Methodologies for Comparing Clustering Algorithms in Wireless Sensor Networks

Stefan Unterschütz
Institute of Telematics
Hamburg University of Technology (TUHH)
Hamburg, Germany
stefan.unterschuetz@tu-harburg.de

***Abstract*— This paper presents general methodologies for comparing distributed algorithms, which are exemplified by clustering algorithms in sensor networks. Significant metrics for evaluating the algorithms are introduced including aspects of a structural, analytical and simulative comparison. Finally a short, exemplary comparison of two clustering algorithms HEED and WCA is made.**

***Index Terms*— clustering algorithm, wireless sensor networks**

I. Introduction

In some practical areas a large-scale, non-invasive observation of physical or ecological state variables like temperature, air pressure or movement is necessary. Wired sensors are an inappropriate technology in such scenarios. This motivates the usage of tiny wireless sensor nodes. These devices are forming large, self-organizing wireless sensor networks (WSN) transferring sensed data to predefined base stations.

To attain nodes of small size and low costs sensors are often equipped with limited resources (e.g., energy, memory). Therefore strategies for administrating this massive distributed system are necessary. Challenges like scalability, fault-tolerance, and robustness have to be met. Additionally, sensor nodes often use batteries as power supply. Above all, the main energy consumer is the radio communication, which implies the need of energy efficient solutions in order to maximize the lifetime of the network.

An adequate strategy to deal with the mentioned challenges, is the grouping of sensor nodes into clusters. For each cluster a leader called cluster-head (CH) is selected, holding the task of coordinating the group. The other nodes of a cluster are referred as cluster-members. In Fig. 1 a clustered network is shown in which sensed data is passed via the cluster-heads to a base station.

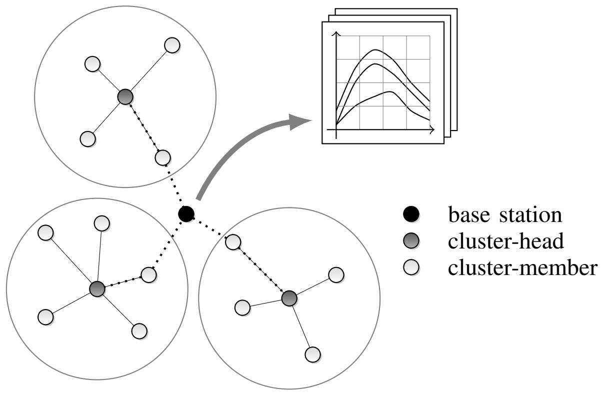


Fig. 1. Clustering in wireless sensor networks

Clustering provides two major advantages: It reduces the complexity of the network and allows an intra-cluster coordination. Concerning the first aspect, routing decisions between clusters become more scalable due to the hierarchical network organization. Moreover, the intra-cluster administration allows the compression of correlated sensor data to reduce the network traffic and makes the establishment of efficient communication schemes more feasible.

Clustering has become a popular area of research, because of its promising advantages. A lot of different algorithms and application scenarios for clustering in sensor networks exist. A main drawback is that current classifications and comparison methods are insufficient for selecting an appropriate algorithm for a given scenario [1], [2].

The aim of this work is to present basic methodologies for comparing clustering algorithms used in wireless sensor networks. First and foremost, metrics have to be defined for measuring the clustering quality. In the following section structural, analytic and simulative methodologies for the comparison are investigated. Afterwards the introduced procedure is applied exemplarily for two clustering algorithms: HEED [3] and WCA [4].

II. Metrics for comparing algorithms

For evaluating and comparing clustering approaches well-defined metrics are necessary. In the following a short summary of different categories of metrics is given.

Regarding scalability the complexity of the algorithm must be examined. In the context of clustering in sensor networks it can be differentiated between runtime, memory-demand, communication cost as well as energy consumption. The latter is important for nodes without renewable energy resources.

Apart from the complexity the quality of the clustered network has to be measured, too. The mean cluster size and number of CHs have an impact on the scalability of protocols based on clusters (e.g., routing). Another metric is the overlap of adjacent clusters. A high overlap produces interferences and packet collisions during internal cluster communication, but is essential for the inter-cluster communication. Concerning this matter an independent set, which means there is no direct link between cluster-heads, is desirable.

Depending on the application scenario further metrics can be defined. Often an energetic efficient clustering is required in absence of renewable energy resources. Here the network lifetime, the time until the nodes run out of energy, must be determined. Note that this metric also depends on other protocol layers. In case of faults, like unreliable communication, the correctness of the clustering procedure has to be examined.

III. Methodologies for comparison

In this paper three domains for evaluating and comparing clustering algorithms are introduced: a structural, analytical, and simulative approach respectively. These methodologies allow to select a suitable algorithm for a deployment of a network in a real application area.

A. Structural approach

A structural comparison enables to group and categorize algorithms with similar properties. For this purpose, concepts are adopted from existing clustering algorithms and different surveys [5], [1], [2]. Fundamental structural properties are identified to create an easy applicable taxonomy. The main categories of this taxonomy are assumptions, goals, and the clustering process itself, which are in the following described in more detail.

In general, each algorithm makes *assumptions* about the configuration and scenario for which it is designed. Concerning the used sensor nodes this can be the hardware equipment as well as the communication medium (e.g., radio). Also the assumed topology and deployment of nodes are important. Topology changes based on mobility, node failures etc. may be allowed. Furthermore, a priori assumptions are often made5. An example is the awareness of the global time and position respectively. Finally assumptions about tolerated faults have to be declared. Typical faults in sensor networks are packet loss and node failures.

Clustering algorithms are designed to reach *goals* like a specific cluster structure and cluster-head distribution respectively. Also load-distribution among CHs, energy saving, high connectivity, and fault-tolerance are often emphasized goals. The achievement of the different objectives can be evaluated by using the metrics described in Section II.

Common patterns of the clustering *process* can be identified. In general an algorithm either runs in a centralized, e.g., controlled by the base station, or distributed manner. The latter one is preferred in terms of scalability. The actual CH selection can be done deterministically or randomly. A further issue is the cluster maintenance, for instance how to cope with topology changes. Common techniques are periodic reclustering of the whole network or local reclustering strategies.

Although the structural approach introduced in this section allows to group algorithms with similar properties, this is insufficient to be able to compare the clustering performance and quality. Here other concepts are necessary.

B. Analytical approach

Clustering algorithms can be compared analytically. This requires an underlying mathematical model. Often the wireless sensor network is represented as a graph with vertices and edges. Additionally, probabilistic theories can be applied for modeling spatial distributions of nodes, traffic rates etc.

Based on the mathematical model an analysis of the clustering algorithm is possible. For instance asymptotic limits are calculable to estimate the complexity. Also estimations concerning the inter- and intra-cluster structure can be given by calculating expectation values.

The mathematical analysis of an algorithm is crucial to prove its correctness, but it is often not suitable to make a representative comparison between different algorithms. The problem is that the constructed assumptions and models may differ. Also the calculation of many of the mentioned metrics is simply not feasible.

C. Simulative approach

Based on simulative results an adequate comparison of algorithms is possible. In this respect the granularity of the simulation is of importance. In the most simple case only message exchange between nodes is simulated, but complex simulations may consider channel characteristics (propagation model), network protocols, energy consumption, mobility, and failures. For this purpose, a network simulator optimized for sensor networks is essential. The simulator should allow an easy implementation, simulation, and evaluation of an algorithm using well-defined metrics.

A simulation enables a comparison of algorithms regarding to a given scenario. Note, that only a marginal part of the possible execution space of clustering algorithms can be covered, but this is sufficient if only a specific scenario is considered.

The final step for comparing algorithms would be the implementation and deployment of a real network. This is associated with high costs. On the other hand modern simulation techniques, like the network simulator NS2 [6], come close to results of real networks [7].

IV. An Exemplary Comparison

The following section gives an example how to use the methodologies explained above. For that purpose two clustering algorithms for sensor networks are compared: A Hybrid, Energy-Efficient, Distributed Clustering Approach (HEED) and an On-Demand Weighted Clustering Algorithm (WCA).

A. A brief specification and structural comparison

Both clustering procedures are distributed algorithms, which terminate after running a finite number of rounds. The clustering process is round based and each data transmission is done by broadcasting a packet. The clustering objectives for the algorithms are the maximization of the cluster-head residual energy as well as the minimization of the number of direct links between cluster-heads (independent set). Both algorithms are designed for error-free environments.

In HEED the probability for becoming a cluster-head, which is indicated by a specific broadcast, rises with the residual energy of a sensor. The other nodes are taking the expected energy and communication costs into account when selecting an available cluster-head in their neighborhood.

In WCA each node computes a combined weight by taking properties like degree (number of neighbors), link quality, residual energy, mobility into account, whereupon the calculation of the weight can vary depending of the network demand. In the clustering process the node with the lowest weight in its neighbourhood and not already member of another cluster is elected as cluster-head.

B. An analytic comparison

For both algorithms the complexity can be estimated. HEED terminates in $O(1)$ rounds with an average packet demand of $O(1)$ for each node. The complexity of WCA is for both cases $O(n)$. The upper limit holds for a chain topology with ordered weights. Comparing both algorithms analytically, HEED clearly outperforms WCA.

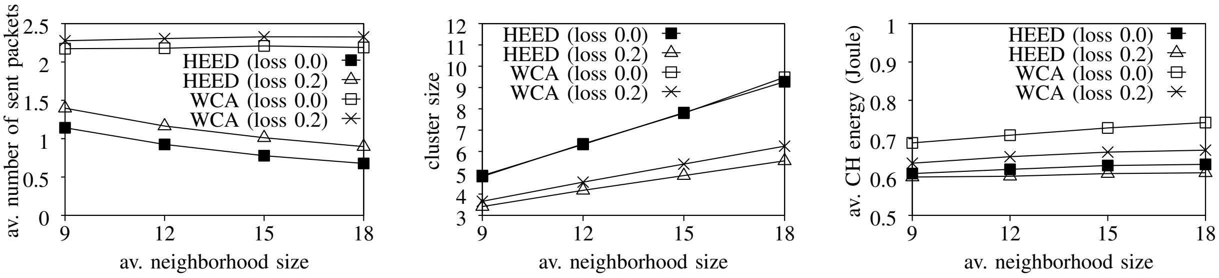
C. A simulative comparison

Finally WCA and HEED are simulated, using a clustering-framework developed in my project work [8]. This framework, which is based on the network simulator NS2, offers an easy to use programming interface, optimized for implementing and evaluating clustering algorithms.

Different settings for the simulations are taken. In all cases random generated topologies, each with 1000 nodes are used. The communication range of the sensor nodes is adapted to achieve an average neighborhood size (node degree) of 9, 12, 15, and 18 respectively. Simulations are done without and with 20 % packet loss. Also the energy of a node is randomly (uniform distributed) assigned to an arbitrary value between 0 to 1 Joule (expectation value is 0.5 Joule). This allows to examine the ability of a clustering algorithm, to prefer nodes of high residual energy as cluster-head.

The WCA algorithm is configured to maximize the residual energy of the CHs. HEED runs with the recommended configuration presented in the corresponding paper. A total number of 3200 simulations are taken to produce the results.

Figure 2(a) shows that the average communication costs (for a node) in HEED are by factor two lower than in WCA. It is observable that the communication is slightly higher under presence of packet



(a) (b) (c)

Fig. 2. Simulation results of HEED and WCA

loss for both algorithms. An interesting result is, that although the packet complexity of WCA is O(n), in most practical scenarios a asymptotic behavior of O(1) can be observed.

The average cluster size is an indicator for the scalability. Figure 2(b) shows that the cluster size increases with the node density. The results for HEED and WCA are similar. In comparison to the simulation with 20 % packet loss, the algorithms, running in an error-free environment, produce clusters of double size.

Both algorithms favor cluster-heads with high residual energy. In WCA this is done in a deterministic manner, whereas in HEED a stochastic approach is taken. For this reason WCA outperforms HEED (Fig. 2(c)). However under presence of packet loss a high drop of the performance is observable for WCA.

D. Evaluation of the methodologies for comparison

The comparison of HEED and WCA demonstrates the importance of an evaluation using different approaches. Only the union of them allows to make meaningful statements about the properties and behavior of a clustering algorithm. Also it was shown that results from the analyses and simulation can differ significantly (e.g., communication costs of WCA).

V. Conclusion

This paper presents different methodologies for comparing clustering algorithms in sensor networks. For this purpose metrics, necessary for the evaluation of the clustering schemes, are defined as well as domains of comparison. The structural approach allows a classification and categorization respectively. The mathematical analysis is used to prove correctness and to calculate expectation values and asymptotic limits for a given model. Finally, the behavior of the algorithms in a specific scenario is examined by simulation. The usage of all three domains allows a broad comparison of clustering algorithms in sensor networks. Even though the methodologies were restricted in this paper for clustering algorithms, they are mostly valid for distributed algorithms in general.

The usage of the introduced techniques were briefly demonstrated for the algorithms HEED and WCA. In my project work [8] as well as in further research activities the simulation of algorithms is performed very extensively. For that purpose popular clustering algorithms were examined under different configurations, by running a total number of over 2 million NS2 simulations.

Acknowledgments

I would like to thank Volker Turau and Christoph Weyer for their helpful comments in respect of writing a scientific paper.

References

[1] Ameer Ahmed Abbasi and Mohamed Younis, “A Survey on Clustering Algorithms for Wireless Sensor Networks,” *Comput. Commun.*, vol. 30, no. 14-15, pp. 2826–2841, October 2007.

[2] Jane Y. Yu and H. J. Peter Chong, “A Survey of Clustering Schemes for Mobile Ad Hoc Network,” *IEEE Commun. Surveys Tuts.*, vol. 7, no. 1, pp. 32–48, 2005.

[3] Ossama Younis and Sonia Fahmy, “HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks,” *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, October 2004.

[4] Mainak Chatterjee, Sajal K. Das, and Damla Turgut, “An On-Demand Weighted Clustering Algorithm (WCA) for Ad hoc Networks,” in *Proceedings of GLOBECOM ’00*, San Fransisco, CA, USA, November 2000.

[5] Kay Römer and Friedemann Mattern, “The Design Space of Wireless Sensor Networks,” *IEEE Wirel. Commun.*, vol. 11, no. 6, pp. 54–61, December 2004.

[6] Kevin Fall and Kannan Varadhan, *The NS Manual*, December 2008, http://www.isi.edu/nsnam/ns/doc/ (last visit: 24.03.09).

[7] Svilen Ivanov, André Herms, and Georg Lukas, “Experimental Validation of the NS-2 Wireless Model using Simulation, Emulation, and Real Network,” in *Proceedings of IEEE WMAN’07*, Bern, Schweiz, March 2007.

[8] Stefan Unterschütz, “Framework zur Simulation und Evaluation von Clustering-Algorithmen,” Project Work, Institute of Telematics, Hamburg University of Technology, 2008.