# A User Configurable Metric for Clustering in Wireless Sensor Networks

Lina Xu, David Lillis, G.M.P OHare and Rem Collier
*Computer Science and Information,University College Dublin, Ireland*
*lina.xu@ucdconnect.ie, {david.lillis, gregory.ohare, rem.collier}@ucd.ie*

Keywords: User configurable, adaptation, clustering, Voronoi

Abstract: Wireless Sensor Networks (WSNs) are comprised of thousands of nodes that are embedded with limited energy resources. Clustering is a well-known technique that can be used to extend the lifetime of such a network. However, user adaption is one criterion that is not taken into account by current clustering algorithms. Here, the term "user" refers to application developer who will adjust their preferences based on the application specific requirements of the service they provide to application users. In this paper, we introduce a novel metric named Communication Distance (ComD), which can be used in clustering algorithms to measure the relative distance between sensors in WSNs. It is tailored by user configuration and its value is computed from real time data. These features allow clustering algorithms based on ComD to adapt to user preferences and dynamic environments. Through experimental and theoretical studies, we seek to deduce a series of formulas to calculate ComD from Time of Flight (ToF), Radio Signal Strength Indicator (RSSI), node density and hop count according to some user profile.

## 1 INTRODUCTION

Clustering is a major approach to energy efficiency in Wireless Sensor Networks (WSNs) (Heo and Varshney, 2005). However, most of the existing work has focuses on energy saving, ignoring diverse requirements from application developers (referred to as "users" in this paper). They expect that the WSN performs in a certain way that can benefit their applications the most. For example, for some applications, the user may want an immediate response in a clustered network rather than simply reducing power consumption. In another case, a WSN application developer may demand transmission quality more than any other criterion. Scenarios such as these involve a tradeoff between competing demands. Power saving is no longer an overriding factor in the system. An advanced WSN should be sufficiently intelligent to understand users' preferences and thereafter adapt as those preferences change. Currently there is little work showing any interest in adapting user preference in a clustering algorithm. (Liu, 2012) concludes that Quality of Service (QoS) is neglected in current research and it needs to be addressed in the future. Besides energy efficiency, transmission quality and network latency are two additional basic metrics used to evaluate the QoS of a system. Voronoi diagrams (Aurenhammer, 1991) form a fundamental data structure for many clustering algorithms. In Voronoi clustering, Euclidean distance needs to be calculated from the coordinates of sensors or from some other mechanism for representing distance.

In this paper, we motivate a novel metric—communication distance (ComD) to substitute the concept of Euclidean distance. A brief implementation and analysis are provided. We also discuss the usability and alternative for ComD in real systems. Since ComD is user configurable, clustering algorithms adopting this metric are also user-configurable. In a network, if cluster heads (CHs) are already preselected randomly or through some CH election algorithm, each sensor can calculate its own ComD to every CH by using a formula that is selected based on user configuration. Thus a sensor joining the closest CH allows CHs to form locally optimal clusters. Using this method, ComD can evaluate and score every single link between a sensor and a CH in the network based on user preference. In the ideal model, every sensor would join the CH that would render network performance closer to the user's preference. ComD can overcome two particular limitations associated with typical Voronoi-based clustering algorithms:

1. Euclidean distance cannot be used to reliably determine the communication quality between sensors. Even though two sensors are close to each other geographically, their radio connection may

prove inadequate due to obstacles.

2. Voronoi-based approaches based on Euclidean distance do not accommodate user preferences. ComD reflects user preferences. Clustering algorithms using ComD therefore will gain adaptive and configurable features.

# 2 RELATED WORK

A wide body of research on clustering algorithms for WSNs is available. In particular, three notable survey papers have been published (Abbasi and Younis, 2007; Boyinbode et al., 2010; Liu, 2012). Clustering algorithms tend to focus on specific objectives, for example, energy efficiency, load balancing or increased scalability. However, the implications of clustering strategies for adapting user preference have received little attention. Individual clustering algorithms use different metrics to decompose the network into connected clusters. Examples of such metrics include distance, hop count and cluster size (Boyinbode et al., 2010). The metrics are pre-fixed in the clustering algorithms and only fulfil certain requirements.

LEACH (Heinzelman et al., 2000) uses Radio Signal Strength Indicator (RSSI) as a metric to cluster sensors. In general, RSSI can be seen as a Link Quality Estimator (LQE) to estimate the transmission reliability. However, rather than using RSSI to guarantee link quality, LEACH uses it from an energy reduction perspective. Although AWARE (Urteaga et al., 2011) concludes that it can achieve a better Packet Receive Rate (PRR), it is not clear why this is the case. Additionally, their clustering algorithm design does not explicitly address the issue of reliable transmission. (Tang and Li, 2006) provide a QoS control scheme on a cluster-based WSN by changing data transmission rates rather than clustering the network from a QoS perspective. (Akkaya and Younis, 2003) introduce a routing protocol that can find a least cost, delay-constrained path based on a clustered WSN. (Saukh et al., 2006) presents a generic metric for tree routing protocols by combing two QoS metrics: 1) end-to-end success rate and 2) resource demand. Experimental results show that it provides considerable energy saving with equivalent end-to-end packets success rate comparing to other metrics.

# 3 MOTIVATION

Clustering algorithms can organise the sensors into clusters to achieve specific objectives. However, little existing work incorporates a user adaptive feature. Since different application developers have their own preferences on the performance of a WSN, supporting user configurable network clustering becomes essential and should be the focus of the future research. Users view the performance of a network from three perspectives: network latency (L), transmission reliability (R) and energy consumption (C). L, R and C are normally treated as independent measurements of a network. However, they are intimately related to each other and optimising for one perspective can impact on the others. For example, if the network reliability is improved, retransmission may be alleviated, so the energy cost and the network latency may correspondingly be reduced.

Our goal is to develop a new metric ComD that accounts for all three perspectives both individually and jointly. It measures the logical distance between sensors and can be used in Voronoi clustering to substitute the concept of physical distance (Euclidean distance). Each user configuration constructs a unique non-Euclidean space. The ComD of the same link in different spaces can have different values. The value of ComD determines how close the performance of a given connection can be to the user's requirements. If a link between two sensors is close to the user requirement, the logical distance is short and vice versa. While the physical sensor network is fixed, the logical structure transforms in different spaces according to user preference. This metric allows a WSN to dynamically re-cluster itself as user requirements change. This will be achieved by designing ComD to be a user-aware measure of the quality of a link between two sensor nodes. In order to achieve this, we must first decompose L, R and C and understand their underlying interrelationships in more detail.

# 4 ANALYSIS AND EXPERIMENT ON ComD

## 4.1 Overview of ComD

Since the value of ComD is entirely determined by user configuration on the QoS criteria (L stands for latency, R stands for reliability and C stands for power consumption), different calculation formulae should be considered for different cases, as shown in Table 1. L can be set to be low to indicate a design for low latency. If there are no constraints on L, it will be set to 0. It is the same situation with R and C. 7 different configurations are available. The configuration (0,0,0) is not considered since it means there is no constraint on the network performance. ComD for

Table 1: User configurations (ComD column reflects the formulae in the following sections).

| Case | L | R | C | ComD |
|---|---|---|---|---|
| $C_1$ | 0 | 0 | Low | (3) |
| $C_2$ | 0 | High | 0 | 1/(2) |
| $C_3$ | 0 | High | Low | TBC |
| $C_4$ | Low | 0 | 0 | (1) |
| $C_5$ | Low | 0 | Low | TBC |
| $C_6$ | Low | High | 0 | TBC |
| $C_7$ | Low | High | Low | TBC |

configuration $C_1$, $C_2$ and $C_4$ are guaranteed fixed formulae that will be presented in the following sections based on experiments. Other cases requiring combinations of criteria are marked "TBC" and will be addressed in future work. Equation (1) (2) and (3) present the calculations for L, R and C. The calculation formula of ComD for configuration $C_2$ is the reciprocal of Equation ( 2). The reason is that the higher transmission quality between two sensors, the shorter ComD should be. For other cases, since there is a tradeoff between different criteria, to simplify our problem, we set equal priority to each of them.
We believe that L, R and C are determined by serval common factors: hop count, radio signal strength (RSSI), time of flight (ToF) and node density. RSSI can indicate link quality between sensors. Since network latency is measured at the user end, ToF in this paper specifically refers to application level communication time rather than hardware signal level. ToF reflects the time delay on a link. A communication path may be formed by several links, therefore hop count is also considered. To reveal the relationship between the three metrics, we need to discover how these factors influence each of the QoS metrics.

## 4.2 Network assumptions

Several assumptions are made:

1. The Base Station (BS) is placed far from the sensing field. It is assumed to exhibit 100% reliability.
2. Each sensor has a fixed location and only belongs to a single cluster at any point. Sensors can connect to their CHs through single or multi-hop.
3. All the sensors, including the CHs, adopt a first-come-first-served (FCFS) processing pattern.
4. When the WSN starts, every sensor broadcasts its own information to its neighbours and records the RSSI, ToF from its neighbours.

Since limited work is focusing on application level ToF, it needs careful investigation. In contrast, a lot of research has examined the relationship between RSSI, transmission power and link quality, so our analysis for this area is partially founded on the existing work (Baccour et al., 2012).

## 4.3 Network latency

Network latency depends on the communication delay between two sensors, which can be influenced by several factors (e.g. distance or sensor processing ability). Application-level ToF not only reveals the communication time spent over the air, but also the processing time on a sensor. Therefore it is highly related to network latency.
To examine network latency, experiments were conducted using SunSPOT nodes (Smith, 2007) embedded with the CC2420 radio chip. All the nodes are time synchronised to simplify the test. Otherwise, round-trip time would be required. To investigate the relationship between application level ToF and distance in an office area, experiments were performed by varying the transmitter's location between 1 metre and 35 metres. The transmitter created a connection with the receiver at 1 m. it was then moved in 1 m interval, with the average ToF value for 10 packets recorded each time. The results from two repeated experiments are shown in Figure 1. This illustrates that there are no major differences for ToF if $distance < 21m$. When $distance > 21m$, ToF is unpredictable and unstable. This is because the link quality is weak and retransmission frequently occurs. Due to the establishment of the connection, the ToF of the first packet is slightly longer. As we can see, when retransmission does not occur, distance will not affect application level ToF. The major reason is that the time spent over air is too slight to be reflected on application level ToF. We conclude therefore that physical distance is not an important factor in determining network latency (L).

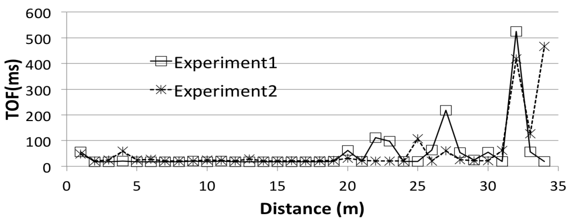


Figure 1: Single hop application level ToF for different distance.

In the above experiments, the measured ToF is based on a single hop. However, a sensor may need more than one hop to communicate with a certain CH. This

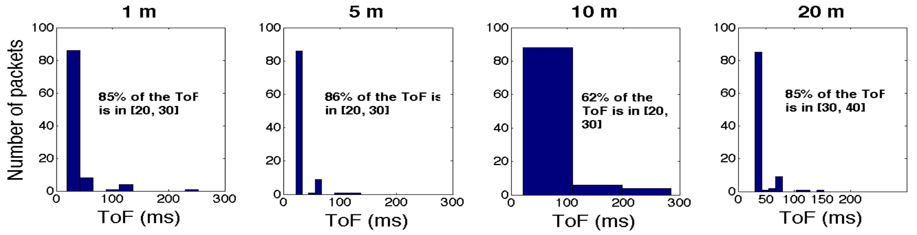


Figure 2: Experiment 1: $S_{1m}$, $S_{5m}$ and $S_{10m}$ need single hop to the receiver, $S_{20m}$ needs 2 hops with the transit from $S_{5m}$.

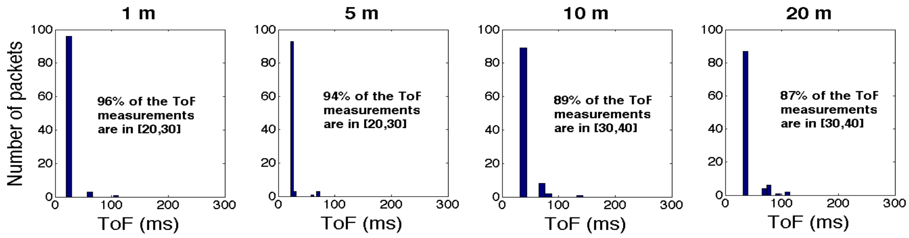


Figure 3: Experiment 2: $S_{10m}$ was changed from single hop to 2 hops with transit from $S_{1m}$. Other conditions keep the same as in experiment 1.

motivates a further experiment in order to discover the relationship between the number of hops and ToF. The relationship helps to estimate the overall ToF of an entire path based on the ToF of each single link and hop counts. Four transmitters ($S_{1m}$, $S_{5m}$, $S_{10m}$ and $S_{20m}$) were placed at fixed locations (1, 5, 10 and 20 meters) and each of them sent 100 packets to the BS, within 1 or 2 hops ($S_{1m}$, $S_{5m}$ and $S_{10m}$ were configured for 1 hop, while $S_{20m}$ required 2 hops). In experiment 1, as shown in Figure 2, $S_{1m}$ and $S_{5m}$'s ToF is mostly in the range $[20,30]$ ms, while $S_{10m}$'s ToF has a somewhat larger range $[20,100]$ ms with a standard deviation (STD) of 126.8ms because its connection quality is not comparable with those at $S_{1m}$ or $S_{5m}$, which necessitates retransmission. However, for single hop communication, the communication time predominantly stays within $[20,30]$ ms. $S_{20m}$'s 2-hop ToF is mostly in $[30,40]$ ms range, longer than a single hop but less than twice of it.

In experiment 2, $S_{10m}$ was changed from a single hop to 2-hop with all other variables being held constant. Following this, its ToF is increased to the range $[30,40]$ ms with the STD of 16.9ms. The reason for lower STD is that the 2-hop connection is more reliable than the single hop and retransmission is alleviated. The ToF of the other sensors stays in a similar range as in experiment 1. If the link quality is poor, retransmission will occur. In such a case, from the user perspective, the required data is delayed. This represents a further reason why we choose application level ToF, since it is related to link quality and can reflect the time cost on retransmission.

The multi-hop experiment confirms the conclusion drawn from the single link experiment namely that distance has limited effect on application level ToF. However, due to that the noise level changed in the same environment, retransmission rate at 10 m is higher than the results that are observed from the single link experiment. The network delay is influenced by multi-hop ToF that is determined by single hop ToF and hop counts. In a real network, sensors can only transmit data when the channel is clear, which means that no other transmitter is using that channel. As a result, the local density of the CH will also affect the network latency (Kim et al., 2012). The network delay between two sensors can be presented as a function of ToF, hop count and node density:

$$L = l\,(ToF, Hops, Density)\,. \qquad (1)$$

## 4.4 Transmission reliability

A packet can reach the receiver successfully only when the over-air connection (the link) between two sensors is reliable and the receiver's buffer is not full (Yousefi et al., 2010). RSSI has been proven to be a good link quality estimator over a reasonable amount of measurements (Baccour et al., 2012). It exhibits a high correlation with PRR and it is more efficient than PRR. In our office environment, the relationship between RSSI and PRR is investigated through experiments and the result is illustrated in Figure 4. In ComD, we can use RSSI as a link quality metric. Adopting the same idea as (Lin et al., 2006), if the RSSI is over a specified threshold, we can assume that the link is transmission reliable. This threshold is related to the environment, hardware and the required quality. From the experiment in Section 4.3 we can see that retransmission can be reflected in the applica-

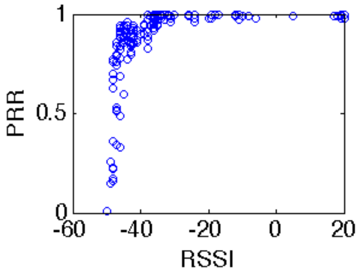


Figure 4: The relationship between RSSI and PRR.

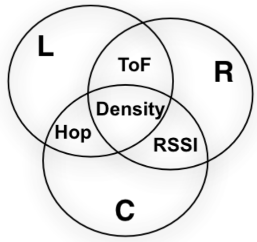


Figure 5: Users view network latency (L), reliability (R) and power consumption (C) as independent objectives, but they are highly coupled.

tion level ToF. Meanwhile retransmission reveals poor link quality. Therefore we believe that ToF can also act as a proxy for link quality. Using the combination of ToF and RSSI can address the link quality better. A packet can be received only when the receiver's buffer is not full. The probability of overflow is determined by data load, which depends on node density (Kleinrock, 1975). However, the model is not practical in real system since the receiving rate from all sources is hard to measure before the network structure is final formed. The idea of predicting receiving rate from local density is used. Hence in order to determine the transmission reliability between two sensors, we use a function that combines ToF, RSSI and node density:

$$R = r\,(RSSI, ToF, Density)\,. \tag{2}$$

## 4.5 Energy cost

A clustering algorithm normally can save power for a network in two distinct ways: One way is through the clustered structure. A CH can assemble the data from each sensor in the cluster and then send the assembled data to the BS. This saves energy through reducing the power consumption of transmission of each individual sensor. However, it increases the overhead of the CHs. In some extreme cases, some clusters are much larger than others. The CHs of these larger clusters will die earlier. As a result, the sensors in a WSN should be distributed to each CH evenly, to balance the power usage. The node density surrounding a CH will dictate the number of sensors that join the cluster. Thus node density is a consideration when determining energy cost. Additionally, (Wang et al., 2006) indicates that multi-hop consumes more energy than single hop under realistic circumstances. Therefore sensors joining a CH through fewer hops can also save more energy. The other way is through transmission power control. If the RSSI between two sensors is higher than a threshold, the transmitter can lower its transmission power to reduce the transmission cost while maintaining the link quality. The threshold erases the problem that when decreasing the transmission power, link quality becomes poor. This technique is called transmission power control. More power can be saved if the RSSI is higher (Lin et al., 2006). For the above reasons, the energy cost for the communication between two sensors can be expressed as:

$$C = c\,(Density, Hops, RSSI)\,. \tag{3}$$

## 4.6 Multi-choice user case

The relationship between network latency (L), reliability (R) and power consumption (C) is shown in Figure 5. Three of them are highly related to each other and they can be determined by four factors including local node density, ToF, hop count and RSSI. However, how significantly the four factors can affect each of the three criteria is unequal. For configuration $C_1$, $C_2$ and $C_4$ in Table 1, the user only have one single requirement. Therefore we can simply assign ComD to $C$, $1/R$ (The higher the transmission reliability is between two sensors, the shorter ComD ought to be.) or $L$. For other multi-selection user configurations, it is necessary to combine two or three metrics to calculate ComD and this needs careful investigation. By default, we can assume that the user puts equal weight on each criterion. One naive proposal is to multiply the corresponding metrics. For example, as $C_3$ combines $C_1$ ($C$) and $C_2$ ($1/R$), it can be represented as $ComD = C/R$. The problem is that the domains of function $l$, $r$, and $c$ may demand the determination and application of some coefficients for normalisation. Furthermore, if RSSI affects C in a linear manner while affecting R in a experimental manner, when RSSI changes, R changes quicker than C. Consequently, R may dominate the combination value. Then the combination configuration will have an uneven priority on C and R. For the above reasons, the multi-choice user cases still need further study.

# 5 DISCUSSION

In this paper, we choose four factors including ToF, hop count, RSSI and node density to characterize network latency (L), reliability (R) and power consumption (C). Although other factors can also be used, there are several reasons supporting our choice:

1. The selected factors are easy to capture. Through the broadcasting in the initial phase, ToF, RSSI and local density can be known. The hop count from a sensor to a CH is available from the packets received from the CH.
2. The protocol only relies on communication information. It is not necessary to know the physical locations of the sensors, which cannot be easily measured by the sensors themselves.
3. The measured values are obtained from real time data, which makes ComD adaptive to a changeable environment.

Currently the user configuration of L, R and C is a binary choice. As we have mentioned in Section 4.1, for the combination cases, we set equal priority to each criterion. In the future, it will be implemented in a manner that an application developer can put different weight on the three metrics.

To support multi-hop communication in a cluster, not only the CHs need to broadcast their information, but also some other sensors. We call this process the second-level broadcast. Deciding the number of sensors that should perform second-level broadcasting is a non-trivial problem. If there are not enough sensors to broadcast, some sensors may be not able to discover a multi-hop route.

# 6 CONCLUSIONS AND FUTURE WORK

This paper provides a novel user-configurable metric that facilitates user adaption in clustering algorithms for WSNs. This metric is influenced by and accommodates three performance objectives that normally exercise users in a WSN, namely: network latency, transmission quality and energy consumption. The underlying relationship between these three operational parameters is revealed. In the future, more work will be undertaken in the analysis of the interrelationship between these aspects so as to construct a formula that balances them within a user configuration. The performance of ComD will be evaluated in both simulation and real time experiments.

# REFERENCES


Abbasi, A. A. and Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer Communications*, 30(1415):2826 – 2841.

Akkaya, K. and Younis, M. (2003). An energy-aware qos routing protocol for wireless sensor networks. In *Distributed Computing Systems Workshops*.

Aurenhammer, F. (1991). Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23.

Baccour, N., Koubâa, A., Mottola, L., Zúñiga, M. A., Youssef, H., Boano, C. A., and Alves, M. (2012). Radio link quality estimation in wireless sensor networks: A survey. *Trans. Sen. Netw.*

Boyinbode, O., Le, H., Mbogho, A., Takizawa, M., and Poliah, R. (2010). A survey on clustering algorithms for wireless sensor networks. In *Network-Based Information Systems*, pages 358–364.

Heinzelman, W., Chandrakasan, A., and Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. In *HICSS*.

Heo, N. and Varshney, P. (2005). Energy-efficient deployment of intelligent mobile sensor networks. *Systems, Man and Cybernetics*, 35(1):78–92.

Kim, D., Abay, B., Uma, R. N., Wu, W., Wang, W., and Tokuta, A. (2012). Minimizing data collection latency in wireless sensor network with multiple mobile elements. In *INFOCOM, 2012 Proceedings IEEE*.

Kleinrock, L. (1975). *Theory, Volume 1, Queueing Systems*. Wiley-Interscience.

Lin, S., Zhang, J., Zhou, G., Gu, L., Stankovic, J. A., and He, T. (2006). Atpc: adaptive transmission power control for wireless sensor networks. In *Embedded networked sensor systems*. ACM.

Liu, X. (2012). A survey on clustering routing protocols in wireless sensor networks. *Sensors*, 12(8).

Saukh, O., Marrn, P., Lachenmann, A., Gauger, M., Minder, D., and Rothermel, K. (2006). Generic routing metric and policies for wsns. In Rmer, K., Karl, H., and Mattern, F., editors, *Wireless Sensor Networks*, volume 3868 of *Lecture Notes in Computer Science*.

Smith, R. B. (2007). Spotworld and the sun spot. In *Proceedings of the 6th international conference on Information processing in sensor networks*, IPSN. ACM.

Tang, S. and Li, W. (2006). Qos supporting and optimal energy allocation for a cluster based wireless sensor network. *Computer Communications*, 29(1314).

Urteaga, I., Yu, N., Hubbell, N., and Han, Q. (2011). Aware: Activity aware network clustering for wireless sensor networks. In *Local Computer Networks (LCN)*.

Wang, Q., Hempstead, M., and Yang, W. (2006). A realistic power consumption model for wireless sensor network devices. In *Sensor and Ad Hoc Communications and Networks*, volume 1.

Yousefi, H., Mizanian, K., and Jahangir, A. (2010). Modeling and evaluating the reliability of cluster-based wireless sensor networks. In *Advanced Information Networking and Applications*.