HoLA Robots: Mitigating Plan-Deviation Attacks in Multi-Robot Systems with Co-Observations and Horizon-Limiting Announcements

Extended Abstract

Kacper Wardega
Boston University
Boston, USA
ktw@bu.edu

Max von Hippel
Northeastern University
Boston, USA
vonhippel.m@northeastern.edu

Roberto Tron
Boston University
Boston, USA
tron@bu.edu

Cristina Nita-Rotaru
Northeastern University
Boston, USA
c.nitarotaru@northeastern.edu

Wenchao Li
Boston University
Boston, USA
wenchao@bu.edu

ABSTRACT

In centralized multi-robot systems, a *central entity* (CE) checks that robots follow their assigned motion plans by comparing their expected location to the location they self-report. We show that this self-reporting monitoring mechanism is vulnerable to *plan-deviation attacks* where compromised robots don't follow their assigned plans while trying to conceal their movement by misreporting their location. We propose a two-pronged mitigation for plan-deviation attacks: (1) an attack detection technique leveraging both the robots' local sensing capabilities to report observations of other robots and *co-observation schedules* generated by the CE, and (2) a prevention technique where the CE issues *horizon-limiting announcements* to the robots, reducing their instantaneous knowledge of forward lookahead steps in the global motion plan. On a large-scale automated warehouse benchmark, we show that our solution enables attack prevention guarantees from a stealthy attacker that has compromised multiple robots.

KEYWORDS

Multi-Robot Systems; Plan-Deviation Attacks; Multi-Robot Security

ACM Reference Format:
Kacper Wardega, Max von Hippel, Roberto Tron, Cristina Nita-Rotaru, and Wenchao Li. 2023. HoLA Robots: Mitigating Plan-Deviation Attacks in Multi-Robot Systems with Co-Observations and Horizon-Limiting Announcements: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023,* IFAAMAS, 3 pages.

1 INTRODUCTION

We study attacks and defenses in multi-robot systems (MRS) following a centralized execution model [4], which is representative of MRS in known, structured environments with centralized management and control. The system consists of an external *application*, the robots achieving the task, and a *central entity* (CE) which is responsible for determining and transmitting the motion plans to each one of the robots. Ideally, unplanned deviations due to malfunctions are detected by the CE by comparing the expected position of the robots to the one they self-report. Unfortunately, compromised robots who deviate from the motion plan and attempt to move through forbidden regions of the environment cannot be detected solely by self-reports of location from robots, as the compromised ones can lie in their reports to remain undetected. We refer to such deliberate deviations as *plan-deviation attacks*.

Our approach to mitigate such attacks is based on two observations about the attackers: (1) they use the motion plan information from the CE to determine how to move towards the forbidden zone, and (2) they lie about their location to try to remain undetected by the CE. The key idea of our approach is a novel mechanism of *horizon-limiting announcements* (HoLA), where we limit how much motion planning information is announced to the robots at any given time in order to stymie the ability of the attacker to plan successful attacks, but still send as many steps as possible. This is achieved through an *efficient* verification algorithm conducted by the CE which checks whether the planned announcements prevent stealthy attackers from moving towards the forbidden zone because of not having enough information; in the worst case only one step will be released.

2 PROBLEM FORMULATION

Assume that an attacker has compromised a subset $A \subseteq R$ of the robots, with the intention to sabotage the system and cause robots in A to violate the CE's safety constraints without being detected. The compromised robots have full information of the motion plan $\alpha(t)$ announced by the CE by time t. Malicious deviations from the nominal plan conducted by a compromised robot are not easily detectable by the CE, since the compromised robot can lie in its self-reports to the CE. We refer to such malicious deviations as *plan-deviation attacks*, and to deviations that in addition seek to move the robot into one of the forbidden areas in $V_{\text{forbidden}}$ as *forbidden plan-deviation attacks*.

Attacker types. *Stealthy attackers* use their knowledge of the currently announced motion plan prefix $\alpha(t)$ to determine whether there exists a different plan $\tilde{x}$ that is guaranteed to be a forbidden

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

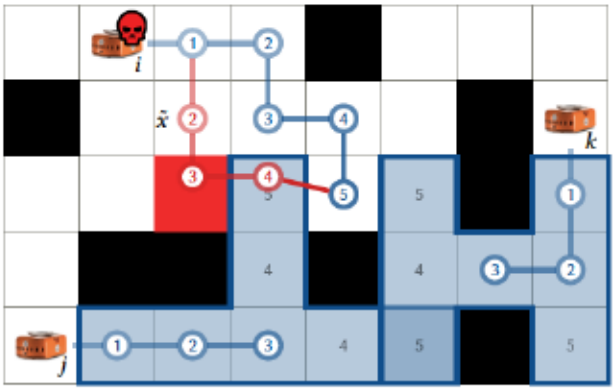

Figure 1: The compromised robot i has computed a forbidden deviation $\bar{x}$ (red paths) on timesteps $(1, 5)$. A stealthy attacker, however, realizes that there is a *possible continuation* (shaded blue region) from the announced portion of the CE's plan (blue paths) that would result in a co-observation-based detection by the CE: if robot j goes north at time step 3, then j would observe i at a location where i is not supposed to be. As a result, the stealthy attacker chooses not to perform the plan-deviation attack.

undetected deviation from the true plan x, $x \geq \alpha(t)$. Attackers that are not stealthy are called *bold*.

Security-aware execution problem. For a variety of reasons, the CE wants to announce as much of the MAPF plan as possible, e.g. due to considerations for network latency, contention, or robustness to network and motion faults [1]. Hence, at each time t, the CE aims to *maximize* $|\alpha(t)|$ subject to the constraint that the unknown compromised subset $A \subseteq R$ of stealthy attackers are not able to perform forbidden plan-deviation attacks.

3 CO-OBSERVATION SCHEDULES

In order to decrease the set of deviations that go undetected by the CE, we propose to include *co-observations* of other robots in the self-reports sent to the CE. Ordinarily, the onboard sensing capabilities of the robots are only used to avoid collisions in fault scenarios. However, we notice that using the sensors to report all inter-robot observations has measurable benefits for security. Our approach is to include in robot i's self-report at time t all observations that i makes of other robots at time t, in addition to i's self-report on action success. We note that this generalizes straightforwardly to environments instrumented with fixed observers (cameras) or fully-trusted agents. If any robot fails to perform an action or does/doesn't observe a robot that it shouldn't/should have, then the self-report it sends to the CE will not match what is expected, triggering a co-observation-based detection in the CE.

4 HORIZON-LIMITING ANNOUNCEMENTS

We now make the attack planning problem against a system with robot co-observation-based mitigation more difficult given a motion plan as input. *The key idea is as follows: the CE can improve the security of the system by preventing the attacker from easily computing forbidden and undetected plan-deviation attacks.* The simplest way to accomplish this is to limit the amount of information available to the attacker about the motion plan, that is, by limiting the amount of future planning information available at every time instant, $\alpha(t)$. In our approach, we ensure for every forbidden deviation for an attacker a from x that there exists a continuation from $\alpha(t)$ that would result in a detection, in which case the stealthy attacker would abstain from deviating from the plan. We term announcements that satisfy this property *horizon-limiting announcements*.

We efficiently compute and verify horizon-limiting announcements using a novel abstraction of multi-agent motion planning that allows non-deterministic movements for the robots. Our abstraction allows non-compromised robots to ignore collisions with other non-compromised robots, allowing us to efficiently explore the co-observation schedules of many plan continuations from the current $\alpha(t)$ simultaneously. Although our abstraction does over-approximate the set of continuations, we can prove that the abstractions preserve the possibility of pairwise co-observation. That is, we are essentially verifying that there is no forbidden deviation through the complement of the observed region under the non-deterministic movement abstraction.

5 EXPERIMENTAL RESULTS

Motion plans are computed using the ECBS algorithm [2], an efficient and bounded sub-optimal graph-based MAPF solver (and so, applicable for centralized MRS), for a set of 100 standard MAPF 4-connected grid benchmark instances [3]. The MAPF instances are solvable (i.e. there exists a MAPF plan that solves the instance), randomly generated 4-connected 32×32 grids with up to 100 robots and ~200 obstacles. We assume each robot has sensing capability within adjacent tiles. In each scenario, we determine whether our approach is able to compute verified horizon-limiting announcements and also the degree to which co-observation-based mitigation detects bold attackers. We find that in 95% of scenarios, our approach computes horizon-limiting announcements that mitigate attacks from stealthy attackers. Furthermore, our results indicate that co-observation-based detections suffice to detect bold attackers in 80% of scenarios. Source code to reproduce the experiments can be found at https://github.com/gitsper/hola-announce

6 CONCLUSION

We introduce the problem of mitigating *plan-deviation attacks* with robot co-observations and incremental plan release. The attacker has two goals: first, to move toward a forbidden zone, and second, to remain undetected by the central entity. We leverage co-observation to mitigate the ability of the attacker to lie about its location; and we limit the size of the incremental plan announcements so that the attacker has limited ability to confidently plan ahead. We describe two types of attackers – "stealthy", and "bold" – based on their desire to remain undetected or not. Our solution prevents attacks for a set of stealthy attackers. For bold attackers we show experimentally that our solution significantly increases the detection of the attacks. Our solution also has a small overhead making it practical for sets of tens to hundreds of robots. The full version of this paper can be found at https://arxiv.org/abs/2301.10704

7 ACKNOWLEDGEMENTS

We gratefully acknowledge the support from the National Science Foundation awards CNS-1932162, CNS-1931997, and GRFP-1938052.

REFERENCES

[1] Dor Atzmon, Roni Stern, Ariel Felner, Glenn Wagner, Roman Barták, and Neng-Fa Zhou. 2020. Robust Multi-Agent Path Finding and Executing. *Journal of Artificial Intelligence Research* 67 (March 2020), 549–579.

[2] Max Barer, Guni Sharon, Roni Stern, and Ariel Felner. 2014. Suboptimal Variants of the Conflict-Based Search Algorithm for the Multi-Agent Pathfinding Problem. *Proceedings of the 7th Annual Symposium on Combinatorial Search* (2014).

[3] Wolfgang Hönig. 2021. libMultiRobotPlanning. https://github.com/whoenig/libMultiRobotPlanning

[4] Wolfgang Hönig, Scott Kiesel, Andrew Tinka, Joseph W Durham, and Nora Ayanian. 2019. Persistent and robust execution of MAPF schedules in warehouses. *IEEE Robotics and Automation Letters* 4, 2 (2019), 1125–1131.