ARTICLE IN PRESS

Computer Communications xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Data proliferation based estimations over distribution factor in heterogeneous wireless sensor networks

Vinod Kumar Verma[a,*], Klimis Ntalianis[b]

[a] Department of Computer Science & Engineering, Sant Longowal Institute of Engineering & Technology, Deemed University, Longowal, 148106, Sangrur, Punjab, India
[b] Athens University of Applied Sciences, Department of Marketing – Online Computing Group, 28, Agiou Spyridonos str., Egaleo, 12210, Athens, Greece

ARTICLE INFO

Keywords:
WSN
DOF
Scalability
Classical flooding
Events
Data proliferation

ABSTRACT

In this paper, a new representation for the chi-squared distribution has been derived over wireless sensor networks (WSN). The underlying correlated data proliferation protocols have strong influence on performance of the deployed system. A proposed model has been deployed to investigate the WSN system over the data proliferation aspect. Initially, the degree of freedom (DOF) factor has been evaluated with respect to scalability issue in the proposal. Further, the factors affecting the outcome of the WSN system assessing data proliferation have been investigated. Moreover, sensor node operations namely: sense count, transmit count and receive redundant count have also been evaluated. Finally, extensive simulation analysis has been carried out to prove the validity of the proposed innovative scheme. However, it has also been investigated that chi-distribution for wireless sensor networks seems intractable with the degree of freedom, when varied with a specific number of nodes.

1. Introduction

In the recent years, a lot of research and development has been reported in the field of wireless sensor networks due to their ubiquitous and flexible nature [1], including real time quality of service (QoS) guarantees [2]. Resource constraints, such as limited computation capability and low battery power, may reduce the extent of guarantee regarding QoS within a wireless sensor network. It is of great importance to exactly measure the position and coverage of a wireless sensor networks. Especially, in the field of military, health, environment and commercial applications such as fire detection in forests, patients monitoring in hospitals, military vehicle tracking, rescuing avalanche victims, cattle herding, inventory control management and car theft detection [3–4]. The global positioning system (GPS) still remains the most suitable for outdoor applications, but it confronts several limitations as in case of harsh propagation environments like within buildings, basements and underground facilities due to the absence of a sight-line. Furthermore, it is extremely difficult to deploy the GPS technology in wireless sensor networks, due to prohibitive power consumptions. Approximately, all the existing methods for localization are based on the Gaussian assumption that either the range estimates or the noise/error are normally distributed, resulting in range distortion [5–10]. Gaussian distribution remains widely used for two reasons: (a) it incorporates a simple model and (b) it enables for a good distribution and clustering around the mean, for several different data sets. Now according to the central limit theorem (CLT), the sum of a random number of variables with finite means and variance can be approximated by a Gaussian distribution [9]. Indeed, a cumulative distribution function of the delay can be used as a probabilistic measure for reliability and timeliness. From the past works in references [11–13], wireless sensor networks have been analyzed from a latency point of view (in terms of means and variance). In the reference [14], concept of network calculus has also been extended for delay measurements in order to derive probabilistic bounds. In the worst case analysis, it remains quite difficult to observe the stochastic behavior. This is because of the randomness and low power nature of communication links in wireless sensor networks. However, authors in reference [15–16] investigated unreliable networks using real time queuing theory. Additionally, probability analysis of the delay in broadcast networks has been reported in [17–21], using different medium access protocols. These results are necessary to fully characterize the sensor node distribution strategies in wireless sensor networks. Previous contributions as reported in references [22–25], have been used to analyze the constraints of specific data sets. The work reported in references [22–25] is based on several different tests. A network discovery approach for the industrial wireless sensor networks using mobile nodes was reported by Montero et al. [26]. The contribution towards an adaptive multi-clustering algorithm with fuzzy logic has been made by Mirzaie et al. [27] in the wireless sensor network domain. Authors in reference [28] highlighted the coverage aspect based survey in wireless sensor

* Corresponding author.
E-mail address: vinod5881@gmal.com (V.K. Verma).

http://dx.doi.org/10.1016/j.comcom.2017.09.017
Received 1 February 2017; Received in revised form 6 September 2017; Accepted 29 September 2017
0140-3664/

network field. The work over the Sybil attack for energy trust system in wireless sensor network domain was proposed by Alsaedi et al. [29]. An innovative strategy for the content delivery in the wireless information-centric network was suggested in reference [30]. A framework for the multi-hop wireless network towards a generalized system was addressed by Saligheh et al. [31]. A more rigorous effort for the overlapping area coverage in the direction of clustered sensor networks has been reported in reference [32]. A newer prediction scheme focusing on the reduction of data transmission was proposed by Dias et al. [33] in wireless sensor networks. Authors in reference [34] presented an itinerary based strategy for the network coding based wireless sensor networks. However, to the best of the authors' knowledge, no work has yet been reported to investigate how the degree of freedom affects the performance of a wireless sensor network system with chi-squared distribution. A newer aspect namely: data proliferation; has been identified and investigated in the WSN domain. Data proliferation may be referred as an aspect of determining the association of sensor node operations with underlying protocol in WSN system. The aim and main novelty of this paper is to determine the influence of the degree of freedom over the scalability of nodes in wireless sensor networks. Additionally, the focus remains on the investigations of data proliferation aspect for the sensor node operation in the said domain.

The rest of this paper is organized as follows: Section 2 reports survey of chi-squared sensor node distributions and classical flooding routing protocols. Section 3 describes the proposed scheme analytically. Section 4, provides the problem definition along with system model, while Section 5 presents the design principles of the proposed experimentation phase. Simulation results and validations are presented and discussed in Section 6. Finally, Section 7 concludes this paper.

2. Chi-squared sensor distributions–related work

This section provides the background and related work regarding chi-squared node distribution strategies and classical flooding data proliferation protocols. The proposed assumptions have been taken into consideration in the following sections.

2.1. Chi-squared distribution

Central and non-central distribution plays a pivotal role in the performance determination of any communication system [35–39]. In the probability context, chi-distribution (also termed as central distribution) represents the sum of squares of k standard random variables with k degrees of freedom [40–45]. Let us assume that $Z_1 \ldots Z_k$ are the independent standard normal random variables. Then the sum of their square chi-distribution with k degrees of freedom can be denoted by Eq. (1).

$$Q = \sum_{i=0}^{k} Z_i^2, \;\; Q \sim \chi^2(k) \; or \; Q \sim \chi_k^2 \tag{1}$$

where k denotes the degrees of freedom with a positive number *i.e.* Z_i. Chi distribution complies with the following probability distribution function of Eq. (2).

$$f(x;k) = \begin{cases} \frac{x^{(k/2)-1} e^{-x/2}}{2^{k/2}\,\Gamma(k/2)}, & x \geq 0; \\ 0, & otherwise. \end{cases} \tag{2}$$

The following derivation represents the chi-distribution's proof for k degrees of freedom with the probability distribution function as shown in Eq. (3):

$$P(Q)dQ = \int_{\mathscr{V}}^{0} \prod_{i=1}^{k} (N(x_i)dx_i) = \int_{\mathscr{V}}^{0} \frac{e^{\frac{x_1^2+x_2^2+\cdots\ldots+x_k^2}{2}}}{(2\pi)^{\frac{k}{2}}} dx_1 dx_2 \ldots dx_k \Bigg) \tag{3}$$

where x_i denotes a single point in the k-dimensional space, $N(x)$ is the standard normal distribution and $\mathscr{V}$ represents the elemental shell volume at $Q(x)$, which is proportional to an one-dimensional space as depicted by Eq. (4).

$$Q = \sum_{i=0}^{k} x_i^2 \tag{4}$$

Alternatively, for an n-sphere where $n = k-1$ with radius $R = \sqrt{Q}$:

$$P(Q)d(Q) = \frac{e^{-Q/2}}{(2\pi)^{\frac{k}{2}}} \int_{\mathscr{V}}^{0} dx_1 dx_2 \ldots dx_k \tag{5}$$

The Eq. (6) becomes the surface area of A of the $(k-1)$-sphere

$$dR = \frac{dQ}{2Q^{\frac{1}{2}}}, \;\; area\ of\ k-1\ sphere\ A = \frac{kR^{k-1}\pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}+1\right)} \tag{6}$$

Now, by making the substitution $(z+1) = z(z)$ and by cancelling the terms, we the Eq. (7) has been obtained.

$$P(Q)dQ = \frac{e^{-Q/2}}{(2\pi)^{k/2}} AdR = \frac{1}{2^{\frac{k}{2}}\Gamma(k/2)} Q^{\frac{k}{2}-1} \;\; Q^{\frac{k}{2}-1} e^{-Q/2} \;\; dQ \tag{7}$$

Hence, the equations for the chi-distribution have been proved. Furthermore in the literature, detailed evaluation of the generalized central chi-squared distribution was presented in [46–49]. A relationship between non-central and generalized Hermite quadratic form distribution was proposed in reference [50]. Even, the diagonal elements of the Wishart matrix follow the chi-squared distribution. The works in references [51–53] has been analysed for the joint probability distribution function with diagonal elements of the Wishart matrix. Blumenson et al. [54] reported another form of multivariate central distribution *i.e.* the generalized Rayleigh probability distribution function. An effort towards the derivation of a closed form solution for the Rayleigh probability distribution function was suggested in references [54–55]. Furthermore, an effort towards improvement of the Nakagami-m distribution using Miller's method [54] along with a correlation matrix was presented in reference [56].

2.2. Classical flooding protocol

In this protocol, each node receives a packet broadcast up to the maximum hop count threshold, provided that it is not the packet destination. This methodology does not require complex constraints like topology maintenance and route discovery in the network. Every node receives a packet and re-transmits it after caching the source id and sequence number of the message to all of its neighbors. However, this process may result in propagation of many unnecessary routing messages. Three major loop holes of the flooding protocol namely implosion, overlap and resource blindness, were identified by Heinzelman et al. [57]. With the CSMA as an implementation mechanism, straightforward broadcast of packets have become costly (in terms of time and energy). Ni et al. [58] observed the seriousness of this broadcast storm problem and reported that it results in redundancy, contention and collisions. In the references [58–59], proposals had been made to reduce the broadcast problem. A lot of energy may be wasted in contention and collision in case, the system is CSMA-based. The data proliferation TDMA-based MAC protocol (named INFUSE) was proposed in reference [60]. Even though, it reduced both the energy and time for flooding, but it also considered implicit acknowledgments for lossy channels. As a result, extra energy is consumed. Recently, time synchronization through flooding has attracted much attention. Elson et al. [61] eliminated the transmitter side non-determinism in the said domain. A hierarchical level structure was proposed in reference [62], where the root node initiates the control phase and the control messages relay from higher level to lower level nodes which further results in synchronization error reduction. Both of these schemes have been reported in references [61–62] for the exchange of large messages;