Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 195–205 DOI: 10.15439/2024F3165 ISSN 2300-5963 ACSIS, Vol. 39

Trust Management Framework for Multi-Robot Systems

Daniel Vojnar, Adela Bierska, and Barbora Buhnova
ORCIDs: 0009-0009-1053-208X, 0009-0009-3119-530X, 0000-0003-4205-101X
Masaryk University, Faculty of Informatics
Brno, Czech Republic
Email: {vojnar, bierska, buhnova}@mail.muni.cz

***Abstract*—As autonomous technologies blend into a variety of industries, the employment of Multi-Robot Systems (MRS) in complex tasks is becoming increasingly prevalent. These systems, characterized by their distributed intelligence and collaborative capabilities, are now largely deployed in executing critical missions ranging from environmental exploration to disaster response. To support the trustworthy execution of such collaborative multi-robot missions, the existence of an underlying trust management framework is becoming imperative, given the rising risks of malicious intruders attempting to join the MRS.**

This paper proposes a Trust Management Framework (TMF) for Multi-Robot Systems, to bridge the gap of scarce trust management support for MRS. By analyzing trust dynamics and integrating direct and indirect trust with contextual data, we address the trust vulnerabilities inherent to MRS. Our primary contribution is the development of a novel TMF significantly supporting the trust, security and reliability of MRS. The paper systematically outlines the evolution of our TMF, from theoretical underpinnings to an integrated solution, and the discussion of its impact on the trustworthiness of cooperative robotic networks.

I. Introduction

THE proliferation of intelligent and independent devices, vehicles, and robots is reshaping our everyday lives [1], [2]. This advancement leads to fully digital environments that can collaborate, compete, or even pose a threat [3], [4], [5]. Networks of multiple robots are advancing to take on human responsibilities in everyday activities and hazardous missions, including exploring underwater terrains, responding to natural disasters, and carrying out military operations [2].

The strength of Multi-Robot Systems (MRS) [6] lies in their distributed nature. Robots with unique capabilities possess different strengths and limitations, which enhances task success and system robustness. Robots have multiple sensors to sense their surroundings and plan the steps based on the mission and actual data. Communication is vital in these systems, enabling them to collaborate and adapt their behavior based on agreement and adaptive decision-making, which at the same time makes these systems vulnerable to robot misbehavior, whether unintentional or underpinned by hidden malicious intents.

Despite the increasing digitalization and interconnectedness of MRS, the integration of Trust Management Frameworks (TMF) [7] tailored for these systems remains largely unexplored. This gap exposes MRS to vulnerabilities, risking their robustness and resilience against attacks. Current trust management approaches, while foundational, fall short in MRS contexts and missing understanding of their applicability and limitations in the context of MRS. These approaches also need to pay more attention to prioritizing individual safety of the members within these systems and the surrounding ecosystems.

To bridge this gap, this paper proposes a tailored Trust Management Framework (TMF) for Multi-Robot Systems (MRS), with the intent to overcome the discussed limitations by proposing a TMF that enhances the trustworthiness and safety of MRS through comprehensive trust inputs, including reputation and peer opinion, thereby ensuring the integrity and resilience of MRS in complex digital environments. To this end, we base our design on a general baseline TMF [7], to which we map the mechanisms necessary to prevent trust attacks in MRS scenarios, as collected in our previous work [8]. Specifically, we focus on the mechanisms necessary to support direct and indirect trust while paying attention to the contextual information necessary in trust decisions. Our main contribution is the creation of a novel Trust Management Framework to prevent vulnerabilities and enhance the safety of multi-robot systems. The proposed TMF is intended for feasible deployment and utilization while ensuring optimal trust and safety measures.

The rest of the paper is structured as follows. After reviewing related work in Section II, the background definition of multi-robot systems and robot capabilities for our specific purpose are provided in Section III. Section IV presents the methodology of the TMF design process, followed with individual components of the TMF design in Section V and VI. The resulting TMF for MRS is presented in Section VII, together with the discussion in Section VIII and conclusion in Section IX.

II. Related Work

Machine-to-machine trust is crucial for successfully implementing fully autonomous systems involving multi-robot systems. The practical implementation of such trust was initially explored by Yang and Parasuraman in 2021 [9], who developed an agent trust model to facilitate cooperation among heterogeneous multi-robots. Their research focuses on trust evaluation based on the robot's needs (safety, basic, teaming, and capability). However, it could be enhanced by other aspects like reputation or peer opinion, which could help provide solutions for unaddressed system vulnerabilities.

Recent studies delve deeper into the theoretical foundations of trust in autonomous systems [10], which focus on robot actions' reliability, predictability, and transparency and allow systems to form relationships. Besides, many studies can be found in the SIoT (Social Internet of Things) [11], which mimic human relationships. The study [12] focuses on applying aerial drone swarms to improve public safety by crowd monitoring and disaster response [13], requiring robust and reliable social interactions among robots. The [14] focused on trust-based mechanisms based on the blockchain model. Although their solution has outperformed various measures such as throughput, latency, accuracy, and block updating limits, it is based only on the local trust of its neighbors. Transitive relationships could improve the algorithm, degrading its performance but providing a more secure approach.

In contrast, the study [15] addresses the usage of Hidden Markov Model (HMM) in trust management for underwater robotics. This model enables the measurement of the trustworthiness of the sensor nodes to combat malicious or internal attacks. In this direction again, peer opinion could extend the robustness of trust management based on the HMM approach.

Ground-based networks, such as Vehicular Ad-hoc Networks (VANETs), allow vehicles to form human-like relationships, improving navigational aids and emergency response capabilities and ensuring safety for drivers and pedestrians [16]. Aslan and Sen designed a dynamic trust management model for vehicular ad-hoc networks [17] where the trustworthiness of vehicles is assessed using the vehicle trust value based on the data trust values of their event messages. This helps to establish a more reliable trust management framework with a combined trust model. While the model is a good source for inspiration, given its different context, it falls short in supporting details necessary in trust management in collaborative robotic missions (integrating the details of the mission, robot role in it, directives of the leader, etc.).

III. Background

This section introduces the reader to multi-robot systems' essential terminology and background.

A. Multi-Robot Systems

The effectiveness of multi-robot systems is attributable to their composition of diverse robots, each enriched with distinct capabilities, enabling a synergistic collaboration that surpasses the abilities of individual robots. This significantly improves the success of a task and relaxes the dependence on a single robot, focusing on the cooperation of many distinct robots instead. As a result, the system becomes more robust and less likely to fail if one of the robots experiences an issue [18]. The distribution of robot types in the system is also essential, as different types of robots have different capabilities and limitations. Communication is a crucial aspect of multi-robot systems, which can be either seamless or introduce challenges, depending on the compatibility of robot communication mechanisms and protocols. Yet, technology exists to facilitate communication between different communication protocols [19]. Efficient communication can significantly enhance the system's capabilities and increase its efficiency. Additionally, robots in multi-robot systems can recognize each other, known as kin recognition [20], and understand peer capabilities and limitations, further contributing to successful task completion.

B. Characteristics of a Robot in MRS

Inspired by [21], robot characteristics can be summarized as follows.

a) Capabilities: Robots are aware of their surroundings. They can generate and store information and combine it in real-time with new knowledge. They can plan their steps based on the mission and actual data. They have a rechargeable battery with limited capacity. They can move, orientate in the environment, and generate a trajectory considering detected obstacles. They are often capable of grabbing and carrying an (appropriately sized) item.

b) Sensors: Robots can sense their position using GPS and IMU. They can have any combination of additional sensors, such as LIDAR, radar, camera, microphone, temperature sensors, or range finders [22]. They can detect and diagnose internal faults and battery level.

c) Communication: Robots can communicate locally with other robots if the communication method is compatible. They can communicate on a private, ciphered channel with other robots with access to it. If possible, they can connect to a shared source of information and exchange data with it, both ways (download and upload). They can also store or backup data on remote storage. However, these remote connections may be impossible due to the bad conditions of the surrounding environment, inaccessible channels, or being too far away.

d) Collaboration: The robots are capable of collaboration with another robot or group of robots. They can consider the actions of other individuals while planning. They can participate in decision-making and adapt their behavior to the agreement. They can process and store information obtained from other robots.

IV. Methodology

The trust management concept is well-defined in various domains such as Internet of Everything or Social Internet of Things [7] but remains unexplored in multi-robot systems. With the ever-increasing digitalization and communication between systems, there is a need to ensure integration of appropriate trust management into such systems to increase their robustness and resilience for their safer integration in our digitalized society. To obtain an applicable solution for trust management in MRS, our Trust-Management Framework (TMF) design process is based on an existing trust framework, which is scrutinized from two directions to (1) validate whether each of its current components is necessary for trust

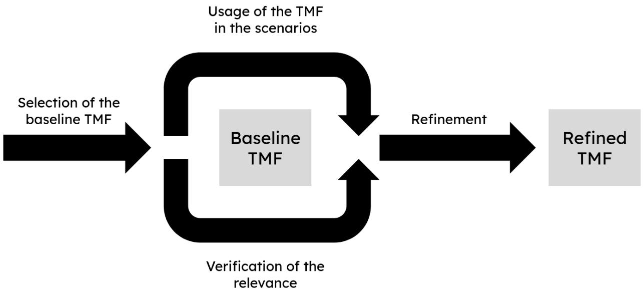

Fig. 1. Schema of the methodology.

management in MRS, and (2) identify missing properties and mechanisms specific to the context of MRS. Based on this analysis, a TMF for MRS solution is designed. These steps are visualized in Figure 1 and described below.

1) Select the baseline TMF: First, we identified a baseline trust management framework, with the intention to lay out a holistic foundation that can be then filtered and extended based on the specifics of MRS. To this end, we have reached for a more general context of the Internet of Everything (and Social Internet of Things) and chose a TMF designed as a collection of concepts from a broad literature review [7], [23]. In the TMF, depicted in Figure 2, trust of one agent (trustor) in another (trustee) is being built via combining the mechanisms of direct trust (based on local and highly context-specific experience with the other agent) and indirect trust (possibly global reputation of the agent) [24]. The two are later combined via the component of trust decision. The TMF components can be described as follows [7].

- *Direct Trust* constitutes an individual assessment by the trustor, derived from direct interaction with the trustee. It specifically stems from a combination of present and past experience (direct observation) that the trustor has with the trustee. Consequently, it is imperative to establish mechanisms that evaluate the experience during real-time interactions between the trustor and the trustee (potentially emulating human cognitive functions) without exposing its vulnerabilities.
- *Indirect Trust* is deduced from propagated opinions across various trust paths to assess a trustee's reputation. The principal sources are the trustor's trusted peers (their opinions and recommendations) and an overseeing authority responsible for monitoring the trustee's reputation within the network. Accordingly, there must be systems to update the reputation and propagate it throughout the ecosystem.
- *Context Information* shapes the trust decision to mirror the trustor's present circumstances during the decision-making process (for instance, the trustor's vulnerability during interactions with the trustee, the risks involved, the accountability of the trustee in the event of malicious actions, and whether any potential damage is reversible or subject to compensation, ensuring reparability). Furthermore, broader contextual information also influences the direct and indirect trust computation.

2) Analysis of the baseline TMF in the context of MRS: To detect the necessary areas for change of the baseline TMF, we investigated its coverage of MRS context, which was done in two steps:

- *Relevance of existing components.* In the first step, we verified that the TMF contains no parts that could be irrelevant to the MRS. For each component in the TMF, we asked: *How can this component (direct trust/indirect trust/context information) be used to promote trust management in MRS?* and compiled answers with the help of knowledge of MRS from existing literature. A sufficient answer to this question denotes that the discussed component has its yield and, therefore, should remain in some form in the final solution.
- *Missing concepts and blind spots.* The second part of the analysis consists of detecting gaps in the coverage of the MRS context, which could cause the TMS to be insufficient. We went through the collection of the scenarios and vulnerabilities in multi-robot systems and its application in swarm robotics [8], and for each scenario and component, we asked: *How can this component be used to prevent or mitigate this type of trust attack in MRS? Is there anything missing in the original TMF that would ease this process?* Exhaustive answers to these questions serve as a base for updating the baseline TMF in the next step.

3) Refined TMF for MRS Design: Ultimately, we use the acquired knowledge to refine the baseline TMF and define the resulting TMF for MRS. The components of the baseline TMF, which were verified as valuable, serve as the base for the refined TMF. From the scenario analysis, we collect observations of the necessary robot support in trustworthy mission execution and cluster them by their source or subject. Each of the clusters gets assigned a component in the original TMF that it shall extend, or an approval to form a new component in the resulting TMF. Using these clusters, we refine the structure of the assigned components to cover them sufficiently.

V. Analysis of the Baseline TMF

This section examines the baseline Trust-Management Framework (TMF), as depicted in Figure 2, to validate whether each of its current components has its role in the context of trust management in Multi-Robot Systems (MRS).

A. Direct Trust

Direct trust is the main component by which the robot can judge other robots. Direct trust consists of two parts: present experience [25] and past experience [26]. Robots can use their current experience in MRS to evaluate their direct interactions with other robots. This can help to give more weight to the ongoing interaction. If a robot only bases its decisions on past interactions with another robot (recorded by past experience)

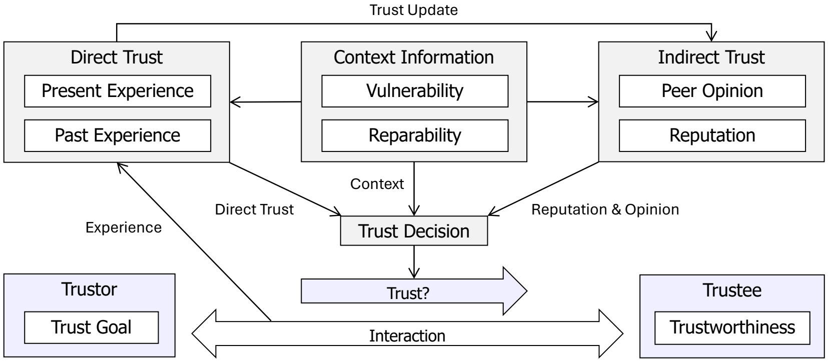

Fig. 2. Baseline Trust Management Framework [7].

and the other robot misbehaves in the current situation, the system might not recognize the misbehavior, and an error could occur. The trap of the experience mentioned above is a crucial component in MRS, as there are many one-to-one interactions between robots, which can provide the robot with significant experience that can help it evaluate trust. This shows that both direct trust components of the baseline TMF are relevant and contribute to correct functioning.

B. Indirect Trust

In the MRS, robots can use direct and indirect trust [27] to evaluate each other. While direct trust comes from personal interaction, indirect trust comes from reports, recommendations, and reputation mediated by others. Robots can leverage indirect trust by relying on peer opinions [28] to decide whether to trust an individual. They can also establish a hierarchy based on individuals' reputations [28], which can increase when a robot performs its tasks well. If a robot completes most of a task, even if not explicitly designed for it, its reputation can reflect it and increase. All baseline TMF aspects are relevant and suitable for the MRS.

C. Context Information

For MRS, context information can carry the same significance as other systems. It can serve as an essential but non-negligible aspect in the process of making trustworthy decisions. Context information can aid in evaluating the suitability, risks [29], and potential advantages of an action or interaction within a specific environment [30] and with specific objects. Environmental factors, information sensitivity [31], and the likelihood of recovering [7] from a risky action can be considered in this context. Additionally, the types of robots in our environment can influence MRS, as each individual may be suited to different tasks, leading to varying levels of security perception. In summary, all the components of the context information baseline TMF remain relevant also in the context of MRS.

VI. Trust Scenarios Mapping to the Baseline TMF

This section presents the analysis of the baseline TMF with the intention to identify missing concepts and blind spots relevant to the context of MRS. It focuses on individual vulnerabilities occurring in multi-robot systems, as collected in [8]. The text is divided into several sections, each explaining the issue and providing a detailed description of the trust-management solution and its mapping to the baseline TMF.

A. Information Manipulation or Ignoring

The improper modification and dissemination of information can cause significant problems during a mission. The transmitted data can be divided into three categories of actions related to erroneous information: individual, external, and passing information [8].

The presence of a Trust Management Framework (TMF) can support the trustor to cross-check information obtained from the trustee with other sources. Therefore, it can point out inconsistencies in the information, which can lead to discarding false information and thereby foiling the attack. The type of mission sets the sensitivity of the data, so it should be considered in the context part of the TMF for MRS. Also, only insiders (trusted peers) may recognize some defects in information as they have the necessary permissions to access, so in the TMF for MRS, they should be differentiated from random passersbys.

1) Individual Information: Let us first examine the TMF robustness against the attack of a robot sharing false individual information. The purpose of this attack is for the robot to share false information about itself that others cannot verify, such as its battery level. The TMF can be very beneficial in this case, as it can use its recorded past interactions with the robot, compare information with other sources [32], and send a test message to determine if the robot is generally truthful. If the system is centralized [33], the robot may have to be asked to answer to a higher authority with more information about itself.

2) External Information: Information manipulation or ignoring attacks on external information are similar to the previous type of attack. In this scenario, the robot observes [34] its surroundings and shares false information about other robots or objects in the environment, which can however be verifiable by others. Since the information is verifiable, other robots can verify the information and update the source robot trust scores accordingly. The impact on behavior depends on environmental conditions and possibly on its changes (e.g., weather, moving objects), which are currently not included in the baseline TMF.

3) Passing the Information: In this type of attack, the robot receives information from other sources, such as another robot or a server, and then passes it along to the next recipient, which is the moment when it can change it to false information. In this scenario, the robot is not the original source of the information but rather a mediator that can modify the transmitted data. An important aspect of robot interactions is the direct experience with a robot that consistently and accurately transmits information without making unwarranted modifications. Trust values from other peers can be received through feedback [32], reputation ratings [35], or recommendations [36]. The authority can select the primary message distributors based on their trust level if the message is not broadcast. When the robot transmits sensitive data, like navigational information, the communication should be more reliable, which can have a different impact on the trust score [31].

B. Manipulation with Communication Channels

Effective communication and information exchange are crucial for mission success. However, vulnerabilities in data passing and storage can be exploited. Even a minor leak of swarm position or resources may lead to significant losses. Insider attacks and changes made to shared information can be significant risks. Also, interference with communication and signals could threaten the mission [8].

Observation of malicious behavior against communication channels by the trustor or their peers, followed by the isolation of the trustee, may stop those attacks from spreading. However, detection of this behavior gets more complex with various permissions and chosen communication channels. Therefore, the TMF for MRS should go into more detail with the context of the mission and roles of the robots.

1) Leak of Information: The trustor needs to observe trustee's history of managing and transmitting sensitive information without exposing it. To detect vulnerable behaviors with the information, storing every robot's action globally and locally is necessary. It will be a marked comparison of these historical actions [37] to see if they are the same or if there are any incorrect things.

Indirect trust can enhance information's credibility by allowing other robots' feedback and reports. This would also enable the implementation of an external security system to monitor unusual communication patterns [37]. Robots involved in previous security breaches or have indirect evidence of such behavior would be considered less trustworthy, even if there is no direct observation of their actions.

The sensitivity of the transferred information is essential for evaluating the manipulation with it as the more sensitive data should be protected better. In the current TMF, this is not completely covered as the context part aims more at the trustor itself than its mission.

2) Changing of Shared Information: Many algorithms often work together using shared information. However, this shared source can be vulnerable to manipulation, leading the group to work with false information.

The trustor should observe the trustee's correct manipulation in updating and handling the information, which enhances the group's problem-solving capabilities. If the trustee is caught making unauthorized or harmful changes in the past, it should take shape in the future evaluations of their trustworthiness.

The shared information should be checked and compared with previous versions, and the changes that were caused to the data should be evaluated. If other robots suspect a change made to the information, they can share that with each other.

3) Restrain Access: The work of MRS relies on distributed communication and information sharing, making interference a real threat. Malicious robots can disrupt signal transmission by overloading the communication channel with irrelevant messages.

There are always multiple communication channels between robots; the higher the trust score of a robot, the more channels are available. The higher the level of the communication channel, the more messages can pass through it with a higher level of confidentiality. The origin and content of messages should be observed to detect overflowing by irrelevant or repeating messages. The evaluation is based on the context of the mission and its presumed communication.

Malicious robots can also disturb communication by noise or physical destruction of transportation medium. Their behavior and sent signals should be observed and evaluated.

The trustor may verify sent messages with peers on other more secure channels to which only they can access.

C. False Performance Promises

In time-critical missions, there is a risk that some individuals may intentionally fail to complete tasks on time. Other robots failing to identify and address this malicious behavior can lead to unnecessary time loss and mission failure. The more time-critical the mission, the more vulnerable it is to this attack. The suitability of the environment and the trustee for the mission can also impact the time limit that the trustor is willing to wait for the mission to be completed, which is connected to the vulnerability of the whole system and the individual waiting for the completion of an action to continue with his task. The time and status of the mission are reflected in the timeout for action. These aspects should be included in the baseline Trust Management Framework (TMF).

The robot's past experience determines if a delay is typical. This past experience can be shared among all system robots, enabling them to share opinions about the targeted individual.

If robots are visibly damaged or noticeably delayed in their work, it has to be factored into the trust decision. Furthermore, the delay can be compared to a pre-estimated time limit.

D. Authority Misusage

In swarm systems with hierarchical leadership, malicious robots in leader roles can manipulate others to do unwanted things, making detection and elimination harder. Malicious leaders may change permissions and force subordinates into unwanted actions, endangering the mission.

This scenario is the least covered by the baseline TMF as it does not recognize the trustee's relation to the mission and the role of the leader as a whole. But it can still detect suspicious behavior of the leader and cause their demotion.

1) Permission Regulations: Each robot should have permission to access resources set according to its role or mission. A malicious robot with authority may change the rights of other robots, denying access to information or resources. If these permissions are not granted or are taken away without clear reason, it should be considered suspicious. These rights assignments could be taken as direct experience within the TMF.

To evaluate their own rights, the trustor should take into account the value and sensitivity of the data used within the mission. It needs to be as objective as possible because even if the robot thinks it needs a particular resource, it does not mean there is no way to solve the mission without it. The baseline TMF covers this sensitivity to some extent through context information.

When introducing the leader hierarchy, the trustor can ask sub or superior leaders for their opinion. Leaders' opinions may have a higher impact on the decision or even be unquestionable. Therefore, this weight could be somehow recognized in the TMF.

Also, the trustor can observe inadequate changes in the permissions of their peers. In that case, it may signify the leader's (trustee's) incompetency or malicious intentions, which are both problematic for the mission's success. The baseline TMF covers this by the peer opinion part.

2) Task Allocation: Getting orders from the leader is usually done via direct communication, and if the trustor knows and understands the mission's aims and plans, they can directly evaluate this kind of interaction. The orders should be consistent with previous ones and any changes should have a reason. The leader should have a plan and reasons for task assignment. If they encourage the trustor to do irrelevant or dangerous tasks, they should provide more details, eventually even evidence that a particular approach does not lead to the destruction of the society.

As the missions of MRSs are, by definition, collaborative, so orders from leaders should take into account the plans of other group members and complement them efficiently. Therefore, references from peers and leaders within the same mission are more valuable for detecting this attack. If the assigned task is irrelevant or even complicates or devalues the work of other group members, it should be considered suspicious. However, similarly to the previous section, the task evaluation should be as objective as possible because the mission may be in a critical state, which requires more extreme solutions.

E. Physical Attacks

In robotics, physical attacks can range from inter-robot communication manipulation to the destruction of robots or even the kidnapping of individual units. These attacks can significantly delay a robot's mission, including stealthy tactics or leading robots into traps. Additionally, changing or destroying the environment can make it difficult for robots to navigate and complete their mission.

Components in the TMF cover all necessary information to detect possible physical attacks. The trustor evaluates their vulnerability and reparability and, based on past or present destructive behavior of the trustee or their bad reputation, may decide to avoid them or even isolate them.

1) Robot Destruction: Aggressivity and destructive intentions may be long-term, especially in the case of trustee malfunction. Therefore, the trustor should consider previous endangering behavior covered by past experience in the TMF. Present experience is also important, the trustor should observe the possible physical superiority of the trustee and their behavior. Even false information or risky instructions obtained from the trustee may mean a deliberate effort to destroy the trustor.

Physical attacks are usually critical and unmistakable with other actions. Any peer can similarly evaluate the situation and refer it to the trustor. Therefore, the opinion of robots with and without knowledge of the mission and context can be considered valuable without much doubt.

The actual state and capabilities of the trustor directly influence its ability to defend itself against the destructive trustee. Also, the trustor's importance for the mission should be considered in the final trust decision. The surrounding environment and mission type influence the view of the trustee's communication. For example, risky instruction may be more tolerable in dangerous environments under time pressure.

2) Kidnapping/Capture: Similarly to the destruction attacks, kidnapping may be predicted by the uneven physical capabilities of the trustor and trustee and by suspicious communication. Exhortation to move to unknown places should especially be critically evaluated.

The risk of being kidnapped impels the trustor to consider their value. They may carry secret information or know-how that may be misused in the wrong hands. Otherwise, the context may be weighted as in case of possible destruction.

The escaped peers will provide information about the kidnapper's behavior. Therefore, the trustor should consider the peer's opinion. Also, the size and strength of the surrounding group can be considered, as the trustor alone is more vulnerable to kidnapping than the whole group.

3) Changing or Destruction of the Environment: Spotting trustee to manipulate their surroundings inadequately may signal their destructive intentions. Also, observing new environmental changes may lead to suspecting nearby robots.

In addition to warnings from witnesses of the malicious behavior, the trustor's peers may also provide information about the environment's last state and the items' positions. Based on this, any changes may be observed without the previous visit to the particular place.

The environment may contain items of various value, and the trustor should evaluate their importance according to the

TABLE I
COLLECTED OBSERVATIONS SORTED BY TMF PART

Direct Trust	Context Information	Indirect Trust
Trustee's motion	State of the environment	Communication with other agents
Communication with trustee	Trustor's experience with similar environment	Feedback loops
Trustor's historical experience with trustee	Static vs dynamic object	Reputation
Accurancy of provided information	Sensitivity of data	Involvement in previous security breaches
Integrity of provided information	Riskiness of the mission	History of shared information source
Trustee's behavior	Discovery of information leak	Other leaders' opinion
Trustee's knowledge	Credibility of provided information	Missing peers
Unauthorized or harmful change of shared information	Trustees's access to information	Leader's opinion
Amount of sent information from trustee	Relevance of information	Destroyed peers
Capabilities of trustee	Author of information	
State of trustee	Time left	
Trustees opinions	Trustor's access to information	
	Threat to other robots	
	Efficiency	
	Mission's state	
	Trustor's state	
	Trustor's capabilities	
	Changes of environment	

mission. Changing or destruction of the environment may not always mean bad intentions. There could be, for example, some obstacles or dangerous objects intended to be displaced or destroyed.

F. Attacks on Internal Intelligence

Machine learning models have vulnerabilities that lead to biased or irrational behavior. To prevent this, it is important to stop false data from reaching the robot's learning model and to verify information with trusted sources. It is also crucial to monitor the robot's behavior for signs of exploitation of known vulnerabilities in its control mechanisms [38]. The robots have past experiences stored in their memory, which includes their encounters with malicious data. The distributed intelligence of the entire group, including the robots' internal intelligence, must be well-protected to prevent attacks by malicious entities. This involves addressing complex decision-making under uncertainty and minimizing information leakage [39], [40].

G. Decision Making Attacks

The decision-making process of robots can be influenced by various types of attacks, such as contrarian behavior, wishy-washy attitudes, following a sect [41], and going along with the majority opinion [8]. It is essential to monitor the trustee's opinions over time to detect any potentially harmful patterns aimed at manipulating decisions. Opinions can be affected by the current context, the mission status, and the trustee's relationship to it, which are essential factors currently not considered. Peer opinions can also carry weight if the peer is trusted and working with the same information and mission. Therefore, understanding the trustor-peer relationship can enhance the effectiveness of the decision-making process. Bots can omit or exclude a trustee from voting if past experiences show the trustee attempted to abstain. The history of individuals plays a significant role in preventing these kinds of attacks. When evaluating opinions, it is crucial to consider the context and recognize that initially controversial opinions may be supported by concrete steps or facts.

H. Summary of Observations

Overall, we have identified 39 observations that characterize the specifics of MRS context in trust management across all the examined scenarios. The collected observations were then clustered by their affiliation to a particular component of the baseline TMF, as presented in Table I.

Main topics missing by baseline TMF are mission information (e.g. its state, plan, kind of data) and different levels inside of the mission hierarchy (leaders, team members, other passerby.

VII. TRUST MANAGEMENT FRAMEWORK FOR MULTI-ROBOT SYSTEMS

This section presents the proposed Trust Management Framework (TMF) for Multi-Robot Systems (MRS), reflecting the observations from Section V and VI, as well as the mapping of the observations to TMF components as presented in Table I. The resulting TMF for MRS is presented in Figure 3.

In summary, the observations led to a more detailed breakdown of the *direct trust* TMF component, where the *present experience* is divided into *communication*, the accuracy and credibility of which should be evaluated by the *trustor* as the *trustee* may easily lie, and *observation*, that does not have to be directly questioned because it comes from a reliable source (*trustor* itself).

Mission information expanded the *context information* component, as the state of the mission directly influences the need for collaboration with other robots and overall safety requirements. The *mission information* consists of three main aspects: *mission state*, *time left*, and *data sensitivity*, described in detail below.

Besides, as the robots may play different roles within the mission with different roles to the *trustor*, the TMF was extended to reflect these roles. The rest of this section presents the TMF components in a structured way.

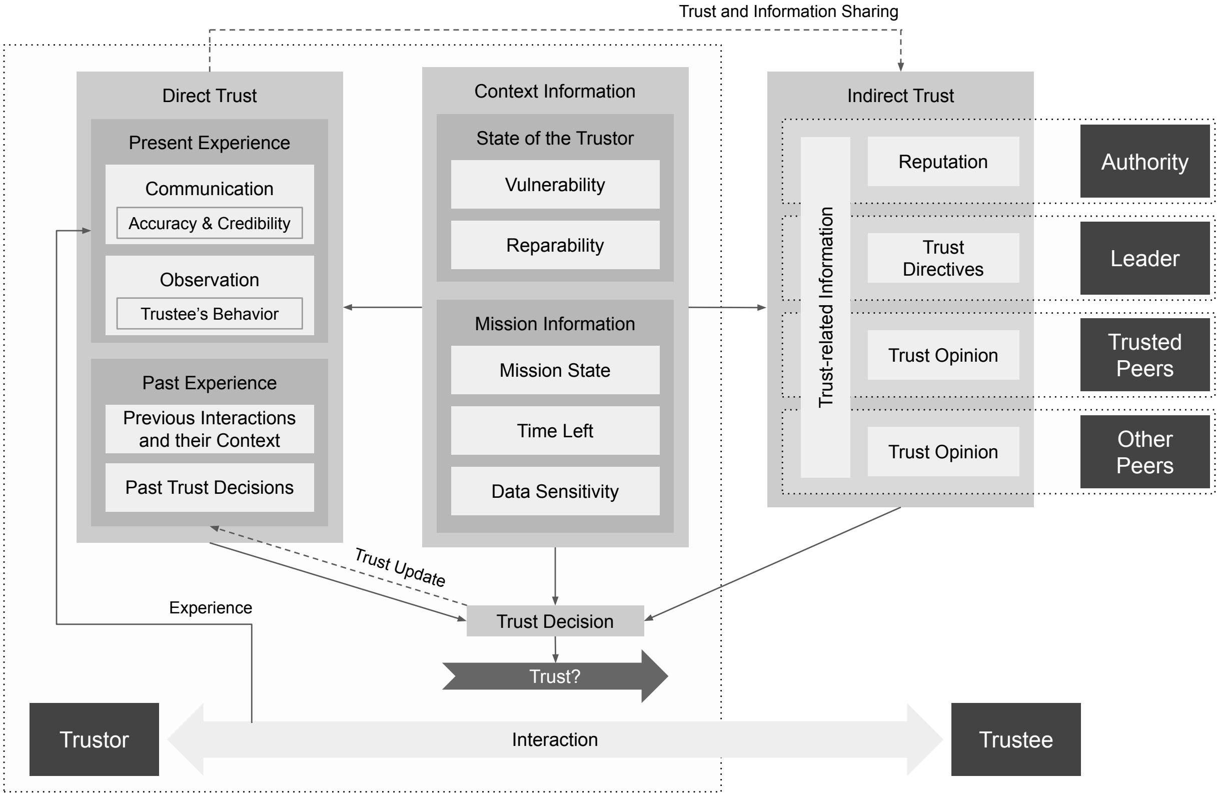

Fig. 3. Trust Management Framework for Multi-Robot Systems.

A. Actors

1) Trustor: represents the robot in question who is in the process of making a decision whether to trust another robot (*trustee*). The TMF is in place to support this *trust decision*.

2) Trustee: represents a robot (typically outside the scope of control by the *trustor*) that is interacting in some way with the *trustor* who has limited certainty about its trustworthiness and might feel vulnerable in its presence.

3) Authority: sits on top of the trust hierarchy and might take form of a set of regulations and directives (does not need to be represented as a robot itself).

4) Leader: is superior to the *trustor* and controls the mission the *trustor* is deployed to complete. In this regard, its trust directives should have higher weight than peer opinion or even be, by definition, unquestionable.

5) Trusted Peers: represent the robots whose trust opinion has higher weight in *trust decision* than that of other robots in the system, as they form a peer (friendship) group with the *trustor*, whose trust they have previously gained.

6) Other Peers: represent other robots who might be relevant in the context of the mission that the *trustor* is aiming to complete but who do not benefit from pre-gained trust of the *trustor*.

B. Direct Trust

The direct trust component consists of the *present* and *past experience* of the *trustor* with the *trustee*, being fed by their ongoing interaction. The *present experience* is influenced by the *communicated* information (its accuracy and credibility) and trustor's direct *observation* of trustee actions and behavior. The *past experience* reflects the experience from *previous interactions*, together with their *context*, such as the role of the trustee in the mission that the past experience relates to (i.e., how dependent the trustor was on the trustee at that point in time to complete its mission). Besides, the *past experience* also records *past trust decisions* made by the *trustor* about the *trustee*.

C. Indirect Trust

The indirect trust component is specific to the role it originates from. While the *authority* governs the global *reputation* of each MRS member, the *leader* gives *trust directives* (e.g., the necessity to trust a certain robot), the peers (both the *trusted peers* and *other peers*) only share their trust opinion, together with supplementary *trust-related information*, which can be shared by any of these actors. The *trust-related information* can, for instance, consist of the elements of peers' subjective *past experience* with the *trustee* in question.

D. Context Information

The *context information* characterizes the context of the *trustor* that influences its *trust decision*. This includes its *vulnerability* and *reparability* as in the baseline TMF, but it also includes *information about the mission* that the *trustor* is deployed to complete (namely the *mission state*, *time left*, and *data sensitivity*). *Mission state* stands for mission progress so far (e.g., environment explored, part of the resources lost) and its future plan (e.g., collecting environmental information, returning to the base). *Time left* records how much time is left to complete the mission. The *mission state* and *time left* directly influence the robot's behavior, decision-making process, the need for collaboration with other robots, and overall security and safety requirements. *Data sensitivity* mainly impacts data transmission and communication—the higher the sensitivity level, the higher trust score of a *trustee/peer* is necessary to approve data communication.

E. Trust Decision

The *trust decision* is made on the basis of the *direct trust* consideration, *indirect trust* inputs, and the current state of the *context* in the particular moment when the *trust decision* is to be made. The result is stored in the *past experience* part of the *direct trust* component (as it is subjective to the *trustor*) and propagated to the rest of the network via *trust and information sharing*.

VIII. Discussion

With the advent of Industry 4.0, Smart Cities, and other advanced contexts [42], [43], MRS are becoming increasingly present in everyday life, where increasingly more tasks rely on them. Therefore, their trustworthiness and safety needs to be taken more seriously. Our work brings ideas on preventing attacks even before they occur by carefully selecting trusted system members. By avoiding suspicious individuals, the whole MRS can become more trustworthy, and hence, it can participate in more critical tasks, including interaction with vulnerable human beings.

The selection of trusted peers cannot be done imprudently as the MRS missions are intended to be done by collaboration between multiple robots. An extensive reduction in the pool size of trusted peers might thus lead to lower efficiency or even failure of the mission.

The proposed TMF for MRS brings structured information and methods to make trust decisions consciously by considering relevant data and the real-time state of the mission and the environment. As the amount of this data may be overwhelming, its processing should be considered in the early phase of the system design. The TMF should be narrowly targeted to MRS to cover all their needs, but it should also be flexible enough to adapt to any of the wide range of MRS types.

A. Threats to Validity

To minimize the threats to validity, the proposed TMF for MRS is built on two main sources, covered by careful methodological design to minimize the risks of compromising its quality. To this end, we have opted for an incremental design, starting from an existing baseline TMF and scrutinized its components in both the direction of the necessity of existing components and their sufficiency to cover possible MRS scenarios.

B. Application of the Proposed Framework

The framework shows which facts should be considered when deciding trust between the trustor and the trustee. When designing a MRS, all of these items should be considered and covered to achieve the best result. Each part of the TMF should be adjusted to the context of the concrete robot, and the collected data should be structured to evaluate the trust effectively. The trustor needs to be assigned an initial trust score and relation to the newly encountered robot and robots from the initial group (considering its leader hierarchy). After that, trust should be evaluated regularly, even with robots deemed trustworthy, due to the trust score erosion over time (governed by the system authority). After each evaluation, the trustor should update records about the trustee in its memory and inform the rest of the system about its findings.

C. Future Directions

There is a multitude of research directions that can take the proposed TMF further, whether in terms of its application or extension. The first intended step on our side is to experiment with the framework in a variety of case studies, exploring the TMF in the context of heterogeneous MRS, where robots with vastly different designs, capabilities, and purposes must work together seamlessly. Next, we aim to explore how the TMF can be scaled to accommodate large and complex MRS, ensuring that trust management remains effective as the number of robots and interactions increases.

In the next phase, we intend to explore the role of the designed TMF in human-robot trust dynamics, especially in scenarios where human intervention or collaboration is necessary. This might include the need to address the ethical implications of trust decisions made by MRS and ensure compliance with emerging regulations in robotics and AI.

IX. Conclusion

In conclusion, this research contributes to the potential of Multi-Robot Systems (MRS) in our digital era, where their distributed intelligence and collaborative capabilities are being harnessed for a multitude of applications, from simple tasks to complex missions. The inherent strength of MRS, derived from the unique abilities and collective resilience of their robotic constituents, is limited by their susceptibility to misbehavior and security threats. Recognizing this challenge, this paper contributes to the solution by introducing a Trust Management Framework (TMF) for MRS designed to fortify the integrity and safety of MRS. Grounded in a baseline TMF, our proposed framework integrates MRS-related mechanisms fed by extensive research on MRS scenarios. The framework's emphasis on both direct and indirect trust, underpinned by contextual awareness, represents a holistic solution towards safeguarding MRS against vulnerabilities and opening to door towards MSR integration in our interconnected world.

Acknowledgment

The work was supported by GAMU project "Forensic Support for Building Trust in Smart Software Ecosystems" (no. MUNI/G/1142/2022).

References

[1] R. Capilla, E. Cioroaica, B. Buhnova, and J. Bosch, "On autonomous dynamic software ecosystems," *IEEE Transactions on Engineering Management*, vol. 69, no. 6, pp. 3633–3647, 2022. doi: 10.1109/TEM.2021.3116873

[2] H. Bangui, B. Buhnova, D. Kusnirakova, and D. Halasz, "Trust management in social internet of things across domains," *Internet of Things*, vol. 23, p. 100833, 2023.

[3] D. Halasz and B. Buhnova, "Rethinking safety in autonomous ecosystems." in *FedCSIS (Position Papers)*, 2022, pp. 81–87.

[4] A. Bierska, B. Buhnova, and H. Bangui, "An integrated checklist for architecture design of critical software systems." in *FedCSIS (Position Papers)*, 2022, pp. 133–140.

[5] M. Macak, S. Bojnak, and B. Buhnova, "Identification of unintentional perpetrator attack vectors using simulation games: A case study," in *2021 16th conference on computer science and intelligence systems (FedCSIS)*. IEEE, 2021, pp. 349–356.

[6] Y. Rizk, M. Awad, and E. W. Tunstel, "Cooperative heterogeneous multi-robot systems: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–31, 2019. doi: 10.1145/3303848

[7] B. Buhnova, "Trust management in the Internet of Everything," in *European Conference on Software Architecture. ECSA 2022 Tracks and Workshops*. Springer, 2023. doi: 10.1007/978-3-031-36889-9_10 pp. 123–137, preprint at http://arxiv.org/abs/2212.14688.

[8] D. Vojnar, A. Bierska, and B. Buhnova, "Scenarios for trust management in swarm robotics," 2024.

[9] Q. Yang and R. Parasuraman, "How can robots trust each other for better cooperation? a relative needs entropy based robot-robot trust assessment model," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2021. doi: 10.1109/SMC52423.2021.9659187 pp. 2656–2663.

[10] P. Gangwani, A. Perez-Pons, and H. Upadhyay, "Evaluating trust management frameworks for wireless sensor networks," *Sensors*, vol. 24, no. 9, p. 2852, 2024. doi: 10.3390/s24092852

[11] F. Amin and G. S. Choi, "Social pal: A combined platform for internet of things and social networks," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 2020. doi: 10.1109/ICCCS49078.2020.9118579 pp. 786–790.

[12] S. H. Alsamhi, O. Ma, M. S. Ansari, and S. K. Gupta, "Collaboration of drone and internet of public safety things in smart cities: An overview of qos and network performance optimization," *Drones*, vol. 3, no. 1, p. 13, 2019. doi: 10.3390/drones3010013

[13] S. M. S. M. Daud, M. Y. P. M. Yusof, C. C. Heo, L. S. Khoo, M. K. C. Singh, M. S. Mahmood, and H. Nawawi, "Applications of drone in disaster management: A scoping review," *Science & Justice*, vol. 62, no. 1, pp. 30–42, 2022. doi: 10.1016/j.scijus.2021.11.002

[14] G. Rathee, A. Kumar, C. A. Kerrache, and R. Iqbal, "A trust-based mechanism for drones in smart cities," *IET Smart Cities*, vol. 4, no. 4, pp. 255–264, 2022. doi: 10.1049/smc2.12039

[15] M. M. Arifeen, D. Bhakta, S. R. H. Remu, M. M. Islam, M. Mahmud, and M. S. Kaiser, "Hidden markov model based trust management model for underwater wireless sensor networks," in *Proceedings Of The International Conference On Computing Advancements*, 2020. doi: 10.1145/3377049.3377054 pp. 1–5.

[16] A. Deshpande, "Review of effective trust management systems in vanet environments," *International Journal of Grid and Distributed Computing*, vol. 14, no. 1, pp. 1771–1780, 2021.

[17] M. Aslan and S. Sen, "A dynamic trust management model for vehicular ad hoc networks," *Vehicular Communications*, vol. 41, p. 100608, 2023. doi: 10.1016/j.vehcom.2023.100608

[18] A. Gautam and S. Mohan, "A review of research in multi-robot systems," in *2012 IEEE 7th international conference on industrial and information systems (ICIIS)*. IEEE, 2012. doi: 10.1109/ICIInfS.2012.6304778 pp. 1–5.

[19] F. Steele Jr and G. Thomas, "Directed stigmergy-based control for multi-robot systems," in *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 2007. doi: 10.1145/1228716.1228747 pp. 223–230.

[20] K. Bolla, T. Kovacs, and G. Fazekas, "Compact image processing based kin recognition, distance measurement and identification method in a robot swarm," in *2010 International Joint Conference on Computational Cybernetics and Technical Informatics*. IEEE, 2010. doi: 10.1109/ICCCYB.2010.5491237 pp. 419–424.

[21] O. Vermesan, R. Bahr, M. Ottella, M. Serrano, T. Karlsen, T. Wahlstrøm, H. E. Sand, M. Ashwathnarayan, and M. T. Gamba, "Internet of robotic things intelligent connectivity and platforms," *Frontiers in Robotics and AI*, vol. 7, p. 104, 2020. doi: 10.3389/frobt.2020.00104

[22] A. Cowley, H.-C. Hsu, and C. J. Taylor, "Distributed sensor databases for multi-robot teams," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 1. IEEE, 2004. doi: 10.1109/ROBOT.2004.1307229 pp. 691–696.

[23] S. Sagar, A. Mahmood, Q. Z. Sheng, J. K. Pabani, and W. E. Zhang, "Understanding the trustworthiness management in the social internet of things: A survey," *arXiv preprint arXiv:2202.03624*, 2022. doi: 10.1016/j.comnet.2024.110611

[24] N. B. Truong, T.-W. Um, B. Zhou, and G. M. Lee, "From personal experience to global reputation for trust evaluation in the social internet of things," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017. doi: 10.1109/GLOCOM.2017.8254523 pp. 1–7.

[25] J. Lee and J. C. Oh, "A node-centric reputation computation algorithm on online social networks," *Applications of Social Media and Social Network Analysis*, pp. 1–22, 2015. doi: 10.1007/978-3-319-19003-7_1

[26] C. Marche and M. Nitti, "Can we trust trust management systems?" *IoT*, vol. 3, no. 2, pp. 262–272, 2022. doi: 10.3390/iot3020015

[27] R. Hussain, J. Lee, and S. Zeadally, "Trust in vanet: A survey of current solutions and future research opportunities," *IEEE transactions on intelligent transportation systems*, vol. 22, no. 5, pp. 2553–2571, 2020. doi: 10.1109/TITS.2020.2973715

[28] B. Zhang, Z. Huang, and Y. Xiang, "A novel multiple-level trust management framework for wireless sensor networks," *Computer Networks*, vol. 72, pp. 45–61, 2014. doi: 10.1016/j.comnet.2014.06.015

[29] H. El-Sayed, H. A. Ignatious, P. Kulkarni, and S. Bouktif, "Machine learning based trust management framework for vehicular networks," *Vehicular Communications*, vol. 25, p. 100256, 2020. doi: 10.1016/j.vehcom.2020.100256

[30] S. A. Ghasempouri and B. T. Ladani, "Modeling trust and reputation systems in hostile environments," *Future Generation Computer Systems*, vol. 99, pp. 571–592, 2019. doi: 10.1016/j.future.2019.05.017

[31] I. Ahmed, T. Nahar, S. S. Urmi, and K. A. Taher, "Protection of sensitive data in zero trust model," in *Proceedings of the international conference on computing advancements*, 2020. doi: 10.1145/3377049.3377114 pp. 1–5.

[32] R. Govindaraj, P. Govindaraj, S. Chowdhury, D. Kim, D.-T. Tran, and A. N. Le, "A review on various applications of reputation based trust management." *International Journal of Interactive Mobile Technologies*, vol. 15, no. 10, 2021. doi: 10.3991/ijim.v15i10.21645

[33] I. U. Din, K. A. Awan, A. Almogren, and B.-S. Kim, "Sharetrust: Centralized trust management mechanism for trustworthy resource sharing in industrial internet of things," *Computers and Electrical Engineering*, vol. 100, p. 108013, 2022. doi: 10.1016/j.compeleceng.2022.108013

[34] F. Ishmanov, A. S. Malik, S. W. Kim, and B. Begalov, "Trust management system in wireless sensor networks: design considerations and research challenges," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 2, pp. 107–130, 2015. doi: 10.1002/ett.2674

[35] H. Yu, Z. Shen, C. Miao, C. Leung, and D. Niyato, "A survey of trust and reputation management systems in wireless communications," *Proceedings of the IEEE*, vol. 98, no. 10, pp. 1755–1772, 2010. doi: 10.1109/JPROC.2010.2059690

[36] A. Abdul-Rahman and S. Hailes, "Supporting trust in virtual communities," in *Proceedings of the 33rd annual Hawaii international conference on system sciences*. IEEE, 2000. doi: 10.1109/HICSS.2000.926814 pp. 9–pp.

[37] S. Ruohomaa and L. Kutvonen, "Trust management survey," in *International Conference on Trust Management*. Springer, 2005. doi: 10.1007/11429760_6 pp. 77–92.

[38] C. Blum and D. Merkle, *Swarm intelligence: introduction and applications*. Springer Science & Business Media, 2008. ISBN 9783540740896, 3540740899

[39] M. H. Nasir, S. A. Khan, M. M. Khan, and M. Fatima, "Swarm intelligence inspired intrusion detection systems—a systematic literature review," *Computer Networks*, vol. 205, p. 108708, 2022. doi: 10.1016/j.comnet.2021.108708

[40] M. Srivatsa, S. Balfe, K. G. Paterson, and P. Rohatgi, "Trust management for secure information flows," in *Proceedings of the 15th ACM conference on Computer and communications security*, 2008. doi: 10.1145/1455770.1455794 pp. 175–188.

[41] F. Canciani, M. S. Talamali, J. A. Marshall, T. Bose, and A. Reina, "Keep calm and vote on: Swarm resiliency in collective decision making," in *Proceedings of workshop resilient robot teams of the 2019 ieee international conference on robotics and automation (ICRA 2019)*, vol. 4, 2019.

[42] L. Walletzky, B. Buhnova, and L. Carrubbo, "Value-driven conceptualization of services in the smart city: a layered approach," *Social dynamics in a systems perspective*, pp. 85–98, 2018.

[43] S. Chren, B. Rossi, B. Bühnova, and T. Pitner, "Reliability data for smart grids: Where the real data can be found," in *2018 smart city symposium prague (scsp)*. IEEE, 2018, pp. 1–6.