Statement of Changes and Answers to Comments

Paper Title:

Distributed Intrusion Detection for the Security of Societies of Robots

Authors: *Adriano Fagiolini, Gianluca Dini, Antonio Bicchi*

Date: May 18, 2017

Editor

✘ **R0-1)** …

A0-1)

Reviewer #1

✔ **R1-1)** *[...] The framework to individually monitor the behavior of a suspicious agent is interesting, particularly Theorem 1 on the 'least-conservative' estimator for an individual robot. The consensus protocol to combine estimates from multiple robots although new, is pretty straightforward and mimics the same steps as in conventional consensus algorithms. [...]*

A1-1) We are glad you appreciate it.

✘ **R1-2)** *I also have an impression that the length of the paper could be reduced by removing a few redundant parts. Is it possible to remove section V, and enrich section VI to illustrate the concepts that were done in section V? The consensus protocol probably does not deserve a whole section, and should be de-emphasized especially since it does not deal with the information corruption. My recommendation is to reduce it to the extent that it is useful for this paper and make it a subsection of Section III. The abstract is too long, and should be reduced too.*

A1-2)

✘ **R1-3)** *The structure of distributed detection protocol apparently requires some initial knowledge about 'suspicion' of a particular agent being non-cooperative. Formally, the choice of subsets of the agents to run the consensus algorithm has to be done carefully to not to include the suspicious agent. In other words, some external knowledge is required in the initial condition of the algorithm. While the authors do mention that they do not consider consensus where information is corrupted by an intruder, the authors should definitely provide arguments in support of the chosen mechanism and how practically feasible it is.*

A1-3)

✘ **R1-4)** *? The idea behind Corollary 1 is to use loose estimates for individual monitors, and then use the intersection between estimates of participating agents to reduce the looseness in estimation of the behavior of the ?suspicious? robot. Alternately, one could have under approximation of the individual estimates and then take union of these individual estimates to form better approximation of the behavior of the suspicious robot.*

I imagine that such an approach would lead to false false positives instead of false negatives for the approach proposed in the paper. The authors might want to comment on why did they choose one approach over the other. ? The proof of Proposition 3 in the appendix has an error, although it does not affect the proposition. The second expression for ci,j in the left column on page 15 does not evaluate to the same quantity as the first expression for ci,j for p?i,k = 1 and p?i,m = 0. This affects the subsequent proof, but my calculations show that it does not affect the proposition. The authors should confirm this.

A1-4)

✔ **R1-5)** *The paper keeps switching between V and $\mathcal{V}$ for the visibility region. For example, see the definition of $\mathcal{P}_i$ in Definition 1 at the bottom of right column on page 1, and the definition of visibility map towards the top of left column on page 3.*

A1-5) Correct. We have decided to use the notation V.

✔ **R1-6)** *The notation m_i used in Fig. 2 is undefined. It is possibly in conflict with m_i used in section IV on consensus.*

A1-6) The equation involving m_i was a refuse from a previous version of the figure. We have removed it.

✘ **R1-7)** *In the definition of nondeterministic automaton ?? in left column on page 4, ??? ? i ?i,? ? ??i? is confusing because, ??i being a subset of ?i, ? ? ??i implies that ? is a subset of ?i which does not go well with ? ? ?i which implies ? is a subset of ?i. This confusion exists in the definitions of all nondeterministic maps throughout the paper.*

A1-7)

✘ **R1-8)**

A1-8) Theorem 1 has several typos: ? the domain of e?i is B?i B?i and not just B?i ? s?i in the definition of e?i is to be replaced with (s?i,vi) ? in the last line of equation (3), ?i,k and ?i,k to be replaced with the corresponding indicator functions as in definition of logical operator ci,j in left column of page 3.

✘ **R1-9)**

A1-9) There seems to be a typo in the definition of ?? ? in left column on page 5, as there is f? i no ??i on the right hand side.

✔ **R1-10)** *The definition of $\bar{I}_i^h$ in the right column on page 5 has two 'such that' vertical bars. Please modify it.*

A1-10) We have corrected it as follows:

$$\bar{I}_i^h = \{\hat{q}_k \in 2^{\mathcal{Q}} \,|\, \hat{q}_k = B_\epsilon(q_k) \text{ for some } q_k \in I_i^h\}\,,$$

✘ **R1-11)**

A1-11) In the definition of ?? in left column on page 5, e in ?? (?, e) to be replaced with e? ? ??i i i

✘ **R1-12)**

A1-12) In the definition of ??i,j in the right column on page 5, for (Ihi , Vh, 1) it is union between a discrete set Ihi ? ?i,j(qi) and a continuous set ?i,j(qi) Vh. It looks a bit strange and the examples do not shed sufficient light on this construction.

✔ **R1-13)** *Please re-emphasize at the beginning of section IV that none of the mi observers are assumed to be malicious.*

A1-13) Thank you for the observation. We've added the following text in the first paragraph of Section IV:
"We also assume that they exchange correct information, whether or not they move according to the nominal motion protocol $\mathcal{P}_i$."

✘ **R1-14)**

A1-14) In the statement of Theorem 2, in Xh(0) = Uh, the quantity Uh is only specified in Fig 3. Please define it somewhere in the text too. ? The subscripts i and h are used for the observing agent interchangeably

throughout the paper creating a lot of confusion. For example, the definition of CVh(p).

✘ **R1-15)**

A1-15) The last sentence in the statement of Corollary 1 ?Moreover, the very same ....? should be removed and probably added as a remark since it is not formal the way it is stated.

✔ **R1-16)** *The last part of the first paragraph in left column on page* 7 *"Note that these rules ... geometry of the lanes and of the vehicles" is opaque and should probably be removed.*

A1-16) Done.

✔ **R1-17)** *The definition of the topologies* $\eta_{i,1}(q_i)$ *are repeated twice in the left column on page* 7*.*

A1-17) Thank you for noticing this. We have corrected this.

✘ **R1-18)** *The authors should provide some description for the 13 events in section V.*

A1-18)

✘ **R1-19)** *The authors should provide some explanation for the output of ?i function in the right column at the bottom of page 7.*

A1-19)

✘ **R1-20)**

A1-20) Section V A mentions that ?the figure (6) shows that all monitors are still unable to decide on the cooperativeness of car 0?. However, figure 6 shows that agent ?03? is able to conclude that agent ?00? is noncooperative. Please clarify this confusion.

✘ **R1-21)** *In section V B, mention explicitly that agent ?01? is going fast. Otherwise, the criteria for judging its cooperativeness is unclear.*

A1-21)

✘ **R1-22)**

A1-22) In figure 10 and especially figure 11, please increase the font size of the labels of the agents. Also, remove the additional ?0? in the labeling to avoid confusion with the text. For example, replace ?00? with ?0?.

✘ **R1-23)** *Section VI C has several typos. Please check it thoroughly: ? In the expression just above ?Consider another agent Ah ...? in the right column on page 11, replace tk with 2T in the exponent. ? Towards the bottom of right column on page 11, replace T1 and T2 with T and 2T respectively.*

A1-23)

✔ **R1-24)** *In the expression for $\tilde{c}_{i,j}$ just before the last display $C = \ldots$ in the left column on page* 16*, $(\Pi_{k=1}^{m} \tilde{s}_{i,k} v_{i,k} + \neg v_{i,k})$ should be replaced with $(\Pi_{k=1}^{m} \neg s_{i,k})$.*

A1-24) You're right! Thank you. We've corrected it.

Reviewer #2

✔ **R2-1)** *This paper proposes a distributed intrusion detection algorithm where a collection of robots try to use their local sensors and information communicated with their neighbors to detect ?misbehaving? robots. [...] An example from a highway system and a real implementation in a factory setting with cooperative forklifts are provided to demonstrate the approach. The application and implementation is interesting and relevant.*

A2-1) Thank you. We are glad of that.

✘ **R2-2)** *The intrusion detection in a robotic network is an interesting problem. The two main results of the paper are a non-conservative local event estimator (theorem 1) and a set-valued consensus protocol (theorem 2). Although the motivation of the paper is clear, some of the technical details are a little vague which makes it difficult to judge the general applicability of the proposed approach. In particular, there is no technical (formal) problem statement. The presentation of the paper can be improved. I think that the following points can be considered/clarified to improve the manuscript.*

A2-2)

✘ **R2-3)** *The notation gets too involved often times and some of the variables are not properly introduced (detailed examples are given below among minor issues). This obscures the main message of the technical sections. In general, it might be useful to verbally explain what each variable is for (in addition to the name of the variable) so that the reader does not need to guess.*

A2-3)

✘ **R2-4)** *Including a technical (formal) problem statement for the intrusion detection problem would improve the paper significantly. It would also be useful to explicitly explain how the hybrid observer and set-valued consensus protocol are put together to deter- mine abnormal/non-cooperative behavior. Again it is not hard to guess but a para- graph summarizing and combining the ideas introduced in sections III and IV and tying them back to a formal problem statement would improve the readability of the paper.*

A2-4)

✘ **R2-5)** *Although the problem of intrusion detection in a robotic setting is well-motivated in the paper, I think it is not clear how an intrusion detection system can be used to eliminate inverse effects of the intruder to overall system performance in such a dynamic setting. That is, it is not clear how the non-cooperative behavior would be accommodated. Is there an additional protocol as to how robots will behave if they detect a non-cooperative agent (i.e., what action the robots should take once they discover this)? If so, does there always exist such a protocol? Maybe some comments along this direction is required. In the supplementary video clip, there is a part where the cooperative forklifts react and resolve the deadlock after they detect the noncooperative one. In the readme file it says ?... In the second part of the video, the remaining two correct forklifts can cooperatively detect the noncooperation of the faulty one by using the proposed IDS, and can thus solve the deadlock. After consenting on the third robots noncooperation, the other two forklifts temporarily exclude it from the cell negotiation and proceed further in their paths...?. Should such reaction be hardcoded?*

A2-5)

✘ **R2-6)** *Another thing that is not clear to me is that in a cooperative setting it is likely that failures will be cascaded. For instance, if a robot misbehaves, for instance by stop- ping and blocking the way of another robot, the other robot might fail to comply with its own discrete automaton. Could you please comment on the possibility of these scenarios within the general framework presented and how one can differentiate be- tween true intruders versus those forced to misbehave in such scenarios?*

A2-6)

✘ **R2-7)** *The framework seems to involve too many combinatorial procedures. Maybe some remarks on computational complexity is warranted.*

A2-7)

✔ **R2-8)** *Is it assumed that discrete deterministic automata are correct and deadlock free if all robots behave correctly according to them?*

A2-8) Yes, although the technique does not rely on this property. We've added the following comment in the introduction:

"Although any considered motion protocol is reasonably correct, i.e. it ensures absence of deadlocks and accidents if all robots correctly follow it, the proposed technique does not rely on such an assumption."

✘ **R2-9)** *Both in theorem 1 and propositions 1-2-3, a ?smallest? approximation is mentioned. I understand what it means but it would be useful to formally define the underlying order (i.e., small in what sense?). Since theorem 1 seems to be one of the main contributions of the paper, I think clarity of its statement is important.*

A2-9)

✘ **R2-10)** *Set-valued observers have been studied in control community (see, for instance, Shamma, Jeff S., and Kuang-Yang Tu. ?Set-valued observers and optimal distur-bance rejection.? Automatic Control, IEEE Transactions on 44.2 (1999): 253-264.). Maybe some comments on connections with this line of research is required.*

A2-10)

✘ **R2-11)** *Some of the ideas in this paper have appeared in reference [9]. In particular theorem 2, and transportation system example look similar to the relevant sections in [9]. Maybe those parts of the paper can be shortened by giving a reference to [9] and more emphasis can be put on the new work on the real implementation with forklifts in the factory setting.*

A2-11)

✔ **R2-12)** *Abstract: "... they can evaluated ..." → "... they can be evaluated ..."; Sentence before section III: "... occupancy status its own... " → "... occupancy status of its own... "*

A2-12) Corrections made.

✘ **R2-13)** *Section III, second paragraph (page 3): It is not clear what is meant by ?uncertain yet complete estimate?, please clarify. Also in the same sentence, where does the ? come from (it has not been introduced up until that point in the paper)?*

A2-13)

✘ **R2-14)** *It might help to be more precise about definitions. I presume Q, the state-space, is assumed to have a manifold structure given that it has a tangent space. However one might lose the manifold structure when set-valued evolutions are considered. It would help to elaborate or give references on how the solutions of the set valued ?differential equation? (page 4, end of left column) evolve, their existence, unique- ness, etc. It seems more general than differential inclusions. Is this the case? Please clarify.*

A2-14)

✘ **R2-15)** *The definition of topologies is also confusing. Initially the map ?i,j : Q ? 2Q is said to be a topology. Later ?i,j(qi) ? 2Q which is an element in the range of the map ?i,j is referred as if it is the topology. Please clarify. Also, the subscript i in the definition of the map seems redundant. Does ?i,j(qk) have a meaning for k?= i?*

A2-15)

✘ **R2-16)** *In equation (1), what are ?, ?, ?? Similarly, right after this equation, what is hi, the last element of the set ?i,j ? ?i,j ? ? Some of the notations and concepts in Figure 2 are not introduced in the text (e.g., reset map, mi).*

A2-16)

✘ **R2-17)** *Theorem 1: is s?i a reduced encoder map or a restricted encoder map?*

A2-17)

✔ **R2-18)** *On page 7, left column, two paragraphs just repeat each other: "... The rules require the introduction of a topology $\eta_{i,1}(q_i)$ representing a region in the immediate front of an agent $\mathcal{A}_i$, a topology $\eta_{i,2}(q_i)$ for a region on its left, a topology $\eta_{i,3}(q_i)$ for a region on its right, and a topology $\eta_{i,4}(q_i)$ for a region on its back (see again Fig. 4)....? and "... Moreover, we need to introduce a topology $\eta_{i,1}(q_i)$ representing a region in the immediate front of the agent, a topology $\eta_{i,2}(q_i)$ for a region on its left,*

a topology $\eta_{i,3}(q_i)$ for a region on its right, and a topology $\eta_{i,4}(q_i)$ for a region on its back (Fig. 4)...?. It is not clear why this repetition is needed.

A2-18) It was an unnecessary repetition. We have replaced the second occurrence of the topologies' definition with the text:
"The above introduced topologies can be formalized as follows:

✔ **R2-19)** *Page 9, end of left column: "... macro-cell if free,..." → "... macro-cell is free,..."*

A2-19) Corrected.

✘ **R2-20)** *In the forklift example (figure 10), are the dotted lines represent the desired paths of the agents? If so, using different colors might make the figure easier to understand. It is not clear how those paths are determined. Also, in equation (6), ? seems to be constant. If so, how can the forklifts rotate? I might have misinterpreted the figure but some clarification along these lines would be useful.*

A2-20)

✘ **R2-21)** *The supplementary video clip can be played faster than 2x.*

A2-21)

Reviewer #3

✘ **R3-1)** *The paper describes a method to detect a misbehaving agent under the hybrid system framework.*

Comments:

(1) Title.

It seems the title is misleading. The proposed method detects faulty or misbehaving agents which diviates from the specified rules. In addition, it is unclear about the term "societies of robots". The title must be revised to correctly represent the work presented in the paper.

(2) Distributed vs. Centralized.

While the paper proposes a distributed algorithm, it does not compare the proposed method against a centralized solution. The comparison can be made about following aspects: communication overhead, processing time or delay before a decision, robustness against faulty communication, time-varying communication graph, performance comparison when agents leave and new agents joins, to name a few.

(3) Heavy notations.

The paper introduces a number of notations while it solves relatively simple problems in simulation and experiment sections. One may wonder if all the notations are necessary. Actually, the correctness of the proposed method depends on the specification of a number of functions, such as dynamics, encoder, and event detector. It seems the generation of those rules will be nontrivial for a more practical problem (which will have a higher order of complexity than examples used in the paper). From specification, can you generate those rules automatically? If not, is there a way to formally verify rules against specification? The authors are encouraged to reduce notation as simple as possible for the readers, which will increase the impact of the paper. Also, it is recommended to motivate the use of formal notations used in the paper by citing available formal verification tools, if there is any.

(4) Computational complexity.

The paper does not mention computational complexity of the proposed method. It seems the proposed method requires heavy computation as it requires, for example, numerical integration and set intersection operations on general sets. How does the proposed method scales as the sizes of state space, events,and topologies increase? If it grows exponentially, it loses the merit of being a distributed approach - since it does not scale.

(5) Theorem 2.

It seems it is an obvious result for operators such as the set intersection operation. As the authors argue that the theorem can be applied to more general cases, what are the other possible candidate set-valued functions other than the intersection operation which can be applied to the problem considered in the paper? If there are no other operations, it can be simplified as a simple fact about the set-intersection operation, instead of making it as a theorem.

(6) Robustness against disturbances and noise.

How does the proposed method handle noise or disturbances? Since the paper is not a theoretic paper and application-oriented, it is necessary to validate how it performs under disturbances and noise.

Minor comments:

p.3, in the definition of $T_i : k_i$ is not defined. p.3, in the definition of T_i: an explanation of the use of multiple topologies can motivate readers (provide an example). p.3, in describing $e_i : gamma_{i,j}, rho_{i,j}, lambda_{i,j}, h_i$ are not defined p.3, Figure 2: m_i and r_i are not defined p.7, dynamic model is not specified. (x, y, θ, v) are not explained. p.7, in specifying $E_i : lambda_{i,j}$ is a set of points as defined above and $s_{i,j}$ is logical-valued. The notation "$c_{i,2} = \not{s}_{i,1}\lambda_{i,2}$" does not make sense.

A3-1)

Reviewer #4

✘ **R4-1)** *In this paper the authors address the problem of detecting "intruders" in a network of mobile robots. Intruders are robots that either work erratically due to failures or act maliciously. The authors model the problem by using the formalism of hybrid systems and propose a heuristic approach that relies on two key ideas. First, the robots individually run a "monitor" algorithm that computes, using locally available information, a set of plausible trajectories for the neighboring robots. The plausible trajectories are then compared with observed trajectories with the goal of detecting a potential (motion) misbehavior. Second, a setvalued consensus algorithm combines the "opinions" of the monitoring robots; this leads to an increase of the detection rate.*

The topic is interesting and the proposed approach appears quite effective. On the other hand, I have some concerns against this paper: 1) Presentation style: the paper is very unclear. The key (rather straightforward) concepts are hidden under layers of unclear definitions and notation. The definitions of the encoder and event detector maps in Section II are especially bad. Also, the presentation of the algorithm is confusing, for example it is not clear how the monitor algorithm is interfaced with the consensus algorithm and several definitions (e.g., at the end of Section III.A) are obscure. 2) Contribution: the intrusion detection algorithm appears quite effective. On the other hand, the algorithm relies on quite strong assumptions (e.g., monitoring robots do not fail, communication is reliable, etc.). Moreover, computational complexity is not characterized. This reduces the contribution of the paper. 3) Grammar: There are many (noticeable) vocabulary, syntactic, and stylistic errors throughout.

A4-1)

✘ **R4-2)** *- The definitions and notation in Section II are difficult to parse and to interpret. - The authors should discuss the intuition behind their modeling choices. - The assumptions should be clearly stated and discussed (currently they are scattered throughout the text). - The problem should be clearly formulated. - In Theorem 1 and in the Appendix the authors claim optimality results (e.g., "smallest event estimator"), but the optimality criterion is unclear. - The quantity h_i is not defined. It should be explained before/near its first use in equation 1. - The definition of the sets λ is very unclear: they seem to be used together with indicator functions in Section II, but are multiplied by scalars elsewhere. This notation, along with calling them "constants," is confusing. - Some of the figures (e.g., Figs. 11 and 12) are too small and are unclear. - The authors*

should discuss the computational complexity of the proposed algorithm. - To improve clarity, the authors should summarize the intrusion detection algorithm by using pseudo-code within an "algorithm" environment. - The term QoS is undefined.

A4-2)

R4-3) *In summary, this paper presents some valid and interesting results, but is not publishable in its current form. The paper needs to be rewritten from scratch: the model should be presented in a clear, organized, and intuitive way. Every definition should be motivated and discussed. Notation should be carefully selected. The assumptions should be carefully stated and the problem should be clearly formulated. The algorithm should be presented in a more structured way and its properties (e.g., computational complexity, termination conditions) should be studied. I suggest to expand Section II with more discussion about the model and to remove Section V (since Section VI is sufficient to illustrate the concepts).*

A4-3)