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:
Kacper Wardega, Roberto Tron, and Wenchao Li. 2019. Masquerade Attack Detection Through Observation Planning for Multi-Robot Systems. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

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* to distinguish this type of attack from masquerade attacks typically considered in the network security literature [21]. In the multi-agent path finding (MAPF) context, this manifests as one of the agents deviating from its pre-planned path and moving into an unauthorized zone. We show that solutions to the traditional MAPF problem are susceptible to this type of attacks.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

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-robot path finding [6, 15, 18, 24–27]. However, relatively scarce literature has taken security into consideration. 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 $\mathbf{x}$ to the MAPF problem [16]. The right column defines solutions $\mathbf{y}$ to the corresponding Attack-MAPF problem. W is the world, U the set of actions, δ the update relation, S_i and G_i are start and goal positions, $\Omega \subset W$ a set obstacles, ϕ the observation relation, and $\Xi \subseteq \Omega$ the safe locations targeted by the attacker.

	MAPF$(W, U, \delta, \{S_i\}_{i=1}^R, \{G_i\}_{i=1}^R, \Omega)$	Attack-MAPF$(W, U, \delta, \Omega, \mathbf{x}, \phi, \Xi)$
`init.`	$(\forall\, i \in \mathbb{N}_R)\,(\mathbf{x}_i^1 = S_i), \quad \mathbb{N}_R = \{1, \ldots, R\}$	$(\exists\, i^* \in \mathbb{N}_R)\,(\mathbf{y}^1 = \mathbf{x}_{i^*}^1),$ call i^* the attacking agent.
`workspace.`	$(\forall\, i \in \mathbb{N}_R, t \in \mathbb{N}_T)\,(\mathbf{x}_i^t \subseteq W), \quad \mathbb{N}_T = \{1, \ldots, T\}$	$(\forall\, t \in \mathbb{N}_T)\,(\mathbf{y}^t \subseteq W)$
`transition.`	$(\forall\, i \in \mathbb{N}_R, t \in \mathbb{N}_{T-1}\, \exists\, u \in U)\,(\delta(\mathbf{x}_i^t, u) = \mathbf{x}_i^{t+1})$	$(\forall\, t \in \mathbb{N}_{T-1}\, \exists\, u \in U)\,(\delta(\mathbf{y}^t, u) = \mathbf{y}^{t+1})$
`collisions.`	$(\forall\, i, j \in \mathbb{N}_R, t \in \mathbb{N}_T)\,(\mathbf{x}_i^t \cap \mathbf{x}_j^t \neq \emptyset \implies i = j)$	$(\forall\, t \in \mathbb{N}_T, j \in \mathbb{N}_R \setminus i^*)\,(\mathbf{x}_i^t \cap \mathbf{y}^t = \emptyset)$
`obstacles.`	$(\forall\, i \in \mathbb{N}_R, t \in \mathbb{N}_T)\,(\mathbf{x}_i^t \cap \Omega = \emptyset)$	$(\forall\, t \in \mathbb{N}_T)\,(\mathbf{y}^t \cap (\Omega \setminus \Xi) = \emptyset)$
`goal.`	$(\forall\, i \in \mathbb{N}_R\, \exists\, t \in \mathbb{N}_T)\,(G_i \in \mathbf{x}_i^t)$	
`attack.`		$(\exists\, t \in \mathbb{N}_T)\,(\mathbf{y}^t \cap \Xi \neq \emptyset)$
`unobserved.`		$(\forall\, t \in \mathbb{N}_T, j \in \mathbb{N}_R \setminus i^*)\,(\phi(\mathbf{x}_j^t, \mathbf{x}_{i^*}^t) \Leftrightarrow \phi(\mathbf{x}_j^t, \mathbf{y}^t))$

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 $\mathbf{x}$, we pose the *Attack-MAPF* problem where an adversarial agent solves for an alternative path that reaches a secure location undetected. The attacker knows that all of the robots are equipped with sensors for inter-robot communication and monitoring such as cameras and radios. Uncompromised agents will be reporting observations to a central controller for verification against the observation plan. The sensor properties are known to the attacker, i.e. the attacker knows which positions relative to uncompromised agents will result in observations being reported to the central controller. If there exists a solution $\mathbf{y}$ to this Attack-MAPF problem, then we say that $\mathbf{x}$ is a vulnerable solution to M, and attack-proof otherwise. Formal definitions of the MAPF and Attack-MAPF problems are outlined in Table 1.

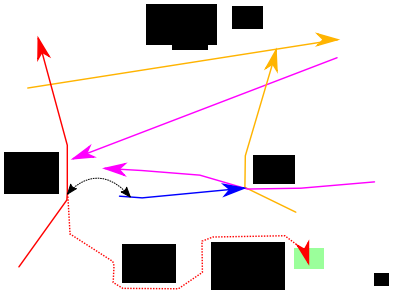

Figure 1: Solution to the MAPF problem (solid lines) for six agents in a continuous workspace. This solution is not attack-proof, since there is a solution to the corresponding Attack-MAPF problem for the red agent (dotted line). The compromised red agent can reach the secure location, shown in green,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

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 are 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

[1] Noa Agmon, Gal A Kaminka, and Sarit Kraus. 2011. Multi-robot Adversarial Patrolling: Facing a Full-knowledge Opponent. *J. Artif. Int. Res.* 42, 1 (Sept. 2011), 887–916. http://dl.acm.org/citation.cfm?id=2208436.2208459

[2] N. Agmon, S. Kraus, and G. A. Kaminka. 2008. Multi-robot perimeter patrol in adversarial settings. In *2008 IEEE International Conference on Robotics and Automation*. 2339–2345. https://doi.org/10.1109/ROBOT.2008.4543563

[3] A. Bicchi, A. Danesi, G. Dini, S. L. Porta, L. Pallottino, I. M. Savino, and R. Schiavi. 2008. Heterogeneous Wireless Multirobot System. *IEEE Robotics Automation Magazine* 15, 1 (March 2008), 62–70. https://doi.org/10.1109/M-RA.2007.914925

[4] Shahriar Bijani and David Robertson. 2014. A review of attacks and security approaches in open multi-agent systems. *Artificial Intelligence Review* 42, 4 (2014), 607–636. https://doi.org/10.1007/s10462-012-9343-1

[5] M. Brunner, H. Hofinger, C. Krauss, C. Roblee, P. Schoo, and S. Todt. 2010. Infiltrating critical infrastructures with next-generation attacks. http://publica.fraunhofer.de/documents/N-151330.html. (2010).

[6] Howie Choset, Kevin M. Lynch, Seth Hutchinson, George A. Kantor, Wolfram Burgard, Lydia E. Kavraki, and Sebastian Thrun. 2005. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press.

[7] A. Fagiolini, M. Pellinacci, G. Valenti, G. Dini, and A. Bicchi. 2008. Consensus-based distributed intrusion detection for multi-robot systems. In *2008 IEEE International Conference on Robotics and Automation*. 120–127. https://doi.org/10.1109/ROBOT.2008.4543196

[8] A. Fagiolini, G. Valenti, L. Pallottino, G. Dini, and A. Bicchi. 2007. Decentralized intrusion detection for secure cooperative multi-agent systems. In *2007 46th IEEE Conference on Decision and Control*. 1553–1558. https://doi.org/10.1109/CDC.2007.4434902

[9] Fetch. [n. d.]. fetchcore: Cloud Robotics Platform. https://fetchrobotics.com/products-technology/fetchcore/. ([n. d.]). Accessed: 2018-05-09.

[10] Conner Forrest. 2017. Robot kills worker on assembly line, raising concerns about human-robot collaboration. https://www.techrepublic.com/article/robot-kills-worker-\on-assembly-line-raising-concerns-about-human-robot-collaboration/. (15 March 2017).

[11] Stephanie Gil, Swarun Kumar, Mark Mazumder, Dina Katabi, and Daniela Rus. 2017. Guaranteeing spoof-resilient multi-robot networks. *Autonomous Robots* 41, 6 (01 Aug 2017), 1383–1400. https://doi.org/10.1007/s10514-017-9621-5

[12] Diksha Gupta, Jared Saia, and Maxwell Young. 2018. Proof of Work Without All the Work. In *Proceedings of the 19th International Conference on Distributed Computing and Networking (ICDCN '18)*. ACM, New York, NY, USA, Article 6, 10 pages. https://doi.org/10.1145/3154273.3154333

[13] A. Y. Javaid, W. Sun, V. K. Devabhaktuni, and M. Alam. 2012. Cyber security threat analysis and modeling of an unmanned aerial vehicle system. In *2012 IEEE Conference on Technologies for Homeland Security (HST)*. 585–590. https://doi.org/10.1109/THS.2012.6459914

[14] Ji Min Kim, Jeong Sik Choi, and Beom Hee Lee. 2008. Multi-agent Coordinated Motion Planning for Monitoring and Controlling the Observed Space in a Security Zone. *IFAC Proceedings Volumes* 41, 2 (2008), 1679–1684. https://doi.org/10.3182/20080706-5-KR-1001.00288 17th IFAC World Congress.

[15] Steven M. LaValle. 2006. *Planning Algorithms*. Cambridge University Press, New York, NY, USA.

[16] Hang Ma, Glenn Wagner, Ariel Felner, Jiaoyang Li, T. K. Satish Kumar, and Sven Koenig. 2018. Multi-Agent Path Finding with Deadlines. July (2018). arXiv:1806.04216 http://arxiv.org/abs/1806.04216

[17] Santiago Morante, Juan G. Victores, and Carlos Balaguer. 2015. Cryptobotics: why robots need cyber safety. https://www.frontiersin.org/articles/10.3389/frobt.2015.00023/full. (29 September 2015).

[18] Aniello Murano, Giuseppe Perelli, and Sasha Rubin. 2015. Multi-agent Path Planning in Known Dynamic Environments. In *PRIMA 2015: Principles and Practice of Multi-Agent Systems*, Qingliang Chen, Paolo Torroni, Serena Villata, Jane Hsu, and Andrea Omicini (Eds.). Springer International Publishing, Cham, 218–231.

[19] Davide Quarta, Marcello Pogliani, Mario Polino, Federico Maggi, Andrea Maria Zanchettin, and Stefano Zanero. 2017. An Experimental Security Analysis of an Industrial Robot Controller. *2017 IEEE Symposium on Security and Privacy (SP)* (2017), 268–286. https://doi.org/10.1109/SP.2017.20

[20] V. Renganathan and T. Summers. 2017. Spoof resilient coordination for distributed multi-robot systems. In *2017 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. 135–141. https://doi.org/10.1109/MRS.2017.8250942

[21] Malek Ben Salem, Shlomo Hershkop, and Salvatore J. Stolfo. 2008. *A Survey of Insider Attack Detection Research*. Springer US, Boston, MA, 69–90. https://doi.org/10.1007/978-0-387-77322-3_5

[22] Yasser Shoukry, Shaunak Mishra, Zutian Luo, and Suhas Diggavi. 2018. Sybil Attack Resilient Traffic Networks: A Physics-based Trust Propagation Approach. In *Proceedings of the 9th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS '18)*. IEEE Press, Piscataway, NJ, USA, 43–54. https://doi.org/10.1109/ICCPS.2018.00013

[23] IEEE Spectrum. [n. d.]. Three Engineers, Hundreds of Robots, One Warehouse. https://www.spectrum.ieee.org/robotics/robotics-software/three-engineers-hundreds-of-robots-one-warehouse. ([n. d.]). Accessed: 2018-04-02.

[24] A. Ulusoy, S. L. Smith, X. C. Ding, and C. Belta. 2012. Robust multi-robot optimal path planning with temporal logic constraints. In *2012 IEEE International Conference on Robotics and Automation*. 4693–4698. https://doi.org/10.1109/ICRA.2012.6224792

[25] A. Ulusoy, S. L. Smith, X. C. Ding, C. Belta, and D. Rus. 2011. Optimal multi-robot path planning with temporal logic constraints. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 3087–3092. https://doi.org/10.1109/IROS.2011.6094884

[26] Ko-Hsin Cindy Wang and Adi Botea. 2009. Tractable Multi-agent Path Planning on Grid Maps. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence (IJCAI'09)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1870–1875. http://dl.acm.org/citation.cfm?id=1661445.1661745

[27] Ko-Hsin Cindy Wang and Adi Botea. 2011. A Scalable Multi-Agent Path Planning Algorithm with Tractability and Completenss Guarantees. *JAIR - Journal of Artificial Intelligence Research* 42 (2011), 55–90. https://doi.org/10.1613/jair.3370