Designing minimal effective normative systems with the help of lightweight formal methods

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1145/2950290.2950307">http://dx.doi.org/10.1145/2950290.2950307</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Association for Computing Machinery</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Feb 09 01:04:43 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/116146">http://hdl.handle.net/1721.1/116146</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
ABSTRACT

Normative systems are an important approach to achieving effective coordination among (often an arbitrary number of) agents in multiagent systems. A normative system should be effective in ensuring the satisfaction of a desirable system property, and minimal (i.e., not containing norms that unnecessarily over-constrain the behaviors of agents). Designing or even automatically synthesizing minimal effective normative systems is highly non-trivial. Previous attempts on synthesizing such systems through simulations often fail to generate normative systems which are both minimal and effective. In this work, we propose a framework that facilitates designing of minimal effective normative systems using lightweight formal methods. Given a minimal effective normative system which coordinates many agents must be minimal and effective for a small number of agents, we start with automatically synthesizing one such system with a few agents using Alloy Analyzer. We then increase the number of agents so as to check whether the same design remains minimal and effective. If it is, we manually establish an induction proof so as to lift the design to an arbitrary number of agents. We show the effectiveness of the framework by using it to design road junction traffic rules and population protocols. The protocols designed using our framework are shown to be as good as those manually designed and published ones.

1. INTRODUCTION

Normative systems have garnered much attention in the multiagent systems (MAS) literature as an effective technique for regulating the behaviors of agents [31, 16, 25, 14]. Roughly, a normative system is a set of rules or norms imposed on an MAS to ensure that a desirable global property is satisfied. Each norm constrains the behaviors of agents by forbidding them from performing one or more actions under certain circumstances. For example, to avoid collision, two autonomous trains traveling through a common tunnel may implement a norm of “if another train is observed inside the tunnel, the action of move should not be allowed”. This paradigm of normative system in MAS naturally maps to the law system implemented in our human system. The normative system paradigm was first proposed by Shoham and Tennenholtz [31, 32], and the problem of norm synthesis has been a major research topic since then. Van der Hoek et al. [16] identified three major computational problems related to normative systems: effectiveness, feasibility and synthesis. To avoid over-regulate the behaviors of agents, Morales et al. [25, 26] later proposed that a desirable normative system should be both effective and necessary.

Designing or synthesizing normative systems which are both effective and necessary are highly non-trivial. For instance, Van der Hoek et al. [16] proposed to solve the computational problems related to normative systems, i.e., effectiveness, feasibility and synthesis, in the context of an Alternating-time Temporal Logic (ATL) [4] model checking problem. Their framework may, however, synthesize unnecessary norms that over-regulate the behaviors of agents. One particular challenge in relying on model checking is that normative systems are often designed for a large or arbitrary number of agents, which often result in state-space explosion. Morales et al. [25, 26] developed a mechanism to automatically synthesize normative systems that are “effective and necessary”. However, effectiveness and necessity are determined in their scheme using simulations, and there is thus no guarantee that the normative system synthesized is effective or necessary.

In this work, we show that the problem can be solved, to certain extent, with techniques developed in the software engineering community. In particular, we propose a framework which aims to facilitate designing minimal effective normative systems based on lightweight formal methods [18]. One simple observation is that if a normative system is minimal effective for a large or arbitrary number of agents, it must be so for a small number of agents as well. This observation matches the underlying principle of lightweight formal methods [18] and automated deduction philosophy [36] in software engineering. Based on the observation, we propose a framework which starts with automatically searching for candidate effective and necessary normative systems for an MAS with as few as 2 agents. Next, we check if the candidate normative systems (or any specialization of them) remain effective and necessary with an increasing number of agents. We keep eliminating candidates until the candidates stabilize. The effectiveness of a candidate normative system is automatically checked using an analysis engine (e.g., Alloy Analyzer [19] or other model checkers) that exhaustively explores all possible behaviors of the MAS. Because we work with a small number of agents, scalability is less an issue in our framework. Once a minimal effective normative system is identified and withstands with an increasing number of agents, we then manually show, often with an induction proof, that the same normative system can be applied to an arbitrary number of agents and remain minimal effective in satisfying the property.

To demonstrate that we can design non-trivial normative systems this way, we apply our framework to a road junction example which
is a typical example (which is often the only example [31, 32, 25, 26]) used to showcase normative system synthesis methods in MAS literature. In addition, we apply our framework to design a number of population protocols, which are a group of self-regulating network protocols designed for agent-like network nodes. We show that the protocols we synthesis are as good as those manually designed and published ones [7, 22, 6]. The remainder of the paper is organized as follows. In Section 2, we present an overview of existing works on norm synthesis proposed in multiagent research community. In Section 3, the background and problem definition are described. Our norm synthesis framework is described in Section 4. In Section 5, we evaluate the effectiveness of our framework using three representative case studies. Conclusion and future work are presented in Section 6.

2. RELATED WORK

Shoham and Tennenholtz [31, 32] first defined the problem of synthesizing useful social laws in multi-agent systems, and investigated the required properties of the synthesis process and its computational complexity. They argued that a social law should not restrict too much of the individual agents’ freedom, and also restrictive enough such that the agents can cooperate towards desirable system goals. Specifically the authors defined the notion of useful social laws, which guarantee that each agent can come up with a plan ensuring him to move from any one focal state to another focal state. They showed that deriving useful social laws in general is NP-complete, and identified conditions under which the problem of deriving useful social laws can be solved in a polynomial time.

Later this idea was generalized in [16]. It was shown that the objective of social laws can be expressed as an Alternating-time Temporal Logic (ATL) [5] formula, and the social law problem is transformed to an ATL model-checking problem. In their framework, the system is modeled as an Action-based Alternating Transition System (AATS) and a social law consists of two components: the objective of the system and the behavioral constraints. The authors identified three major computational problems related to social laws, i.e., effectiveness, feasibility and synthesis, and investigated the corresponding computational complexities. The authors showed that the complexity of the feasibility problem is no more complex than the complexity of the corresponding problem defined by Shoham and Tennenholtz [31], which is NP-complete. Christelis and Rovatsos [9] proposed a novel norm synthesis approach using traditional AI planning techniques to synthesize effective norms in the planning domains. However, no notion of optimality of a social law is considered in both of the work and an effective social law may over regulate the behaviors of the agents.

The problem of synthesizing effective social laws was further refined in [2] which takes into consideration the fact that implementing social laws may have cost, and that the system designer may have multiple goals of varying value. In their framework, the system is modeled using Kripke structures and the desirable objectives are expressed using Computational Tree Logic (CTL). Social laws are modeled as restrictions on Kripke structures, and the costs on edges are used to model the cost of implementing different norms. The optimality of a social law is defined as the benefit of the desirable objective brought by this law minus the cost of implementing it. The optimal social law is the one that owns the highest utility. The authors solved the problem of designing an optimal social law by formulating it as an integer linear program and also investigated possible ways of reducing the computational complexity. A common problem of the above approaches are that due to the state space explosion problem which is often observed when model checking techniques are employed, these methods are often limited to synthesize MAS for a small number of agents.

Morales et al. [25, 26] proposed that a normative system should not only be effective in guaranteeing the global property, but also avoid over-regulating the behaviors of the agents. They developed a simulation-based mechanism to synthesize both effective and necessary normative systems, and applied their mechanism to the road junction example to illustrate its effectiveness. However, their approach does not provide a theoretical guarantee that the synthesized normative system is effective and minimal. Furthermore, their approach may result in non-effective normative systems that cannot be detected through extensive simulation. In Section 5, we provide a more detailed comparison of our approach against theirs using the same benchmark system.

The work in [1] is remotely related to ours. Ágoston et al. [1] proposed the notion of conservative social laws, which are those effective social laws making minimal degree of changes to the original system based on the criterion of the distance metric. The concept of conservative social laws is to model the least change principle inspired from social laws in human society, where the laws with minimal change effect on people’s habits would be easily accepted by the public. While this idea of minimal social laws has been explored previously, most efforts have been focused on exploring the theoretical boundaries of this notion and not on designing practical algorithms for synthesizing minimal norms.

Compared to the above work, our approach is complementary as we not only formally define minimal effective normative systems, but also provide a practical algorithm based on lightweight formal methods to automatically synthesize them. Wamberto et al. [35] proposed an approach using first-order constraint solving techniques for norm conflicts detection and resolution in a normative system, however, their goal is simply to resolve conflicts among norms, i.e., there is no global properties to be satisfied or the notion of minimality and effectiveness of norms.

This work is also related to work on using formal method based synthesis techniques to solve goal operationalisation problem in requirement engineering [23, 3]. In [15], Degiovanni et al. proposed an approach using interpolation and SAT solving techniques to automatically computes required preconditions and required triggering conditions for operations such that the resulting operations establish the (safety) goals. Their idea of iteratively refining operational specifications based on the counterexamples from the model checker is similar to our iterative approach of synthesizing norms from small-size MAS. However, our work is different from theirs as we solve a different problem, e.g., we do not consider liveness properties whilst we do consider the minimality requirement of norms. As far as the authors know, this work is the first on automatically synthesizing norms using formal methods, complementary to the state-of-the-art simulation-based approaches [26].

3. PROBLEM DEFINITION

In this section, We define our problem formally.

Normative Systems A multi-agent system (MAS) is composed of a set of interacting agents, where each agent \(i \in G\) can perform actions from a finite, non-empty set \(A_i\) of actions. Formally, a multi-agent system \(\mathcal{M}\) is a tuple \(\langle S, G, \{A_i\}, R, \Psi, \pi \rangle\), where \(S\) is a finite set of global states; \(G\) is a finite set of agents; \(A_i\) for each \(i \in G\) is a finite set of actions of agent \(i\); \(\Psi\) is a finite set of atomic propositions; and

- \(R : S \times J_G \rightarrow S\) is a partial system transition function, which defines how the system state changes given the set of actions that the agents choose to perform. Here, \(J_G\) is the set
of joint actions of the set $G$ of agents, i.e., $J_G = \Pi_{i \in G} A_i$.

- $\pi : S \to 2^P$ is an interpretation function, which maps each system state to the set of primitive propositions that hold in that state. That is, $p \in \pi(q)$ where $q \in S$ denotes that the atomic proposition $p$ is true in state $q$.

In this work, we assume that the agents share the same set of actions, i.e., $A_i = A_j$ for any agent $i \in G$. We do not distinguish the identities of agents and thus the system can be considered as being agent-symmetric. Theoretically, this assumption can be lifted.

**Example** We use a traffic cross system as a running example to explain each step of our approach. As shown in Figure 1(a), a traffic network is composed of two orthogonal, intersecting roads. Each road has two 7-cell lanes, and the direction of each lane is indicated by the arrows in the figure. Each car travels in the indicated direction, and stays within the lane. However, once it arrives at the junction, it is free to switch to any one of the four lanes. We assume that cars have the same speed, and can perform one of two possible actions; move one cell in the indicated direction or stay in the same cell (i.e., $A = \{\text{Move}, \text{Stop}\}$). At any given state, a car is described to be in one of four orientations, depending on the direction that is headed (east, west, north, and south). We assume that each car can observe the orientation of cars in three of its neighboring cells, as shown in Figure 1(b): top-left, top, and top-right cells.

Given an MAS $\mathcal{M}$, we can define a set of desirable states in $C$ as a property $\omega$, which can be represented as a propositional logic formula constituted by propositions in $\Psi$. For example, in the traffic cross system, one desirable property is the non-collision property: the system never enters a state in which one or more cars collide. One way to make sure that the system remains in a desirable state is to regulate the actions of the agents using a set of norms. A norm is a constraint on the actions that an agent is not allowed to perform under certain condition. Formally,

**Definition 1.** A norm is a tuple $\langle \psi, A' \rangle$, where $\psi$ is the precondition of the norm, and $A' \subseteq A$ is the subset of actions that an agent is forbidden to perform when $\psi$ is satisfied.

A norm is enforced on an agent only under certain circumstances, which are characterized by the precondition of the norm. In practice, each individual agent may not observe the global state of the system due to physical or communication constraints. Thus $\psi$ should not refer to the global state of the system. Instead, we assume that each agent $g$ can observe the states of agents within its neighborhood. The neighborhood of an agent is problem-specific and the neighborhood of a car agent in the running example is illustrated in Figure 1(b). We define the local view $\rho_g(s)$ of a global state $s$ from the perspective of agent $g$ as the set of predicates that are evaluated over the neighboring agents that agent $g$ can observe. We assume that the local view of an agent is consistent with the global state of the system. Thus, the norm $n = \langle \psi, A' \rangle$ is enforced on an agent $g$ whenever the agent’s local view satisfies the precondition $\psi$, which is formally denoted as $\rho_g(s) \models \psi$.

In the following, we present the exact syntax for norms, as shown in Figure 2. Without loss of generality, we restrict the set $A'$ to be a singleton, which is the action that the agent is forbidden to perform under the norm. Note that norms consisting of multiple actions can be naturally represented as multiple norms of single action. The precondition $\psi$ is a conjunction of atomic predicates that describe the characteristics of a single agent or relationship between multiple agents. Every $n$-ary predicate, $p^n \in P$, is applied over $n$ terms, each of which represents a particular agent, an integer value, or a user-defined symbol. In our running example, consider a norm $n = \langle \psi, \{\text{Move}\} \rangle$, where $\psi$ is defined as: $\text{dirNorth}(g) \land \text{topLeft}(h, g) \land \text{dirEast}(h)$ where $g$ and $h$ are variables representing agents. Enforcing this norm on all agents means that any car $g$ is not allowed to move if it is north-heading (i.e., $\text{dirNorth}(g)$ is true), and if there is another car $h$ which is located diagonally left of it (i.e., $\text{topLeft}(h, g)$ is true) and heading towards east (i.e., $\text{dirEast}(h)$ is true).

Following previous work [31, 25], a normative system is defined as a set of norms. Given an MAS $\mathcal{M}$, the system designer may apply a normative system to regulate the behaviors of agents. When a normative system $\Phi$ is applied to an MAS, each norm in $\Phi$ is applied on all agents in $\mathcal{M}$ (a.k.a. all agents are norm-abiding). We write $\mathcal{M} \oplus \Phi$ to denote the system which is the result of implementing the normative system $\Phi$ on $\mathcal{M}$ (i.e., by ruling out all actions forbidden by the normative system). Formally,

**Definition 2.** Let $\mathcal{M} = \langle S, G, \{A_i\}, R, \Psi, \pi \rangle$ be an MAS and $\Phi = \{\langle \psi_k, A'_k \rangle\}$ be a normative system. The implementation of $\Phi$ over $\mathcal{M}$, written as $\mathcal{M} \oplus \Phi$, is a Kripke structure $(S', S_0, R', \Psi, L')$ such that $S' \subseteq S$ is a set of states; $S_0$ is the initial state; $L' : S' \to 2^P$ such that $L'(s) = \pi(s)$ for all $s \in S'$ and $R'$ satisfies the following: $(s, s') \in R' \iff (s, \vec{a}, \vec{a'}) \in R$ where $\vec{a} = \Pi_{i \in G} A_i$, and for all $k$, $\vec{a}_i \not\in A'_k$ for all $i$ if $s$ satisfies $\psi_k$.

The problem of determining whether $\mathcal{M} \oplus \Phi$ satisfies the property $\omega$ can be expressed as a Computation Tree Logic (CTL) [11] satisfiability problem as follows.

$$\mathcal{M} \oplus \Phi \models \Box \omega$$  \hspace{1cm} (1)

where $\mathcal{M}$ means “along all paths” and $\Box$ is the temporal operator meaning “now and forever more”. We skip the formal definition of $\models$ as it is standard [12]. Intuitively, the above is satisfied if the system $\mathcal{M}$ always remains at a desirable state if the agents behave according to the norms. A normative system $\Phi$ (for an MAS $\mathcal{M}$) that satisfies Equation 1 is called effective. For example, in the traffic cross system, given the non-collision property, a normative system is effective if it never enters a state in which cars collide.

**Minimal Effective Normative System** A normative system can be effective but useless. For instance, in the traffic cross system, a trivially effective normative system would be one which disallows car movement altogether. Thus, in this work, we are interested in effective normative systems which are also minimally constraining (hereafter minimal). A normative system that is effective but not minimal over-constrains the behavior of the agents by disallowing innocuous actions, and represents a non-optimal design. We
normative system to be between pairs of normative systems to define what it means for a
specializes another norm

For example, consider a pair of norms

\( n_1 = \langle \psi_1, A_1 \rangle \) and \( n_2 = \langle \psi_2, A_2 \rangle \) be a pair of norms. Norm \( n_1 \) strictly specializes norm \( n_2 \) iff (1) \( \psi_1 \Rightarrow \psi_2 \), and (2) \( A_1 = A_2 \).

For example, consider a pair of norms \( n = \langle \psi, \{\text{Move}\} \rangle \), and \( n' = \langle \psi', \{\text{Move}\} \rangle \) in the traffic cross system, where

- \( \psi = \\text{dirNorth}(g) \land \text{topLeft}(h,g) \land \text{dirEast}(h) \)
- \( \psi' = \\text{dirNorth}(g) \land \text{topLeft}(h,g) \land \text{dirEast}(h) \land \text{top}(h',g) \)

We can see norm \( n' \) strictly specializes \( n \), since \( n' \) is applicable in a more restricted set of circumstances than \( n \) is, namely, \( n' \) requires the presence of a third car \( h' \) front of \( g \). We can similarly define the weak specialization relation.

**Definition 4.** Let \( n_1 = \langle \psi_1, A_1 \rangle \) and \( n_2 = \langle \psi_2, A_2 \rangle \) be a pair of norms. Norm \( n_1 \) weakly specializes norm \( n_2 \) iff the following are satisfied: (1) \( \psi_1 \Rightarrow \psi_2 \), (2) \( A_1 = A_2 \).

If a norm \( n \) strictly specializes another norm \( n' \), we can equivalently say \( n' \) strictly generalizes \( n \). Similarly, if a norm \( n \) weakly specializes another norm \( n' \), we say \( n' \) weakly generalizes \( n \). Next we define the specialization relation between two normative systems based on the specialization relation between norms. We remark that we assume that there are no duplicated norms, i.e., no multiple semantically equivalent norms coexist in the same normative system. Intuitively, if the normative system \( \Phi_1 \) specializes another normative system \( \Phi_2 \), \( \Phi_2 \) puts less constraints on the behaviors of the agents in the system than \( \Phi_2 \).

**Definition 5.** Let \( \Phi_1 \) and \( \Phi_2 \) be two normative systems. \( \Phi_1 \) strictly specializes \( \Phi_2 \) if the following two conditions are satisfied: (1) for each norm \( n_1^* \in \Phi_1 \), there exists a corresponding norm \( n_2^* \in \Phi_2 \) such that norm \( n_1^* \) (weakly) specializes norm \( n_2^* \) (2) strict specialization holds for at least one norm in \( \Phi_1 \) or for at least one norm \( n_1^* \in \Phi_2 \), no norm in \( \Phi_1 \) specializes \( n_1^* \).

We remark that the specialization relation is transitive, i.e., if \( \Phi_1 \) specializes \( \Phi_2 \) and \( \Phi_2 \) specializes \( \Phi_3 \), then \( \Phi_1 \) also specializes \( \Phi_3 \). We can also define generalization relation between normative systems as the inverse of specialization. If \( \Phi_1 \) strictly specializes \( \Phi_2 \), we say that \( \Phi_2 \) is a strict generalization of \( \Phi_1 \) (or we say \( \Phi_2 \) strictly generalizes \( \Phi_1 \)). Finally, we use the strict specialization relation between pairs of normative systems to define what it means for a normative system to be minimal effective.

**Definition 6.** Let \( \mathcal{M} \) be an MAS; \( \Phi \) be a normative system; \( \omega \) be a property. \( \Phi \) is minimal effective with respect to property \( \omega \) iff (1) \( \mathcal{M} \models \Phi \Rightarrow \mathcal{M} \models \omega \) and (2) there does not exist any \( \Phi' \) such that \( \mathcal{M} \models \Phi' \Rightarrow \mathcal{M} \models \omega \) and \( \Phi' \) strictly specializes \( \Phi \).

Intuitively, \( \Phi \) is minimal effective if and only if it is effective and there does not exist any strictly specialized normative system \( \Phi' \) which is effective. Our problem in this work is defined as follows: given an MAS \( \mathcal{M} \) and a property \( \Phi \), how do we systematically synthesize a normative system which is both effective and minimal? For instance, in our running example, intuitively there should be at least two set of equivalent minimal effective normative systems which correspond to the “give way to car coming from left (or right)” principle. We present the details of the minimal effective normative systems for this example in Section 5.1. Note that there may be multiple minimal effective normative systems in general for a given property and our framework can compute all of them.

4. OUR FRAMEWORK

In this section, we present the details of our framework for designing minimal effective normative systems using lightweight formal methods. The particular lightweight formal method we employ in this work is the Alloy modeling language and its associated Alloy Analyzer. In the following, we start with a brief introduction of Alloy. We refer interested readers to [19] for details.

4.1 Alloy

Alloy [20] is a modeling language based on first-order relational logic with transitive closure. It has been used to model and analyze a wide range of systems, including ones from the MAS domain [29, 34, 33]. Alloy Analyzer takes an Alloy model and generates valid system traces or verifies the model against a desired property. Informally speaking, Alloy Analyzer translates an input model into a Boolean constraint by finitizing the size of each domain, according to user-provided bounds, and uses off-the-shelf SAT solvers for automatic model finding, i.e., finding a model within the bound which would satisfy all the constraints in the model. The analysis is exhaustive; it explores every possible configuration (which could be the traces) of the system up to the given bounds, and is guaranteed to find a model (which could be a counterexample), if there is any. Alloy and Alloy Analyzer are designed based on the principle of lightweight formal methods. That is, if a counterexample is present in a complicated system, it is likely that some corresponding counterexample exists in a scaled-down version of the system and thus it is often important and sufficient to focus on finding (and fixing) counterexamples within a relative small bound. We remark this assumption is echoed in our setting. MAS often contain a large or even arbitrary number of agents which are often similar or even identical. A set of social norms which would work with a large number of agents must work for a small number of them and furthermore, as we demonstrate in our case studies later, a stable set of social norms work for a small number of agents often apply to a large number of agents as well. The underlying reason is that agents in normative systems are inherently symmetric.

In this work, we use Alloy to model the normative systems, the candidate norms, as well as the properties, and use Alloy Analyzer to automatically identify minimal effective norms, for a bounded number of agents. It is important to note that our synthesis framework itself does not prescribe the use of a particular language or tool. Any other modeling tool (e.g., NuSMV model checker [10]) could be substituted for Alloy so long as: (1) its language is expressive enough for specifying first-order predicates, (2) it allows an exhaustive analysis of the system against a property, and (3) it
produces a concrete counterexample, which is used for synthesizing and refining norms.

4.2 Norm Synthesis Process

We now outline a process for automatically synthesizing minimal effective normative systems. Given an MAS, without loss of generality, we assume that each agent can observe the states of \( m \) agents (except itself) within its neighborhood. For instance, in the traffic cross system, \( m \) is set to be 3. The number of possible normative systems is exponential in the number of all possible norms, which is exponential in the number of agents constituting the preconditions. To efficiently search this huge space for a minimal effective normative system, we propose an approach in which we first synthesize normative systems involving only the most generalized norms for a minimal number of agents, and then incrementally refine them by taking more agents into consideration. We believe that this strategy is justified as we deliberately searching for simple norms (e.g., an agent would need to observe only a small number of his/her neighbors) first, which in practice often works effectively.

The workflow of the norm synthesis framework is shown in Figure 3. The synthesis process is divided into two phases: (1) generating an initial candidate set of normative systems, and (2) refining them until one of them qualifies as minimal effective. In the following, we present the details of the two phases.

**Initial Norm Synthesis** The top three components in Figure 3 are responsible for synthesizing an initial set of candidate normative systems: Norm Synthesis, Norm Interface and the Alloy Analyzer. The designer starts by modeling the behaviors of an MAS and a desired property \( \omega \) in the Alloy Analyzer\(^3\). The MAS model is parameterized with three inputs: a desired property \( (\omega) \), the number of agents \( (k) \) in the MAS, and a normative system to be enforced (\( \Phi \)). Given these inputs, the analyzer automatically checks whether the MAS satisfies \( \omega \); if the MAS violates \( \omega \), then the analyzer returns a counterexample that shows the sequence of states and actions that leads to the violation.

We start the norm synthesis for MAS involving two agents only \((k = 2)\). Since agents are assumed to be identical, and thus there is no need to distinguish the agents. Since \( \Phi \) is initially empty, the analyzer will generate a counterexample \( c \) that shows how \( \omega \) is violated (unless \( \omega \) is trivially true without norms). The norm interface translates and passes this counterexample into the SynthInitNorms function (Algorithm 1). Note that only the first generated counterexample is considered if there are multiple counterexamples.

\[\text{Algorithm 1 SynthInitNorms}(\omega, \Phi, \Omega, c)\]

1: \( \Omega = \text{ExtractNS}(c) \) \hspace{1cm} // all norms that would prevent \( c \)
2: for \( \Phi_e \in \Omega \) do
3: if \( \Phi_e \) not in \( \Phi \) then
4: \( \Phi = \Phi \cup \Phi_e \)
5: end if
6: \( c' = \text{Verify}(\omega, 2, \Phi) \) \hspace{1cm} // analyze \( \Phi \) against \( \omega \) for 2 agents
7: if \( c' = \emptyset \) then
8: \( \Omega = \Omega \cup \{ \Phi \} \) \hspace{1cm} // no counterexample; \( \omega \) satisfied
9: else
10: \( \Omega = \text{SynthInitNorms}(\omega, \Phi, \Omega, c') \)
11: end if
12: end for
13: return \( \Omega \)

The algorithm works as follows. We begin by extracting from counterexample \( c \) the set of norms that would have prevented \( c \) (line 1). The ExtractNS function is constructed as follows. The precondition is simply the current local view of an agent involved in the counterexample \( c \), and its action to be forbidden is the action that this agent chooses right before the occurrence of \( c \). Note that it might return multiple norms since multiple agents are involved in the counterexample. The current normative system \( \Phi \) is expanded to include each possible \( \Phi_e \in \Phi \) (line 3), and is analyzed for its effectiveness by re-running the analyzer (Verify on line 6). If the analyzer fails to find a counterexample, \( \Phi \) must be effective for all possible scenarios in 2-agent system, and thus is included in \( \Omega \) to pass onto the refinement phase (line 8). If \( \Phi \) leads to another counterexample, \( c' \), then we repeat the process (line 10). When the algorithm completes, the resulting \( \Omega \) contains the set of all feasible normative systems that always guarantee property \( \omega \) for the MAS with two agents (satisfying Formula 1). In the traffic cross system, suppose the first collision counterexample returned by Alloy Analyzer is that car \( g \) heads North while car \( h \) is heading East and located at the TopLeft position of car \( g \), this counterexample can be mapped to either of the following two norms:

- \( \{ \psi_1, \{ \text{Move} \} \} \), where \( \psi_1 = \text{dirNorth}(g) \land \text{topLeft}(h, g) \land \text{dirEast}(h) \)
- \( \{ \psi_2, \{ \text{Move} \} \} \), where \( \psi_2 = \text{dirEast}(g) \land \text{topRight}(h, g) \land \text{dirNorth}(h) \)

After that, we enforce either of them on all agents’ behaviors and rerun Alloy to generate possible counterexamples separately. This procedure is repeated until no new counterexample is identified. The set of normative systems returned after the initial norm synthesis step are passed to the next phase: norm refinement phase.

**Norm Refinement** The normative systems in \( \Omega \) are composed of the most general and effective norms for the 2-agent MAS, i.e., the precondition of each norm in the normative systems in \( \Omega \) is defined as a conjunction of atomic predicates that describes the relationship between two agents at most. The normative systems in \( \Omega \) are over-constrained the behaviors of the agents. The norm refine phase explores all possible specializations of each normative system \( \Phi \in \Omega \), in a scope in which \( k \) has been increased by one, discarding ineffective ones. When \( k \) reaches \( m + 1 \), the algorithm terminates and returns a set of minimal effective normative systems for the MAS with \( m + 1 \) agents.

The space of specializations is exponential in the number of atomic predicates used to describe the local state of an agent. In practice, however, many of those specializations are ineffective, and can be pruned away using a top-down searching approach. The key intuition is as follows. Conceptually, we can order all possible norma-
Let us assume that the normative system $NS$ is not effective for a property $\omega$, then implementing any normative system which generalizes the existing normative system instead also ensures that the same property always holds in the MAS.

**Lemma 1.** Let $\mathcal{M}$ be an MAS with an arbitrary number of agents, $\omega$ be a property, and $\Phi$ be a normative system such that $\mathcal{M} \uplus \Phi \models A \Box \omega$. We have $\mathcal{M} \uplus \Phi' \models A \Box \omega$ for any normative system $\Phi'$ that generalizes $\Phi$. □

The above lemma can be easily proved based on a property of ATL which has been proved in [16], and can be expressed intuitively as follows: “Implementing behavioral constraints on a multiagent system guarantees to preserve the universal properties of the system.” We omit the proof of this property and simply use the result here. The overall formula $A \Box \omega$ is one kind of universal property [16], and also based on definition of generalization relation (Section 3), enforcing any generalized normative system is equivalent with implementing additional behavioral constraints on agents. Therefore, this lemma holds.

Another useful property is that if a normative system ensures the satisfaction of a property in a $k$-agent MAS, then the same property can also be satisfied in any MAS with fewer agents by enforcing the same normative system. Given an MAS $\mathcal{M}$, we write $\mathcal{M}_n$ to denote the MAS with exactly $n$-agents.

**Lemma 2.** Let $\mathcal{M}$ be an MAS; $\omega$ be a property; and $\Phi$ be a normative system such that $\mathcal{M}_k \uplus \Phi \models A \Box \omega$. Then $\mathcal{M}_n \uplus \Phi \models A \Box \omega$ for any $n \leq k$. □

We sketch the proof of this lemma using contradiction as follows. Let us assume that the normative system $\Phi$ is not effective for a
multiagent system $\mathcal{M}'$ with less than $k$ agents. It means that the system must be able to evolve to certain state $s$ which violates property $\omega$ under $\Phi$. We can construct a $k$-agent system by letting one agent $i$ always be outside the neighborhood of the rest of agents and not choose any action each round. Thus this additional agent $i$ does not affect the state transition of the rest of agents each round, and the dynamics of state transitions for the rest of agents is actually the same for the system with less than $k$ agents. Therefore, the agents must also be able to reach a state that violates property $\omega$. This leads to a contradiction.

Now we are ready to establish the correctness of our approach, i.e., every normative system synthesized from the framework is minimal effective in always satisfying property $\omega$ for any MAS with $m + 1$ agents (Formula 1).

**Theorem 1.** Let $\mathcal{M}$ be an MAS and $\omega$ be a property. If any agent $\mathcal{M}$ can only observe the states of $m$ neighboring agents, any normative system $\Phi$ in the final set $\Omega$ synthesized from the framework is minimal effective with respects to $\omega$ for $\mathcal{M}_{m+1}$.

The proof of the theorem is available in Appendix. Although the above theorem guarantees that $\Phi \in \Omega$ is minimal effective for the MAS with $m + 1$ agents, it may not be for a larger number of agents. In contrast, the objective of designing normative systems is often to have a system which is minimal effective for many or arbitrary number of agents. Though our framework fulfills the objective directly, as we show in our case studies, since the normative systems synthesized using our framework often could be proved to be minima effective for an arbitrary number of agents, with a manual induction proof. The following theorem further reduces the effort required for manual proving by showing that as long as we prove that $\Phi$ is effective for the MAS with more agents (i.e., $\mathcal{M}_k$ where $k > m + 1$), $\Phi$ is guaranteed to be minimal as well.

**Theorem 2.** Let $\mathcal{M}$ be an MAS. Let $\phi \in \Omega$ be a minimal effective normative system for $\mathcal{M}_{m+1}$ with respects to property $\omega$ synthesized using our approach. If $\Phi$ is effective with respects to property $\omega$ for $\mathcal{M}_k$ where $k > m + 1$, it is minimal effective with respects to property $\omega$ for $\mathcal{M}_k$.

**Proof.** We prove this theorem by contradiction. Let us assume that there exists another normative system $\Phi_1$ which specializes $\Phi$ and is minimal effective for the $k$-agent system ($k > m + 1$). Based on Lemma 2, we know that $\Phi_1$ is also effective for the ($m + 1$)-agent system, which also specializes $\Phi$. This contradicts with the fact that $\Phi$ is minimal effective for the ($m + 1$)-agent system.

**5. Evaluation**

In this section, we evaluate our synthesis framework with multiple MASs in order to answer the following two research questions.

- **RQ1:** How efficient is it to use our framework to synthesize a set of minimal effective normative systems?
- **RQ2:** Are our underlying assumption justified, i.e., are the normative systems we synthesized based on a small number of agents generalizes to an arbitrary number of agents?

**5.1 Traffic Junction Model**

In the following, we apply our synthesis framework to a practical and well-studied multi-agent coordination problem, i.e., a traffic junction network [25, 26, 21]. We present minimal effective normative systems synthesized from our framework for ensuring a non-collision property, and compare our results to the state-of-the-art simulation-based synthesis approach. Note that this property is also applicable to other similar domains such as prey-predator problem [37] or other multi-agent coordination problems [27].

Our representation of the traffic network is based on the model by Morales et al. [26]. We have already introduced this model in Section 3 and the traffic network is shown in Figure 1(a). Assuming the absence of a central controller that co-ordinates the movement of the cars within the junction (i.e., no traffic lights), our goal is to automatically synthesize minimal effective norm for the following safety property: (non-collision) no two cars should ever occupy the single cell on the traffic network anytime.

**Predicates:** in each local view, there are 4 predicates (e.g., dirEast($g_i$) describing the direction of a car in each cell (i.e., 4$^3$ possibilities), along with 3 predicates (e.g., topLeft) that describe a relationship between any two cars in their local views (i.e., 2$^6$ possibilities). Thus, we can estimate there are $4^4 \times 2^6 = 2048$ possible norms, leading to potentially $2^{2048}$ normative systems.

**Experiment Results** Our framework generated a minimal effective normative system $\Omega$ for the non-collision property with $m = 3$ and $k = 5$. The overall synthesis process invoked the Alloy Analyzer (i.e., Verify from the synthesis algorithm) 108 times to check the effectiveness of candidate normative systems, which pruned away lots of candidate normative systems. The reason is that any specialization when $k > 2$ (i.e., any norm with a precondition that is expressed over more than 2 agents) turned out to be ineffective, thus there is no need to further specialize any more. We ran our experiments on a Windows 7 machine with a 2.4GHZ CPU and 4GB RAM, and it took approximately 4 minutes to finish the synthesis.

The generated normative system $\Omega$ contains 8 norms. Many of these norms are symmetric, and cover the equivalent scenario under different orientations of the car. For example, $\Omega$ requires a car $g$ to stop when it is directly behind another car $h$; $\Omega$ covers this scenario using four separate norms corresponding to norm $n_2$, $n_4$, $n_6$ and $n_8$ in Table 1 respectively, depending on the direction of $g$. The rest of four norms in Table 1 cover another equivalent scenario where car $g$ is not allowed to move and yield to another car $h$ that is approaching $g$ from the left (i.e., priority to the left).

We compare the results of our experiment ($\Phi_A$) to the normative system ($\Phi_B$) produced for the same non-collision problem using the state-of-the-art simulation-based synthesis framework in [26]. The authors of [26] used a different set of atomic predicates to model the traffic junction, though the two models are equivalent in terms of agent behaviors. Thus, we first translate their normative system into our representation. Their normative system contains 16 norms, as listed in Table 2. The norms can be divided into four sets of symmetric norms ($n_1 - n_4$, $n_5 - n_8$, $n_9 - n_{12}$, $n_{13} - n_{16}$) that differ only in the direction of car $g$.

First we can observe that $\Phi_A$ and $\Phi_B$ share one common set of norms: namely, priority to the left norms ($n_1$ and $n_3$). However, norm $n_2$ is more constraining than the combination of the norms $n_2^1$, $n_2^5$, and $n_2^6$, since $n_2$ prevents $g$ from moving forward when

| Table 1: The Norms Generated from Our Framework |
|-----------------|-----------------|-----------------|
| Norm | Precondition | Actions |
| n1 | dirNorth($g$) $\land$ topLeft($h$, $g$) $\land$ dirEast($h$) | Move |
| n2 | dirNorth($g$) $\land$ topLeft($h$, $g$) | Move |
| n3 | dirSouth($g$) $\land$ topLeft($h$, $g$) $\land$ dirEast($h$) | Move |
| n4 | dirSouth($g$) $\land$ top($h$, $g$) | Move |
| n5 | dirWest($g$) $\land$ topLeft($h$, $g$) $\land$ dirEast($h$) | Move |
| n6 | dirWest($g$) $\land$ top($h$, $g$) | Move |
| n7 | dirEast($g$) $\land$ topLeft($h$, $g$) $\land$ dirEast($h$) | Move |
| n8 | dirEast($g$) $\land$ top($h$, $g$) | Move |
5.2 Two-hop Coloring Problem

In the following, we apply our approach to synthesize population protocols. The population protocol model has become a very important computation paradigm for modeling distributed systems such as mobile ad hoc networks [17] and researchers are actively designing and publishing such protocols [7, 22]. The challenge is that a population protocol must work with an arbitrary number of agents and an arbitrary initial configuration of the agents. Intuitively, a population protocol can be considered as a set of rules specifying the local behaviors of each agent in a population such that certain desirable property can be satisfied eventually. For instance, in the two-hop coloring problem [6], a population protocol is required for a ring of agents so that eventually always each pair of neighboring agents should have different colors. For a ring network, we can see that three colors should be sufficient. Thus the problem is to design a protocol to regulate the behaviors of each agent such that the goal can be eventually and always satisfied through finite local interactions (computations) among each pair of neighboring agents. Existing research work [28, 24, 13] mainly focuses on investigating how to effectively and automatically check whether a given two-hop coloring protocol can eventually and always satisfy the goal. Here we take a different perspective and consider the question of how an effective two-hop coloring protocol can be automatically synthesized using our framework.

To synthesize a two-hop coloring protocol, we observe that a protocol corresponds to the opposite of a normative system. For example, given a state $s$, if a protocol specifies that an agent should perform an action from a subset $A' \subseteq A$, then the corresponding norm should specify that this agent should not perform any action from the subset $A \setminus A'$. Based on this observation, the original question of synthesizing an effective protocol can be translated into an equivalent question of how to synthesize a minimal effective normative system such that the desirable property is satisfied. Formally we aim at synthesizing a minimal effective normative system $\Phi$ such that the following property can be satisfied:

\[
\forall g, h \in V, \forall A \subseteq V, \exists n, m : (g, h) \in N(n, m) \Rightarrow \text{there exists a state } s \text{ of } (V, E) \text{ such that } s \models n.A \land s \models m.A \land s \models A \land \text{there exists a state } s' \text{ of } (V, E) \text{ such that } s' \models \neg n.A \land s' \models \neg m.A \land s' \models \neg A.
\]

In the following, we intuitively show that the synthesized norms can be easily generalized to an arbitrary number of agents. That is, the minimal effective normative system synthesized for the $7 \times 7$ road junction with 4 cars is also minimal effective for the general case of $n \times n$ road junction with $m$ cars. Based on the semantics of the normative system we synthesized in Table 1, the normative system is essentially equivalent with implementing the following general rule: an arbitrary car $c$ should stop if there is a car $c'$ right in front of $c$ or if there is another car $c'$ on its left-hand side and heading towards its right-side. It can be proved that, for any $n \times n$ road junction with $m$ cars, if there is a collision, there must be a violation of this rule. Thus the normative system in Table 1 always guarantees no collision for any general case.

### Table 2: The Normative System Generated in [11]

<table>
<thead>
<tr>
<th>Norm</th>
<th>Precondition</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1$</td>
<td>$\text{dirNorth}(g) \land \text{topLeft}(h, g) \land \text{dirEast}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_2$</td>
<td>$\text{dirNorth}(g) \land \text{top}(h_1, g) \land \text{dirWest}(h_1) \land \text{topLeft}(h_2, g) \land \text{dirEast}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_3$</td>
<td>$\text{dirNorth}(g) \land \text{top}(h_1, g) \land \text{dirEast}(h_1) \land \text{topRight}(h_2, g) \land \text{dirEast}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_4$</td>
<td>$\text{dirNorth}(g) \land \text{top}(h, g) \land \text{dirNorth}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_5$</td>
<td>$\text{dirSouth}(g) \land \text{topLeft}(h, g) \land \text{dirWest}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_6$</td>
<td>$\text{dirSouth}(g) \land \text{top}(h_1, g) \land \text{dirEast}(h_1) \land \text{topLeft}(h_2, g) \land \text{dirEast}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_7$</td>
<td>$\text{dirSouth}(g) \land \text{top}(h_1, g) \land \text{dirWest}(h_1) \land \text{topRight}(h_2, g) \land \text{dirEast}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_8$</td>
<td>$\text{dirSouth}(g) \land \text{top}(h, g) \land \text{dirSouth}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_9$</td>
<td>$\text{dirWest}(g) \land \text{topLeft}(h, g) \land \text{dirNorth}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{10}$</td>
<td>$\text{dirWest}(g) \land \text{top}(h_1, g) \land \text{dirSouth}(h_1) \land \text{topLeft}(h_2, g) \land \text{dirNorth}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{11}$</td>
<td>$\text{dirWest}(g) \land \text{top}(h_1, g) \land \text{dirNorth}(h_1) \land \text{topRight}(h_2, g) \land \text{dirNorth}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{12}$</td>
<td>$\text{dirWest}(g) \land \text{top}(h, g) \land \text{dirWest}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{13}$</td>
<td>$\text{dirEast}(g) \land \text{topLeft}(h, g) \land \text{dirSouth}(h)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{14}$</td>
<td>$\text{dirEast}(g) \land \text{topLeft}(h, g) \land \text{dirNorth}(h) \land \text{topLeft}(h_2, g) \land \text{dirNorth}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{15}$</td>
<td>$\text{dirEast}(g) \land \text{top}(h_1, g) \land \text{dirNorth}(h_1) \land \text{topRight}(h_2, g) \land \text{dirNorth}(h_2)$</td>
<td>$\text{Move}$</td>
</tr>
<tr>
<td>$n_{16}$</td>
<td>$\text{dirEast}(g) \land \text{top}(h, g) \land \text{dirEast}(h)$</td>
<td>$\text{Move}$</td>
</tr>
</tbody>
</table>

None of the norms in $\Phi_B$ is more specialized than $\Phi_A$, which might seem more desirable than $\Phi_A$. However, $\Phi_B$ is not effective in ensuring that the MAS satisfies the non-collision property. To demonstrate this, we encoded $\Phi_B$ in our Alloy model, and checked their effectiveness. The Alloy Analyzer returned a counterexample that demonstrates a violation of this rule. Thus the normative system in Table 1 always guarantees no collision for any general case.
Experimental Results Our framework first generated a minimal effective normative system for the two-hop coloring property with a 3-node ring. Alloy Analyzer was invoked 1028 times to check the effectiveness of candidate normative systems. We ran our experiments on the same machine and it took approximately 50 minutes to synthesize the norms. The normative system synthesized from our framework is shown in Table 3, which specifies the actions that are not allowed in different states. We then obtain the most “general” protocol for this two-hop coloring problem by “negating” the normative system. The two-hop coloring protocol synthesized is shown in Table 4. Specifically, rule 1 and 2 specify how the values of \( F[u][\text{color}[j]] \) and \( F[j][\text{color}[j]] \) should be updated if they are synchronized; rule 3 specifies that the value of \( \text{color[i]} \) can be updated to any integer value between 0 and 2 if they are unsynchronized; rule 4 specifies how the value of \( F[i][\text{color}[j]] \) should be updated if they are unsynchronized.

Next, we show that the protocol synthesized above for a 3-node ring works for a network of an arbitrary number of nodes. Formally, we would like to show that, for any \( n \)-node ring, the protocol in Table 4 guarantees that eventually always each pair of neighboring agents have different colors (hereafter the property). We sketch the proof based on mathematical induction as follows. We already know that the protocol works for a 3-node ring. For induction, we assume that it works for a \( (n+1) \)-node ring and proves that it works for a \( (n+1) \)-node ring as well. Given a \( (n+1) \)-node ring, for any node \( u \) connecting another two neighboring nodes \( v \) and \( w \), we can first reduce it into a \( n \)-node ring by connecting nodes \( v \) and \( v' \) and ignoring node \( u \) first. Then the reduced \( n \)-node ring can always satisfy the two-hop coloring property. After that, we insert node \( u \) back into its original position in the ring and apply the same protocol again. Next we show that after a finite steps, the \( (n+1) \)-node ring satisfies the property. First, based on the protocol in Table 4, we can see that if the ring satisfies the property and all nodes are synchronized, nothing changes in the ring and the property is satisfied. Now let us consider all possible scenarios when node \( u \) is inserted back into the ring.

If node \( u \) violates the property, and without loss of generality, let us assume \( \text{color}[u] = \text{color}[v] \) where node \( w \) is the other neighbor of node \( u \). We consider the following three cases.

Case 1 \( (F[u][\text{color}[v]] \neq F[v][\text{color}[u]]) \): let the protocol work on the pair of nodes \( (u, v) \). There must be one color \( c \) which can be assigned to node \( u \) such that property is satisfied. Let \( \text{color}[u] = c \). This operation removes the color from node \( u \). Then by updating \( F[u][\text{color}[v]] = F[v][\text{color}[u]], u \) and \( v \) are synchronized.

Case 2 \( (F[u][\text{color}[v]] = F[v][\text{color}[u]] \neq F[w][\text{color}[v]]) \): Let the protocol work on the pair of nodes \( (u, w) \). Then the values of \( F[u][\text{color}[v]] \) and \( F[v][\text{color}[u]] \) will be flipped. Node \( u \) and \( v \) will be unsynchronized. Next, it is resolved as Case 1.

Case 3 \( (F[u][\text{color}[v]] = F[v][\text{color}[u]] = F[w][\text{color}[v]]) \): Let the protocol work on the pair of \( (v, w) \) and \( (u, v) \) sequentially. This will cause the values of \( F[v][\text{color}[w]] \) and \( F[w][\text{color}[v]] \) flip once and the value of \( F[u][\text{color}[v]] \) flip twice. Thus node \( u \) and \( v \) become unsynchronized. Next, it is resolved as Case 1.

If the original color of node \( u \) satisfies the property, we need to make sure it is synchronized with its neighbors \( v \) and \( w \). Suppose that they are not synchronized initially. Let the color update operation be \( \text{color}[u] = \text{color}[v] \) (unchanged), and the update operation on function \( F \) will synchronize the value of \( F[u][\text{color}[v]] \) and \( F[v][\text{color}[w]] \). Based on the previous result, we conclude the protocol works for the an arbitrary number of nodes.

5.3 Orienting Undirected Rings Problem

In the following, we apply our approach to synthesize a population protocol for orienting undirected rings and compare the result with the one published in [6]. Given a ring that is two-hop colored already, we are interested in reaching a state such that all nodes are well-oriented in the sense that (1) each node has exactly one predecessor and one successor and the predecessor and successor should not be the same; (2) for each pair of nodes \( u \) and \( v \), if \( u \) is the predecessor of \( v \), then \( u \) must be the successor of \( v \); (3) for each pair of neighboring nodes \( u \) and \( v \), either \( u \) is the predecessor or successor of \( v \). We are interested in finding a protocol such that the previous orientation property can be eventually satisfied in the ring as long as all nodes (agents) follow the protocol to update their states in a pairwise manner. Similar to the previous example, to synthesize such a protocol, we start with synthesizing the corresponding effective normative system using our framework. Formally we aim to synthesize a minimal effective normative system \( N \) such that the following property can be satisfied: \( M \models N \models \Phi \models \langle \mathcal{O}, \omega \rangle \), where \( \omega \) denotes the targeted state satisfying the above orientation property.

Naturally, we represent the state of a node \( u \) using the following three components: its current color \( \text{color}[u] \), its successor node color \( \text{successor}[u] \) and its predecessor node color \( \text{predecessor}[u] \). During an interaction between a pair of nodes \( (u, v) \), without loss of generality, node \( u \) is assumed to be the initiator and node \( v \) is the responder. By considering whether the initiator node \( u \) is the successor or predecessor of responder node \( v \), we distinguish the states between a pair of nodes into the following four states:

- \( u \) is the successor but not the predecessor of \( v \): \( \text{successor}[u] = \text{color}[v] \) and \( \text{predecessor}[u] \neq \text{color}[v] \)
- \( u \) is the predecessor but not the successor of \( v \): \( \text{predecessor}[u] = \text{color}[v] \) and \( \text{successor}[u] \neq \text{color}[v] \)


ever, our approach shows the potential of the lightweight formal
theory. How-...published ones. In general, manual generalization of the syn-
alized to cases with arbitrary number of agents, and are as good as
those published ones. In general, manual generalization of the syn-
alized to cases with arbitrary number of agents, and are as good as

denote its neighboring nodes as $w_0$, $w_1$, ..., $w_n$, and $w_{n+1}$.

The two-hop coloring property, we know that only one possibility
tated is satisfied in always satisfying property $\forall k+1$ agents, and $\forall k$
e effective for $(k+1)$-agent system and also not in $\Omega_{k+1}$. We can
always construct a corresponding generalized normative system $\Phi''$
of $\Phi'$, such that the precondition of each norm in $\Phi''$ is defined as a
conjunction of atomic predicates that describes the relationship
between $k$ agents at most. Based on Lemma 1, we know that $\Phi''$
is also effective for $(k+1)$-agent system. Further, $\Phi''$ is also
effective for $k$-agent system based on Lemma 2. Thus $\Phi''$ must be
in $\Omega_k$, which leads to a contradiction. Therefore, there does not exist
any normative system that does not belong to $\Omega_{k+1}$ and also
is effective for $(k+1)$-agent system.

Based on previous induction, finally we have the set $\Omega_{m+1}$ con-
tains all the effective normative systems in always satisfying prop-
erty $\omega$ for the $(m+1)$-agent system. Since the normative systems
in $\Omega_{m+1}$ are not comparable in terms of specialization relation-
ship (all generalized normative systems are excluded in Algorithm
2 (Line 6)), any normative system in $\Omega_{m+1}$ is minimal effective
in always satisfying property $\omega$ for $(m+1)$-agent system.

### Table 5: The Protocol Generated from Our Framework for a Pair of Neighboring Agents $i$ and $j$

<table>
<thead>
<tr>
<th>Norm</th>
<th>Precondition</th>
<th>(allowed) Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1$</td>
<td>$\text{Equal}(\text{succolor}[u], \text{color}[v])$ and $\text{NotEqual}(\text{predcolor}[u], \text{color}[v])$</td>
<td>{action a1}</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>$\text{NotEqual}(\text{predcolor}[u], \text{color}[v])$ and $\text{Equal}(\text{predcolor}[u], \text{color}[v])$</td>
<td>{action a2}</td>
</tr>
<tr>
<td>$\tau_3$</td>
<td>$\text{Equal}(\text{succolor}[u], \text{color}[v])$ and $\text{Equal}(\text{predcolor}[u], \text{color}[v])$</td>
<td>{action a2}</td>
</tr>
<tr>
<td>$\tau_4$</td>
<td>$\text{NotEqual}(\text{predcolor}[u], \text{color}[v])$ and $\text{Equal}(\text{predcolor}[u], \text{color}[v])$</td>
<td>{action a2}</td>
</tr>
</tbody>
</table>

### Experimental Results

Our framework first generated a minimal

effective normative system for the orienting property under a 3-
node ring. The overall synthesis process involves 124 times of
 checking to validate the effectiveness of candidate normative sys-
tems. We ran our experiments on the same machine and it took
approximately 5 minutes to synthesize the norms. The synthesized
protocol for this orienting undirected ring problem can be obtained
by taking the opposite of the synthesized normative system simi-
larly to the previous example, which is shown in Table 5. The
protocol consists of four rules, each of which specifies the action
to be performed under each condition (state).

In the following, we generalize the protocol to any $n$-node ring.
We again prove it works $n$-node ring using mathematical induc-
tion as follows. Let us first assume that the protocol works for a
$m$-node ring. Next we construct the corresponding $(m+1)$-node
ring by inserting another node $v$ into an arbitrary position. Let us
denote its neighboring nodes as $u$ and $w$ and their neighbors are $u'$
and $w'$ respectively. Without loss of generality, let us assume that
$\text{color}[u] = \text{predcolor}[u']$ and $\text{color}[w] = \text{succolor}[w']$.

Assume that node $u$ and $v$ interact first. Since the $(m+1)$-node
ring also satisfies the two-hop coloring property, there is only one
possibility after this interaction, i.e., $\text{predcolor}[u] = \text{color}[v]$ and
$\text{succolor}[v] = \text{color}[u]$. Next nodes $v$ and $w$ interact, based on the
two-hop coloring property, we know that only one possibility
happens as follows: $\text{color}[v] = \text{succolor}[w]$ and $\text{predcolor}[v] = 
\text{color}[w]$. After that all nodes have been well-oriented and the well-
oriented configuration remains thereafter since always action a2 is
selected following the protocol. Thus the theorem holds.

### Remarks

We note that for both case studies in Section 5.2 and 5.3, the
synthesized protocols are equivalent with the ones proposed by
Angluin et al. [6]. In response to the research questions we intro-
duced at the beginning of Section 5, we have demonstrated in the
above examples that 1) we can efficiently synthesize non-trivial
normative systems; 2) the synthesized results indeed can be gen-
eralized to cases with arbitrary number of agents, and are as good as
those published ones. In general, manual generalization of the syn-
thesized norms for complex problems could be challenging. How-
ever, our approach shows the potential of the lightweight formal
method: it can be useful for automatically generating candidate
norms which could be generalized. Furthermore, while it is true
the formal methods employed in this work suffer from scalability
issues, it is less a concern in our framework as in all these cases,
we find the right norms with as few as three agents.

### 6. CONCLUSION AND FUTURE WORK

We proposed an approach based on lightweight formal methods
and tools to automatically synthesize minimal effective normative
systems and protocols for multi-agent systems. Complementary
to a simulation-based approach, our approach provides a theoret-
cal guarantee on the effectiveness or minimality within the given
bounds of the analysis. As future work, one natural direction is
to apply the framework to other multi-agent coordination problems
[8] or other multi-agent domains [30].

### Appendix: Proof sketch of Theorem 1

For the sake of convenience, we denote the set of normative sys-
tems returned from $\text{SynthInitNorms}$ and each round of norm refine-
ment process ($\text{RefineNorms}$) with $k = 3, \ldots, m+1$ agents as $\Omega_2$
and $\Omega_k$ respectively. We prove this theorem by induction as follows.

Initially, when the system has only two agents ($k = 2$), from the
initial norm synthesis process in Section 4.2, we know that $\Omega_2$
satisfied the following properties: 1) it contains all effective normative
system for 2-agent system where each norm precondition is defined
in terms of at most $k = 2$ agents. 2) each norm in $\Omega_k$ can not be
further specialized given that the norm precondition is defined in
terms of at most $k = 2$ agents.

When we increase the number of agents and synthesize special-
ized normative systems during the norm refinement process, we are
actually exploring all the possible specialized normative systems
by specializing the precondition of each norm in each normative
system obtained from the initial norm synthesis process. Next we
prove that for any case of $k$ agents, if the set of $\Omega_k$ of normative
systems satisfied the above properties, then the above two proper-
ties also hold for the set $\Omega_{k+1}$. Suppose that this is not true, then
there must exist another normative system $\Phi'$ in which each norm
precondition is defined in terms of at most $k+1$ agents, and $\Phi'$ is
effective for $(k+1)$-agent system and also not in $\Omega_{k+1}$. We can
always construct a corresponding generalized normative system $\Phi''$
of $\Phi'$, such that the precondition of each norm in $\Phi''$ is defined as a
conjunction of atomic predicates that describes the relationship
between $k$ agents at most. Based on Lemma 1, we know that $\Phi''$
is also effective for $(k+1)$-agent system. Further, $\Phi''$ is also
effective for $k$-agent system based on Lemma 2. Thus $\Phi''$ must be
in $\Omega_k$, which leads to a contradiction. Therefore, there does not exist
any normative system that does not belong to $\Omega_{k+1}$ and also
is effective for $(k+1)$-agent system.

Based on previous induction, finally we have the set $\Omega_{m+1}$ con-
tains all the effective normative systems in always satisfying prop-
erty $\omega$ for the $(m+1)$-agent system. Since the normative systems
in $\Omega_{m+1}$ are not comparable in terms of specialization relation-
ship (all generalized normative systems are excluded in Algorithm
2 (Line 6)), any normative system in $\Omega_{m+1}$ is minimal effective
in always satisfying property $\omega$ for $(m+1)$-agent system.
7. REFERENCES


