Article

Evaluation of Epidemic-Based Information Dissemination in a Wireless Network Testbed

Andreana Stylidou [1],* [ID], Alexandros Zervopoulos [1] [ID], Aikaterini Georgia Alvanou [1] [ID], George Koufoudakis [1] [ID], Georgios Tsoumanis [2] [ID] and Konstantinos Oikonomou [1],* [ID]

[1] Department of Informatics, Ionian University, GR-49100 Corfu, Greece; azervop@ionio.gr (A.Z.); akorina@ionio.gr (A.G.A.); gkoufoud@ionio.gr (G.K.)

[2] Department of Informatics and Telecommunications, University of Ioannina, GR-47100 Arta, Greece; gtsoum@uoi.gr

* Correspondence: astylidou@ionio.gr (A.S.); okon@ionio.gr (K.O.)

check for updates

Abstract: Information dissemination is an integral part of modern networking environments, such as Wireless Sensor Networks (WSNs). Probabilistic flooding, a common epidemic-based approach, is used as an efficient alternative to traditional blind flooding as it minimizes redundant transmissions and energy consumption. It shares some similarities with the Susceptible-Infected-Recovered (SIR) epidemic model, in the sense that the dissemination process and the epidemic thresholds, which achieve maximum coverage with the minimum required transmissions, have been found to be common in certain cases. In this paper, some of these similarities between probabilistic flooding and the SIR epidemic model are identified, particularly with respect to the epidemic thresholds. Both of these epidemic algorithms are experimentally evaluated on a university campus testbed, where a low-cost WSN, consisting of 25 nodes, is deployed. Both algorithm implementations are shown to be efficient at covering a large portion of the network's nodes, with probabilistic flooding behaving largely in accordance with the considered epidemic thresholds. On the other hand, the implementation of the SIR epidemic model behaves quite unexpectedly, as the epidemic thresholds underestimate sufficient network coverage, a fact that can be attributed to implementation limitations.

Keywords: probabilistic flooding; coverage; epidemic information dissemination; epidemic threshold; wireless sensor networks; Internet of Things

1. Introduction

Modern technological environments are overloaded with large volumes of data, requiring simultaneously high speeds for the transfer and exchange of information, while maximizing energy savings [1–3]. In this context comes the spreading of the Internet of Things (IoT) [4], which is continuously invading everyday life and harmonizing with human needs and requirements. Due to the technological advances in micro-electro-mechanical systems and wireless communications, including the IoT, the development of Wireless Sensor Networks (WSNs), which monitor certain environmental parameters and intelligently organize the collected data in central locations, has become a reality [5–7]. This class of networks usually involves a large number of such small low-powered wireless devices that require simple information dissemination algorithms in order to achieve global outreach. With regard to their spontaneous formation and wireless nature, WSNs bare many similarities to the wireless ad hoc networks.

Information dissemination aims to spread information to multiple destinations, typically through the use of broadcast techniques. It is an integral part of basic network operations, including routing, network discovery, and finding the optimum-path, and it has been studied extensively to provide

the ultimate efficiency and effectiveness of systems in such environments [8]. The goal of broadcast techniques usually involves global outreach, whereby all of the network's nodes are covered by propagated information. However, it is often sufficient in real-world applications for the majority of the nodes, i.e., a large percentage of them, to receive the information.

Alternative dissemination methods to traditional blind flooding [9] have been investigated to address the unnecessarily large number of transmitted messages and the storm problem [10], one of which is probabilistic flooding. In probabilistic flooding, message transmissions take place based on a predefined forwarding probability. This forwarding probability is set by the initiator node and gets carried along with the message throughout the whole network. In this way, the overall dissemination relies on a much smaller number of messages; ergo, making it less costly. Therefore, there is a clear compromise between reducing the total number of messages and the network's coverage with probabilistic flooding; by setting the forwarding probability considerably small, a large mount of the network's overhead is avoided; however, many of the network's nodes never receive the message.

Subsequently, probabilistic flooding is among the simplest examples of an epidemic algorithm. The term "epidemic" refers to algorithms that allow fast dissemination of information from an initial node (source) through purely local interactions between nodes, where the neighboring ones rebroadcast the dissemination message. Significant research work is focused on epidemic-based approaches for information dissemination in large-scale systems [11], such as peer-to-peer systems deployed on Internet or ad hoc networks, due to the fact that they are easy to implement, robust, and error resilient. In addition, this allows for convenient adjustment of the epidemic algorithm's parameters to achieve high reliability in dynamic network topologies [12].

A common approach for modeling the spread of an infection in a system is achieved through the use of compartmental models, first introduced by Kermack and McKendrick [13], in order to study the individuals of an infected population divided into groups, according to their disease-related state. A simple such model is the SIR epidemic model, which groups the individuals into three compartments: Susceptible (S), Infected (I), and Recovered (R). All nodes are initially in the susceptible state, with the exception of the initiator node, which is infected. Infected nodes infect neighbors in the susceptible state through interactions (transmissions, in this case). Infected nodes eventually recover, becoming immune, so they can no longer infect their neighbors. The transition from one state to another takes place based on the corresponding infection rate and the infection's recovery rate.

It is noteworthy that probabilistic flooding can be considered as a special case of the SIR model, under certain conditions. In both cases, information dissemination initiates from a specific node and propagates in a probabilistic fashion through the network. Similarly, some time after receiving the disseminated information, a node may stop propagating information. Furthermore, certain parameters, which govern how capable both models are at providing sufficient information dissemination, are similar in nature, such as the initiator node [14].

Most importantly, for the focus of this work, the mechanisms that maximize coverage with minimum resource use, i.e., epidemic thresholds, have been found to be identical in certain cases in the literature [15–17]. For both models, these epidemic thresholds generally depend on various properties of a graph, including its average degree and its spectrum [15]. However, the analysis needed to derive these results is heavily probabilistic, while often involving abstract notions that cannot be directly translated to real environments, such as network models or discrete time-steps. Therefore, it is worth investigating whether some of the assumptions made still hold true in practice, which can only be achieved through experimentation under realistic, yet controlled, conditions.

In this light, this paper focuses on identifying similarities between probabilistic flooding and the SIR model, especially with respect to their coverage and epidemic thresholds. These models are implemented and experimentally evaluated to compare their behavior and effectiveness at disseminating information. The experimental evaluation takes place on CaBIUs [18], a university campus testbed, where a WSN is deployed. This relatively small-scale WSN, consisting of 25 low-cost nodes, should be adequate to reliably acquire results in a controlled environment, despite the

models' probabilistic nature. The acquired results indicate that the models perform similarly, with the performance of probabilistic flooding being in accordance with previous analytical models, while the SIR epidemic model relatively differs. The considered thresholds for probabilistic flooding successfully predict the network's coverage, whereas the thresholds for the SIR epidemic model seem to overestimate the network's coverage for a given set of parameters. This fact can mainly be attributed to the selected SIR implementation, whereby a transition to the recovered state may take place, prior to the transmission from an infected node to another. Therefore, despite commonalities between the thresholds for the two models, the resulting coverage differs for the considered implementations.

The remainder of this paper is structured as follows: In Section 2, an overview of related work is provided focusing on the discussion of the algorithms' epidemic thresholds. Section 3 describes the relevant models and epidemic thresholds used in the paper, and in Section 4, the system used for the experimentation in the real wireless network environment is described, along with the implementation of the two information dissemination approaches. Section 5 thoroughly describes the conducted experiments and presents the results, while in Section 6, conclusions are drawn and future work is outlined.

2. Past Related Work

A large number of studies has focused on the efficiency of flooding. The broadcast storm problem pointed out by Ni et al. [10] is known to hold flooding responsible for causing a significant amount of network traffic. They proposed various alternative approaches to eliminate the problem. William et al. conducted a study [19], comparing different broadcast algorithms in ad hoc networks, one of which was probabilistic flooding. They showed that flooding was inadequate for congested networks as the increased number of nodes in static networks leads to performance loss due to the large number of failed message re-transmissions.

Efforts have been made to optimize probabilistic flooding's efficiency. Fixed probabilistic flooding has been mainly studied in random networks, using percolation theory and phase transition phenomenon [20–22]. Crisóstomo et al. [23] derived values for the threshold probability, considering various stochastic topologies. Additionally, they randomly sampled the graph's total node set, forming an induced subgraph, based on the forwarding probability of probabilistic flooding. They proved that the event of this subgraph being a connected dominating set is probabilistically equal to the event of global outreach by probabilistic flooding. In [24], Oikonomou et al. derived asymptotic expressions for probabilistic flooding with underlying networks that are random graphs. Later, in [14], Koufoudakis et al. studied probabilistic flooding using algebraic graph theory elements, and an algorithm was introduced to estimate the threshold probability, whereas in [25], a lower bound of the forwarding probability was proposed, allowing for global network outreach. It was shown that for large values of time, a forwarding probability sufficient to provide global outreach is inversely proportional to the average number of neighbors in the network.

However, results observed for random graphs are not always applicable to ad hoc networks. Consequently, it is a hard task to determine an optimum forwarding probability, as it may depend on various network parameters, such as density, distance among nodes, and speed. An alternative approach to fixed probabilistic flooding is to use adaptive probabilistic schemes [26], where the forwarding probability is determined according to the aforementioned parameters.

Probabilistic flooding shares many of its features with epidemic-based dissemination approaches with respect to the network's coverage [17]. The rise of theoretical epidemiology began in the early 20th Century with the mathematical modeling of infectious diseases and has been mainly a subject of biologists and health physicians such as Kermack-McKendrick [13], Anderson-May [27], and Ross [28]. Their main purpose was to predict the spreading properties of an infectious disease, and they conventionally have relied on the use of compartmental models. As early as of the 1980s, the mathematics underlying the propagation of epidemic approaches has been shown to be related to communication protocols [29]. In the years following, epidemic models and their applications have

been widely investigated in communication systems [12,30]. More recent approaches include epidemic information dissemination for mobile molecular communication systems [31] and WSNs [32], the optimal control study of information epidemics in social networks for campaigning strategies [33?] and even for evaluating the dissemination of information in urban scenarios with different crowd densities and renewal rates [34].

Cohen's pioneering work [35] on the subject of computer viruses suggested that the spread of an infection follows the transitive closure of information flow in a system. He was also the first to point out that there is no algorithm that can perfectly detect all possible viruses. Later, he and Murray [36] indicated a link between the spread of computer viruses and biological epidemiology; however, they did not develop an explicit model linking the two together. The first serious attempt at adapting mathematical epidemiology to the subject of computer virus replication within computing systems was presented by Kephart and White [37]. They showed that the epidemic threshold is related to the average node degree of the network (average number of neighbors in a network) and stated that the absolute prevention of a computer virus is indeed impossible; yet, controlling the rate at which a a detected virus is removed from the system may be sufficient for the prevention of an epidemic.

Epidemic thresholds have been a subject of research for various network topologies. Pastor-Satorras and Vespignani [38–41] focused on the epidemic spread in scale-free networks where they proposed an analytical model for power-law graphs and respectively derived an epidemic threshold. Scale-free networks exhibit a highly skewed connectivity distribution making them virally more susceptible, as computer viruses are more likely to propagate among the nodes of such high connectivity variance networks, thus making the infection harder to extinguish. Eguiluz and Klemm [42] proposed an even tighter epidemic threshold for real graphs and provided confirming results on several Internet graphs. Serrano and Boguñá [43] derived the epidemic threshold for clustered networks.

Later, Boguñá et al. [44] studied epidemic spreading in correlated complex networks, where the connectivity of a particular individual is dependent on its neighbors' connectivity. Moreno et al. [45] also studied the epidemic modeling in complex networks of Watts-Strogatz and Barabasi-Albert and related the epidemic threshold to the moments of the connectivity distribution. They also confirmed the absence of any finite threshold for particular types of connectivity distributions. Wang et al. [16] designed a general epidemic spreading model suitable for any network and proposed an epidemic threshold condition, which applies to any arbitrary graph. This threshold is inversely related to the largest eigenvalue of the network's adjacency matrix. The number of infected nodes in a network exponentially decay for values below this epidemic threshold.

In [14,17], Oikonoumou et al. and Koufoudakis et al. studied probabilistic flooding under the light of epidemic-based dissemination, revealing the close relation of the two approaches regarding the epidemic threshold. The network's coverage was transformed into a coverage polynomial of which the largest root was a lower bound for the threshold probability, and it was tighter than the epidemic thresholds proposed by Kephart-White and Wang et al. [16,37].

3. Models' Definition

The exploration of the models takes place in a network topology, which is modeled by an undirected graph $G(N, L)$, with N and L being the sets of nodes and links (or edges), respectively. Furthermore, A is an $|N| \times |N|$ adjacency matrix of $G(N, L)$, with element $A_{i,j}$ being one, if nodes i and j are adjacent, and zero otherwise. The largest eigenvalue of matrix A is denoted by λ_1. It is also assumed that the implemented algorithms' termination takes place for $t = T$, where t is a given point in time and T is referred to as termination time.

3.1. Probabilistic Flooding

Information dissemination in probabilistic flooding begins to spread from a particular node, namely the source node of the network. The message re-transmission from a node that receives the

information message for the first time takes place, based on a predefined forwarding probability q. Any messages that are received by the nodes after the first time are ignored. The fraction of the nodes covered by the propagated information under probabilistic flooding is referred to as the network's coverage and is denoted by $C_{PF}(t)$ for the time t.

The network's coverage in probabilistic flooding depends on the value of q. The smaller the forwarding probability, the more difficult it becomes for a message to travel across the network, thus the greater the uncertainty for the network's global outreach. The optimum forwarding probability that allows high network coverage and at the same time minimizes the message transmission is referred to as the threshold probability $\hat{q}$.

3.2. SIR Epidemic Model

In a typical SIR epidemic model, which includes the aspect of recovering from the infected state, the spread of an infection from an infected individual takes place based on a fixed infection rate β, and the process of a node recovering from an infection takes place at an infection recovery rate γ. The dynamics of an infection transmission are defined by the basic reproduction number $\tau = \beta/\gamma$. Thus, τ implies the average number of cases generated by a primary infection case in a susceptible population, throughout its duration [46].

Based on the SIR epidemic model, a susceptible individual can, upon meeting an infectious individual, become infectious, and after some time, the infectious individual recovers. It is assumed that the total number of the population is given by:

$$N = S_t \cup I_t \cup R_t, \ or \ |N| = |S_t| + |I_t| + |R_t| \tag{1}$$

where S_t, I_t, R_t are the sets of susceptible, infected, and recovered nodes for time t, respectively.

In a network of $|N|$ nodes, each node represents an individual subject, and the edges L specify the interactions between the different individuals. The spreading of the infection starts from a source node, and the rate at which an infected node i can infect a susceptible neighbor j is defined as the infection rate β. The recovery process that transfers a node from the infected state to its following state (recovered) is an internal Markovian (memory-less) process, which takes place at a constant rate γ. An individual becomes fully immune to a particular disease after recovering from its infected state and therefore, if reinfected, cannot re-transmit the disease to its neighbors.

An important assumption that has to be made regarding the spreading of the disease to a susceptible individual is that the infection occurs with a fixed probability β, for each link between a susceptible and an infected individual. Similarly, transitioning from the infected to the recovered state occurs based on a constant probability γ per unit time. These assumptions are introduced to simplify the mathematical modeling, especially when it comes to differential equations, while also simplifying the implementation, as discussed in Section 4.2.

The critical value of τ that determines whether an infection will die out or will be able to spread through a population is referred to as the epidemic threshold and denoted by $\hat{\tau}$. This means that for values of $\tau \geq \hat{\tau}$, the infection rate is high enough, such that the recovery rate is insufficient and the disease spreads out through an entire susceptible population, whereas for values of τ below this threshold, either the generated outbreak is insignificant or the disease does not occur at all. The value of this threshold helps to develop measures for an epidemic's prevention and control [47]. The fraction of the nodes covered by the SIR infection message is referred to as the network's infection coverage and is denoted by $C_{SIR}(t)$ for the time t, that is:

$$C_{SIR}(t) = \frac{|I_t| + |R_t|}{|N|}. \tag{2}$$

This formulation is selected due to the fact that information dissemination is the topic at hand in the current work.

3.3. Epidemic Thresholds

Both probabilistic flooding and the SIR epidemic model function in a relatively similar manner. Probabilistic flooding comes up if each node were to operate according to the SIR epidemic model and become recovered after infecting its neighbors exactly once. Much work has focused on identifying the epidemic thresholds for both of these models, some of which share a few similarities. In this work, due to the fact that the focus is on information dissemination, an infection is considered to be an epidemic if it has affected most of the network. The network's average degree $\bar{d}$ and the highest eigenvalue λ_1 of its adjacency matrix are most prominent in epidemic thresholds.

The epidemic thresholds are values for which sufficient coverage is achieved. For probabilistic flooding, an epidemic threshold $\hat{q}$ is such that $C_{PF}(T)$ is expected to be close to one for $q \geq \hat{q}$. Similarly, for the SIR epidemic model, $\bar{\tau}$ is an epidemic threshold if $C_{SIR}(T)$ is, again, expected to be close to one, when $\tau = \beta/\gamma \geq \bar{\tau}$.

When it comes to probabilistic flooding, the relevant thresholds that are being focused on in the current study include $\hat{q}_1 = 1/\bar{d}$ and $\hat{q}_2 = 1/\lambda_1$. It was also argued in [25] that $\hat{q}_3 = 4/\bar{d}$ and $\hat{q}_4 = 4/\lambda_1$ are tighter bounds for sufficient coverage. In general, the thresholds that involve eigenvalues are regarded as more precise, as $\lambda_1 \geq \bar{d}$, but they can be harder to estimate in real scenarios, due to requiring knowledge of the entire topology.

Regarding the SIR epidemic model, identifying tight epidemic thresholds is much more challenging, with multiple ones being proposed in the literature. Interestingly, few of these thresholds are identical to the ones detailed for probabilistic flooding. In particular, $\hat{\tau}_1 = 1/\bar{d}$ and $\hat{\tau}_2 = 1/\lambda_1$ once again come up for some networks, although they are based on somewhat "naive" expectations [15].

Therefore, probabilistic flooding and the SIR epidemic model exhibit various similarities, both in terms of functionality and epidemic thresholds. In order to investigate how similar the capabilities of both mechanisms are at disseminating information in a real networking environment, this paper focuses on their experimental evaluation.
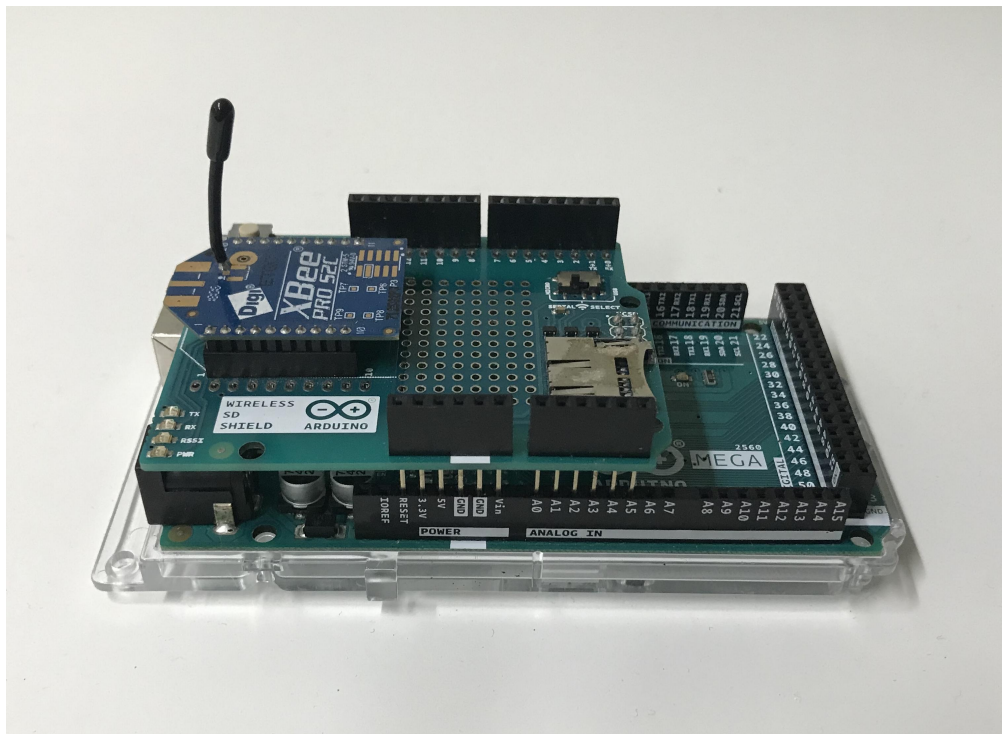
4. Implementation Description

In this section, the implementation of the system that is used for the algorithms' evaluation is discussed with respect to its hardware and software components. In Section 4.1, the hardware constituting the nodes in the utilized WSN is described, and in Section 4.2, the implementation description of the probabilistic flooding and the SIR epidemic algorithms takes place.

4.1. Hardware

The implementation and the experimental approach of the proposed algorithms took place on the CAmpus TestBed of the Ionian University (CaBIUs) [18], where the network nodes comprised open-source Arduino Mega microcontroller boards [48] that were equipped with an XBee S2C ZigBee module [49]. The successful implementation of a fully-fledged node was achieved with the help of a wireless SD shield [50]. The shield allowed the placement of the module onto the Arduino microcontroller board and created an interface between the two components. Such a node is shown in Figure 1. Arduino Mega boards were chosen for this specific application, due to the large volume of the measurements that needed to be collected, as they have high storage capabilities.

Regarding the data transmission medium, the XBee S2C ZigBee module was selected, as its high functionality specifications were considered vital for the efficiency of the network. It can support device types such as ZigBee coordinator, which starts the network and is responsible for its overall management, and ZigBee routers for which a network can contain one or more depending on the size and topology [51]. The XBee devices are also capable of operating in two modes, transparent and Application Programming Interface (API) [52]. For the purposes of this paper, API mode was considered, which transmits data in well defined frames, carrying additional information related

to either the configuration or communication required for transmission between multiple devices, and also supports a programming interface for application implementation.

Figure 1. A picture of a CaBIUs node.

4.2. Software

In the sequel, the software developed for this study is described, consisting of the algorithms' implementation, which was uploaded to each of the network's nodes (microcontrollers).

The realistic characteristics of the setup presented various difficulties that were related to the environmental factors, such as weather conditions, and thus made it difficult to establish signal stability. Additionally, the XBee S2C ZigBee module and the protocols it features do not provide a reliable means of establishing and acquiring links between nodes, as outlined in [18]. Therefore, conducting the experiments required following a certain procedure to establish the network's topology. This needed to take place in such a way that each of the network's links was symmetrical and achieved by transmitting messages from each node to all of its one-hop neighbors and accordingly waiting for replies. The previous was repeated three times after which the nodes exchanged the number of successful transmissions with each other. Based on the sum of the successful transmissions, the decision whether a node was listed as a neighbor or not was made. At the end of this procedure, the coordinator node collected the networks topology by combining the network nodes' replies accordingly into an adjacency matrix. The replies were collected through a simple echo algorithm at the end of a predefined time period, which was considered sufficient for the end of all the algorithm message exchanges.

However, since it was important to be able to correlate the specific time interval at which an event occurred, the dissemination message contained a value that acted as a step (hop) counter, which was initialized by the source node and increased with each transmission. The first step in both the probabilistic and the epidemic algorithm included a consensus clock synchronization [53]. The initiator node, which corresponded to the source node and coordinator node, based on the model's and ZigBee's description, respectively, broadcast a message to the entire network in which it declared its clock's value. The nodes then saved the time difference between their own clock and that of the initiator. In this way, when they replied with their time measurement value, they could adapt it,

accordingly, to the clock of the initiator. This method did not take into account the time difference between the network's nodes, as it was assumed that it did not drastically affect the measurements.

4.2.1. Probabilistic Flooding Implementation

To achieve the probabilistic flooding among the network, a flooding message was sent by the initiator node to its neighbor nodes, in which the forwarding probability for the current experiment round was specified. Upon receiving the flooding message for the first time, the nodes of the network saved their time of coverage and added a unit to the current flooding message's total number of hops. They, then, sent a result message to the initiator node, in which they specified the time and message hops at the moment of their coverage.

The nodes, then, attempted to forward the flooding message to their neighbors, based on the probability of the received flooding message. The attempt was successful for each of the node's different neighbor based on $rand()$ mod $100 < a$, where rand()is a built-in Arduino random number generator and a is a positive integer, which was set to $q * 100$. The expression $rand()$ mod 100 yields a random number between zero and 99, resulting in the probability of it being less than a or equal to q. If a node received the flooding message more than one time during the same run, it disregarded it.

After the reception of any result message, the initiator node saved its content in its internal memory. The flooding process terminated after a predetermined amount of time, that is the maximum time required for a blind flooding to reach all of the network's nodes.

4.2.2. Epidemic Implementation

Similar to probabilistic flooding, the SIR epidemic model implementation initiated when the initiator node sent the infection message to its neighbor nodes. Initially, every node was in the susceptible state, with the exception of the initiator node, who was infected. As stated in 3.2, the recovery rate from an infection was based on a constant recovery probability per unit time. This recovery probability γ was specified in the infection message along with the infection probability β.

The analytical expressions that described the behavior of the dissemination algorithms presented in Section 3 implied the presence of discrete time. In order to approximate the notion of time-steps, each node maintained a local countdown, starting from when an infection message was received. Thus, these "time-steps" were counted differently for each node; they were not globally synchronized. Once the countdown had ended, it reset, and the node first checked whether a recovery occurred. If a recovery did not occur, the node proceeded to infect some of its neighbors; both recovery and infections were determined by the probabilities β and γ, respectively, and were computed identically to the probabilistic flooding mechanism. The countdown length, corresponding to a time-step, was set to be significantly larger than the synchronization inaccuracy (five seconds), in order to minimize the latter's effects on the acquired results.

Upon receiving the first infection message, the susceptible nodes of the network entered the infectious state. They saved their time of coverage and added a unit to the current infection message's total number of hops. They, then, sent a result message to the initiator node, in which they specified the time and message hops during their infection. An additional result message was sent to the initiator node, when an infected node decided to become recovered, which contained the time and time-step of the recovery. After the reception of any infected or recovered message, the initiator node saved its content in its internal memory. The flooding process terminated after a predetermined amount of time, which was the maximum time required for an infection to reach all of the network's nodes.

5. Experimental Evaluation

The contribution of the work presented in this paper was the experimental evaluation of the dynamics of both the probabilistic flooding and the SIR epidemic model. In particular, the experiments aimed to evaluate the information dissemination capabilities of both algorithms under real conditions, with respect to the presented epidemic thresholds, as well as to examine similarities between them.

The experiments were conducted on CaBIUs, a university campus testbed using a 25 node WSN, as described in Section 4. Across all experiments, a total of 10 runs were conducted for any given selection of parameter values regarding q, β, and γ, with meant and 95% confidence intervals being reported, where appropriate. It should be noted that the initiator node was located at the same position across all experiments.

Based on previous experiments on CaBIUs [18], the value of $\bar{d}$ was expected to be approximately equal to 5.76. Hence, the corresponding thresholds $\hat{q}_1$ and $\hat{\tau}_1$ were expected to be equal to $1/5.76 = 0.17$. Due to the nodes' limited computational capabilities and the fact that the network's links changed dynamically, eigenvalues and average degrees were not computed during runtime. A set of predefined probabilities was used instead, with the probability q and reproduction number τ being set at the start of each experiment.

Probabilistic flooding was tested for each of the forwarding probabilities $q \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$. For these five experiments, the topologies that occurred resulted in averages of $\bar{d} = 4.83$ and $\lambda_1 = 6.19$. Thus, the corresponding thresholds were on average $\hat{q}_1 = 1/\bar{d} = 0.21$ and $\hat{q}_2 = 1/\lambda_1 = 0.16$, while $\hat{q}_3 = 4 \times \hat{q}_1 = 0.84$ and $\hat{q}_4 = 4 \times \hat{q}_2 = 0.64$. Concerning the SIR epidemic model, a total of nine experiments were conducted, with the ranges of infection probabilities and recovery probabilities used being $\beta \in \{0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7\}$ for $\gamma = 0.6$ and $\beta \in \{0.7, 0.8\}$ for $\gamma = 0.5$. The range of these probabilities yielded different values for the reproduction number τ. The experiments conducted for the SIR epidemic model yielded $\bar{d} = 5.00$ and $\lambda_1 = 6.30$, with the corresponding thresholds being $\hat{\tau}_1 = 0.20$ and $\hat{\tau}_2 = 0.16$.

The total coverage for different values of q is depicted for the case of probabilistic flooding in Figure 2a. For $q = 0.2$, the total coverage $C_{PF}(T)$ was exceedingly low, with less than 10% of the network being covered by information. It was, therefore, evident that thresholds $\hat{q}_1$ and $\hat{q}_2$ severely underestimated the probability q required to cover most of the network's nodes. On the other hand, for $q = 0.6$, a much larger percentage of the network was covered by the disseminated information, being 85.2%. Thus, the acquired results were more in accordance with thresholds $\hat{q}_3$ and $\hat{q}_4$.
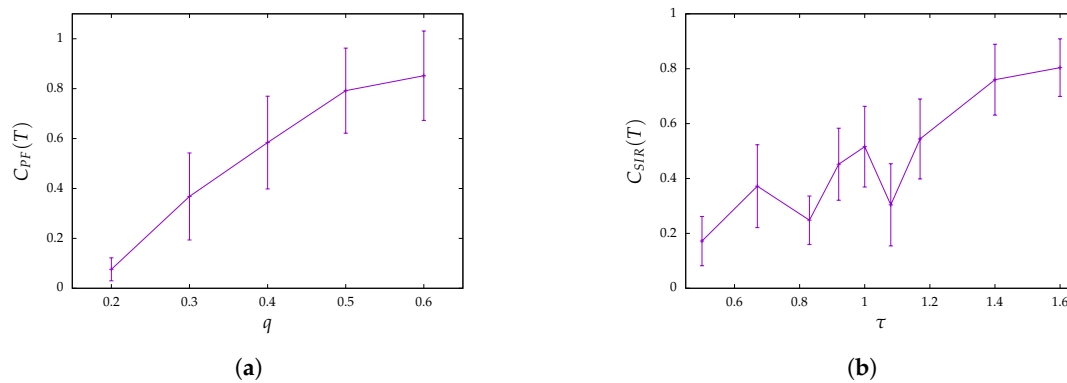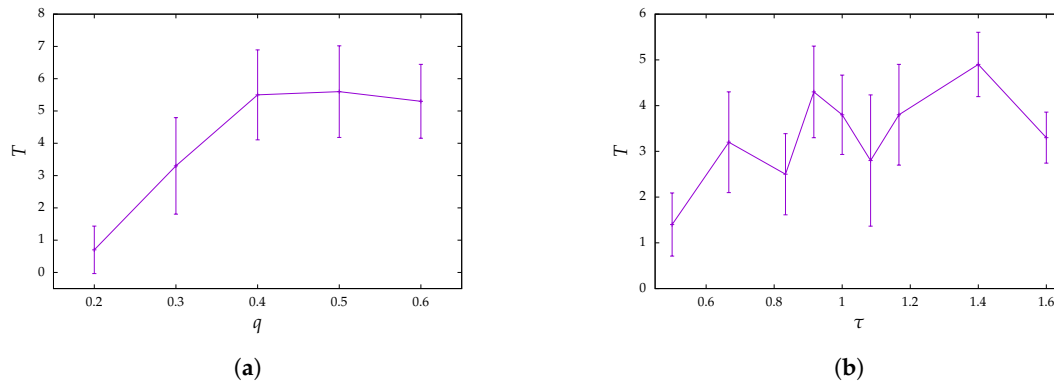

Figure 2. Comparison of total coverage between probabilistic flooding and the SIR epidemic: (**a**) probabilistic flooding coverage $C_{PF}(T)$ by probability q ; (**b**) SIR epidemic model coverage $C_{SIR}(T)$ by reproduction number τ.

Similarly, Figure 2b showcases the total coverage $C_{SIR}(T)$ for the SIR epidemic model. The lowest value of τ for which experiments were conducted, which was 0.5, was considerably higher than the one indicated by thresholds $\hat{\tau}_1$ and $\hat{\tau}_2$. However, the network coverage even for such comparatively high values was only around 20%. In contrast, the highest coverage (80.4%) was achieved for $\tau = 1.6$ ($\beta = 0.8, \gamma = 0.5$). This highlighted a significant difference between the theoretical model and the implementation investigated here and might be attributed to the fact that according to the implementation, a node may recover before ever infecting any of its neighbors.

Regarding termination times in terms of hops, they are displayed in Figure 3 with respect to probabilities q and reproduction numbers τ. In the case of probabilistic flooding, the lowest values

for termination time occurred for $q = 0.2$, as the message being disseminated did not travel very far. For high values of q, probabilistic flooding functioned similarly to blind flooding, so the termination time should be closer to the source node's centrality (the maximum distance between the source node and any other node of the network). For intermediate values of q, messages did not always reach their destination through the most direct path, so the termination time could be higher. A similar, although less stable, behavior was observed for the SIR epidemic model, with termination times being lower in most cases. Due to the overall less stable behavior evident in both the coverage and termination time of the SIR epidemic model, it could be argued that probabilistic flooding was more suitable in applications that did not favor high variance in the information dissemination process.

(a)   **(b)**

Figure 3. Comparison of termination time T in hops between probabilistic flooding and the SIR epidemic model: (**a**) probabilistic flooding termination time T by probability q; (**b**) SIR epidemic model termination time T by reproduction number τ.

The network's coverage with respect to time in hops is depicted in Figure 4. For probabilistic flooding, the behavior of which is shown in Figure 4a, a large portion of the network's nodes were covered within a couple of time steps (e.g., in the range of $t \in [2, 4]$), which is often referred to as the transition phase. This range also seemed to be slightly wider the higher the value of q was. For the SIR epidemic model, on the other hand, the transition phases did not seem to occur at the same time or last as long, for different values of τ, as can be observed in Figure 4b. It is worth noting here that in a few cases, coverage did not monotonically increase with τ. This unexpected behavior could be attributed to a number of factors, such as the size of the network used for the evaluation and the difficulties encountered in networks that were operating under the conditions of a real environment (topology changes across experiments, transmission failures).

(a)   **(b)**
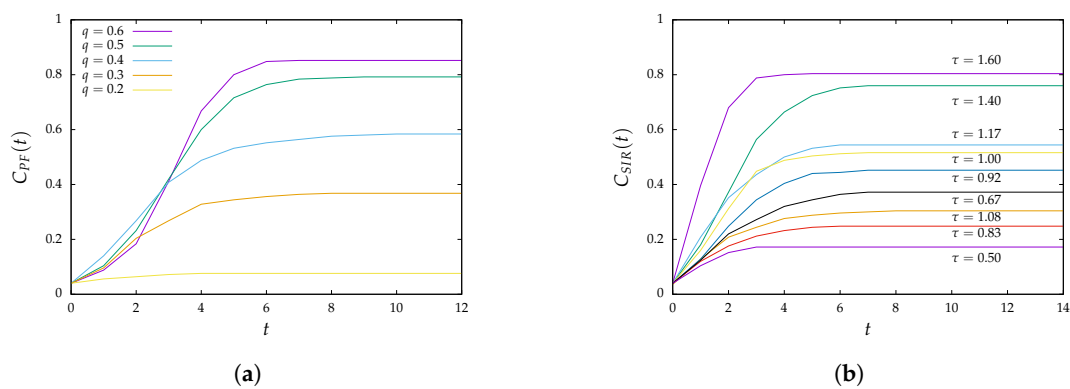
Figure 4. Comparison of coverage with respect to time t in hops between probabilistic flooding and the SIR epidemic model: (**a**) probabilistic flooding coverage $C_{PF}(t)$; (**b**) SIR epidemic model coverage $C_{SIR}(t)$.

6. Conclusions and Future Work

The progress in the development of the IoT has ignited scientific interest for the last decade. The success of the IoT is defined by the effectiveness of WSNs, which IoT is based on, in order to reinforce human comfort. As a result, a primary role is played by the information dissemination across such sensors, a process that was extensively studied and discussed and that could be implemented with a variety of techniques.

In this paper, the dissemination of information was achieved via probabilistic flooding and the SIR epidemic model by implementing the corresponding distributed algorithms. The similarities between probabilistic flooding and the SIR epidemic model were pinpointed, regarding both their behavior and their epidemic thresholds. The main contribution concerned the experimental evaluation of their similarities in CaBIUs, a university campus testbed. This network consisted of low-cost wireless sensor nodes that were installed across the campus of the Ionian University.

The experimental evaluation showed that the implementation was generally in agreement with previous analytical models. Especially in the case of probabilistic flooding, some of the considered epidemic thresholds could accurately predict the sufficient coverage of the network's nodes. For the implementation of the SIR epidemic model, the considered thresholds were shown to overestimate the network's coverage for a given set of parameters. This might be attributed to the fact that, according to the implementation, a node may enter the recovered state before infecting any of its neighbors. Therefore, despite commonalities between the thresholds of probabilistic flooding and the SIR epidemic model, the resulting coverage differed for the implementations described in this paper.

While the acquired results were promising, future work could aid in establishing their validity. More specifically: (i) a different implementation of SIR could be examined, to investigate whether recovery checks should occur only after an infection check has taken place; (ii) more thresholds could be considered in the conduction of additional experiments; (iii) different compartmental models could be compared, such as susceptible-infected-susceptible (SIS); and (iv) a larger number of nodes may be considered for experimentation, to capture the probabilistic nature of the two algorithms more reliably.

Author Contributions: All authors contributed to this project equally. All authors read and agreed to the published version of the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

WSN	Wireless Sensor Network
SIR	Susceptible-Infected-Recovered
IoT	Internet of Things
CaBIUs	CAmpus TestBed of the Ionian University
API	Application Programming Interface
SIS	Susceptible-Infected-Recovered

References

1. Oikonomou, K.; Aissa, S. Dynamic sink assignment for efficient energy consumption in wireless sensor networks. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Paris, France, 1–4 April **2012**; pp. 1876–1881.

2. Angelopoulos, C.M.; Nikoletseas, S.; Raptis, T.P. Wireless energy transfer in sensor networks with adaptive, limited knowledge protocols. *Comput. Netw.* **2014**, *70*, 113–141.

3. Han, G.; Qian, A.; Jiang, J.; Sun, N.; Liu, L. A grid-based joint routing and charging algorithm for industrial wireless rechargeable sensor networks. *Comput. Netw.* **2016**, *101*, 19–28.

4. Xia, F.; Yang, L.T.; Wang, L.; Vinel, A. Internet of things. *Int. J. Commun. Syst.* **2012**, *25*, 1101.

5. Akyildiz, I.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. Wireless sensor networks: A survey. *Comput. Netw.* **2002**, *38*, 393–422, doi:10.1016/S1389-1286(01)00302-4.

6. Yin, C.; Xiong, Z.; Chen, H.; Wang, J.; Cooper, D.; David, B. A literature survey on smart cities. *Sci. China Inf. Sci.* **2015**, *58*, 1–18.

7. Arasteh, H.; Hosseinnezhad, V.; Loia, V.; Tommasetti, A.; Troisi, O.; Shafie-khah, M.; Siano, P. Iot-based smart cities: A survey. In Proceedings of the 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), Florence, Italy, 7–10 June **2016**; pp. 1–6.

8. Betoule, C.; Bonald, T.; Clavier, R.; Rossi, D.; Rossini, G.; Thouenon, G. Adaptive Probabilistic Flooding for Multipath Routing. In Proceedings of the 2012 5th International Conference on New Technologies, Mobility and Security (NTMS), Istanbul, Turkey, 7–10 May **2012**; pp. 1–6.

9. Segall, A. Distributed network protocols. *IEEE Trans. Inf. Theory* **1983**, *29*, 23–35.

10. Ni, S.Y.; Tseng, Y.C.; Chen, Y.S.; Sheu, J.P. The broadcast storm problem in a mobile ad hoc network. In Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, Seattle, WA, USA, August 15–19, **1999**; pp. 151–162.

11. Birman, K.P.; Hayden, M.; Ozkasap, O.; Xiao, Z.; Budiu, M.; Minsky, Y. Bimodal multicast. *ACM Trans. Comput. Syst. (TOCS)* **1999**, *17*, 41–88.

12. Eugster, P.T.; Guerraoui, R.; Kermarrec, A.M.; Massoulié, L. Epidemic information dissemination in distributed systems. *Computer* **2004**, *37*, 60–67.

13. Kermack, W.O.; McKendrick, A.G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London. Ser. A Contain. Pap. A Math. Phys. Character* **1927**, *115*, 700–721.

14. Koufoudakis, G.; Oikonomou, K.; Giannakis, K.; Aïssa, S. Probabilistic flooding coverage analysis for efficient information dissemination in wireless networks. *Comput. Netw.* **2018**, *140*, 51–61.

15. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925.

16. Wang, Y.; Chakrabarti, D.; Wang, C.; Faloutsos, C. Epidemic spreading in real networks: An eigenvalue viewpoint. In Proceedings of the 22nd International Symposium on Reliable Distributed Systems, Florence, Italy, 6–8 October **2003**; pp. 25–34.

17. Oikonomou, K.; Koufoudakis, G.; Aissa, S. Probabilistic flooding coverage analysis in large scale wireless networks. In Proceedings of the 2012 19th International Conference on Telecommunications (ICT), Jounieh, Lebanon, 23–25 April **2012**; pp. 1–6.

18. Alvanou, A.G.; Zervopoulos, A.; Papamichail, A.; Bezas, K.; Vergis, S.; Stylidou, A.; Tsipis, A.; Komianos, V.; Tsoumanis, G.; Koufoudakis, G.; et al. CaBIUs: Description of the Enhanced Wireless Campus Testbed of the Ionian University. *Electronics* **2020**, *9*, 454.

19. Williams, B.; Camp, T. Comparison of broadcasting techniques for mobile ad hoc networks. In Proceedings of the 3rd ACM International Symposium on Mobile ad hoc Networking & Computing, Lausanne, Switzerland, 9–11 June **2002**; pp. 194–205.

20. Sasson, Y.; Cavin, D.; Schiper, A. Probabilistic broadcast for flooding in wireless mobile ad hoc networks. In Proceedings of the 2003 IEEE Wireless Communications and Networking (WCNC), New Orleans, LA, USA, 16–20 March **2003**; Volume 2, pp. 1124–1130.

21. Haas, Z.J.; Halpern, J.Y.; Li, L. Gossip-based ad hoc routing. *IEEE/ACM Trans. Netw.* **2006**, *14*, 479–491.

22. Shen, C.C.; Huang, Z.; Jaikaeo, C. Directional broadcast for mobile ad hoc networks with percolation theory. *IEEE Trans. Mob. Comput.* **2006**, *5*, 317–332.

23. Crisóstomo, S.; Schilcher, U.; Bettstetter, C.; Barros, J. Probabilistic flooding in stochastic networks: Analysis of global information outreach. *Comput. Netw.* **2012**, *56*, 142–156, doi:10.1016/j.comnet.2011.08.014.

24. Oikonomou, K.; Stavrakakis, I. Performance analysis of probabilistic flooding using random graphs. In Proceedings of the 2007 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Espoo, Finland, 18–21 June **2007**; pp. 1–6.

25. Koufoudakis, G.; Oikonomou, K.; Aïssa, S.; Stavrakakis, I. Analysis of Spectral Properties for Efficient Coverage Under Probabilistic Flooding. In Proceedings of the 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Chania, Greece, 12–15 June **2018**; pp. 1–9.

26. Reina, D.; Toral, S.; Johnson, P.; Barrero, F. A survey on probabilistic broadcast schemes for wireless ad hoc networks. *Ad Hoc Netw.* **2015**, *25*, 263–292.

27. Anderson, R.M.; May, R.M. Understanding the AIDS pandemic. *Sci. Am.* **1992**, *266*, 58–67.

28. Ross, R. *Memoirs: With a Full Account of the Great Malaria Problem and Its Solution*; J. Murray: Stevens Point, WI, USA, **1923**.

29. Demers, A.; Greene, D.; Hauser, C.; Irish, W.; Larson, J.; Shenker, S.; Sturgis, H.; Swinehart, D.; Terry, D. Epidemic algorithms for replicated database maintenance. In Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing, Vancouver, British Columbia, Canada, August 10–12, **1987**; pp. 1–12.

30. Lloyd, A.L.; May, R.M. How viruses spread among computers and people. *Science* **2001**, *292*, 1316–1317.

31. Ishiyama, S.; Nakano, T.; Okaie, Y.; Hara, T. Epidemic Information Dissemination in Mobile Molecular Communication Systems. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December **2018**; pp. 1–7.

32. Anagnostopoulos, C.; Sekkas, O.; Hadjiefthymiades, S. An adaptive epidemic information dissemination model for wireless sensor networks. *Pervasive Mob. Comput.* **2012**, *8*, 751–763, doi:10.1016/j.pmcj.2011.06.005.

33. Kandhway, K.; Kuri, J. Optimal Resource Allocation Over Time and Degree Classes for Maximizing Information Dissemination in Social Networks. *IEEE/ACM Trans. Netw.* **2016**, *24*, 3204–3217.

34. Chancay-García, L.; Hernández-Orallo, E.; Manzoni, P.; Calafate, C.T.; Cano, J. Evaluating and Enhancing Information Dissemination in Urban Areas of Interest Using Opportunistic Networks. *IEEE Access* **2018**, *6*, 32514–32531.

35. Cohen, F. Computer viruses: Theory and experiments. *Comput. Secur.* **1987**, *6*, 22–35.

36. Murray, W.H. The application of epidemiology to computer viruses. *Comput. Secur.* **1988**, *7*, 139–145.

37. Kephart, J.O.; White, S.R. Directed-graph epidemiological models of computer viruses. In *Computation: The Micro and the Macro View*; World Scientific: Singapore **1992**; pp. 71–102.

38. Pastor-Satorras, R.; Vespignani, A. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* **2001**, *63*, 066117.

39. Pastor-Satorras, R.; Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **2001**, *86*, 3200.

40. Pastor-Satorras, R.; Vázquez, A.; Vespignani, A. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **2001**, *87*, 258701.

41. Pastor-Satorras, R.; Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **2002**, *65*, 036104.

42. Eguiluz, V.M.; Klemm, K. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.* **2002**, *89*, 108701.

43. Serrano, M.Á.; Boguná, M. Percolation and epidemic thresholds in clustered networks. *Phys. Rev. Lett.* **2006**, *97*, 088701.

44. Boguna, M.; Pastor-Satorras, R.; Vespignani, A. Epidemic Spreading in Complex Networks with Degree Correlations. *Lect. Notes Phys.* **2003**, *625*, doi:10.1007/978-3-540-44943-0_8.

45. Moreno, Y.; Pastor-Satorras, R.; Vespignani, A. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B-Condens. Matter Complex Syst.* **2002**, *26*, 521–529.

46. Diekmann, O.; Heesterbeek, J.A.P. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*; John Wiley & Sons: Hoboken, NJ, USA, **2000**; Volume 5.

47. Anderson, R.M.; Anderson, B.; May, R.M. *Infectious Diseases of Humans: Dynamics and Control*; Oxford University Press: Oxford, UK, **1992**.

48. Arduino Mega. Available online: https://store.arduino.cc/arduino-mega-2560-rev3 (accessed on 27 June 2020).

49. Xbee Pro S2C—RF Module. Available online: https://www.digi.com/products/embedded-systems/rf-modules/2-4-ghz-modules/xbee-zigbee (accessed on 4 November 2019).

50. Arduino Xbee Shield. Available online: https://www.arduino.cc/en/Main/ArduinoXbeeShield (accessed on 4 November 2019).

51. Baronti, P.; Pillai, P.; Chook, V.W.; Chessa, S.; Gotta, A.; Hu, Y.F. Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and ZigBee standards. *Comput. Commun.* **2007**, *30*, 1655–1695.

52. XBee®/XBee-PRO S2C Zigbee®. Available online: https://www.digi.com/resources/documentation/dig idocs/pdfs/90002002.pdf (accessed on 4 November 2019).

53. Zervopoulos, A.; Tsipis, A.; Alvanou, A.G.; Bezas, K.; Papamichail, A.; Vergis, S.; Stylidou, A.; Tsoumanis, G.; Komianos, V.; Koufoudakis, G.; others. Wireless Sensor Network Synchronization for Precision Agriculture Applications. *Agriculture* **2020**, *10*, 89.