frontiers | Frontiers in Blockchain

TYPE Perspective
PUBLISHED 16 May 2023
DOI 10.3389/fbloc.2023.1181820

OPEN ACCESS

EDITED BY
Adedoyin Ahmed Hussain,
KIbris Bati University, Cyprus

REVIEWED BY
Ayan Dutta,
University of North Florida, United States

*CORRESPONDENCE
Eduardo Castelló Ferrer,
ecstll@mit.edu

RECEIVED 07 March 2023
ACCEPTED 13 April 2023
PUBLISHED 16 May 2023

CITATION
Ferrer EC (2023), If blockchain is the solution, robot security is the problem. *Front. Blockchain* 6:1181820. doi: 10.3389/fbloc.2023.1181820

If blockchain is the solution, robot security is the problem

Eduardo Castelló Ferrer*

Massachusetts Institute of Technology, Cambridge, MA, United States

Robotics systems of all types are revolutionizing a wide variety of industries—transportation, manufacturing, and even healthcare—and yet, many essential ingredients for robotics systems in the real world are not technologically ready for deployment. Currently, robots lack the protocols and standards required to be safe and secure outside factories. In an attempt to close this gap, recent research has demonstrated the security benefits of combining robotics systems with blockchain-based and related technologies (e.g., smart contracts, zero-knowledge proofs, Merkle trees). In this perspective article, I argue that blockchain-based robotics is starting to provide innovative solutions (e.g., secure data sharing, consensus mechanisms, and new interaction methods) to urgent problems of robot security. I list the most important takeaways so far from this emerging field of research that I helped establish together with a growing community. I close the article by discussing the implications of the security challenges that the robotics research community is facing, and possible ways for us to move forward.

KEYWORDS

blockchain, robotics, safety, security, cryptography, smart contracts, decentralized autonomous organizations (DAOs)

Introduction

Imagine in a non-distant future, you are inside a self-driving car reading your science magazine, when something pops up in the car's dashboard. It is a message in harsh red characters: "This self-driving car has been compromised, all your data has been encrypted, and the car will not break unless you transfer 1000 USD in Bitcoin to the following address: 0×141F. . ." Even if you do not own any Bitcoin, I am sure that will be your first transaction. If you are reading this article, you might already agree with me that robotic systems—from fleets of autonomous vehicles to surgical precision devices—are revolutionizing a wide variety of industries, including manufacturing, transport, and even healthcare. Boosted by technical breakthroughs such as orders-of-magnitude more capable hardware, AI-driven computer vision, and software for control systems, the emergence of robotics is expected to be one of the main socioeconomic disruptions of upcoming decades (Yang et al., 2018). Despite this expectation, many ingredients that are essential for robotics in the real world are not yet technologically ready. Consider that, when a PC is hacked, the direct consequences typically remain virtual or economic. By contrast, when a vulnerability of a robot (e.g., your future self-driving car) is exploited, not only can there be privacy violations, data breaches, or economic losses, there can also be direct physical consequences.

From robot safety to robot security

Perhaps the earliest mention of safety in robotics was not from a researcher but a science fiction writer: Isaac Asimov's three laws of robotics. When his first law declared "A robot may not

injure a human being or, through inaction, allow a human being to come to harm," (Asimov and Robot 1950) it established the idea that a robot could potentially bring harm to its environment (and to humans) if strict rules and measures were not followed. This paved the way for the robot safety research field (Haddadin et al., 2009) to formally start in the early 70s, when companies, especially in the automotive sector, were deploying the first high-power robots in their assembly lines side-by-side with human workers. Safety devices such as cages, kill switch mechanisms, and intention recognition methods were developed over the years to keep humans safe from their mechanical co-workers. Nowadays, robotic technology is being quickly diversified and utilized in new sectors, mostly outside factories. The primary future user of robots will be the general consumer and an increasing number of powerful robots will become basic household items. But unlike our history with industrial robots, we have a new problem: We are not ready, and we urgently need to overcome the common misconception that robot safety is only about preventing a robot from harming its environment, like in Asimov's laws.

If we sit back in our self-driving car, we can be sure that its manufacturer invested significant effort developing advanced safety measures such as autonomous breaking, cruise control, and connectivity with traffic signs. However, if a non-authorized third-party gains access to the robot and turns off these measures, they become useless. Research is beginning to show that hacked robots can be reprogrammed to manifest unwanted or even dangerous behaviors, breaking any safety measures included in them (Mayoral-Vilches et al., 2020). Moreover, deceitful information in a decentralized robotic system (e.g., a swarm of self-driving cars operating in a city) can accumulate in a dangerous way through cascades of robot-robot interactions (Vivek et al., 2019). Although these robots are safe when isolated, they are not secure in open environments. Counterintuitively, security deals with the opposite of safety: unlike safety, which ensures a robot does not conflict with its environment, security ensures the environment does not conflict with the robot's programmed behavior. Still, there is an intrinsic connection between the two, because functional safety standards for robotic systems cannot be relied upon without their guaranteed security. In other words, there is no real robot safety without robot security (Figure 1A–G).

"The blockchain"

"The blockchain" (Nakamoto, 2008), which is quickly becoming a phrase that could describe everything or nothing, is meant to refer to a set of cryptographic methods that allow transactions between agents to be recorded securely without any centralized control. "The blockchain" has been the subject of much hype in the past decade, revolving around a great deal of cryptocurrency speculation and many disreputable projects. Still, it has shown a great deal of durability (Bitcoin has been running for 15 years without interruptions or breakdowns). On one hand, its technological drawbacks (e.g., long wait times between when transactions are sent and received, environmental toll depending on the consensus algorithms used, and weight of a continuously growing database replicated almost everywhere) still make "the blockchain" overkill for most applications. On the other hand, "the blockchain" includes the complete set of components needed for a secure network of robots: namely, message authentication and resolution of conflicting states, as well as a tamper-proof decentralized database (which can be public or closed, depending on the needs of the application). To put it in another way, blockchain technology brings a unique combination of characteristics (Figure 1H-J), including secure data sharing, data logging, and new incentive mechanisms, which makes it an ideal candidate to provide security solutions for robotic systems, before, during and after something goes wrong.

More than a possibility, the problem of our hacked self-driving car is a practical inevitability—the complexity of modern programming frameworks makes it nearly impossible for robot manufacturers to keep up with the speed of third-party actors uncovering new vulnerabilities. Therefore, we are bound to ask ourselves: *How can we reduce the number of security breaches we have to deal with in the first place?* Modern blockchain technology provides us with tools such as smart contracts (i.e., computer code embedded in the blockchain that directly controls the transfer of digital assets between parties). These tools can bootstrap new types of organizations (e.g., Decentralized Autonomous Organizations, a.k.a. DAOs) that offer economic incentives to third-party actors to find, report, and fix vulnerabilities rather than exploit them, similar to Bitcoin miners providing their resources (e.g., CPU, disk space, electricity) for validating the network instead of hacking it. Still, *when a problem does appear, how do we minimize the harm that a robot can cause?* Based on current research, we know that if robots use cryptographic techniques (e.g., zero-knowledge proofs) as part of their communication substrate, they can verify data authenticity without accessing the raw data itself. Sensitive information (e.g., datasets, maps, personal records) can therefore be used by robots without them actually having access to it (e.g., your self-driving car can verify you own a certain parking spot without knowing your name or other personal information). Also, because all robot interactions are automatically registered in the ledger, when a problem appears (e.g., accidental hardware malfunction or malicious software interference), the erratic behavior can easily be flagged, and the robot quickly isolated, before more harm can be caused. Used in this way, blockchain technology can both minimize the exposure of sensitive data during a breach and isolate the harm of a compromised robot away from the rest of the system. However, if a security breach does happen, our next question will likely be: *How can we analyze past incidents to prevent this same problem from happening again?* Because blockchain technology is an append-only, non-counterfeitable ledger, its use as robot communication substrate also guarantees that any malicious interference is attributable and that all robot decisions and actions are auditable (Ferrer et al., 2018). This is of great societal importance whenever robots make decisions about humans or humans must understand the decision-making processes of robots. In short, blockchain technology's unique combination of capabilities offers several new ways to increase the security of robotic systems. I will now explain how my research has demonstrated this potential.

Securing robots with blockchain technology

In 2014, when I was working on distributed robotic systems intended to operate in complex and dangerous environments, I was

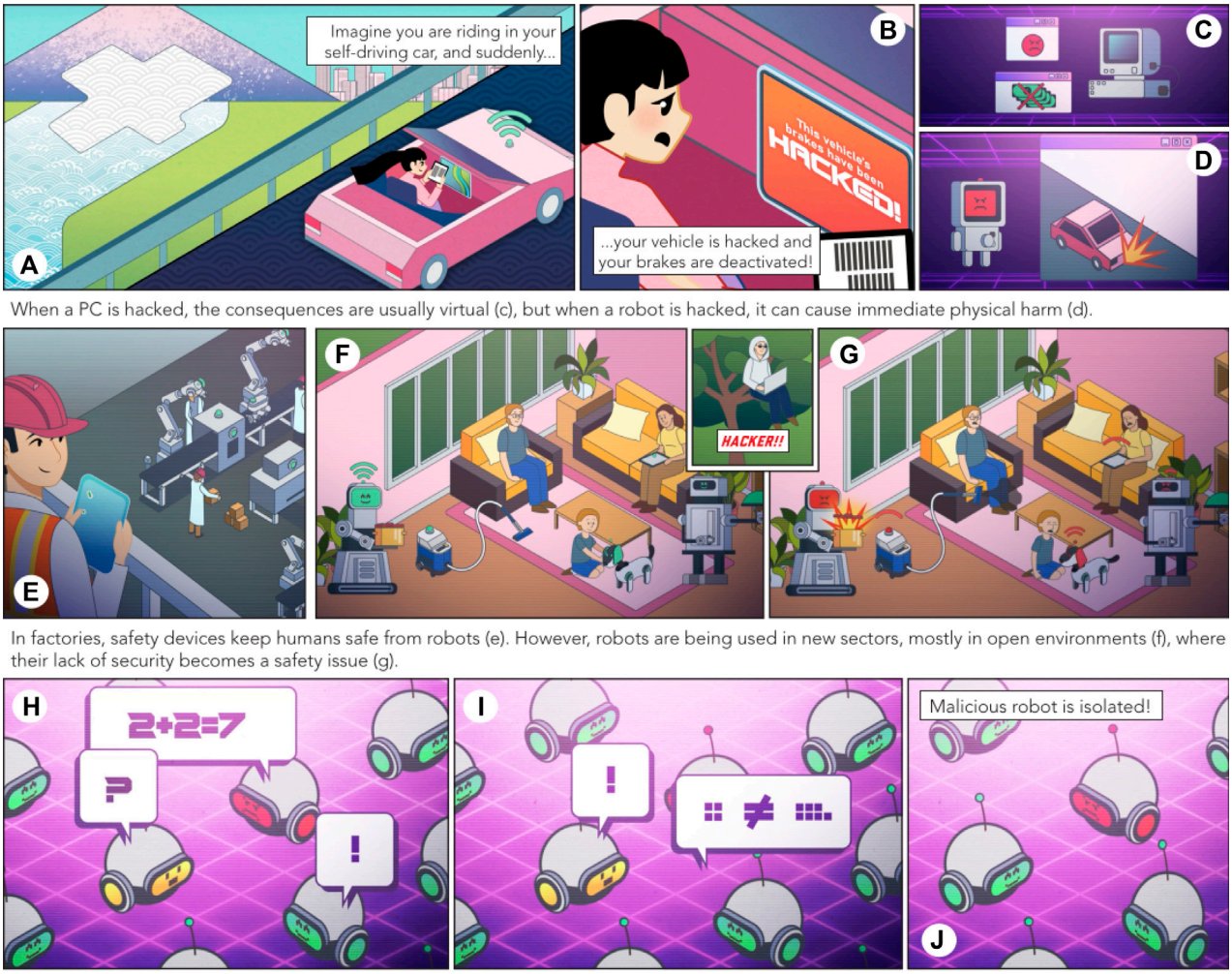

FIGURE 1
Scenes from an animation video describing the key takeaways of this article. The video is available here: https://youtu.be/HCyomzw4tdU.

alarmed to notice how easy it would be to compromise their security and integrity—in contrast to the predominant opinion. This worried me, so the next year I began to dig into their security vulnerabilities. I found that blockchain technology provides some solutions, not as a hammer that could nail every problem, but as a young Swiss knife: a versatile collection of tools and principles that, through maturation and research, could be used to tackle urgent problems in emerging technology fields. From blockchain technology's unique combination of capabilities, I found that its secure data sharing established cryptographic transactions as the basis for communication, its secure data logging allowed timestamping of information for future attestation, and its secure decentralized consensus algorithms allowed agreements to be reached under a wide array of circumstances. I realized this technology to be capable of securing robotic systems in new and general ways compared to existing security methods (e.g., intrusion detection systems) which are typically centralized and target only one step of the process. More precisely, I discovered that blockchain technology can protect against types of attacks that are particular to robotic systems, such as new robots being programmed maliciously (i.e., byzantine robots) to enter an existing network and broadcast deceitful information, causing even uncompromised peers to act in compromised and dangerous ways.

When I realized these features would be highly relevant to pressing robotics security problems, I published my findings and initiated a research field combining blockchain technology and robotic systems (Ferrer, 2016). Then, step-by-step we started verifying these findings by building the first robotic systems based on blockchain-based smart contracts (Strobel et al., 2018; Strobel et al., 2020). As we expected, smart contracts were indeed able to establish secure coordination of a group of robots: "bad bots" (i.e., robots programmed to breach consensus with misleading information) were successfully "self-neutralized" by the robots themselves without human intervention. The robots (even with very limited CPU resources) were able to protect themselves by finding inconsistencies in the trail of transactions left in the ledger. As an example: first, robots used their reputations to vouch for the information they shared with the group. If the group (e.g., a robot

swarm) found the information shared by individual robots to be misleading or inaccurate (e.g., deviated too much from the swarm's mean), those robots gradually lost their individual resources (i.e., their reputation), to the point where they became untrustworthy and were subsequently ignored by the rest of the group. In other words, inconsistent behavior (e.g., software or hardware malfunction or malicious activity) costs a robot the resources it requires to take actions in the system and its impact is therefore intrinsically limited.

The combination of these two fields—blockchain technology and robotics—allowed us to define new interaction methods. For example, when we used blockchain technology as a secure data record tool in a multirobot follow-the-leader mission, we found that although robots could temporarily be misled by compromised peers, they could always undo their misinformed actions by analyzing and reverting the trail of transactions (Ferrer et al., 2021a). Of course, current blockchain technology might not be the correct security solution for all robotic systems, so other types of cryptographic methods are also crucial to investigate (Queralta et al., 2022). As an example, when we encapsulated cooperative robot missions in Merkle trees (i.e., a cryptographic hash-based tree data structure) (Ferrer et al., 2021b), we found that secure and secret robot missions could indeed be guaranteed. Operators could fully encrypt and provide the "blueprint" of a collective robot mission without disclosing its raw data. Robots discovered the tasks to complete by exploring the environment, generating cryptographic hashes of possible actions (e.g., moving object A to location B has a resultant hash identifier of 0×131Da. . .), and comparing them to the encrypted plan they had received from the operator (e.g., 0×131Da. . . is a leaf in the tree, thus I should complete the task even though I do not know what the other tasks in the mission are). In other words, data verification was able to be separated from the data itself. Under this framework, to cooperate among themselves, robots had to "prove" their integrity to their peers by exchanging cryptographic proofs, pointing to new interaction methods for robots. For instance, instead of blindly relying on the information coming from other agents (robots or humans), a robot can now tell you: for security reasons, do not show me your data, but instead prove to me (cryptographicaly) that you have it and then I will trust you.

The challenges and a possible way forward

Robotics must confront the challenges of its significant security gap. We all expect many new ways for people, machines, and organizations to be interconnected, communicate, and exchange and share information. However, as happens in many industries, robot manufacturers have been taking a "rush to market" approach, with the collateral damage that fundamental security aspects are being overlooked in early stages of product design. Complementarily, lack of security in academic environments is a widespread reality (DeMarinis et al., 2019). As we have learned from previous technological upheavals (e.g., social media or smartphones), the right time to be thinking about security was yesterday. Therefore, what are the grand challenges that we collectively should address to enter this fourth industrial revolution with some security guarantees? What should be done before, during, and after security vulnerabilities in robots potentially wreak havoc and cause real harm, like in a hacked self-driving car?

A prevention strategy would suggest that security vulnerabilities be fixed before things can go wrong. In order to do that, for instance, how can vulnerabilities be discovered and broadcast securely to all connected robots? Without exposing sensitive data, how do we generate technology that allows us to trust robots and, crucially, robots to trust us? In case things go wrong, how can forensic investigations generate credible evidence as to who did what, where, when, and why? Disregarding the hype around a technology that is developing and maturing, "the blockchain" brings a unique combination of characteristics that cannot be found elsewhere. Blockchain technology might provide tools to build new cyber-physical institutions (Ferrer et al., 2023), where even agents with disputable intentions might contribute to securing the system, with the right incentives and disincentives. No matter which tools are used to face it, robot security is set to be an unavoidable challenge. It might be the unsolved problem that "the blockchain" did not expect to find.

Data availability statement

The original contributions presented are included in the article. In addition, an animated video describing the key takeaways of the article can be found in the following Youtube URL: https://youtu.be/HCyomzw4tdU. Further inquiries can be directed to the corresponding author.

Author contributions

ECF wrote the manuscript and directed the project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbloc.2023.1181820/full#supplementary-material

References

Asimov, I., and Robot, I. (1950). *Doubleday science fiction*. New York: Bantam Books.

DeMarinis, N., Tellex, S., and Vasileios, P. (2019). "Kemerlis, george konidaris, and rodrigo fonseca. Scanning the internet for ros: A view of security in robotics research," in *2019 international conference on robotics and automation* (ICRA), 8514–8521.

Ferrer, E. C., Berman, I., Kapitonov, A., Manaenko, V., Chernyaev, M., and Tarasov, P. (2023). *Gaka-chu: a self-employed autonomous robot artist*

Ferrer, E. C., Hardjono, T., Pentland, A., and Dorigo, M. (2021). Secure and secret cooperation in robot swarms. *Sci. Robotics* 6 (56), eabf1538. doi:10.1126/scirobotics.abf1538

Ferrer, E. C., Jiménez, E., Luis Lopez-Presa, J., and Martín-Rueda, J. (2021). Following leaders in byzantine multirobot systems by using blockchain technology. *IEEE Trans. Robotics*, 1–17.

Ferrer, E. C., Rudovic, O., Hardjono, T., and Pentland, A. (2018). Robochain: A secure data-sharing framework for human-robot interaction. Corr. *abs/1802*, 04480.

Ferrer, E. C. (2016). *The blockchain: A new framework for robotic swarm systems*.

Haddadin, S., Albu-Schäffer, A., and Hirzinger, G. (2009). Requirements for safe robots: Measurements, analysis and new insights. *Int. J. Robotics Res.* 28 (11-12), 1507–1527. doi:10.1177/0278364909343970

Mayoral-Vilches, V., Carbajo, U. A., and Gil-Uriarte, E. (2020). "Industrial robot ransomware: Akerbeltz," in *2020 fourth IEEE international conference on robotic computing* (IRC), 432–435.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*.

Queralta, J. P., Li, Q., Ferrer, E. C., and Westerlund, T. (2022). Secure encoded instruction graphs for end-to-end data validation in autonomous robots. *IEEE Internet Things J.* 9 (18), 18028–18040. doi:10.1109/jiot.2022.3164545

Strobel, V., Ferrer, E. C., and Dorigo, M. (2020). Blockchain technology secures robot swarms: A comparison of consensus protocols and their resilience to byzantine robots. *Front. Robotics AI* 7, 54. doi:10.3389/frobt.2020.00054

Strobel, V., Ferrer, E. C., and Dorigo, M. (2018). "Managing byzantine robots via blockchain technology in a swarm robotics collective decision making scenario," in Proceedings of the 17th international Conference on autonomous Agents and MultiAgent systems, *AAMAS '18* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 541–549.

Vivek, S., Yanni, D., Yunker, P. J., and Silverberg, J. L. (2019). Cyberphysical risks of hacked internet-connected vehicles. *Phys. Rev. E* 100, 012316. doi:10.1103/physreve.100.012316

Yang, G-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., et al. (2018). The grand challenges of science robotics. *Sci. Robotics* 3 (14), eaar7650. doi:10.1126/scirobotics.aar7650