Masquerade Attack Detection Through Observation Planning for Multi-Robot Systems

Extended Abstract

Kacper Wardega
Boston University
ktw@bu.edu

Roberto Tron
Boston University
tron@bu.edu

Wenchao Li
Boston University
wenchao@bu.edu

ABSTRACT
The increasing adoption of autonomous mobile robots comes with a rising concern over the security of these systems. In this work, we examine the dangers that an adversary could pose in a multi-agent robot system. We show that conventional multi-agent plans are vulnerable to strong attackers masquerading as a properly functioning agent. We propose a novel technique to incorporate attack detection into the multi-agent path-finding problem through the simultaneous synthesis of observation plans. We show that by specially crafting the multi-agent plan, the induced inter-agent observations can provide introspective monitoring guarantees; we achieve guarantees that any adversarial agent that plans to break the system-wide security specification must necessarily violate the induced observation plan.

KEYWORDS
Multi-robot systems; Multi-agent pathfinding; Masquerade attacks; Observation planning

ACM Reference Format:

1 INTRODUCTION
Recent trends in industrial production automation indicate ever-increasing adoption of autonomous mobile robots. Systems from Fetch Robotics and Amazon Robotics are prime examples [9, 23]. These robots, distributed across a factory floor, aid production efficiency and lower human effort, but the security research community has begun to raise alarm over the security of these systems [19]. The factory floor is at risk from malicious actors aiming towards production shutdown [5] or causing human injury [10] through manipulation of the robots in the environment. These threats also extend to multi-agent systems in a less structured environment such as unmanned aerial vehicles [13]. It is therefore important to devise new strategies that can preemptively address these threats.

We consider a novel class of attacks called physical masquerade attacks – a compromised insider (robot) masquerading as a properly functioning robot and attempting to gain access into unauthorized locations without being noticed. We use the term physical

2 RELATED WORK
Autonomous agents are increasingly used to manage various physical systems. This has introduced a number of vulnerabilities. Quarta et al. explore the vulnerabilities in robotic arms of the type used in factory assembly lines and also give a review of some notable exploits such as in automated blast furnaces and nuclear plants [19].

The interconnection and interaction of industrial robots with the physical world can also open up new attack surfaces. Bijani and Robertson provide a taxonomical treatment of attacks on multi-agent systems [4]. The common theme of these studies is that interconnected autonomous agents suffer from lack of effective monitoring. Our work provides introspective monitoring guarantees by crafting a multi-agent plan in a way that requires an agent to be seen by other agents at specific locations and at specific times.

There is a large body of work on multi-agent path finding [6, 15, 18, 24–27]. However, relatively scarce literature has taken security consideration into account. Among those that consider security, existing works are primarily limited to patrol strategies for intrusion detection [1, 2, 7, 8, 14], secure communication [3, 17] and attack-resilient network protocols [12, 20]. More recently, approaches that leverage the physics of the environment to counter cyberattacks began to emerge. In [11], the authors propose an algorithm that uses the physics of wireless signals to defend against Sybil attacks in multi-robot networks. In [22], the authors propose a Sybil attack-resilient traffic estimation and routing algorithm that uses information from sensing infrastructure and the dynamics and proximities of vehicles. Our work is similar to these in spirit in the use of physical channels. In addition, we consider novel attacker models that not only involve insider attacks but also involve maneuvers in the physical space.

3 OBSERVATION PLANNING
We reformulate the multi-agent path finding problem to directly incorporate security requirements. The main idea is that by scheduling the robots’ paths concurrently with an observation plan, the overall system is able to detect when specific robots are not at assigned locations at predetermined times. We call this sort of multi-agent path finding multi-agent observation planning. A multi-agent observation plan entails sequences of planned observations

Table 1: The left column defines solutions $x$ to the MAPF problem [16]. The right column defines solutions $y$ to the corresponding Attack-MAPF problem. $W$ is the world, $U$ the set of actions, $\delta$ the update relation, $S_i$ and $G_i$ are start and goal positions, $\Omega \subset W$ a set obstacles, $\phi$ the observation relation, and $\Xi \subset \Omega$ the safe locations targeted by the attacker.

<table>
<thead>
<tr>
<th>MAPF($W, U, \delta, {S_i}<em>{i=1}^R, {G_i}</em>{i=1}^R, \Omega$)</th>
<th>Attack-MAPF($W, U, \delta, \Omega, x, \phi, \Xi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>init.</strong></td>
<td>$(\forall i \in \mathbb{N}_R)(x_i^1 = S_i), \ N_R = {1, \ldots, R}$</td>
</tr>
<tr>
<td><strong>workspace.</strong></td>
<td>$(\forall i \in \mathbb{N}_R, t \in \mathbb{N}_T)(x_i^t \subseteq W), \ N_T = {1, \ldots, T}$</td>
</tr>
<tr>
<td><strong>transition.</strong></td>
<td>$(\forall i \in \mathbb{N}_R, t \in \mathbb{N}_T, \exists u \in U)(\delta(x_i^t, u) = x_i^{t+1})$</td>
</tr>
<tr>
<td><strong>collisions.</strong></td>
<td>$(\forall i, j \in \mathbb{N}_R, t \in \mathbb{N}_T)(x_i^t \cap x_j^t \neq \emptyset \implies i = j)$</td>
</tr>
<tr>
<td><strong>obstacles.</strong></td>
<td>$(\forall i \in \mathbb{N}_R, t \in \mathbb{N}_T)(x_i^t \cap \Omega = \emptyset)$</td>
</tr>
<tr>
<td><strong>goal.</strong></td>
<td>$(\forall i \in \mathbb{N}_R \ \exists t \in \mathbb{N}_T)(G_i \in x_i^t)$</td>
</tr>
<tr>
<td><strong>attack.</strong></td>
<td>$(\exists i^* \in \mathbb{N}<em>R)(y^1 = x</em>{i^<em>}^1), \ \text{call } i^</em> \text{ the attacking agent.}$</td>
</tr>
<tr>
<td><strong>unobserved.</strong></td>
<td>$(\forall t \in \mathbb{N}_T)(y^t \subseteq W)$</td>
</tr>
</tbody>
</table>

Figure 1: Solution to the MAPF problem (solid lines) for six agents in a continuous workspace. This solution is not attack-proof, since there is a solution to the corresponding Attack-MAPF problem for the red agent (dotted line). The compromised red agent can reach the secure location, shown in green, after being appropriately observed by the blue agent as in the original plan (double-headed black line) without creating any unplanned observations.

between robots. By carefully constructing this multi-agent observation plan, the system can detect attacks (and faults) by detecting any difference between the planned observations and the actual observations reported by the robots. In fact, we would like to construct the multi-agent observation plan in a way that if a faulty or attacking agent breaks the security specification then that agent would necessarily violate the observation plan.

Given a MAPF problem instance $M$ with solution $x$, we pose the Attack-MAPF problem where an adversarial agent solves for a vulnerable solution to the MAPF problem for the red agent (dotted line). The compromised red agent can reach the secure location, shown in green, after being appropriately observed by the blue agent as in the original plan (double-headed black line) without creating any unplanned observations.

### 4 RESULTS

The key metric for evaluating the danger posed by physical masquerade attacks is the percentage of time that conventionally obtained MAPF solutions are vulnerable to the corresponding Attack-MAPF problem. We studied a 4-connected grid environment through an encoding to Satisfiability Modulo Theory (SMT) and a continuous-space/action environment through an encoding to a Mixed-Integer Quadratically-Constrained Program. We found that overwhelmingly (in excess of 90% on average) conventionally-obtained MAPF solutions on random problem instances gave vulnerable solutions. An Attack-MAPF solution is shown along with corresponding MAPF solution in Figure 1. We also developed a complete and optimal approach to computing attack-proof MAPF solutions via an encoding to Exists Forall SMT in the 4-connected grid case. The encoding effectively requires that no attack-MAPF solution exists for the multi-agent plan regardless of which agent is compromised. The details of our encodings are omitted due to space constraints.

### 5 CONCLUSION & FUTURE WORK

This paper introduces a new class of attacks for multi-robot systems where a compromised robot can masquerade as a properly functioning agent and conduct clandestine maneuvers without being detected by other agents. We indicate that solutions to purely MAPF problems are susceptible to this type of attack. Further, we propose a novel mechanism for detecting these physical masquerade attacks by simultaneously synthesizing observation constraints during path planning. In the future, we plan to study weaker attacker models such as attackers knowing only part of the plan and the security implication of these models. In the case where more than one agent are compromised, collusion between these agents is possible and new strategies (likely consensus-based) will need to developed to detect and defend against masquerade attacks. Computationally, MAPF problems are in general NP-hard and attack-proof MAPF additionally requires the absence of potential attack paths in the solutions to the MAPF problems. A subject of current investigation is the exact complexity characterization of attack MAPF and attack-proof MAPF. In addition, since the EFSMT-based approach is effectively centralized planning and these types of approaches often face scalability issues, we plan to investigate decoupled and/or decentralized approaches to compute attack-proof MAPF solutions.
REFERENCES


