electronics

MDPI

Article

An Enhanced Trust Mechanism with Consensus-Based False Information Filtering Algorithm against Bad-Mouthing Attacks and False-Praise Attacks in WSNs

Taisuk Suh and Youngho Cho *

Department of Computer Engineering, Graduate School of Defense Management, Korean National Defense University, Nonsan 33021, Korea; corneli1202@gmail.com

* Correspondence: youngho@kndu.ac.kr

Received: 4 October 2019; Accepted: 14 November 2019; Published: 16 November 2019

Abstract: To defend against insider attacks in wireless sensor networks (WSNs), trust mechanisms (TMs) using the notion of trust in human society have been proposed and are still actively researched. In the WSN with a trust mechanism (TM), each sensor node evaluates the trustworthiness of its neighbor sensors based on their behaviors, for example packet forwarding, and collaborates only with trustworthy neighbors while removing untrustworthy neighbor from its neighbor list. The reputation system (RS) is an advanced type of trust mechanism that evaluates the trustworthiness of a node by additionally considering neighbor nodes' observations or evaluations about it. However, intelligent inside attackers in WSNs can discover the security vulnerabilities of trust mechanisms by examining the operations of TM (or RS), because the software modules of the TM (or RS) are installed and operating in their local storage and memory, and thus, they can avoid detection by the trust mechanisms. Bad-mouthing attacks and false-praise attacks are well-known examples of such intelligent insider attacks. We observed that existing trust mechanisms do not have effective countermeasures to defend against such attacks. In this paper, we propose an enhanced trust mechanism with a consensus-based false information filtering algorithm (TM-CFIFA) that can effectively defend against bad-mouthing attacks and false-praise attacks. According to our experiment results, compared with an existing representative RS model, our TM-CFIFA shortened the detection time of a packet drop attacker, which is supported by a false-praise attacker by at least 83%, and also extended the lifetime of a victim sensor node that is under bad-mouthing attacks by at least 15.8%.

Keywords: trust mechanism; insider attacks; bad-mouthing attack; false-praise attack; consensus observation; false information filtering; wireless sensor network

1. Introduction

With recent advancements in Internet-of-Things (IoT) technologies, it is expected that tens of billions of IoT devices will be interconnected by 2022 [1], and thus the usage of WSNs will also grow quickly in various industry areas [2–5] as well as in military fields [6]. Due to many WSN characteristics, such as it is a wireless medium and the limited resources of sensors (low battery, storage, and computing speed), security is one of the most important design factors of WSNs. WSNs are considered more unsafe than other types of networks, and are especially vulnerable to insider threats [7–9]. In addition, energy-efficiency is another critical design factor to maximize the lifetime of WSNs [10,11].

To defend against insider attacks in WSNs, trust mechanisms (TMs) have been proposed and studied as a promising defense method [12–14]. In general, a basic TM works in three phases as follows: (1) it observes its neighbor nodes' behaviors (direct observations); (2) evaluates the neighbor sensor's

trustworthiness based on monitored behaviors; and (3) detects inside attackers (or untrustworthy sensors). In addition, as an advanced type of TM, reputation systems (RSs) have been proposed which improves TM's second phase such that a sensor node with RS evaluates the trustworthiness of its neighbor sensors by additionally considering information (indirect observations or indirect trust evaluations) from its other neighbor nodes. However, intelligent inside attackers in WSNs may be able to discover security vulnerabilities in trust mechanisms by investigating their operations, and thus, they can avoid the detection of trust mechanisms. Bad-mouthing attacks and false-praise attacks are well-known as intelligent insider attacks [9,15,16]. In these attacks, attackers provide an evaluating sensor with false information to hamper accurate trust evaluation. This is possible because existing TMs and RSs simply receive such false information from its neighbor nodes with high trust values above a certain threshold and mistakenly calculate the final trust value based on such false information. Moreover, according to our extensive survey, we observed that existing TMs and RSs do not have effective countermeasures to defend against bad-mouthing attacks and false-praise attacks.

In this paper, we advance existing trust mechanisms by eliminating such false information from bad-mouthing attackers and false-praise attackers by using our consensus-based false information filtering algorithm (CFIFA). Our contributions can be summarized as the following:

- We propose an enhanced trust mechanism with a consensus-based false information filtering algorithm (TM-CFIFA) that can effectively defend against bad-mouthing attacks and false-praise attacks.
- We conduct experiments that show our TM-CFIFA can better defend against two attacks by comparing it with a representative RS used in various trust-aware routing algorithms including light-weight trust aware routing protocol (LTRP) [17–20]. The results show that our TM-CFIFA not only better defends against bad-mouthing attacks and false-praise attacks but also extends the network lifetime of WSNs by at least 15.8% in our experimental setups.

The rest of this paper is organized as follows. In Section 2, we give a brief overview of insider attack problems in WSNs, trust mechanisms and reputation systems, two intelligent insider attacks, and existing defense methods. In Section 3, we discuss the proposed design of our TM-CFIFA. In Section 4, we describe the experiments that show the performance of TM-CFIFA compared to a representative RS model. Finally, we make our conclusions in Section 5.

2. Background and Related Works

2.1. Insider Attacks in WSNs

In WSNs, each sensor node sends its data packets toward the destination node by means of multi-hop collaboration. For example, as shown in Figure 1a, when source node A wants to send its packet to the destination node D, node A cannot directly send it to node D due to its limited energy or hardware capability [21]. Instead, node A first forwards it to node B, hoping that its data packet can reach node D via a routing path A → B → C → D. That is, A needs the help of two intermediate nodes B and C. Consequently, establishing mutual trust among inside nodes in WSNs are essential to guarantee that WSNs work correctly according to their design goals.

However, it is possible for these intermediate nodes to become inside attackers (or traitors) due to various reasons (e.g., hacking by adversaries) [7,9]. Moreover, what makes this problem more serious is that sensors may not have the same advanced heavy security mechanisms used in other networks due to their unique limitations, as mentioned above. For this reason, various inside attack problems in WSNs, such as attack models and defense mechanisms have been actively discussed and studied [17,19,20,22], and also trust mechanisms have been researched as promising defense mechanisms against inside attackers in WSNs [12,18,23].

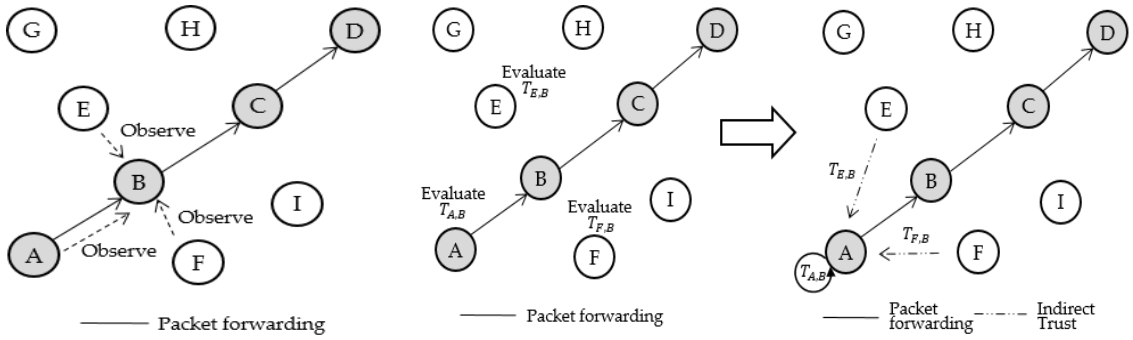


Figure 1. Working phase 1 and 2 of general trust mechanism; (**a**) Phase 1(monitoring/recording) and (**b**) Phase 2 (trust measurement).

2.2. Trust Mechanism (TM) and Reputation System (RS)

When a TM is deployed in WSNs, every sensor node has TM in its local memory. Each sensor node can evaluate the trustworthiness of its neighboring nodes according to their behaviors or operations in WSNs such that if a neighbor node's behavior is observed as successful or cooperative, the trust value of the node will increase or otherwise the trust value of the node will decrease. To do this, TMs generally work in three phases as follows [9].

- **Phase 1 (Monitoring/Recording):** Each sensor node monitors its neighbor nodes' behaviors, for example, packet forwarding/relaying, and then records whether their behaviors are performed successfully and cooperatively (see Figure 1a). Watchdog is a representative, widely adopted monitoring mechanism for this phase [24,25]. Basically, Watchdog uses two counters such as a success counter and a failure counter and these counters increase and are recorded according to the observed behaviors of neighbor nodes.
- **Phase 2 (Trust Measurement):** Based on the observation results in Phase 1, each node evaluates the trustworthiness of its neighbor nodes. For trust evaluation, various mathematical trust models have been proposed [18,26]. The Beta trust model [27] is representative of trust models for WSNs because it is lightweight and mathematically sound. When node i evaluates node j's trust value, the Beta trust model calculates the trust value $T_{i,j}$ by Equation (1).

$$T_{i,j}(as,\ af) = \frac{as + 1}{as + af + 2} \tag{1}$$

  where ***as*** is the accumulated number of successes and ***af*** is the accumulated number of failures. $T_{i,j}$ has a value between 0 and 1, and the higher $T_{i,j}$, the more trustworthy the evaluated node is. As we can see in (1), the Beta trust model uses only two parameters (***as*** and ***af***), and thus the combined implementation of the Beta trust model and Watchdog are widely used for Phase 1 and Phase 2 in WSNs [28].
- **Phase 3 (Attack Detection):** In this phase, a sensor node determines whether its neighbor nodes are trustworthy for cooperation. That is, if a certain neighbor node's measured trust value is lower than a certain trust threshold (θ_T), then it is detected as an inside attacker and removed from the WSN.

The reputation system (RS) is an advanced form of trust mechanism that also considers information from neighbor sensor nodes for more accurate trust evaluation [23,29]. That is, when the above example is considered, node i evaluates node j's trust value by using not only its direct observations on node j, but also its neighbor nodes' observations on node j (indirect observations).

For example, as shown in Figure 1b, after node A forwards its packet to node B, hoping that node B will forward the packet toward C → D, node A with Watchdog will monitor B's behaviors and also A's neighbor nodes E and F may be able to observe B's behaviors by using their Watchdog mechanism. Next, to evaluate node B's trust value, node A with RS can use both its direct trust value $DT_{A,B}$ and indirect trust values $IT_{E,B}$ and $IT_{F,B}$ from node E and F, respectively. Then, the final trust value $T_{R:A\rightarrow B}$ can be obtained by Equation (2).

$$T_{R:A\rightarrow B} = w_1 DT_{A,B} + w_2 f(IT_{E,B}, IT_{F,B}) \quad (2)$$

where w_1 is the weight value for direct trust, w_2 is the weight value for indirect trust values and $w_1 + w_2 = 1$ and $f(IT_{E,B}, IT_{F,B})$ is a function that combines indirect trust values; f can be implemented in various ways.

Algorithm 1 describes the basic pseudocodes of the general reputation system that we have explained above, and we used this for the experiments described in Section 5.

Algorithm 1 Reputation System (RS)

Input:
Num. of neighbor nodes which provided node i with their indirect observations: n
Weight factor : $w_1, w_2 \in [0, 1]$
Node i's direct observation to node j: $DO_{i,j} \in \{s, f\}$ # s: success, f: failure
Node k's indirect observation to node j: $IO_{k,j} \in \{s, f\}$
Output:
Direct trust value: $DT_{i,j}$
Indirect trust value: $IT_{i,j}$
Overall trust value: $T_{i,j}$

Begin
  # Direct trust calculation
  if $DO_{i,j}$ == s**:**
    as = as + 1 # increase accumulated success count (***as***) by 1
  else:
    af = af + 1 # increase accumulated failure count (***af***) by 1

  $DT_{i,j} = \frac{as+1}{as+af+2}$

  # Indirect trust calculation
  for each neighbor node k where $1 \leq k \leq n$**:**
    if $IO_{k,j}$ == s**:**
      $as_{k,j} = as_{k,j} + 1$
    else:
      $af_{k,j} = af_{k,j} + 1$

    $IT_{k,j} = \frac{as_{k,j}+1}{as_{k,j}+af_{k,j}+2}$

  $IT_{i,j} = \frac{1}{n}\sum_{k=1}^{n} IT_{k,j}$

  # Overall trust calculation
  $T_{i,j} = w_1 DT_{i,j} + w_2 IT_{i,j}$
End

2.3. Intelligent Insider Attacks: Bad-Mouthing Attack and False-Praise Attack

Here, we introduce two intelligent attacks (bad mouthing attack and false-praise attack) that exploit the design and operational characteristics of reputation systems used in WSNs, and thus, they can hamper the correct operation of trust mechanisms.

- **Bad-Mouthing Attack**: As shown in Figure 2a, the bad-mouthing attacker (node F or I) intentionally provides the evaluating node (node A) with false information about the evaluated node (node B) such that B does not forward A's packets correctly, although B forwards A's packets to node C correctly. If the bad-mouthing attacker continues to launch attacks, B's trust value will become lower than a trust threshold and eventually B will be removed from its neighbor list. Once B is removed, A will find another neighbor node (node E) as its next hop and then A's packets will be routed along the path E → H → D, which is less optimal than the original optimal routing path A→B→C→D in terms of energy efficiency or routing distance. Consequently, bad-mouthing attacks can degrade the entire network performance by eliminating many normal nodes in WSNs.
- **False-Praise Attack**: As shown in Figure 2b, this attacker (node F or I) deliberately increases the trust value of an evaluated node (node B); in this example, B is a packet drop attacker and is collaborating with these false-praise attackers. As the attack name shows, the false-praise attackers (node F and I) continue to provide node A with false information such that node B behaves correctly although B drops all packets from node A. As a result, node B's trust in A may not significantly decrease due to false observations or indirect trust values from the two false-praise attackers (see Equation (2)).

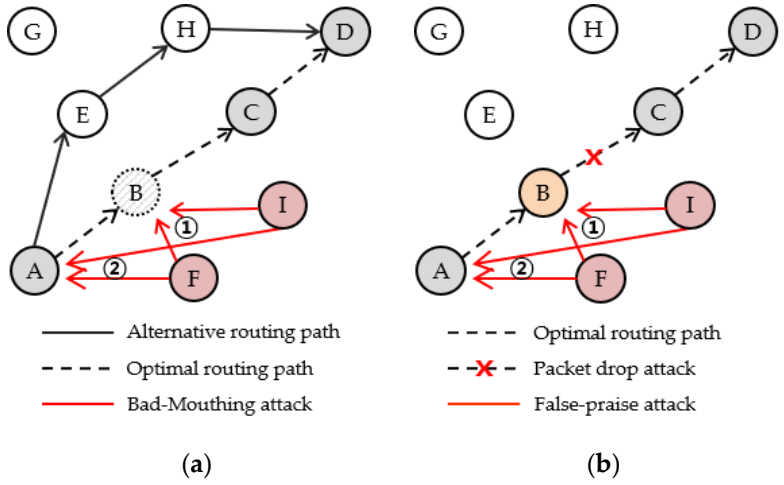


Figure 2. Two intelligent attacks against reputation systems; (**a**) Bad-mouthing attack and (**b**) False-praise attack.

2.4. Existing Defense Approaches against Bad-Mouthing Attacks and False-Praise Attacks

There are a number of review studies that overview the defense capabilities of existing reputation systems against inside attacks including bad-mouthing attacks and false-praise attacks.

Khalid et al. [18] compared various trust and reputation systems in WSNs. They examined them in terms of network initialization, trust computation, security attack prevention, and so on. In particular, they reported that CORE, ATSR, DETM, CONFIDANT, and RRS can defend against bad-mouthing attacks and false-praise attacks, and these models adopt a reputation system framework. Similarly, Ahmed et al. [23] examined existing trust models and mechanisms in terms of trust evidence, trust evaluation, attack model, routing protocol, and so on. They introduced several reputation systems that can effectively defend against bad-mouthing attacks.

Reputation system-based secure routings have also been studied to counter misbehaving nodes in WSNs.

Duan et al. [19] used the trust-aware secure routing framework (TSRF) to defend against misbehaving nodes. TSRF uses trust and QoS metrics together to find optimal routes from the source node to the destination node before packet transmission. In this case, nodes send and receive recommendation requests to find such optimal routes. Their experiments show that when TSRF is used in WSNs, the effect of bad-mouthing attacks diminishes.

Tornos et al. [20] proposed trust authenticated dynamic source routing (TADSR) in MANETs to detect rogue nodes and improve the routing performance. The basic concept of TADSR is to mix secure routing and trust management. They used bad-mouthing attack models to verify TADSR's defense performance against inside attacks.

Ahmed et al. [17] used light-weight trust aware routing protocol (LTRP) to detect misbehaving nodes and isolate them. LTRP considers various metrics such as trust, remaining energy, and hop count to defend against malicious nodes.

The above three models (TSRF, TADSR, and LTRP) use their own features such as hop count, QoS metrics and remaining energy to find routing paths. They have a common feature, the reputation system framework, which uses direct and indirect trust to defend against misbehaving nodes. All three models use the basic equation of reputation systems first, and then consider some other metrics to improve routing performance.

According to our survey, the existing reputation system-based approaches have a critical limitation, that is, they receive indirect information from neighboring nodes and then simply use them for trust evaluation without examining whether they are true or false.

Consequently, in this paper, we propose an enhanced trust mechanism based on a consensus-based false information filtering algorithm (TM-CFIFA) that can improve the trust evaluation process of existing trust mechanisms by using a false information filtering algorithm.

3. Proposed Trust Mechanism with Consensus-Based False Information Filtering Algorithm

In this section, we first describe a critical weakness in the existing reputation systems that bad-mouthing attackers and false-praise attackers can exploit, explain our idea to enhance existing trust mechanisms to better defend against such attacks, and outline the design of our proposed mechanism, the TM-CFIFA.

3.1. Weakness in Existing Reputation Systems

To defend against insider attacks in WSN, using neighbor nodes' help is very useful and that is why reputation systems have been proposed in this research area. In the WSN with reputation systems, an evaluating node will receive indirect information only from trustworthy neighbors with high trust value above a predetermined trust threshold. However, it may not be safe to assume that nodes with high trust value are not inside attackers, because such nodes with high trust value may turn into insider attackers for various reasons such as hacking by adversaries; these kind of insider attackers with high trust value are called traitors.

However, the existing reputation systems do not recognize these inside attackers with high trust value, and thus they simply receive the false information provided by them. As a result, bad-mouthing and false-praise attackers can easily achieve their intended goals by disguising the evaluating nodes. For example, Figure 3a shows the trust evaluation phase of a general reputation system. In the figure, evaluating node I will calculate the final value (or reputation value) $T_{R:I \rightarrow J}$ by using both direct trust values (DT) and indirect trust values (IT_1, IT_2, ... , IT_n). In this case, if IT_2 is an indirect trust value provided from an inside attacker (e.g., bad-mouthing attacker), RS will use IT_2 as one of the input values for evaluating the aggregated indirect trust value and overall trust value, which negatively affects the correct evaluation of the evaluated node J. That is, there is no countermeasure that removes such false information (IT_2) before the overall trust evaluation phase is conducted.

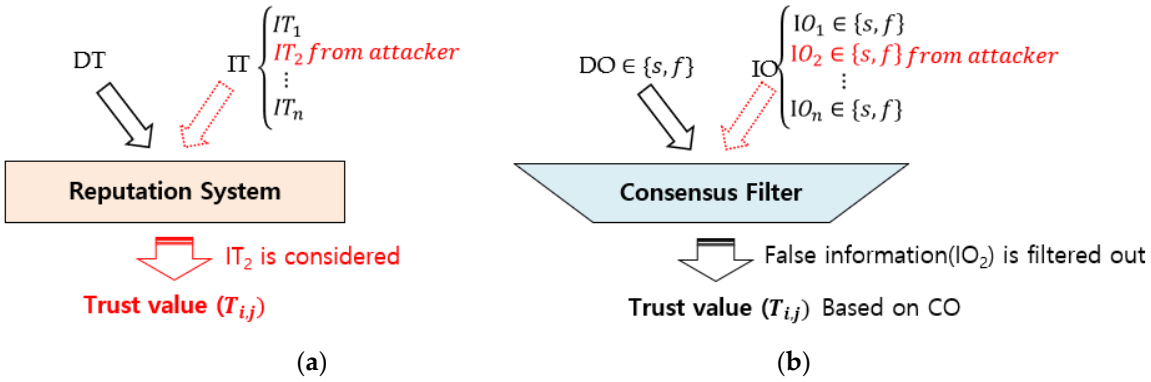


Figure 3. Handling false information provided from an inside attacker in existing reputation systems and our approach, a trust mechanism with a consensus-based false information filtering algorithm (TM-CFIFA); (**a**) Reputation System (RS) and (**b**) TM-CFIFA.

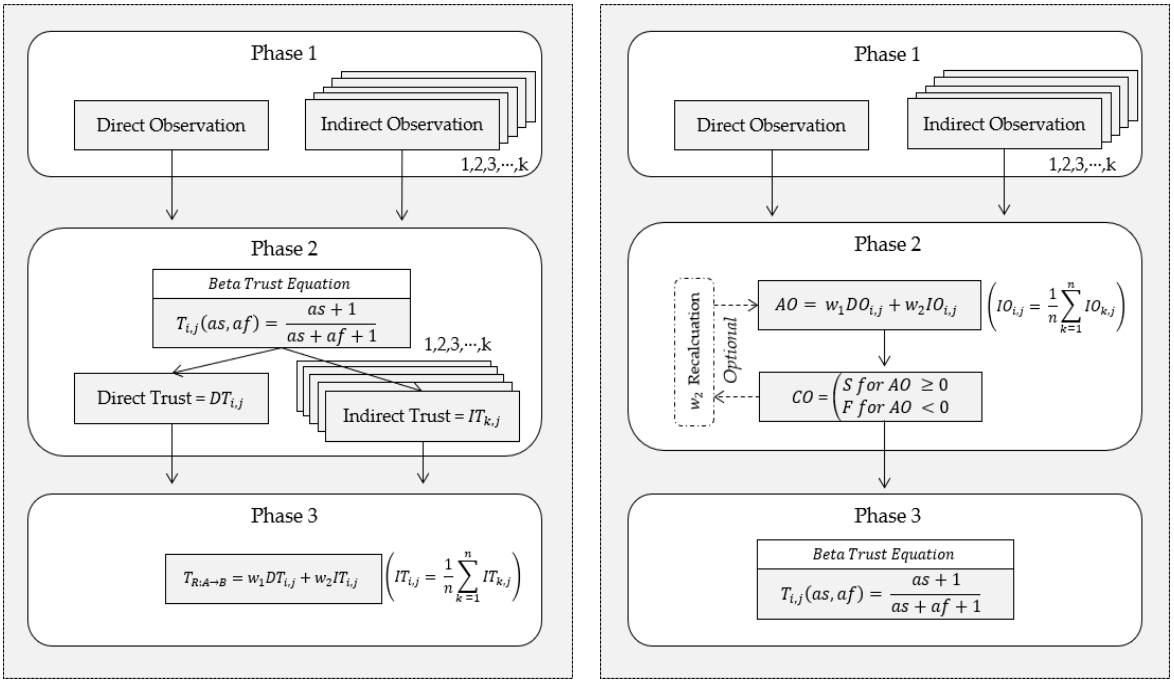
3.2. Our Idea: Filter False Information Based on Consensus Observations about Evaluated Nodes' Behaviors

To resolve the above-mentioned weakness in the existing reputation systems, our approach is to filter out false information from inside attackers by consensus among nodes, and thus such false information can be removed and cannot be used in trust evaluation, if there are more than half of the good nodes participating in the consensus process.

Figure 3b shows how our proposed TM-CFIFA can remove false information from inside attackers even in a situation where we do not know which of the neighbor nodes are inside attackers. In this figure, like the RS, our TM-CFIFA first receives both direct observation (DO) and indirect observations (IOs) from neighbor nodes and then the consensus filtering algorithm of TM-CFIFA produces consensus observation (CO), which is either "success" or "failure". One of the common ways to make a consensus is to use a majority voting method. Consequently, if we assume that more than half of the nodes are good in the WSN, the consensus observation will be a true observation according to the concept of majority voting. In this manner, we believe our TM-CFIFA will correctly eliminate false information from inside attackers in WSNs. Based on this rationale, in this paper, we propose an enhanced trust mechanism that uses a consensus-based false information filtering algorithm (TM-CFIFA) to defend against bad-mouth attackers and false-praise attackers in WSNs.

3.3. Design of TM-CFIFA

We designed our proposed trust mechanism with a consensus-based false information filtering algorithm (TM-CFIFA) as follows. First, we explain how our TM-CFIFA evaluates the final trust value in the presence of false information provided by an inside attacker, and then we compare TM-CFIFA with a general reputation system in terms of algorithm time complexity. For simplicity, we used a wireless sensor network model with nine nodes as shown in Figure 1. In this WSN, node A (source node) wants to deliver its packets to node D (destination node) with the help of intermediate nodes B and C in the routing path A → B → C → D. Each packet that A forwarded to node B will be monitored by A's two neighbor nodes E and F as well as A itself. Whenever A sends a packet to B, A will evaluate B's trust value, $T_{A,B}$ by using both A's observation about B's packet forwarding behavior and neighbor nodes' (E and F) observation (or trust evaluation) about B. Based on the above description, to evaluate the final trust value $T_{A,B}$, our TM-CFIFA in node A uses the following steps (see Algorithm 2); for comparison, the working steps for the existing reputation system and our TM-CFIFA are shown in Figure 4.



(**a**) (**b**)

Figure 4. Comparison of working steps in the reputation system (RS) and TM-CFIFA; (**a**) Reputation System (RS) and (**b**) TM-CFIFA.

Step 1. Node A records its direct observations ($DO_{A,B}$) and receives indirect observations ($IO_{E,B}$ and $IO_{F,B}$) from neighbor nodes (node E and F) after monitoring node B's behavior; each observation is recorded as either **s** (for success) or **f** (for failure) in A's local memory.

Step 2. TM-CFIFA calculates the aggregated observation ($AO_{A,B}$) by using $DO_{A,B}$, $IO_{E,B}$ and $IO_{F,B}$ by Equation (3); $AO_{A,B}$ will be used later to generate the consensus observation in Step 3.

$$AO_{A,B} = w_1 DO_{A,B} + w_2 IO_{A,B} \tag{3}$$

where w_1 and w_2 are weight factors for DO and IO, respectively, and $w_1 + w_2 = 1$. In addition, we define $IO_{A,B}$ as the aggregated indirect observation by considering A's neighbor's observations on node B and $IO_{A,B}$ is calculated by Equation (4).

$$IO_{A,B} = \frac{1}{n}\sum_{k \in NS_A} IO_{k,B} \tag{4}$$

where NS_A is the neighbor set of node A, and NS_A = {E, F} in this example, and n is the number of A's neighbor nodes, and n = 2 in this example. To ease the calculation of $IO_{A,B}$, we used 1 for s (success) and -1 for f (fail). For example, if $w_1 = w_2 = 0.5$, $DO_{A,B}$ = s, $IO_{E,B}$ = s, and $IO_{F,B}$ = f, then $IO_{A,B} = 0$ and $AO_{A,B} = 0.5$ by (4) and (3), respectively. We will explain how $AO_{A,B}$ can be used for generating consensus observations in Step 3. Meanwhile, although we set the initial weight factors w_1 and w_2 to 0.5, these weights can be updated periodically by using reinforcement learning techniques [30,31] by considering them after each trust evaluation process ends.

Step 3. Based on $AO_{A,B}$, TM-CFIFA generates consensus observation (CO) by using (5).

$$CO_{A,B} = \begin{cases} \text{s (success) if } AO_{A,B} \geq 0 \\ \text{f (failure) if } AO_{A,B} < 0 \end{cases} \tag{5}$$

Algorithm 2 TM-CFIFA

Input:
Num. of neighbor nodes which provided node i with their indirect observations: n
Weight factor : $w_1, w_2 \in [0, 1]$
Node i's direct observation to node j: $DO_{i,j} \in \{s, f\}$
Node k's indirect observation to node j: $IO_{k,j} \in \{s, f\}$
Output:
Aggregation observation: $AO \in [-1, 1]$
Consensus observation: CO
Overall trust value: $T_{i,j}$

Begin
  # For ease calculation, set DO = 1 for success (s) and DO = -1 for failure (f)
  if $DO_{i,j}$ == s**:**
    $DO_{i,j} = 1$
  else:
    $DO_{i,j} = -1$
  # For ease calculation, set $IO_{k,j}$ = 1 for success (s) and $IO_{k,j}$ = -1 for failure (f)
  for each neighbor node k where $1 \leq k \leq n$ (n: the number of neighbor nodes)
    if $IO_{k,j}$ == s:
      $IO_{k,j} = 1$
    else:
      $IO_{k,j} = -1$
  # Calculate AO by using DO and IO
  $AO_{i,j} = w_1 DO_{i,j} + w_2 \frac{1}{n} \sum_{k=1}^{n} IO_{k,j}$
  # Determine CO according to AO
  if $AO_{i,j} \geq 0$ **:**
    CO = s
    as = as + 1
  else:
    CO = f
    af = af + 1
  # Final trust calculation
  $T_{i,j} = \frac{as+1}{as+af+2}$
End

Next, we conducted an algorithm time complexity analysis by comparing RS (Algorithm 1) with our TM-CFIFA (Algorithm 2) and the analysis results are shown in Table 1. For complexity analysis, we did not consider Phase 1, in which both RS and our TM-CFIFA use the Watchdog mechanism.

Table 1. Comparison of algorithmic time complexity.

Step	Reputation System (RS)	TM-CFIFA
Phase2	O(n)	O(n)
Phase3	O(1)	O(1)
Overall	O(n)	O(n)

First, our algorithm works in O(n), because Phase 2 of Algorithm 2 is the most time-consuming part and Phase 2 has only one single for loop and one summation calculation. Thus, given the input size is n (the number of neighbor nodes), its computational cost will grow linearly as the input size n grows. Therefore, we do not expect our algorithm will introduce huge computational cost when it is used in large-scale WSN with many sensor nodes.

Next, as can be seen in Algorithm 1, the existing reputation system (RS) also works in O(n). Consequently, we claim that our TM-CFIFA will be feasible in large-scale WSNs where the existing reputation system (RS) are used, because our TM-CFIFA work similarly to RS in terms of time complexity (see Table 1). In general, the existing reputation system is used in many parts, including WSNs, because of its lightweight design [27,32].

Therefore, because our TM-CFIFA does not have huge additional computation cost compared with RS, it can better defend against false-praise attacks and bad-mouthing attacks as we will discuss later in Section 4.

4. Experiment and Analysis

4.1. Experimental Environment and Methods

The main purpose of this experiment was to show that our proposed TM-CFIFA, which is an advanced implementation of a trust mechanism, can better defend against bad-mouthing attacks and false-praise attacks compared with an existing reputation system. For this purpose, with Python 3 programming language, we implemented our TM-CFIFA according to Algorithm 2. In addition, for comparative analysis with an existing reputation system, according to Algorithm 1, we implemented a reputation system (RS) that is used in many trust models and trust-aware routing algorithms such as LTRP [17], CORE [33], ATSR [34], TADSR [20], and so on.

We used the following experimental methods and assumptions.

- **Wireless Network Model:** We considered a simple WSN with nine sensor nodes as shown in Figure 5. In this WSN, node A (source node) generates packets and wants to deliver them to the destination node D. As depicted in Figure 5, we assume that the optimal routing path from A to D is determined as A → B → C → D by a routing algorithm in A. Considering natural packet losses in WSNs, the packet forwarding success rate is set to 70%. Each node can monitor its neighbor nodes' packet forwarding behaviors by using the Watchdog mechanism. In this network topology, node A's neighbor nodes are B, E, and F which means that the observation of nodes E and F will be provided to node A.
- **Attack Models**
  (1) Bad-mouthing attack model: Node F (red-colored) launches bad-mouthing attacks to node B (see Figure 5a). That is, F will send false information about B to A such that even though node B successfully forwards A's packets to C, the bad-mouthing attacker F will falsely say B did not send A's packet to C in order to let A mistakenly decrease B's trust value.
  (2) False-praise attack model: Unlike the bad-mouthing attack model, as shown in Figure 5b, node B and F are inside attackers and collaborate with each other; B is a packet drop attacker and F is a false-praise attacker. In this attack model, when node A sends its packet to node B, the packet drop attacker B randomly drops the packet with a drop rate of 70%. However, the false-praise attacker F sends false information to node A such that node B correctly forwarded A's packet to node C in order to let A mistakenly increase B's trust value.

We conducted two types of experiments (Experiment 1 and Experiment 2) to compare the defense performance of the existing reputation system and our proposed TM-CFIFA as follows. In Experiment 1, by using a bad-mouthing attack model, we could compare how long the victim node B stayed in the WSN when RS and our TM-CFIFA were used. In Experiment 2, by using a false-praise attack model, we could compare how quickly the packet drop attacker was captured while a false-praise attacker is helping the packet drop attacker when RS and our TM-CFIFA were used. For both RS and our TM-CFIFA, the weight factor w_1 and w_2 were set to 0.5. In addition, the initial trust value was set to 0.99 in our experiment because we used very high trust thresholds such as 0.9, and thus if the initial trust value was as low as 0.5, then most nodes would be eliminated soon after the simulation

starts. For this reason, high initial trust values have been used in experiments in many studies in the literature [35,36]. We explain each experiment in detail in Section 4.2.

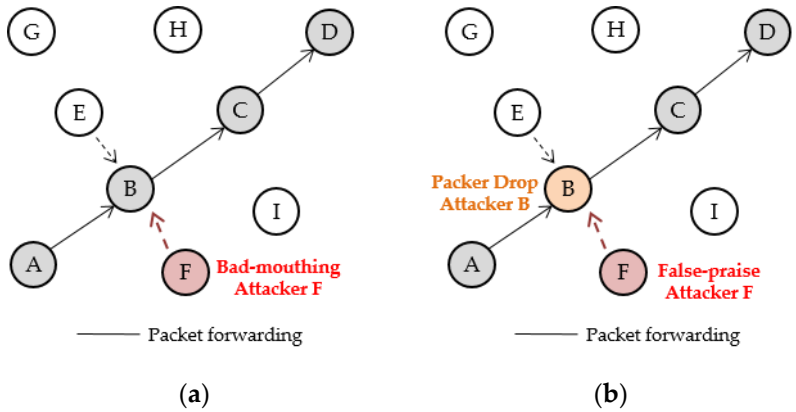


Figure 5. Attack models; (**a**) Bad-mouthing attack and (**b**) False-praise attack.

4.2. Experiment Results and Analysis

4.2.1. Experiment 1: Comparison of Defense Performance in the Presence of Bad-Mouthing Attacks

- Experimental Purpose, Metric and Methods

In Experiment 1, we compared how a trust mechanism in node A accurately evaluates the trust value of a victim node B, and thus lets the victim node B stay in the WSN without being mistakenly eliminated by node A, even in the presence of bad-mouthing attacks (by the attacker F). To this end, we used a metric lifetime (LT), which is defined as the time when node B is falsely detected by a trust mechanism (RS or our TM-CFIFA). For Experiment 1, we used the parameter values shown in Table 2.

Table 2. Experiment parameters.

Parameters	Values	
	Experiment 1 (Bad-Mouthing)	**Experiment 2 (False-Praise)**
Max simulation time	20 min	20 min
Number of Attackers	1 bad-mouthing attacker	1 packet drop attacker 1 false-praise attacker
Initial trust value	0.99	0.99
Trust threshold (θ_T)	0.3~0.9	0.3~0.9
Packet forwarding rate	70%	70%

We conducted Experiment 1 as follows. First, as shown in Figure 5a, node A creates a packet and then sends it to B. When B receives a packet from node A, B forwards it to the next hop node C randomly with a packet forwarding rate = 70%. After that, A collects indirect observations (for our TM-CFIFA) or indirect trust values (for RS) from its neighbor nodes E and F. Next, node A calculates the final trust values by TM-CFIFA and RS. Finally, we check whether the victim node B is falsely detected by TM-CFIFA and RS. We used various detection threshold values in 0.3, 0.9). We set the initial trust value of each node to 0.99. We terminated each experiment either when both TM-CFIFA and RS detected the victim node B or when the simulation time reached 20 min. We conducted 500 experiments and then measured the average LT by TM-CFIFA and RS.

- Results and Analysis

Figure 6 and Table 3 show the results of Experiment 1. According to our experimental results, we can see that in the presence of a bad-mouthing attacker, node B can stay much longer when our

TM-CFIFA is used compared with when RS is used. For example, when θ_T = 0.85, TM-CFIFA falsely detected node B as a packet drop attacker when t = 130 s while RS falsely detected node B when t = 40 s. That is, when our TM-CFIFA is used, node B can continue to stay and participate in the WSN about 325% longer than when RS is used. Table 2 shows the LT of RS and TM-CFIFA according to various θ_T values. We can see that as θ_T grows, the increment of lifetime by our TM-CFIFA also grows. In addition, when $0.75 \leq \theta_T \leq 0.8$, node B was not detected when our TM-CFIFA was used while node B was removed when RS was used. This means that node B can continue to stay and participate in the WSN even in the presence of bad-mouthing attacker since our TM-CFIFA eliminates the false information by the attacker, and thus evaluates node B's trust value correctly. Meanwhile, when $\theta_T \leq 0.7$, both RS and TM-CFIFA could not detect node B in our experiments. This is not surprising because in our experimental WSN, about 30% of packets can be dropped naturally, and thus it is unlikely that node B's trust value will be less than 0.7.

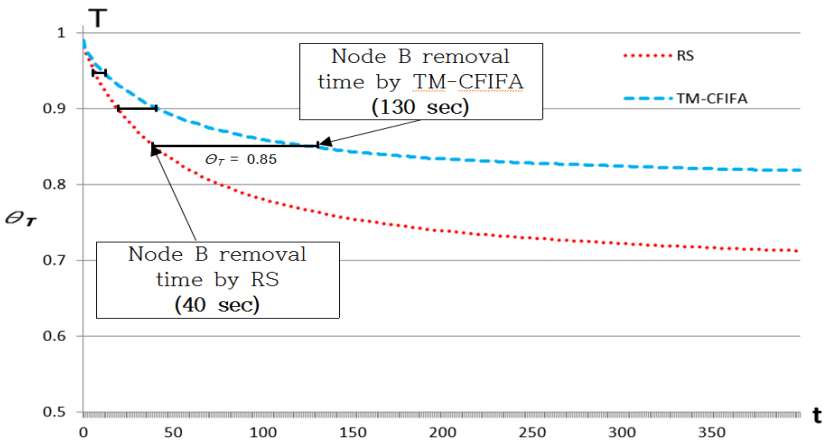


Figure 6. Lifetime (LT) of the victim node B given θ_T (RS vs TM-CFIFA).

Table 3. Experimental results of Experiment 1.

θ_T	DT (Detection Time of Node B)		Comparison Result	
	RS	TM-CFIFA	Lifetime (Δt)	Improvement (%)
0.95	6	11	+ 5	83
0.9	20	42	+ 22	210
0.85	40	130	+ 90	325
0.8	78	Active (not removed)	-	-
0.75	165	Active (not removed)	-	-
0.70	Active (not removed)	Active (not removed)	-	-

4.2.2. Experiment 2: Comparison of Defense Performance in the Presence of False-Praise Attacks

- Experimental Purpose, Metric and Methods

In Experiment 2, we compared how RS and TM-CFIFA evaluate the trustworthiness of a packet drop attacker even in the presence of a false-praise attacker. That is, the false-praise attacker (node F) will keep telling the evaluating node (node A) that the packet drop attacker (node B) forwards its packet correctly towards the destination. To this end, we use a metric detection time (DT) which is defined as the time when the packet drop attack is detected by a trust mechanism. For Experiment 2, we used the parameter values shown in Table 2.

We conducted Experiment 2 as follows. As shown in Figure 5b, node A sends packets to node B, and B forwards it to node C towards the destination node D. In this scenario, B is a packet drop attacker and node F is a false-praise attacker, and B and F are collaborating with each other. Like Experiment 1, we used various trust threshold values in [0.3, 0.9] and set the initial trust value of each node to 0.99.

We terminated each experiment when both RS and TM-CFIFA detected the false-praise attacker node B or when the simulation time reached 20 min. We conducted 500 experiments and then measured average DT of the false-praise attacker by RS and our TM-CFIFA.

- Results and Analysis

Figure 7 and Table 4 show the results of Experiment 2. According to our experimental results, we can see that our proposed TM-CFIFA detected the packet drop attacker much faster than RS, even in the presence of a false-praise attacker. Specifically, TM-CFIFA lowered the detection time (DT) by 15.8~53% compared to RS, according to various θ_T values. For example, when θ_T = 0.6, TM-CFIFA detected the packet drop attacker when t = 134 s, while RS detected the attacker when t = 206 s. That is, TM-CFIFA detected the packet drop attacker (node B) and then removed it 35% faster than RS. Table 4 shows the measured DTs when RS and TM-CFIFA are used given various θ_T. We can see that as θ_T decreases, the improvement in the detection time of our TM-CFIFA also grows. Moreover, when θ_T = 0.4, only our TM-CFIFA could detect the packet drop attacker while the packet drop attacker continues to stay and attack the network when RS is used. However, when we used very low values of θ_T such as 0.3, both RS and our TM-CFIFA were unable to detect the packet drop attacker with a packet drop rate = 70%, within the maximum simulation time (20 min).

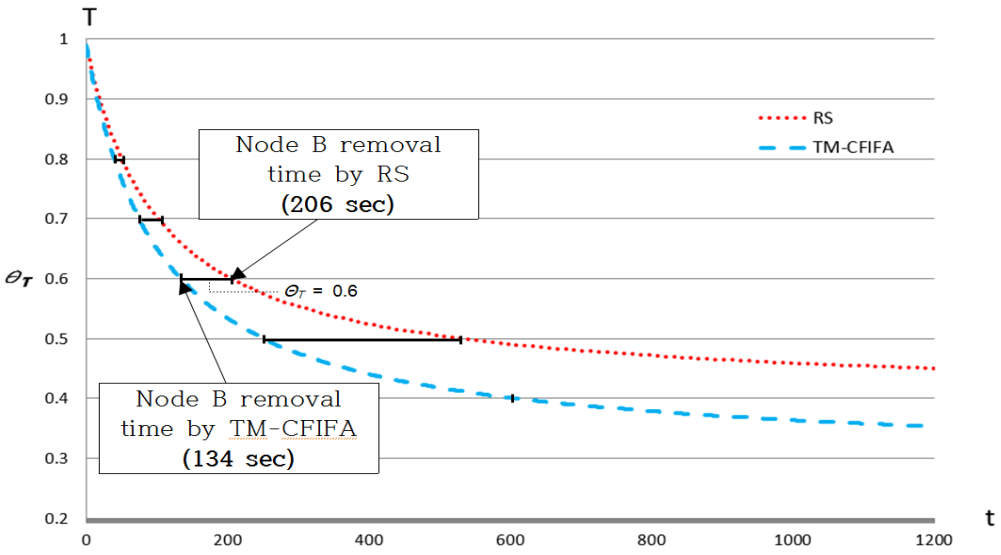


Figure 7. Detection time (DT) of the packet drop attacker given θ_T (RS vs. TM-CFIFA).

Table 4. Experimental results of Experiment 2.

θ_T	DT (Detection Time of Node B)		Comparison Result	
	RS	TM-CFIFA	Detection Time (Δt)	Improvement (%)
0.9	19	16	−3	15.8%
0.8	50	40	−10	20%
0.7	102	77	−25	24.5%
0.6	206	134	−72	35%
0.5	533	251	−282	53%
0.4	Active (Not removed)	607	-	-
0.3	Active (Not removed)	Active (Not removed)	-	-

5. Conclusions and Future Works

In this paper, we proposed an enhanced trust mechanism based on a consensus-based false information filtering algorithm (TM-CFIFA) to effectively defend against bad-mouthing attacks and false-praise attacks in WSNs. Since existing trust mechanisms, including reputation systems,

simply use all or parts of the false information provided by attackers, we proposed and designed the consensus-based false information filtering algorithm (CFIFA) and combined it with the generic architecture of trust mechanisms. According to the results of our experiment, our TM-CFIFA showed a better defense performance against two attack models (bad-mouthing attacks and false-praise attacks) compared with an existing reputation system (RS). Specifically, in our experimental setups, our TM-CFIFA shortened the detection time of a packet drop attacker supported by a false-praise attacker by at least 83% and also extended the lifetime of a victim sensor node that was under bad-mouthing attacks by at least 15.8%.

Future research directions are as follows. First, we will study an insider attack prevention mechanism based on trust mechanisms and blockchain technologies. Specifically, once a trust mechanism detects inside attackers, the identified attackers' identities can be stored in blockchains and then safely spread over the entire sensor nodes, even in the presence of inside attackers in WSNs. Second, we will further investigate the potential limitations and vulnerabilities of current trust mechanisms and reputation systems in the presence of multiple collaborative attackers in WSNs, and thus, we will devise advanced countermeasures that can improve the defense capabilities of existing trust mechanisms and reputation systems to better defend against inside attackers in WSNs. Last, our consensus approach may be vulnerable if Sybil attackers can generate fake identities for more than half of the sensor nodes and can successfully participate in our proposed consensus process. We would like to further investigate Sybil attacks to existing reputation systems in terms of valid attack techniques and their defense methods.

Author Contributions: Conceptualization, T.S. and Y.C.; methodology, Y.C.; software, T.S.; validation, T.S.; formal analysis, T.S. and Y.C.; investigation, T.S.; writing—original draft preparation, T.S.; writing—review and editing, Y.C.; visualization, T.S.; supervision, Y.C.; project administration, Y.C.; funding acquisition, Y.C.

Funding: This study was supported by the Republic of Korea Air Force Academy Research Fund (ROKAFA 19-A-2).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ericson Home Page. Available online: https://www.ericsson.com/en/mobility-report/internet-of-things-forecast (accessed on 3 October 2019).
2. Amin, R.; Islam, S.K.H.; Biswas, G.P.; Khan, M.K. A robust and anonymous patient monitoring system using wireless medical sensor networks. *Future Gener. Comput. Syst.* **2018**, *80*, 483–495. [CrossRef]
3. Cheung, W.F.; Lin, T.H.; Lin, Y.C. A real-time construction safety monitoring system for hazardous gas integrating wireless sensor network and building information modeling technologies. *Sensors* **2018**, *18*, 436. [CrossRef] [PubMed]
4. Ahmedi, F.; Ahmedi, L.; O'Flynn, B.; Kurti, A. InWaterSense: An Intelligent Wireless sensor network for monitoring surface water quality to a river in Kosovo. In *Innovations and Trends in Environmental and Agricultural Informatics*; IGI Global: Hershey, PA, USA, 2018; pp. 58–85.
5. Patil, D.; Thanuja, T.C.; Melinamath, B.C. Air Pollution Monitoring System Using Wireless Sensor Network (WSN). In Proceedings of the 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Pune, India, 19–21 December 2016; pp. 391–400.
6. Ismail, M.N.; Shukran, M.A.; Isa, M.R.M.; Adib, M. Establishing a soldier wireless sensor network (WSN) communication for military operation monitoring. *Int. J. Inf. Commun. Technol.* **2018**, *7*, 89–95. [CrossRef]
7. Shi, E.; Perrig, A. Designing secure sensor networks. *IEEE Wirel. Commun.* **2004**, *11*, 38–43.
8. Ali, S.; Bulushi, T.A.; Nadir, Z. Improving the resilience of Wireless Sensor Networks against security threats: A survey and open research issues. *Int. J. Technol.* **2018**, *9*, 828–839. [CrossRef]
9. Daia, A.S.A.; Ramadan, R.A.; Fayek, M.B. Sensor Networks Attacks Classifications and Mitigation. *Ann. Emerg. Technol. Comput. (AETiC)* **2018**, *2*, 28–43. [CrossRef]
10. Singh, O.; Rishiwal, V.; Kumar, L. Secure Energy Aware Routing in Wireless Sensor Networks. In Proceedings of the 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, India, 18–19 April 2019.

11. Rehman, E.; Sher, M.; Naqvi, S.H.A. Energy efficient secure trust based clustering algorithm for mobile wireless sensor network. *J. Comput. Netw. Commun.* **2017**, *2017*, 1630673. [CrossRef]
12. Yu, Y.; Li, K.; Zhou, W.; Li, P. Trust mechanisms in wireless sensor networks: Attack analysis and countermeasures. *J. Netw. Comput. App.* **2012**, *35*, 867–880. [CrossRef]
13. Garg, S.; Varshney, M.; Nailwal, A. Insider Threats in Wireless Sensor Networks and Their Countermeasures. *Mon. J. Comput. Sci. Inf. Technol.* **2016**, *5*, 476–486.
14. Ishmanov, F.; Bin Zikria, Y. Trust mechanisms to secure routing in wireless sensor networks: Current state of the research and open research issues. *J. Sens.* **2017**, *2017*, 4724852. [CrossRef]
15. Kang, S.; Wu, Y. A trust-based pollution attack prevention scheme in peer-to-peer streaming networks. *Comput. Netw.* **2014**, *72*, 62–73. [CrossRef]
16. Esposito, C.; Castiglione, A.; Palmieri, F. Information theoretic-based detection and removal of slander and/or false-praise attacks for robust trust management with Dempster-Shafer combination of linguistic fuzzy terms. *Concurr. Comput. Pract. Exp.* **2018**, *30*, e4302. [CrossRef]
17. Ahmed, A.; Haseeb, K.; Khokhar, S. A light-weight trust aware routing protocol for wireless sensor network. *Gomal. Univ. J. Res. (Sci.)* **2017**, *33*, 102–112.
18. Khalid, O.; Khan, S.U.; Madani, S.A.; Hayat, K. Comparative study of trust and reputation systems for wireless sensor networks. *Secur. Commun. Netw.* **2013**, *6*, 669–688. [CrossRef]
19. Duan, J.; Yang, D.; Zhu, H.; Zhang, S.; Zhao, J. TSRF: A trust-aware secure routing framework in wireless sensor networks. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 209436. [CrossRef]
20. Tornos, J.L.; Salazar, J.L.; Piles, J.J. Secure Trust Management with Source Routing Protocol for MANETs. *Netw. Protoc. Algorithms* **2015**, *7*, 42–59.
21. Tsiropoulou, E.E.; Baras, J.S.; Papavassiliou, S. On the Mitigation of Interference Imposed by Intruders in Passive RFID Networks. In Proceedings of the International Conference on Decision and Game Theory for Security, New York, NY, USA, 2–4 November 2016.
22. Cho, Y.; Qu, G. Enhancing Trust-Aware Routing by False Alarm Detection and Recovery. In Proceedings of the 2014 IEEE Military Communications Conference, Baltimore, MD, USA, 6–8 October 2014.
23. Ahmed, A.; Bakar, K.A.; Channa, M.I.; Haseeb, K. A survey on trust based detection and isolation of malicious nodes in ad-hoc and sensor networks. *Front. Comput. Sci.* **2015**, *9*, 280–296. [CrossRef]
24. Reddy, V.B.; Negi, A.; Venkataraman, S. A Similarity based Trust Model to Mitigate Badmouthing Attacks in Internet of Things (IoT). In Proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), Limerick, Ireland, 15–18 April 2019.
25. Subba, B.; Biswas, S.; Karmakar, S. Intrusion detection in Mobile Ad-hoc Networks: Bayesian game formulation. *Eng. Sci. Technol. Int. J.* **2016**, *19*, 782–799. [CrossRef]
26. Wu, Y.; Zhao, Y.; Riguidel, M.; Wang, G. Security and trust management in opportunistic networks: A survey. *Secur. Commun. Netw.* **2015**, *8*, 1812–1827. [CrossRef]
27. Oracevic, A.; Akbas, S.; Ozdemir, S. Secure and reliable object tracking in wireless sensor networks. *Comput. Secur.* **2017**, *70*, 307–318. [CrossRef]
28. Cho, Y.; Qu, G.; Wu, Y. Insider threats against trust mechanism with watchdog and defending approaches in wireless sensor networks. In Proceedings of the 2012 IEEE Symposium on Security and Privacy Workshops, San Francisco, CA, USA, 24–25 May 2012; pp. 134–141.
29. Kiefhaber, R. Calculating and Aggregating Direct Trust and Reputation in Organic Computing Systems. Ph.D. Thesis, University of Augsburg, Augsburg, Germany, 2014.
30. Ghobaei-Arani, M.; Jabbehdari, S.; Pourmina, M.A. An autonomic resource provisioning approach for service-based cloud applications: A hybrid approach. *Future Gener. Comput. Syst.* **2018**, *78*, 191–210. [CrossRef]
31. Chu, Y.; Kosunalp, S.; Mitchell, P.D.; Grace, D. Application of reinforcement learning to medium access control for wireless sensor networks. *Eng. Appl. Artif. Intell.* **2015**, *46*, 23–32. [CrossRef]
32. Azad, M.A.; Bag, S.; Hao, F.; Salah, K. M2m-rep: Reputation system for machines in the internet of things. *Comput. Secur.* **2018**, *79*, 1–16. [CrossRef]
33. Michiardi, M.; Molva, R. Core: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In *Advanced Communications and Multimedia Security*; Springer: Boston, MA, USA, 2002; pp. 107–121.

34. Zahariadis, T.; Leligou, H.; Karkazis, P. Design and implementation of a trust-aware routing protocol for large WSNs. *Int. J. Netw. Secur. Its Appl. (IJNSA)* **2010**, *2*, 52–68. [CrossRef]
35. Chen, Z.; He, M.; Liang, W.; Chen, K. Trust-aware and low energy consumption security topology protocol of wireless sensor network. *J. Sens.* **2015**, *2015*, 716468. [CrossRef]
36. Labraoui, N.; Gueroui, M.; Sekhri, L. On-off attacks mitigation against trust systems in wireless sensor networks. In Proceedings of the IFIP International Conference on Computer Science and Its Applications, Saida, Algeria, 20–21 May 2015; Springer: Cham, Switzerland, 2015.