2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)
Virtual Conference, July 8-10, 2021

The Role of Cybersecurity and HPC in the Explainability of Autonomous Robots Behavior

Vicente Matellán,
Fundación Centro de Supercomputación
de Castilla y León (SCAYLE)
24071 León (Spain)
e-mail: vicente.matellan@scayle.es

Francisco-J.-Rodríguez-Lera and
Ángel-M.-Guerrero-Higueras
Robotics Group
Universidad de León (ULE)
24007 León (Spain)

Francisco-Martín Rico
and Jonatan Ginés
Intelligent Robotics Lab
Universidad Rey Juan Carlos (URJC)
28933 Fuenlabrada, (Spain)

2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO) | 978-1-6654-4953-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ARSO51874.2021.9542829

Abstract—**Autonomous robots are increasingly widespread in our society. These robots need to be safe, reliable, respectful of privacy, not manipulable by external agents, and capable of offering explanations of their behavior in order to be accountable and acceptable in our societies. Companies offering robotic services will need to provide mechanisms to address these issues using High Performance Computing (HPC) facilities, where logs and off-line forensic analysis could be addressed if required, but these solutions are still not available in software development frameworks for robots. The aim of this paper is to discuss the implications and interactions among cybersecurity, safety, and explainability with the goal of making autonomous robots more trustworthy.**

Index Terms—**Explainability, cybersecurity, HPC, robotics, forensic, ROS.**

I. Introduction

Unexpected *events* involving robots have been registered since the 1960s, with their introduction in manufacturing. It is assumed that the first human casualty caused by a robot was Mr. Robert Williams in January 1979 while working in a Ford Motor Company casting plant. That accident, after a long investigation, resulted in 15-million-dollar compensation payment, and the increase of safety measurements for the use of robots in the industry.

Current spreading on the use of robots is not centered on industrial robots kept in protected cages. Nowadays, there are mobile robots controlled by systems based on artificial intelligence (AI) sharing environments in our homes, shops, in the streets, etc. Besides collaborative robots (cobots), there are personal transportation systems, security robots, social robots…

Human-robot interaction is exponentially increasing, and the frequency of the "events" involving robots is going to keep growing, requiring, in addition to safety and security dimension, the inclusion of explicability systems that let us understand what has happened, and why, in order to let us trust these systems [4] and to make the entities that have built them accountable. Explainability will not guarantee [17] higher trust, but it will be more than a legal requirement.

It is unavoidable that some incidents will be caused by cyberattacks [9]. Cybersecurity is an increasingly important issue in technology and cyberattacks are widespread. They are becoming a serious problem for all kinds of infrastructures and organizations. The risk of cyberattacks targeting robots that can cause physical damage (intentionally or not) is very high. Nevertheless, robotics has been established in a "happy naivety" [10] for a long time. In the near future, safety, security and accountability mechanisms will be mandatory for autonomous robotic systems operating autonomous or semi-autonomously. New security tools will be required to provide confidentiality, authentication and integrity, to detect intrusions and to mitigate attacks in robotic systems.

Ubiquitous *cloud services* are already increasing their role in the management of robotics fleets. Cloud-controlled robots could have additional problems due to networking problems, real-time requirements, etc. and also on the privacy of the communications, or the cybersecurity of the communications. These issues have also to be considered for the safely deployment of services robots.

Cybersecurity is not only a potential cause of incidents. It also has to play a crucial role for assuring the integrity of the evidences in any incident and also for guaranteeing the privacy of the data gather for robots, both locally or in cloud-based systems.

Figure 1 summarizes the aim of this conceptual paper. There are three different disciplines we are considering in it. First, the area of explainable artificial intelligence (XAI), understood as the theories and systems for making AI systems' decisions understandable. In this paper we focus on *post hoc* explainability, that is, we are not interested on the robots explaining their future actions to users, but to be able to make a forensics analysis ofevents. A second area of interest is the cybersecurity of the robots. Once again, we are more interested in this paper on the role and the implications of possible attacks to the logs, than on the security of the robots themselves. And the third one is the role of the growing use of cloud services for controlling robots.

Next section discuses the relationship between cybersecurity and explainability, making a travel for the forensic analysis and preservation of logs, to the accountability of autonomous systems and their explainability. The third one focuses on the additional problems introduced by the use of cloud solutions, and also with computational requirements that could be required both form implementing those solu-

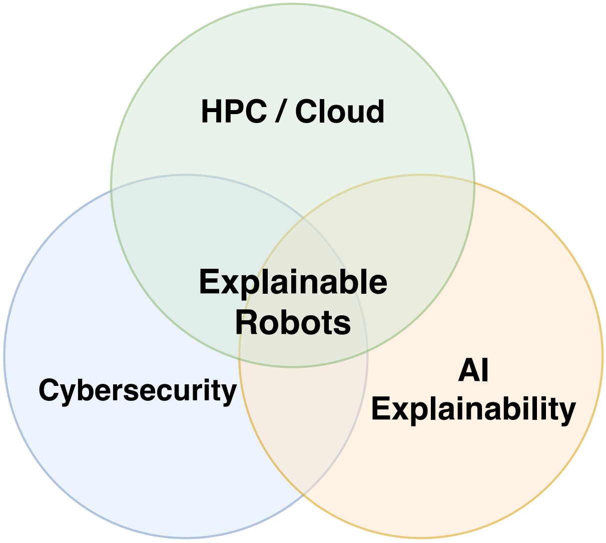

Fig. 1. Relationship among different concepts

tions and form the for investigation of the incidents, that could require high performance computing capabilities. In the fourth one, a glimpse of the architecture incorporating all these issues is described. Finally, the last section tries to offer some conclusions on these topics.

II. Cybersecurity and Explainability

Let's imagine a situation where a robot has had a problem (for instance a safety or privacy issue involving hurting or spying on a toddler it was taking care of). How can it be investigated? Logs of the robotic system would be a good starting point for forensic analysis, but this means that artificial intelligence systems should be able to generate rationales for their decisions, which leads to a hot research topic, the explainability of AI-based systems. Additionally, these logs could have been tampered with, so mechanisms to ensure their integrity should be included, as well as for guaranteeing that the cybersecurity of the robot has not been compromised. And if that problem is due to a malicious intrusion into the system that affects the robot's decision making? Could it have been detected? How could it have been avoided? How could the privacy and safety of people and the environment have been preserved?

Assuming that events will be unavoidable in autonomous robots, and that forensic investigation of the incidents will be very complex, it is crucial to put in place mechanisms to guarantee the accountability of the robots' and the explicability of the decisions taken by the software systems controlling them, based on artificial intelligence or in more deterministic algorithms.

This concept of "explicability" [15] is one of the more relevant open problems for the deployment of autonomous robots both in industry and service sectors, and in general for the use of artificial intelligence (XAI - Explainable Artificial Intelligence) [7]. In particular, systems based on machine learning are particularly challenging, even more, if they are based on neural networks.

And if that was not enough, explainability capabilities may be compromised by cyberattacks. Most explainability systems and almost every accountability system are based on *a posteriori* analysis of log files. But when a system is compromised, attackers usually try to forge the log files in order to delete or counterfeit logged events that could provide evidence of the attack (e.g. network addresses, vulnerable accounts, attack vectors, and so on).

We have proposed [12] an approach for arranging low-level knowledge from logs generated by ROS tools in the in order to be useful for developers and regulators. This approach, however, does not provide tamper-evident logs. We think that tamper-evident logs are fundamental for digital forensic triage and explainability. We propose the generation of authentication codes in a way that even if the code-generator is compromised, it is not possible to forge data pertaining to the past. The use of the Forward Integrity Model [3] will ensure that the explainability data generated before a possible exploitation is tamper-evident.

In the same way, we also think that Operational Modes will be useful to establish access to the certain sensitive privacy information and the tasks that a robot can perform based on the robot's level of security. During the operation of a robot, if it considers that its safety may have been compromised, it can establish operating modes in which the camera is not used or that the robot cannot access certain areas of the environment until the appropriate security recovery level is reached.

In summary, we think that tools and libraries for enforcing cybersecurity and accountability will have to be included in any middleware for autonomous robots as a primary brick. In the same way, the good practices for software for robots has to take into account not only the cybersecurity of the system, but also the preservation of the logs of the system, in the same was as the physical access to the robot or the security of communications have to be guranteed.

III. Cloud robotics and Explainability

Cloud robotics are becoming mainstream. Cloud-based robots can access remote resources to perform computational intensive tasks, and taking advantage of parallel processing in HPC facilities, big data analysis, etc. and also improving the performance of the on-site hardware, for instance by extending battery life. Also, many companies do not want to deal with the complexity of managing their robots, and they prefer to outsource the robotic services for their logistic or manufacturing needs.

Outsourcing robotics services is already usual in industrial companies, particularly for managing large robot fleets. It is also spreading in service robotics, where robots interacting with customers are remotely controlled, letting companies gather and analyze data. For instance, companies as Human-

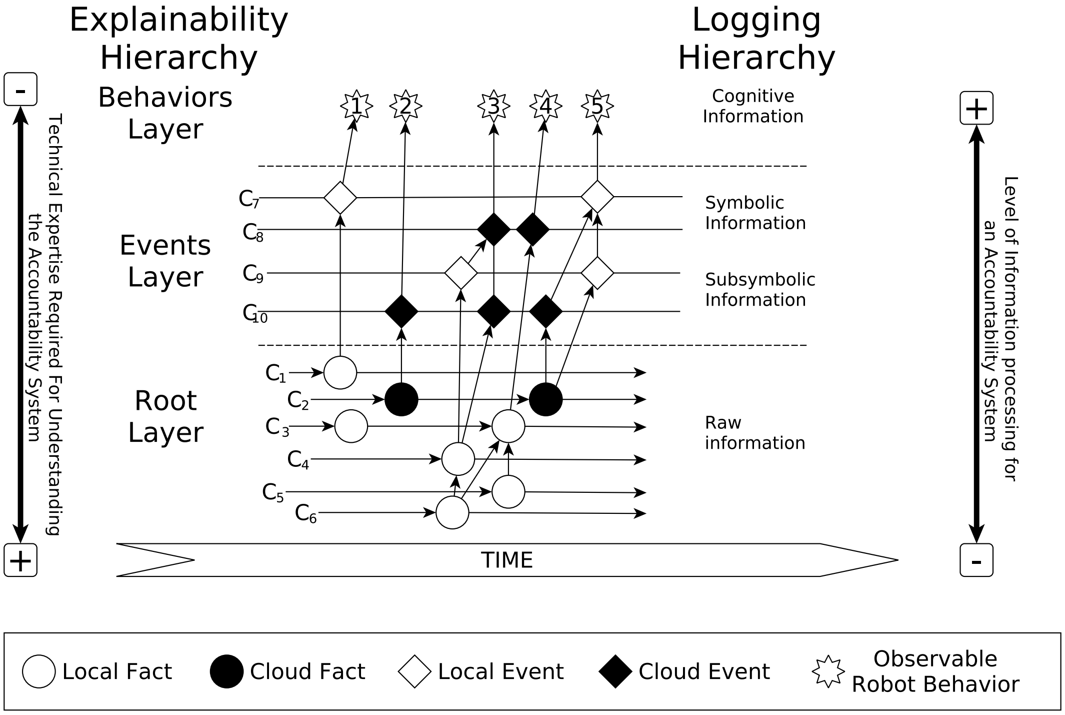

Fig. 2. Three level model of explainability

izing [1] offer management platforms for controlling social robots in retail, to work as receptionists, etc.

One of the problems of these cloud systems is that many robotic systems operate on the edge, where Internet access is not reliable or consistent, resulting in interruptions to network connectivity to the cloud. But we think that these problems, related with latencies or reliability are expected to be tackled with the spreading of 5G and edge computing technologies. But, on the other hand, the same problems discussed in the previous section applies to cloud-controlled robots: forensic analysis of incidents related to robots will required.

The need of preserving evidences of the decisions process made by robots are a more evident requirement in cloud environments. Not only forensic analysis could be needed in case of an event, but also because the companies hiring the outsource service wants to analyze the performance, the interaction with the clients, and many other parameters of the robots operating in their facilities.

There are already cloud logging systems commercially available for robotic systems, provided even by the cloud tycoons. For example, AWS RoboMaker‘s CloudWatch [2] extensions are ROS packages that enable log and metric data uploads to the cloud from remote robots.

But in cloud managed robotics other issues arise. In the first place, the trust that has to be place in the cloud operator. Trustworthiness on the provider has several dimensions. In the case of incidents, cloud providers should guarantee that evidences cannot be not tampered with, as discussed in the previous seciton, because these evidences are not in the physical robot. Providers should give not only principles declaration[2], but legal binding assurances to their clients. This is even more relevant when the physical robot and some of the services are not even in the same jurisdiction. In this way, we think that cloud services for robotics logging should be based on distributed blockchain, as the ones proposed in [16].

Another dimension in the cloud robotics market is the management of the privacy of the information involved in the human-robot interactions. Users may not even be aware that the robot is remotely controlled. These issues are analysed from different perspectives and in different domains. For instance, it is specially relevant in health-care cloud robotics [6].

Another issue to be considered are the robot learning processes that can be exploited in cloud environments. How could be ensured that learnt behaviors from different robots are appropriate for other domains? It will have to be a different level of accountability for learnt behaviors to ensure robot’s safety.

IV. Explainability Tools for Robotics

Explainability has been widely faced in the literature. In [1] it is proposed and discussed a taxonomy of the contributions related to the explainability of different machine learning models. But beyond these theoretical studies, and

[1] https://www.humanizing.com/software/

[2] https://ai.google/principles/

assuming that accidents in robotic systems will be unavoidable, and that forensic investigation of the incidents will be very complex, it is crucial to put in place mechanisms to guarantee the explicability of the robots' actions and the artificial intelligence systems controlling them.

Nevertheless, explainability capabilities may be useless in case of cyberattacks. When a system is compromised, attackers usually try to forge the log files in order to delete or counterfeit logged events that could provide evidence of the attack (e.g. network addresses, vulnerable accounts, attack vectors and so on). Tamper-evident logs are fundamental for digital forensic triage and explainability.

Explainable Security was proposed in [14] as an extension of XAI to security. This idea consider that explainability, when including the cybersecurity aspect, is multi-faceted domain that requires reasoning about system model, threat model and properties of security, privacy and trust.

However, there are no similar tools for robotics. We have focused our efforts on ROS because it has become the *de facto* standard framework for robotic software design, both in academia, in the industry, even in highly demanding environments as space robotics (NASA Viper robot will be using ROS[3]).

There are many modules, libraries, services, etc. built on ROS that provide solutions to the classic robotics problems: SLAM, navigation, manipulation, object recognition, etc. But, there are no tools for explainability available in ROS. ROS2, the new version of this standard, offers new features such as secure and real-time communications, predictability or portability, which were weaknesses of ROS that had prevented it from being adopted by the industry in critical systems, but it still lacks explainability libraries and tools.

We propose a three-dimensional model for the explainability of the behavior of autonomous robots. This model comprises three different elements that deal with the different levels of information, from the low-level, the $RootLayer$, which includes elements such as the messages exchanged by the components provided by ROS; to the highest level, where the complex and ah-hoc behaviors of the robot are implemented.

This hierarchy of explainability is shown on the left side of figure 2. It includes the basic analysis of logs previously proposed in [11] for accountability. Notwithstanding, the proposal presented here faces levels of abstraction associated with the explainability hierarchy: the logging hierarchy, the time, and the location where robot intelligence has place.

The example shown in figure 2 presents five observable robot behaviors, marked at the top of the diagram as stars 1-5. These behaviors have been generated by ten software components (named as C_i, where i is the id of the component). The figure also presents the concept of facts (circles) and events (diamonds). A set of 1 or n facts can generate an event. All these facts and events happen along time and should be concatenated in one way or another for generating events, which will finally produce an observable behavior (star).

An observable behavior is one that we can perceive as an ”output” produced by the robot (verbal or non-verbal). The five robot behaviors represented in figure 2 were generated by a set of software components that triggered an event (for instance recognizing a request from an individual or triggering a scheduled predefined behavior). These components, which can include very sophisticated AI-based systems, can run on-board or can be transferred to the cloud, hindering the process of explainability.

Software components implement controllers at different abstraction levels of the hierarchy, each one requiring a specific apparatus for explaining the different levels of a fact, an event, or a robot behavior. These apparatus will demand engineering techniques at $RootLayer$ (software or hardware) and AI approaches at event and behavioral layers.

On the other hand, the logging hierarchy (right side of figure 2) represents different levels of abstraction attending to the produced outputs of the logging system. A new level of complexity it is added to interact with the logging hierarchy in mixed or pure cloud based solutions.

According to this proposal, for instance, in a control architecture based on behaviors, each software component triggers different events, and the logging mapping is performed through an auditing process of the raw information from the logs. In this process, it is necessary to cross check the events generated by each component. Then, it is possible to link the behavior generated as a result of a set of these events.

V. Conclusion

Rephrasing what Robyn Speer wrote about the ConceptNet ambitions [13], we want to avoid letting robots be awful to people, just because people are awful to people. Besides from implementing ethic rules in robots' control systems, they have to be accountable, so regulations and tools have to be put in place to let forensic investigators understand why autonomous robot have taken. New methodologies and tools are required, as well as their implementation in software to guarantee explicability. We have described in this paper some basic tools for ROS that we have developed and that need to be extended, as well as the roadmap to improve them.

Regarding the regulations, some governments are starting to create registries for autonomous systems. For instance, the FAA in the USA has established a registry for UAVs, or the Dept. of Transportation in San Jose [8] has a created a registry for autonomous mobile robots, in particular for kiwibot[4]. The European Parlament [5] in its ”Civil Law Rules on Robotics” recommends a comprehensive Union system of registration of advanced robots should be introduced within the Union's internal market.

[3] https://www.nasa.gov/feature/ames/vipers-many-brains-are-better-than-one

[4] https://www.kiwibot.com/

We have also presented why cloud-based robotics is completely different from regular cloud computing. Robots can move around by themselves, can interact with people, gather information about the environment, etc. Offering *robot as a service (RaaS* by well established cloud providers as the one mentioned (Amazon, Google, etc.) would required a clear definition of the roles and specific regulation.

Acknowledgment

This work has been partially funded by the Kingdom of Spain under grant RTI2018-100683-B-I00, and by INCIBE SA under grant Adenda 4.

References

[1] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Richard, B. Chatila, R., Herrera, F. (2020). *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information Fusion, Vol. 58, pp. 82-115.

[2] Devin Bonnie and Camilo Buscaron. (2019). *AWS RoboMaker's CloudWatch ROS nodes with offline support*. In AWS Open Source Blog. Last checked December 2020.

[3] M. Bellare and B. S. Yee, (1997). *Forward integrity for secure audit logs*. University of California at San Diego, Technical Report.

[4] Devitt, S. K. (2018). *Trustworthiness of autonomous systems*. In Foundations of trusted autonomy (pp. 161-184). Springer, Cham

[5] Resolution on Civil Law Rules on Robotics European Parliament. resolution with recommendations to the EC (2015/2103(INL) (2017). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

[6] Fosch, E., Felzman, H. Mahler, T., and Ramos M. (2018) *Cloud services for robotic nurses? Assessing legal and ethical issues in the use of cloud services for healthcare robots*. Proc. 2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), pp. 290-296. doi: 10.1109/IROS.2018.8593591

[7] Gunning, D. Stefik, M. Choi, J., Miller, T. Stumpf, S. y Yang G.Z. (2019) *XAI—Explainable artificial intelligence*.In Science Robotics 18. Vol. 4, Issue 37. doi: 10.1126/scirobotics.aay7120

[8] Higginbotham, Stacey (2020). *Who's Behind that robot?*. In IEEE Spectrum, Vol. 20, Num. 12, pag. 24. D

[9] Matellán, V. Bonaci, T. and Sabaliauskaite, G. (2018). *Cyber-security in robotics and autonomous systems*. Robotics and Autonomous Systems. Volume 100. doi: 10.1016/j.robot.2017.10.020.

[10] Morante, S., Victores, J. G. and Balager, C. (2015). *Cryptobotics: why robots need safety*. Frontiers in Robotics and AI, 29. doi: 10.3389/frobt.2015.00023.

[11] Rodríguez-Lera F.J., Guerrero-Higueras Á.M., Martín-Rico F., Gines J., Sierra J.F.G., Matellán V. (2020) *Adapting ROS Logs to Facilitate Transparency and Accountability in Service Robotics*. In: Silva M., Luís Lima J., Reis L., Sanfeliu A., Tardioli D. (eds) Robot 2019: Fourth Iberian Robotics Conference. ROBOT 2019. Advances in Intelligent Systems and Computing, vol 1093. Springer, Cham. doi: 10.1007/978-3-030-36150-1_48

[12] Rodríguez-Lera F.J., González Santamarta M.Á., Guerrero Á.M., Martín F., Matellán V. (2021). *Traceability and Accountability in Autonomous Agents*. In the 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2020). CISIS 2019. Advances in Intelligent Systems and Computing, vol 1267. Springer, Cham. doi: 10.1007/978-3-030-57805-3_28

[13] Robyn Speer. *ConceptNet Numberbatch 17.04: better, less-stereotyped word vectors*. ConceptNet Blog, last checked, December 2020.

[14] Vigano, L., and Magazzeni, D. 2018). *Explainable Security*. In IJCAI/ECAI 2018 Workshop on Explainable Artificial Intelligence (XAI).

[15] Wachter, S., Mittelstadt, B., and Floridi, L. (2017). *Transparent, explainable, and accountable AI for robotics*. Science Robotics, 2(6). doi: 10.1126/scirobotics.aan6080.

[16] White, R. Caiazza, G. Cortesi, A. Cho Y.I. and Christensen H.I. *Black Block Recorder: Immutable Black Box Logging for Robots via Blockchain*, in IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 3812-3819. doi: 10.1109/LRA.2019.2928780.

[17] Alan R. Wagner, and Paul Robinette. *An explanation is not an excuse: Trust calibration in an age of transparent robots.* In ”Trust in Human-Robot Interaction”, pp. 197-208. Chang S. Nam, Joseph B. Lyons editors. Academic Press, 2021. doi: 10.1016/B978-0-12-819472-0.00009-5.