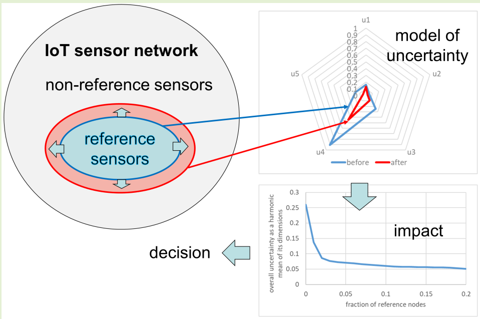
Impact of Reference Nodes on Uncertainty in Hybrid Ad-Hoc Sensor Networks

Piotr Cofta, *Senior Member, IEEE*, and Beata Marciniak

***Abstract*—The uncertainty introduced by the use of low-cost sensors (LCSs) in ad-hoc sensor networks is an ongoing concern that can be alleviated at the network level through a hybrid solution that relies on the use of reference nodes and reputation-based trust management. As reference nodes impose a significant expense, it is important to minimize their number while maximizing their impact on the reduction of uncertainty. This article presents a multidimensional analytical model developed through simulations that helps in predicting the extent of the decrease in uncertainty caused by an increase in the fraction of reference nodes in hybrid ad-hoc networks. The model shows that the marginal benefit of introducing reference nodes is much higher for networks with small fractions of such nodes, and quickly reaches near-saturation at about 5% of reference nodes, across all aspects of uncertainty.**



***Index Terms*—Hybrid ad-hoc sensor network, reference node, trust management, uncertainty.**

Nomenclature

s_a	Sensor a.
v_a	Current value reported by the sensor s_a.
ω	Opinion, as described in [32]. Constants representing opinions are denoted either as (b, d, u) or as (x, u).
f	Fraction of reference nodes in the network.
$\omega^i_{j,f}$	Reputation of node j at step i of the simulation, for a fraction f of reference nodes.
ω_{ab}	Opinion of sensor a about the trustworthiness of sensor b.
g_{ab}	Distance between sensors a and b.
$g_{\max}$	Maximum distance between sensors in the network.
$e_{ab} \in E$	Inbound edge from sensor a to sensor b; each edge represents an opinion.
d	Decay factor that controls the gradual decay of previous values of the reputation by increasing their uncertainty.
$\oplus, \oplus\sum$	Consensus operator, in its normal or group form. The operator produces the opinion that is the result of consensus in its arguments.
$\otimes$	Discounting operator that takes two opinions and produces an opinion which represents the opinion on the right-hand side discounted by (e.g., conditioned on the reputation of) the one on the left-hand side.
$\circledast$	Decay operator that takes the opinion and the decay factor and produces an opinion with uncertainty that is increased by the decay factor, with both its belief and disbelief corrected accordingly.

Manuscript received 9 July 2023; revised 2 October 2023; accepted 15 October 2023. Date of publication 26 October 2023; date of current version 30 November 2023. This work was supported in part by the Bydgoszcz University of Science and Technology. The associate editor coordinating the review of this article and approving it for publication was Prof. Nitaigour P. Mahalik. *(Corresponding author: Piotr Cofta.)*

The authors are with the Department of Telecommunication, Computer Science, and Electrical Engineering, Bydgoszcz University of Science and Technology, 85-796 Bydgoszcz, Poland (e-mail: piotr.cofta@pbs.edu.pl; beata.marciniak@pbs.edu.pl).

Digital Object Identifier 10.1109/JSEN.2023.3326200

I. Introduction

AD-HOC sensor networks, also known as citizen or resident networks, are gaining in popularity due to both growing environmental concerns and the increased availability of low-cost sensors (LCSs) [1]. They use opportunistic configurations, and their maintenance is often carried out by untrained volunteers. Compared with supervised, professional sensor networks, they are less expensive, but the measurements they provide may involve a heightened level of uncertainty, which forms one of the most important challenges to the network.

Since these measurements are needed to make important decisions, higher uncertainty may render them (and the network) useless. Insights into how to make these measurements more certain can lead to more rational investments in networks, and to their wider adoption.

One feasible solution is to use a reputation-based scheme, which can effectively exploit the over-provisioning of sensors to decrease the impact of faulty or uncalibrated ones. However, algorithms that rely solely on opinions may eventually drift from the ground truth (e.g., [2]). Reference nodes can be used to limit this drift.

This article focuses on hybrid networks that combine the use of reference nodes with "type C" reputation-based algorithms [3]. A network of this sort consists of a mix of LCSs and reference nodes, which act as trust anchors. In contrast to LCS nodes, reference nodes are better designed and benefit from regular, professional maintenance, including frequent recalibration. Unfortunately, reference nodes can be also much more expensive, meaning that the network must minimize their usage while maximizing the benefits they offer.

The direct inspiration for this research was the development of a hybrid urban ad-hoc sensor network dedicated to the measurement of air quality [4] that used available reference nodes as a backbone, with LCSs to help provide better spatial resolution at an acceptable cost [5], [6].

The research question addressed in this article concerns the relationship between the fraction of reference nodes and the reduction in uncertainty they offer, in the expectation that not all investments in reference nodes are warranted. We concentrate on the use of reference nodes in reputation-based networks. While both the use of reputation-based algorithms and the use of reference nodes are popular (see Section III), the relationship between the fraction of reference nodes and the reduction in uncertainty has not yet been investigated.

This article presents an analytical model that links changes in the fraction of reference nodes with the level of uncertainty. The model was developed on the basis of simulations, and takes into account five aspects of uncertainty [7]. We show that although the marginal benefit of adding the first few reference nodes is significant, this benefit quickly reaches its saturation point, so that further additions may have a negligible effect.

The aim of this research was to provide a tool that could help a network manager to make decisions regarding the development of the network. More specifically, a decision on whether to invest in reference nodes may benefit from a better understanding of benefits that they provide.

The novel aspects of this article are threefold: 1) we model the relationship between the fraction of reference nodes and the level of uncertainty in hybrid ad-hoc sensor networks; 2) we demonstrate the existence of a near-saturation point beyond which further investments in reference nodes may not be warranted; and 3) we demonstrate that various aspects of uncertainty respond differently to an increase in the number of reference nodes.

The article is structured as follows. Section II introduces some terminology, and Section III presents a review of the literature, with a focus on solutions used in sensor networks. Our methodology is discussed in Section IV, which includes the structure of the problem as well as a description of the simulation algorithm. The simulation and its results are introduced in Section V, while Section VI presents the proposed model. A discussion and conclusions in Section VII close the article.

II. Terminology

A. Hybrid Ad-Hoc Sensor Network

Also known as a citizen or resident network, this is a type of sensor network that predominantly contains LCSs, supported by reference nodes. It is often operated by volunteers or untrained staff, with planning and management performed in an opportunistic way [7].

B. Reference Node

This is a sensor node that is specifically designed to carry out measurements with the minimum achievable uncertainty. Reference nodes tend to be more complex and more expensive compared to LCSs [8]. In the scheme proposed in this article, a reference node is one that permanently holds the highest possible reputation. We note that the term "reference node" may have a different meaning in wireless ad-hoc networks [9].

C. Uncertainty in Measurement

Uncertainty refers to epistemic situations where information, as received by the observer, is imperfect or may be unknown [10]. The construct of uncertainty encompasses several aspects and components [11], and various taxonomies have provided some structure in this area (e.g., [12]). In this article, uncertainty is considered to be an attribute of an opinion, of a reputation, and of the network itself.

D. Trustworthiness and Opinions

Trustworthiness is the quality of a node in which it performs as it should across a wide range of situations [13]. In this article, we assume that trustworthiness cannot be directly measured, although various entities may provide their opinions of a node's trustworthiness.

E. Reputation

Reputation is the extent of the trustworthiness of a node, as agreed upon by the population of nodes [13] based on opinions about trustworthiness. In this article, reputation is calculated iteratively from opinions, the reputations of the nodes that provide opinions, and previous reputations.

F. Trust Management

Trust management is a framework that determines how a group of nodes (i.e., a network) utilizes the computational versions of trustworthiness and reputation. It describes an algorithm that is used to collect opinions and process them to determine trustworthiness and reputation.

III. Literature Review

Through this literature review, we attempt to establish the relevance of this research as well as its novelty in relation to the extant body of knowledge. To this end, we consider four questions: 1) how popular are reputation-based algorithms in the literature, and more specifically for sensor networks? 2) what is the most typical variant of the reputation-based algorithms that is used? 3) how popular is the use of reference nodes in a network (i.e., the use of a “C”-type network, as opposed to types “A” and “B”’)? and 4) is there any established knowledge or model that links the use of reference nodes with changes in uncertainty?

Sensor networks encounter situations in which the uncertainty in the data they provide is increased. These situations include [7] the unavoidable presence of measurement noise, the failure of a node, leading to incorrect or no data being sent, malicious behavior of nodes, long term-drift, etc. If they are left untreated, the network may eventually become useless.

The containment of uncertainty in sensor networks can be addressed in three complementary ways. First, this can be achieved by improving the quality of individual nodes and sensors, usually through a process of calibration (i.e., through a “micro-calibration” approach [14]) Second, the process used for data reconstruction can be improved to eliminate the impact of sensors of lower quality, or which are malicious or faulty, at the network level (i.e., a “macro-calibration” approach). Our research focuses on this approach. Finally, the problem of uncertainty can be addressed at the network design level [15], [16].

The use of both calibration and network design approaches may be problematic in ad-hoc sensor networks, and particularly in citizen networks. The relaxed approach to maintenance means that regular calibration of all nodes may be hard to achieve. The opportunistic growth of the network also makes a design-based approach infeasible. This leaves only the second option, i.e., network-level improvements to data reconstruction, as a viable one [17], [18]. The use of trust and reputation is of particular interest in wireless sensor networks [19], mobile networks in general [20], vehicular mobile networks [21], and flying ad-hoc networks [22], to name only a few areas of application.

In addition to the more popular reputation-based schemes discussed below, other approaches can be used. For example, Rizwanullah et al. [23] used fuzzy logic to mimic human perceptions of trust and reputation for “things” in the IoT networks. Another example is the direct application of machine learning in [24] to an “Industry 4.0” sensor network to detect sensors of dubious quality. Note, that although the latter paper addressed a very similar problem to the former, it used a very different scheme, with different semantics, and defined “reputation” in a different way.

Reputation-based approaches that are implemented at the network level and employ various trust management schemes have been widely researched and used (see [25], while [26] provides more insights into representative implementations). In such schemes, nodes can issue opinions about each other, and these opinions, disseminated by some form of gossip protocol [27], can be fused to create a reputation. Data delivered by a node are associated with the reputation of that node; this reputation is taken into account in further calculations, and is often used as a metric of the generalized quality or certainty of data.

Three types of reputation-based schemes are in use [3]. Type A is a flat scheme, where the reputation of a node that provides an opinion is not considered. In schemes of type B, this reputation is considered, but the initial reputations of all nodes are identical. Finally, in schemes of type C, certain root nodes have an immutably high reputation, while the reputations of other nodes (as well as the importance of their opinions) are determined by those root nodes, or by nodes that have been endorsed by them.

Types A and B can operate without reference nodes (i.e., a priori trusted nodes), but have shown certain weaknesses [28]. A network may eventually drift from the ground truth if there are enough sensors that can outvote the correct ones, and this problem is particularly visible when it comes to long-term drift, where the majority of sensors gradually deliver incorrect data [2].

Hence, when a reputation-based scheme is used in a sensor network, it tends to be of type C, where the root nodes ensure at least some ground truth. A typical architecture [29] may split sensors into a group of observers, which act as roots of trust (also known as super-nodes, supervisors, or agents), and the remaining nodes, whose reputation is calculated by those roots of trust, or by nodes already verified and endorsed by them. The architecture researched in this article follows this model, with reference nodes used as roots of trust.

We note that these schemes do not assume any particular method of formulating opinions. The literature includes examples of the use of a simple distance/loss function [2], rule-based reasoning [30], artificial neural networks [31], and supervised learning [22], among others.

The fusion algorithm used for trust management is responsible for converting opinions into reputations. It needs to take into account two aspects: the computation of opinions (i.e., how several opinions are fused into a single reputation value), and the decay in reputation (i.e., how a reputation changes over time). Several algorithms can be used to fuse opinions into reputations [13], and the choice depends on the ability of nodes to issue opinions, their computational capability, bandwidth constraints, etc.

In reference to the four questions formulated at the beginning of this section, we note the following.

1) The use of reputation and reputation-based algorithms is popular. The review in [19] reported that 8 out of 20 schemes used some form of reputation, while the review in [20] recorded 16 schemes out of 48 as using reputation, and the review in [21] indicated that of 111 schemes, 63 used reputation.
2) There is no typical algorithm that is used to calculate reputation. The most popular approach is based on averaging, with the next most popular employing Bayesian inference. It does not seem, however, that the performance of the scheme significantly depends on the algorithm itself.
3) Reference nodes (also known as super-nodes, observers, or agents) are the most popular solution. The reviews in [19], and [20] report the use of reference nodes in 70% and 35% of schemes, respectively. In another review of reputation-based schemes [32], 12 out of 18 schemes were found to use reference nodes.
4) No paper could be found that focused on the relationship between the fraction of reference nodes and the level of

uncertainty. The existence of reference nodes seems to be taken for granted.

In conclusion, both reputation-based schemes and the use of reference nodes are popular enough to warrant research. In view of the lack of research on the relationship between reference nodes and uncertainty, our approach can be considered a novel one.

IV. Methodology

The objective of this research was to establish a predictive model to link the fraction of reference nodes in a network with the level of uncertainty it provides. It addresses the question of what could be the benefit (in terms of uncertainty reduction) of increasing the fraction of reference nodes in a network.

The methodology employed in this article is based on structurization and simulation, leading to the formulation of a model. In terms of structurization, the problem considered here is structured into five aspects based on the taxonomy of uncertainty in sensor networks [7]. Each aspect is analyzed using an expanded version of the simulator originally used in [2].

The parameters of the simulation were determined on the basis of preliminary runs, which were carried out to ascertain the overall changes in the uncertainty and to determine the sensitivity of the results to changes in the values of these parameters. In each case, this sensitivity was insignificant. Simulations were therefore performed for the most representative selection of parameters.

A regression analysis was applied to each aspect separately. The best fit from a linear regression, a polynomial regression (up to third degree), and exponential, logarithmic, power, and hyperbolic ones was chosen to describe the relationship between the fraction of reference nodes and the value of uncertainty. The findings were used to construct a model that covers five aspects of uncertainty.

A. Taxonomy of Uncertainty

In [7], a framework was proposed that represented six aspects of uncertainty in a sensor network. These are briefly described below.

1) Aleatory uncertainty (which in this case consists of noise) relates to the physicality of the measurement itself, i.e., the irreducible randomness attributable to the physical operation of the sensor or its environment.
2) Completeness uncertainty (which in this case corresponds to inoperability) relates to the fact that the sensor network is a complex system in which elements may be corrupted or inoperable.
3) Logical uncertainty (represented here by the response to a transition) relates to algorithmic data processing, which may introduce various distortions to data reconstruction.
4) Utilitarian authority (represented in this case by long-term drift) is concerned with the long-term benefits that the network delivers with changes in its operating environment and in its own operation.
5) Ethical uncertainty (corresponding here to the failure of nodes) is concerned with the impact of failed nodes, regardless of the reasons for failure.
6) Epistemic uncertainty relates to the social reflection on the uncertainty and the operation of the network. This form of uncertainty, unlike the others, is not addressed by the simulation, but by the existence of this article.

We note that these aspects are not directly comparable, as the semantics differ. For this reason, our model covers five separate aspects, and delivers a vector of five values of uncertainty.

B. Logic for Uncertain Probabilities

As there is no dominant algorithm that is used by reputation-based schemes, the proposed simulator uses logic for uncertain probabilities [33] for all calculations related to opinions, trustworthiness, and reputations. This logic is based on the Dempster–Shaffer theory of evidence (e.g., [34]). Within the frame of discernment (i.e., an opinion), three components are considered: belief (b), disbelief (d), and uncertainty (u). Thus, an opinion, trustworthiness and reputation can be uniformly described by a tuple (b, d, u).

The main benefit of introducing this notation lies in the way it treats uncertainty as an equally important part of the opinion, thereby simplifying the operations on opinions and removing some of the paradoxes arising from alternative solutions [13].

A preliminary run of otherwise identical networks was carried out using the earlier simulator from [2], which used a weighted average and the exponentially weighted moving average (EWMA) decay, to see whether the introduction of the logic of uncertain probabilities would significantly affect the results. These simulations demonstrated that the differences between results were not significant, although the variant that used the logic was slightly more responsive to changes. As the logic for uncertain probabilities uses a variant of Bayesian inference, it is not expected that the results will significantly differ also from networks that use Bayesian approach.

C. Simulated Network

The simulator used in this research allowed for the creation of sensor networks with varying fractions of sensors of different types. Each type defines the behavior of a sensor, for example whether it is operational or faulty, whether it is subject to certain measurement distortions, whether its behavior changes at a particular step, etc.

We attempted to mimic the structure of a network that could be used for environmental monitoring. For this reason, it was structured as a square grid, with the locations of the reference, regular and faulty nodes selected randomly.

The phenomenon monitored by the network had a fixed value across the simulated area covered by the network, meaning that all the nodes were expected to give the same value; however, the readings were also subjected to simulated distortions or failures, in order to introduce uncertainty into the simulation.

At each simulation step, a node could express its opinion about the trustworthiness of another node. The selection of pairs of nodes for an opinion was random, with the proviso that a node was not allowed to provide an opinion on itself. Each opinion was expressed as a pair (x, u). The value of x

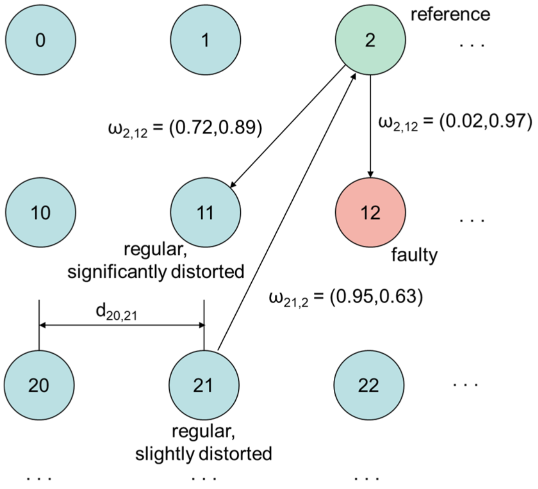

0
1
2
reference
10
11
12
faulty
20
21
22
$\omega_{2,12}$ = (0.72,0.89)
$\omega_{2,12}$ = (0.02,0.97)
$\omega_{21,2}$ = (0.95,0.63)
$d_{20,21}$
regular, significantly distorted
regular, slightly distorted

Fig. 1. Fragment of a grid network.

(the expectation) was calculated with the help of a nonlinear function of the difference between the value measured by the node expressing an opinion and the value reported by the other nodes as the measurement. The lower this difference, the higher the expectation.

The value of the uncertainty (u) depended on a nonlinear function that took as input the distance between the nodes. Opinions about neighboring nodes had much lower degree of uncertainty than opinions about nodes located far away.

An example of a fragment of a grid network is shown in Fig. 1. Node 2 is a reference node, node 12 is a faulty node (returning zero), and the remainder are regular nodes, reporting a correct value (one), but with distortion. Three opinions are shown: two are provided by the reference node, and one by node 21 about reference node 2. The expectation depends on the difference between the readings, while the uncertainty depends on the distance between nodes. The values given here are for illustration purposes only.

The parameters common to all simulations are listed below. The values of these parameters were set according to previous experience with the simulator, and to reflect those typically used in actual networks.

1) *Type of Network (Random, With Constraints):* At any iteration, an opinion could be provided by any node about any other node except itself.
2) *Number of Nodes (Sensors) in the Network:* 100, structured as a 10 × 10 rectangular grid. This value was determined experimentally. An increase in the number of nodes to above 100 did not affect the results.
3) *Graph Density (3):* This parameter emulates the behavior of a network with a rectangular layout, which can be used, e.g., for atmospheric measurements. In such networks, most reliable opinions are expressed about nodes located upwind and downwind, on average affecting three neighbors.
4) *Decay Factor (0.2):* This parameter controls the gradual decay of the reputation. The reputation from past steps is expected to be less relevant than the current value, so that the uncertainty increases accordingly. For details, see the algorithm in Section IV-D.
5) *Lambda (2.4):* This parameter controls the nonlinearity in the calculation of the opinion. For details, see the algorithm in the next section.
6) *Number of Simulation Runs for Each Case (50):* Due to the random distribution of the network, several runs of simulations for different networks with the same parameters were conducted, and the results were averaged.
7) *Number of Steps in Each Simulation (200):* This value was determined experimentally. As a minimum, the network required about 20 steps (under stable conditions) to settle. The remaining steps allowed for the introduction of changes in the measured phenomenon and the response of the network.

In terms of the operation of nodes, there were no differences between the reference nodes and the standard ones, except as follows.

1) The initial value of the reputation of the reference node was set to (1.0, 0.0, 0.0), i.e., to full belief without any uncertainty; for the other nodes, it was set to (0.0, 0.0, 1.0), i.e., to complete uncertainty.
2) The reputation of the reference node remained unchanged throughout the simulation, while for the other nodes, it was recalculated from the available opinions.
3) The reference node was assumed to always report the exact value of the phenomenon; it never failed, was not affected by noise or drift, and reacted immediately to any change in the value of the phenomenon.

D. Algorithm

The algorithm executed by the simulator was an iterative version of a popular reputation-based algorithm that uses opinions about nodes provided by other nodes, and fuses them to give reputations by applying the logic for uncertain probabilities. The notation used by the algorithm is shown in Nomenclature.

The algorithm used in the simulation is shown in Table I, in a form of pseudocode. In lines 1 and 2, the algorithm generates a graph and initializes the nodes. A loop is then used (line 3) to process the subsequent simulation steps. In each step, measurements are taken (lines 4 and 5), and opinions are formulated (lines 7 and 8). Finally the reputation of each node is changed based on the consolidated opinions (lines 10–15).

This algorithm uses a fusion function to combine opinions to obtain the value of reputation. It was demonstrated in [32] that there is no one fusion function that is objectively the best; instead, this function is an expression of the objective of the fusion. The fusion function used here allows highly trustworthy nodes to have more impact on the value of trustworthiness, even if their assessment does not correspond to the general opinion. Reference nodes are treated as highly trustworthy, and this function lets them have a higher impact on the outcome.

E. Unit Response

In this research, we used the network unit response as a primary metric. The network unit response (unit response)

TABLE I
ALGORITHM USED IN THE SIMULATION

```
Step   Operation
1.     Generate $n*f$ reference and $n*(1-f)$ non-reference nodes for the required mix
2.     For each node $j=1,..,.n$
           set $\omega_{j,f}^{0}$ to either $(1.0, 0.0, 0.0)$ (reference node)
               or to $(0.0, 0.0, 1.0)$ (otherwise)
3.     For each step, $i=1,..,.m$
4.         Calculate the expected value of the phenomenon for this step
5.         For each node $j = 1,\ldots,n$ simulate its measurement $v_{j,f}$,
               according to the parameters
6.         Remove old edges and generate new $n*g$ edges to satisfy
               the required distribution
7.         For each edge from $s_a$ to $s_b$, decorate the edge with
8.             the opinion, calculated as
                   $\omega_{ab} = \omega_{a,f}^{0} \otimes (e^{-\lambda|v_a-v_b|}, e^{-\lambda(g_{ab}/g_{max})})$
9.         For each node $j=1,\ldots,n$:
10.            Identify the set of inbound edges $e_{aj} \in E$
11.            For all non-reference nodes:
12.                set $\omega_{j,f}^{i} = \oplus \sum_{e_{aj} \in E} \omega_{aj} \oplus \omega_{j,f}^{i-1} \circledast d$
13.            For all reference nodes:
14.                set $\omega_{j,f}^{i} = \omega_{j,f}^{i-1}$
15.        Report current values of $\omega_{j,f}^{i}$ for all nodes, $j=1,\ldots,n$
```

is a utility function that assesses the extent of the lack of uncertainty at the network level, i.e., the ability of the network to determine a value that is close to the correct value of the phenomenon from the current measurements, despite the fact that some nodes introduce more uncertainty than others.

The value of the unit response ranges from zero to one, where one indicates a perfect network and lower values indicate that the network is unable to fully handle the uncertainties. A network consisting solely of reference nodes will always return a value of one.

This metric assumes a situation where all nodes of the network are expected to measure the same value of the phenomenon. The unit response is calculated as a weighted average of the values from all nodes, where each weight is proportional to the reputation of the node [2].

V. SIMULATION AND RESULTS

The outcome of the simulation modeling is discussed in this section in regard to the various aspects of uncertainty identified earlier in this article. For each of these, a description of the simulation is provided together with its results and associated conclusions. A regression model is also presented that links the level of uncertainty with the fraction of reference nodes.

The aim of the research was to simulate and construct a model of the behavior of the network over a usable range of zero to 20% reference nodes. It is unlikely that a network will employ more than 20% reference nodes, not only because the cost would be prohibitive, but because this might invalidate the concept of a low-cost volunteer-operated network. Whenever possible, the simulation (and the model) was run for a slightly wider range, to allow for more precise modeling.

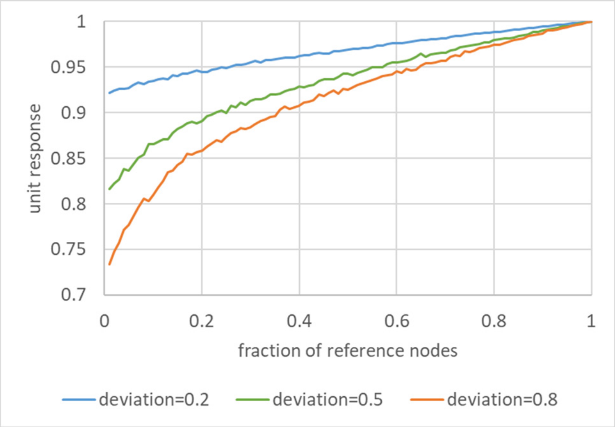


Fig. 2. Unit response as a function of the fraction of reference nodes, with deviation as a parameter.

A. Noise (Aleatory Uncertainty)

The noise was simulated by allowing the readings of all non-reference nodes to be distorted by one-sided Gaussian noise with a mean value of zero and standard deviations ranging from 0.01 to 0.8.The impact of the deviation always decreased the value of the unit response.

Fig. 2 shows the unit response as a function of the fraction of reference nodes for various deviations. The impact of reference nodes is super-linear across the whole range of values of the standard deviation.

From a regression analysis, it was found that a third-order polynomial provided the best fit, and explained 0.992 of the variability. The coefficients of this polynomial were in turn linearly dependent on the expected deviation, and were estimated through linear regression. The experimentally obtained values of the coefficients gave the expression in the following equation:

$$\begin{aligned} u_1 &= 1 - \left(a_1 r^3 + b_1 r^2 + c_1 r + d_1\right) \\ a_1 &= 0.6157 * v - 0.0781 \\ b_1 &= -1.2223 * v + 0.1452 \\ c_1 &= 0.9057 * v - 0.0444 \\ d_1 &= -0.2829 * v + 0.9757 \end{aligned} \tag{1}$$

where

u_1	first aspect of the uncertainty (noise);
a_1, b_1, c_1, d_1	coefficients;
v	expected deviation [0–1];
r	fraction of reference nodes (0.0–1.0].

B. Inoperativity (Completeness Uncertainty)

This uncertainty is caused by nodes that from the outset do not deliver correct data (or do not deliver data at all), as they are inoperable or malicious. This differs from the situation where the operative node eventually fails, as in this case, other nodes can obtain information about its reputation.

The simulation used varying mixes of regular, reference and inoperable nodes. Regular and reference nodes reliably reported the value of the phenomenon, while the inoperative nodes continuously reported zero.

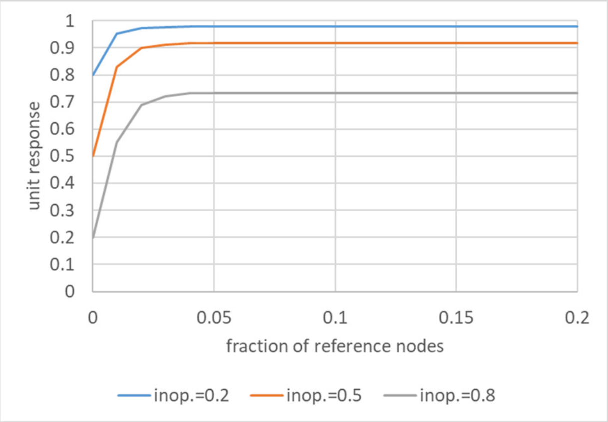


Fig. 3. Unit response as a function of a fraction of reference nodes, with the fraction of inoperative nodes as a parameter.

The results of the simulation are shown in Fig. 3, and it can be seen that the impact of the reference nodes is super-linear. Small fractions of reference nodes show a particularly significant difference.

A regression analysis showed that this family of hyperbolic tangent functions could be estimated as shown in the following equation, explaining 0.992 of the variability:

$$u_2 = 1 - (\tanh(122.5 * r) * 0.873 * z + (1 - z)) \tag{2}$$

where

u_2 second aspect of the uncertainty (inoperativity);
r fraction of reference nodes;
z fraction of permanently faulty nodes.

C. Transition (Logical Uncertainty)

The algorithm itself can also be a source of uncertainty. This situation may arise during a change in the value of the phenomenon, where the change gradually affects subsequent groups of nodes; for example, a cloud may gradually cover the area over which the network operates.

In situations such as these, a node may downgrade the reputations of other nodes simply because they are reliably reporting the new value. This situation was simulated by letting the network stabilize for a particular value of the phenomenon, and then changing the value of the phenomenon. Fig. 4 shows the unit response as a function of the fraction of reference nodes, for different speeds of this transition (i.e., the per-step increment in the fraction of nodes that receive a new value, described here as the "slope").

The simulation shows that in the presence of reference nodes, the response improves, as the network more rapidly and correctly reports new values, thus decreasing the overall uncertainty associated with the transition.

The resulting family of functions is hyperbolic tangent functions. The empirical formula for calculating the uncertainty, shown in the following equation, explains 0.992 of the variability:

$$u_3 = 1 - (0.4 * \tanh(r/s * 0.715) + 0.51) \tag{3}$$

where

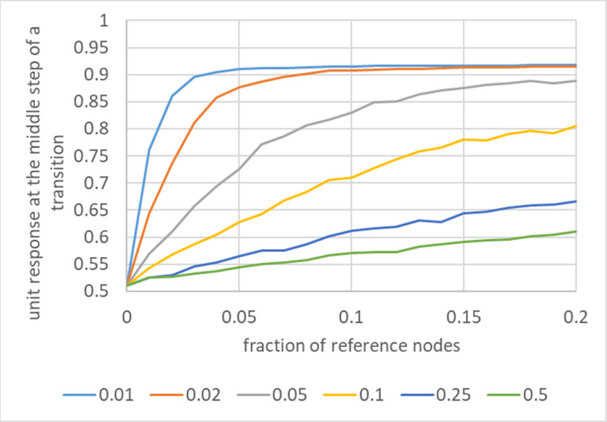


Fig. 4. Unit response as a function of the fraction of reference nodes, with the slope of the transition as a parameter.

u_3 third aspect of the uncertainty (transition);
r fraction of reference nodes (0, 1];
s slope, i.e., the fraction of nodes that are exposed to the change in the phenomenon in one simulation step (0, 1).

D. Drift (Utilitarian Uncertainty)

Sensors are subject to wear and tear, which may affect their readings. For example, particulate matter (PM) sensors based on scatter laser technology [35] require the ambient air to be passed through the measurement chamber with the aid of a fan. There are two factors that may cause drift: the wear of the fan, which may affect the speed of the air flow, and the accumulation of dust in the chamber itself. The problem of drift is particularly acute, as it may affect all the sensors in the whole network at the same rate.

Long-term drift was simulated by gradually altering the simulated readings of non-reference nodes, so that during the simulation, the values read by non-reference nodes gradually decreased from the correct value of 1.0 down to 0.0. Reference nodes were free from drift, and always reported the correct value. The network unit response after 100 steps following the moment at which the drift started was used to determine the uncertainty.

The relationship between the fraction of reference nodes and the unit response is shown in Fig. 5. This relationship is practically identical, with differences not exceeding 0.002, regardless of the speed of drift, which was varied from 0.01 to 0.1.

A regression analysis gave the expression in the following equation, a third-order polynomial that explains the 0.9992 of the variability, with coefficients that were determined experimentally:

$$u_4 = 1 - \left(102.13 * r^3 - 49.134 * r^2 + 9.4406 * r + 0.0012\right) \tag{4}$$

where

u_4 fourth aspect of the uncertainty (drift);
r fraction of reference nodes (0, 1].

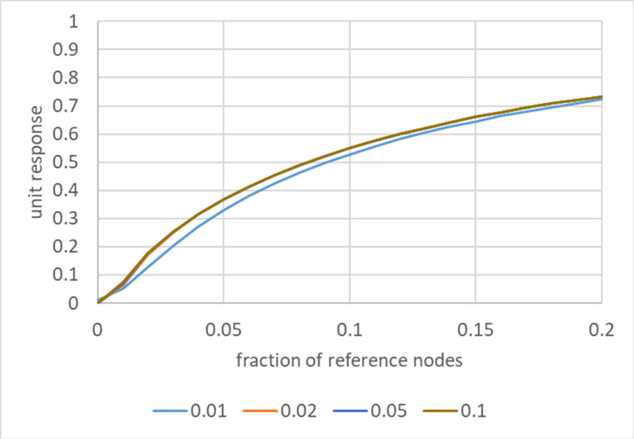


Fig. 5. Unit response as a function of the fraction of reference nodes, with the value of the drift as a parameter.

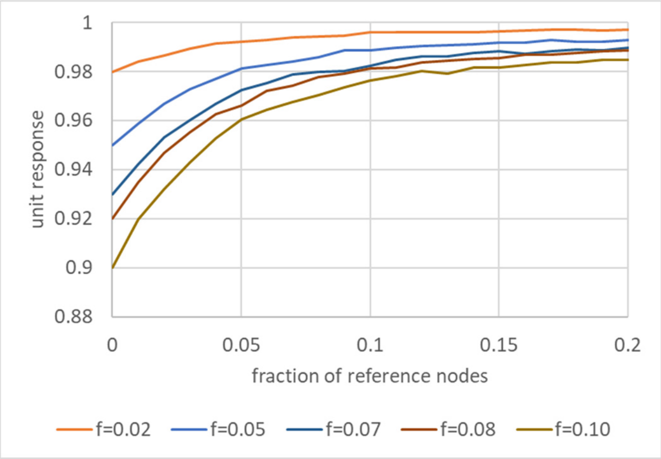


Fig. 6. Unit response as a function of the fraction of reference nodes, following the failure of a fraction of the nodes *f*.

E. Failure (Ethical Uncertainty)

The failure of a node during operation is an example of ethical uncertainty. This uncertainty is caused by the faulty node being incorrectly reassessed, or a correct reassessment being significantly delayed.

In the simulation, a single failure scenario was considered in which a set fraction of nodes failed at certain point, and started reporting incorrect data. This event was followed by a transition period in which the remaining nodes re-assessed the reputation of these nodes.

The uncertainty was calculated at the fourth step after the one in which the nodes failed. This choice was made on the basis of additional simulations.

Fig. 6 illustrates the changes in the unit response as a function of the fraction of reference nodes. The impact of reference nodes is positive and super-linear.

The regression analysis indicated that a third-order polynomial provided the best fit, explaining 0.95 of the variability. Experimentally obtained values of these coefficients gave the expression in the following equation:

$$\begin{aligned}u_5 &= 1-\left(a_5r^3+b_5r^2+c_5r+d_5\right)\\ a_5 &= 223.89*f+0.4523\\ b_5 &= -100*f-0.1851\\ c_5 &= 15.09*f+0.0202\\ d_5 &= -0.9601*f+1\end{aligned}\tag{5}$$

where

u_5	fifth aspect of the uncertainty (failure);
a_5, b_5, c_5, d_5	coefficients;
f	failure rate, i.e., the fraction of nodes that failed at the same time [0, 1];
r	fraction of reference nodes (0, 1].

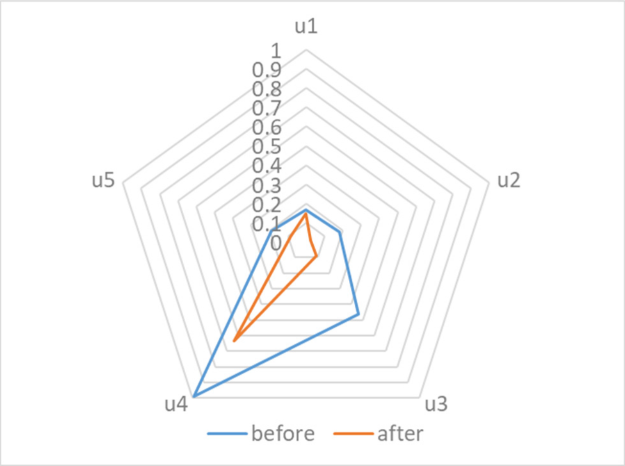


Fig. 7. Output of the model, showing the change in uncertainty across five aspects, for an increase in the fraction of reference nodes from 0.001 to 0.05 (u1: noise, u2: inoperability, u3: transition, u4: drift; u5: failure).

VI. Model

The objective of this research was to construct a model that linked the changes in the fraction of reference nodes to the changes in the level of uncertainty, to provide information for decisions on the replacement of some of the regular nodes with reference nodes. As already stated, different aspects of uncertainty may not be easily comparable. The model therefore deals with these uncertainties separately, using equations developed on the basis of simulations.

The model developed here can be used to compare the structure of the uncertainty across five aspects, before and after a planned increase in the fraction of reference nodes. The graphical representation of the outcome of the model is shown as a radar plot in Fig. 7.

In order to use the model, it is necessary to input the current and planned fractions of reference nodes, and to set the four parameters used in the equations: the deviation, the fraction of faulty nodes, the slope of the change, and the failure rate. The values of these parameters can usually be determined from the network requirements or planning documents. In the example provided below, the values used are as follows: deviation: 0.5, faulty nodes: 0.2, slope: 0.01, failure rate: 0.2.

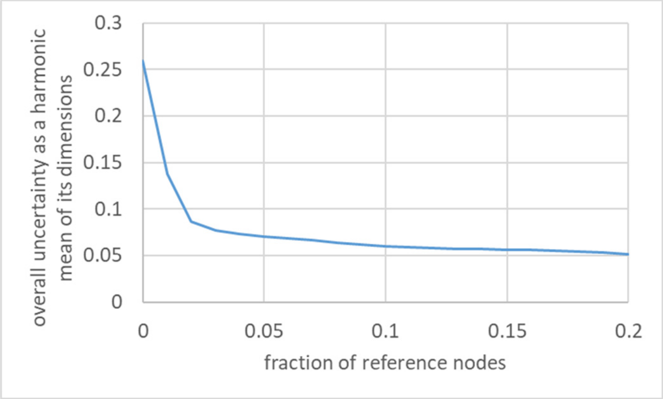


Fig. 8. Overall uncertainty (calculated as a harmonic mean of its aspects) as a function of the fraction of reference nodes.

Fig. 7. shows the impact of replacing the first 5% of regular nodes with reference nodes on a network that contains only 0.1% of reference nodes. It can be seen that the area associated with the original uncertainty shrinks significantly.

While it is not generally advisable to combine the different aspects of uncertainty, for illustration purposes only, Fig. 8 shows the relationship between the fraction of reference nodes and the harmonic average of all five aspects of uncertainty, for set values of the parameters.

It can be seen that the marginal benefit (in terms of the reduction in uncertainty) of adding the initial few percentages of reference nodes significantly outweighs any further investment in additional reference nodes.

VII. Discussion and Conclusion

This article presents a model of the impact of reference nodes on the uncertainty in hybrid ad-hoc sensor networks that use popular reference-based schemes of type C [3]. The model allows for a comparison of the uncertainty before and after a planned change, enabling the user to assess the benefit of introducing additional reference nodes across five aspects of uncertainty [6], for low fractions of reference nodes.

The model was developed on the basis of a series of simulations. The algorithm and the parameters were chosen to reflect typical, expected conditions of the network.

The results showed that an increase in the fraction of reference nodes had a positive impact on the containment of uncertainty across all aspects, although to different extents. Furthermore, the marginal benefit of adding a reference node was much higher for networks with smaller fractions of existing reference nodes, and quickly diminished once the fraction of reference nodes reached 5%–10%. In view of the significant price difference between regular and reference nodes, this indicates that it is economically feasible to create a quality ad-hoc sensor network with a relatively small additional outlay.

Compared to reputation-based schemes with no reference nodes (as described in [2]), the proposed model dealt with long-term drift particularly well. Long-term drift is generally challenging for any network that uses some form of reputation, as a gradual increase in uncertainty may not be picked up by opinions and reputations, leading to reference nodes being out-voted from the network. In this case, even a very significant drift was successfully counteracted by relatively small fractions of reference nodes.

There are some inherent limitations on the applicability of this approach. The choice of simulation as a method required several decisions about the implementation of the fusion algorithm, the architecture of the network, and the details of the metric. This may mean that the results presented here are not directly comparable. However, it was not the objective of this article to develop a better algorithm; instead, we aimed to demonstrate that an excessive investment in reference nodes is not necessary as only a small fraction of reference nodes may be all that the network requires to contain its uncertainty.

References

[1] F. Mao, K. Khamis, S. Krause, J. Clark, and D. M. Hannah, "Low-cost environmental sensor networks: Recent advances and future directions," *Frontiers Earth Sci.*, vol. 7, p. 221, Sep. 2019, doi: 10.3389/feart.2019.00221.

[2] P. Cofta, C. Orlowski, and J. Lebiedz, "Trust-based model for the assessment of the uncertainty of measurements in hybrid IoT networks," *Sensors*, vol. 20, no. 23, p. 6956, Dec. 2020, doi: 10.3390/s20236956.

[3] A. Gutscher, J. Heesen, and O. Siemoneit, "Possibilities and limitations of modeling trust and reputation," in *Proc. 5th Int. Workshop Philosophy Informat.*, vol. 332, Apr. 2008, pp. 50–61.

[4] C. Orlowski, P. Cofta, M. Wasik, P. Welfler, and J. Pastuszka, "The use of group decision-making to improve the monitoring of air quality," in *Transactions on Computational Collective Intelligence XXXIV*, vol. 11890, N. T. Nguyen, R. Kowalczyk, J. Mercik, A. Motylska-Kuzma, Eds. Berlin, Germany: Springer, 2019, pp. 127–145, doi: 10.1007/978-3-662-60555-4_9.

[5] M. V. Narayana, D. Jalihal, and S. M. S. Nagendra, "Establishing a sustainable low-cost air quality monitoring setup: A survey of the state-of-the-art," *Sensors*, vol. 22, no. 1, p. 394, Jan. 2022, doi: 10.3390/s22010394.

[6] N. Zimmerman, "Tutorial: Guidelines for implementing low-cost sensor networks for aerosol monitoring," *J. Aerosol Sci.*, vol. 159, Jan. 2022, Art. no. 105872, doi: 10.1016/j.jaerosci.2021.105872.

[7] P. Cofta, K. Karatzas, and C. Orlowski, "A conceptual model of measurement uncertainty in IoT sensor networks," *Sensors*, vol. 21, no. 5, p. 1827, Mar. 2021, doi: 10.3390/s21051827.

[8] E. Adamiec et al., "Using medium-cost sensors to estimate air quality in remote locations. Case study of Niedzica, Southern Poland," *Atmosphere*, vol. 10, no. 7, p. 393, Jul. 2019, doi: 10.3390/atmos10070393.

[9] M. Simek, P. Moravek, D. Komosny, and M. Dusik, "Distributed recognition of reference nodes for wireless sensor network localization," *Radioengineering*, vol. 21, no. 1, pp. 89–98, Apr. 2012.

[10] K. Ghuman, "Some theory and an experiment on the fundamentals of Hirschman uncertainty," Ph.D. dissertation, Dept. Elect. Comput. Eng., Florida State Univ., FL, USA, 2015. [Online]. Available: http://purl.flvc.org/fsu/fd/FSU_2015fall_ Ghuman_fsu _0071E_12257

[11] B. E. Lovell, "A taxonomy of types of uncertainty," Ph.D. dissertation, Dept. Syst. Sci., Portland State Univ., Portland, OR, USA, 1995. [Online]. Available: https://pdxscholar.library.pdx.edu/open_access_etds/1396/

[12] P. V. Varde, M. G. Pecht, and H. Pham, "Uncertainty modeling," in *Risk-Based Engineering: An Integrated Approach to Complex Systems-Special Reference to Nuclear Plants* (Springer Series in Reliability Engineering). Singapore: Springer, 2018, pp. 291–311.

[13] P. Cofta, *Trust, Complexity and Control: Confidence in a Convergent World*. Hoboken, NJ, USA: Wiley, 2007, doi: 10.1002/9780470517857.

[14] J. M. Barcelo-Ordinas, M. Doudou, J. Garcia-Vidal, and N. Badache, "Self-calibration methods for uncontrolled environments in sensor networks: A reference survey," *Ad Hoc Netw.*, vol. 88, pp. 142–159, May 2019, doi: 10.1016/j.adhoc.2019.01.008.

[15] M. R. Senouci, A. Mellouk, L. Oukhellou, and A. Aissani, "Uncertainty-aware sensor network deployment," in *Proc. IEEE Global Telecommun. Conf.*, Houston, TX, USA, Dec. 2011, pp. 1–5, doi: 0.1109/GLOCOM.2011.6134363.

[16] B. Placzek and M. Bernas, "Uncertainty-based information extraction in wireless sensor networks for control applications," *Ad Hoc Netw.*, vol. 14, pp. 106–117, Mar. 2014, doi: 10.1016/j.adhoc.2013.11.009.

[17] S. Guo, H. Zhang, Z. Zhong, J. Chen, Q. Cao, and T. He, "Detecting faulty nodes with data errors for wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 10, no. 3, pp. 1–27, Apr. 2014, doi: 10.1145/2594773.

[18] A. Mahapatro and P. M. Khilar, "Detection of node failure in wireless image sensor networks," *Int. Scholarly Res. Notices*, vol. 7, no. 1, 2012, Art. no. 342514, doi: 10.5402/2012/342514.

[19] S. Abdelwahab, T. Gaber, and M. Wahed, "Trust-based security models in wireless sensor networks: A survey," *Int. J. Comput. Intell. Stud.*, vol. 6, nos. 2–3, pp. 245–266, 2017, doi: 10.1504/IJCISTUDIES.2017.089057.

[20] J.-H. Cho, A. Swami, and I.-R. Chen, "A survey on trust management for mobile ad hoc networks," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 562–583, 4th Quart., 2011, doi: 10.1109/SURV.2011.092110.00088.

[21] S. A. Soleymani et al., "Trust management in vehicular ad hoc network: A systematic review," *EURASIP J. Wireless Commun. Netw.*, vol. 2015, no. 1, pp. 1–22, Dec. 2015, doi: 10.1186/s13638-015-0353-y.

[22] S. Sun, Z. Ma, L. Liu, H. Gao, and J. Peng, "Detection of malicious nodes in drone ad-hoc network based on supervised learning and clustering algorithms," in *Proc. 16th Int. Conf. Mobility, Sens. Netw. (MSN)*, Dec. 2020, pp. 145–152, doi: 10.1109/MSN50589.2020.00037.

[23] M. Rizwanullah et al., "Development of a model for trust management in the social Internet of Things," *Electronics*, vol. 12, no. 1, p. 41, Dec. 2022, doi: 10.3390/electronics12010041.

[24] H. Darvishi, D. Ciuonzo, and P. S. Rossi, "A machine-learning architecture for sensor fault detection, isolation, and accommodation in digital twins," *IEEE Sensors J.*, vol. 23, no. 3, pp. 2522–2538, Feb. 2023, doi: 10.1109/JSEN.2022.3227713.

[25] F. Azzedin and M. Ghaleb, "Internet-of-Things and information fusion: Trust perspective survey," *Sensors*, vol. 19, no. 8, p. 1929, Apr. 2019, doi: 10.3390/s19081929.

[26] V. U. Rani and K. S. Sundaram, "Review of trust models in wireless sensor networks," *Int. J. Comput. Inf. Syst. Control Eng.*, vol. 8, pp. 371–377, Apr. 2014.

[27] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006, doi: 10.1109/TIT.2006.874516.

[28] S. S. Babu, A. Raha, and M. K. Naskar, "A direct trust dependent link state routing protocol using route trusts for WSNs (DTLSRP)," *Wireless Sensor Netw.*, vol. 3, no. 4, pp. 125–134, 2011, doi: 10.4236/wsn.2011.34015.

[29] P. Shi and H. Chen, "RASN: Resist on-off attack for wireless sensor networks," in *Proc. 2nd Int. Conf. Comput. Appl. Syst. Model.*, 2012, pp. 690–693, doi: 10.2991/iccasm.2012.175.

[30] Y. Wang, K.-J. Lin, D. S. Wong, and V. Varadharajan, "The design of a rule-based and event-driven trust management framework," in *Proc. IEEE Int. Conf. e-Business Eng. (ICEBE)*, Hong Kong, Oct. 2007, pp. 97–104, doi: 10.1109/ICEBE.2007.28.

[31] K. A. Awan, I. U. Din, A. Almogren, H. Almajed, I. Mohiuddin, and M. Guizani, "NeuroTrust—Artificial-neural-network-based intelligent trust management mechanism for large-scale Internet of Medical Things," *IEEE Internet Things J.*, vol. 8, no. 21, pp. 15672–15682, Nov. 2021, doi: 10.1109/JIOT.2020.3029221.

[32] H. Alzaid, M. Alfaraj, S. Ries, A. Jøsang, M. Albabtain, and A. Abuhaimed, "Reputation-based trust systems for wireless sensor networks: A comprehensive review," in *Trust Management VII*, vol. 401, C. Fernandez-Gago, F. Martinelli, S. Pearson, and I. Agudo, Eds. Berlin, Germany: Springer, 2013, doi: 10.1007/978-3-642-38323-6_5.

[33] A. Jøsang, "A logic for uncertain probabilities," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 9, no. 3, pp. 279–311, Jun. 2001, doi: 10.1142/S0218488501000831.

[34] K. Sentz and S. Ferson. (2002). *Combination of Evidence in Dempster-Shafer Theory*. [Online]. Available: https://www.stat.berkeley.edu/~aldous/Real_World/dempster_shafer.pdf

[35] J. Whalley and S. Zandi, "Particulate matter sampling techniques and data modelling methods," in *Air Quality—Measurement and Modeling*, P. Sallis, Ed. London, U.K.: IntechOpen, 2016.

Piotr Cofta (Senior Member, IEEE) is a Professor with the Telecommunications, Computer Science and Electrical Engineering, Bydgoszcz University of Science and Technology, Bydgoszcz, Poland. He has authored or coauthored over 60 peer-reviewed publications and several patents. His research interests include uncertainty reduction in heterogeneous ad-hoc sensor networks with the use of computational trust and artificial neural networks.

Mr. Cofta was a Fellow Member of BCS.

Beata Marciniak is an Assistant Professor with the Telecommunications, Computer Science and Electrical Engineering, Bydgoszcz University of Science and Technology, Bydgoszcz, Poland. She has authored or coauthored over 45 peer-reviewed publications and for over ten years. she has been working in the organizing committees of international conferences organized/co-organized by the Faculty of Telecommunications, Computer Science and Electrical Engineering. Her research interests include the study of throughput in ICT networks, decision making in multi-agent systems, and the use of artificial intelligence in various aspects of image processing.