with $N^{(1)}$ =220.

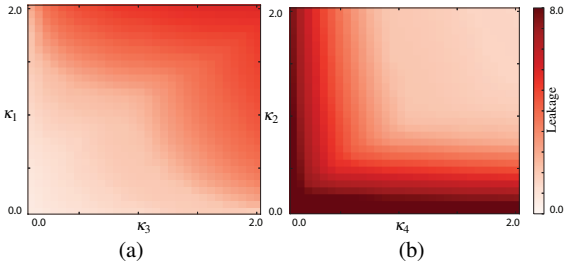

Fig. 6 Leakage for varying aggregation rates with robot populations fixed at $N^{(1)} = N^{(2)} = 220$, $N^{(3)} = 200$. In (a) we vary the rates κ_1, κ_3 at which aggregates form, while fixing $\kappa_2 = \kappa_4 = 1$. In (b) we show the reverse (varying the rates at which aggregates decompose).

which only depends on variable $\bar{\mathrm{x}}^{\{1,2\}}$. The solution to the remaining variables $\bar{\mathrm{x}}^{\{1\}}$, $\bar{\mathrm{x}}^{\{2\}}$, $\bar{\mathrm{x}}^{\{3\}}$, $\bar{\mathrm{x}}^{\{1,2,3\}}$ of this system can be found by substitution. Of the four possible solutions, there is only a single all-positive solution (which corresponds to the single equilibrium of the complex-balanced system). Finally, we can compute the leakage $\mathscr{L}(\mathscr{D})$ of the observed system according to Corollary 1.

5.1.2 Evaluation

The observable state is a function of the system-level state, and is defined by the number of robots per species $N^{(s)}$, and by reaction rates $\boldsymbol{\kappa}$. Hence, we vary these values to identify their relation to privacy. We note that this relationship is made mathematically evident in Eq. (16). We compute the leakage for two settings, shown in Fig. 5 and Fig. 6. In the first setting, we fix the reaction rates and vary the robot populations of two species, while keeping the third species fixed. In Fig. 5(a), we observe a clear "valley" of minimum leakage values for an equal number of species 1 and 2. The overall minimum is at $N^{(1)} = N^{(2)} = 220, N^{(3)} = 200$. The plot also reveals that increasing the total number of robots increases privacy, as shown by the expansion of the valley in the upper right corner. Panel Fig. 5(b) shows the resulting leakage for varying species 2 and 3. We observe a sharp drop in privacy as the population $N^{(2)}$ deviates from $N^{(1)}$, with a minimum at $N^{(1)} = N^{(2)} = 220, N^{(3)} = 200$ (as previously observed in Fig. 5(a)). We conclude that species that are topologically equivalent (species 1 and 2, as visible in Fig. 2) should have a balanced number of robots for increased privacy. In the second setting (Fig. 6), we vary the reaction rates

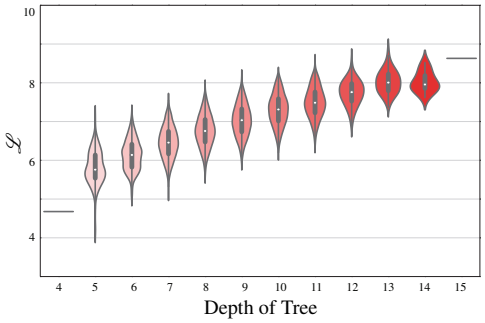

Fig. 7 Leakage for binary aggregation trees of varying topology (with 16 leaves). Trees of same depth are assembled by one violin plot that features a kernel density estimation of the underlying distribution. The Pearson correlation coefficient evaluated on this data is 0.74.

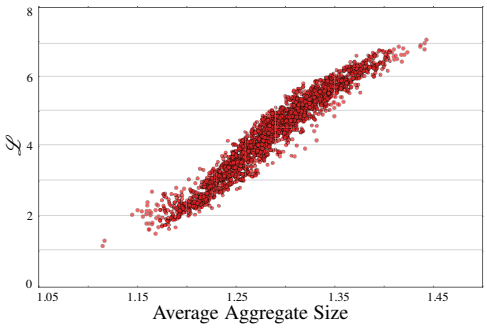

Fig. 8 Leakage for 2000 trees with identical topology (symmetric binary tree, with 8 leaves), and with all 7 aggregation rates varied uniformly and randomly in the range $[0.1, 2]$ (decomposition rates are held constant, equal to 1). The Pearson correlation coefficient evaluated on this data is 0.97.

while keeping the robot populations fixed. Figures 5(b) and 6(a) indicate that if we increase the probability of reaching larger sized aggregates, either by increasing the number $N^{(3)}$, or by increasing the aggregation rates, we decrease privacy. Indeed, in this setting, the 2-aggregates and 3-aggregates are unique, hence, they expose more information about the system.

5.2 Evaluating the Impact of Topology and Parameters

The results of the preceding example indicate that both the topology and parameters of the CRN affect privacy. To expose the impact of a CRN's topology on the leakage, we proceed by considering aggregation-DAGs that can be represented by binary trees. For a system composed of 16 species, we evaluate the leakage for each of the possible 10905 unlabeled binary rooted trees with 16 leaves (which corresponds to the Wedderburn-Etherington number). Fig. 7 shows the leakage as a function of the depth of the tree. We see a clear correlation between irregular, unbalanced topologies (with greater depth) and high leakage values, and between more balanced, symmetric topologies (with smaller depth) and low leakage values.

Next, to expose the impact of a CRN's parameters (i.e., aggregation rates) on the leakage, we proceed by considering a symmetric binary tree with 8 leaves (of depth 3), and we vary the aggregation rates uniformly and randomly in the range $[0.1, 2]$,

gathering 2000 datapoints. Fig 8 shows the leakage as a function of the average aggregate size (at steady-state) for each set of rates. We see that as the average size increases, so does the leakage.

These results together indicate that privacy can be increased by *(i)* designing aggregation mechanisms that are balanced (asymmetric aggregation mechanisms create more unique aggregates, and hence, reveal more information about the system), or by *(ii)* throttling the aggregation of larger aggregates (which tend to be more unique).

6 Conclusion

In this work, we showed how to analyze the privacy of aggregation mechanisms in heterogeneous robot swarms. Our main contribution consists of a closed-form expression that quantifies the leakage of dynamic swarms that can be modeled as complex-balanced reaction networks. We demonstrated that aggregation mechanisms are complex-balanced, and hence, were able to use our formula to efficiently evaluate various settings. The reported results showed how privacy levels vary, as the topology as well as the parameters of the aggregation mechanism are varied. This means that we are able to identify the relation between privacy loss at a macroscopic level, and swarm design parameters (such as reaction rates, reaction topology, and swarm composition). As a consequence, our framework paves the way for methods that control the swarm, while guaranteeing bounds on privacy loss. We intend to further this line of work by developing active privacy mechanisms that are able control the loss of privacy, while maintaining the overall performance of the underlying swarm system.

Acknowledgements The authors would like to thank the anonymous referees for their constructive feedback. We gratefully acknowledge the support of ONR grants N00014-15-1-2115 and N00014-14-1-0510, ARL grant W911NF-08-2-0004, NSF grant IIS-1426840, and TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

References

[1] Agrawal R, Srikant R (2000) Privacy-preserving data mining. ACM Sigmod Record 29(2):439–450

[2] Anderson DF, Craciun G, Kurtz TG (2010) Product-form stationary distributions for deficiency zero chemical reaction networks. Bulletin of Mathematical Biology pp 1–23

[3] Cheng J, Cheng W, Nagpal R (2005) Robust and self-repairing formation control for swarms of mobile agents. AAAI

[4] Correll N, Martinoli A (2007) Modeling Self-Organized Aggregation in a Swarm of Miniature Robots. IEEE International Conference Robotics and Automation (ICRA)

[5] Dorigo M, Floreano D, Gambardella LM, Mondada F, Nolfi S, et al (2013) Swarmanoid: A Novel Concept for the Study of Heterogeneous Robotic Swarms. IEEE Robotics & Automation Magazine 20(4):60–71
[6] Dwork C (2008) Differential Privacy: A Survey of Results. In: Theory and Applications of Models of Computation, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–19
[7] Dwork C (2011) Differential Privacy. Encyclopedia of Cryptography and Security pp 338–340
[8] Evfimievski A, Gehrke J, Srikant R (2003) Limiting privacy breaches in privacy preserving data mining. In: the twenty-second ACM SIGMOD-SIGACT-SIGART symposium, ACM Press, New York, New York, USA, pp 211–222
[9] Feinberg M (1991) Some Recent Results in Chemical Reaction Network Theory. In: Patterns and Dynamics in Reactive Media, Springer New York, New York, NY, pp 43–70
[10] Garnier S, Jost C, Gautrais J, Asadpour M, Caprari G, Jeanson R, Grimal A, Theraulaz G (2008) The Embodiment of Cockroach Aggregation Behavior in a Group of Micro-robots. Artificial Life 14(4):387–408
[11] Groß R, Dorigo M (2008) Self-Assembly at the Macroscopic Scale. Proceedings of the IEEE 96(9):1490–1508
[12] Halloy J, Sempo G, Caprari G, Rivault C, Asadpour M, Tache F, Said I, Durier V, Canonge S, Ame JM, Detrain C, Correll N, Martinoli A, Mondada F, Siegwart R, Deneubourg JL (2007) Social Integration of Robots into Groups of Cockroaches to Control Self-Organized Choices. Science 318(5853):1155–1158
[13] Hosokawa K, Shimoyama I, Miura H (2010) Dynamics of Self-Assembling Systems: Analogy with Chemical Kinetics. dxdoiorg 1(4):413–427
[14] Kasiviswanathan SP, Smith A (2008) A note on differential privacy: Defining resistance to arbitrary side information. CoRR abs
[15] Klavins E, Burden S, Napp N (2006) Optimal rules for programmed stochastic self-assembly. Robotics: Science and Systems
[16] Martinoli A, Ijspeert AJ, Gambardella LM (1999) A Probabilistic Model for Understanding and Comparing Collective Aggregation Mechanisms. In: Advances in Artificial Life, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 575–584
[17] Matthey L, Berman S, Kumar V (2009) Stochastic strategies for a swarm robotic assembly system. In: IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp 1953–1958
[18] Parrish JK, Edelstein-Keshet L (1999) Complexity, Pattern, and Evolutionary Trade-Offs in Animal Aggregation. Science 284(5411):99–101
[19] Prorok A, Kumar V (2016) A Macroscopic Privacy Model for Heterogeneous Robot Swarms. In: International Conference on Swarm Intelligence
[20] Prorok A, Hsieh AM, Kumar V (2016) Formalizing the Impact of Diversity on Performance in a Heterogeneous Swarm of Robots. In: IEEE International Conference on Robotics and Automation (ICRA)
[21] Siegel D, MacLean D (2000) Global stability of complex balanced mechanisms. J of Mathematical Chemistry 27:89–110