Data Aggregation, Spatio-Temporal Correlation and Energy-Aware Solutions to Perform Data Collection in Wireless Sensor Networks

Leandro A Villas[1,2], **Regina B. de Araujo**[3], **Antonio A. F. Loureiro**[1]

[1] UWL - Research Laboratory, Federal University of Minas Gerais
Belo Horizonte – MG – Brazil

[2] Institute of Computing, State University of Campinas
Campinas –SP – Brazil

[3] WINDIS - Research Laboratory, Federal University of Sao Carlos
Sao Carlos –SP – Brazil

leandro@ic.unicamp.br, regina@dc.ufscar.br, loureiro@dcc.ufmg.br

Resumo. *Este trabalho discute diferentes abordagens para agregação de dados e a exploração da correlação espaço-temporal dos dados em Redes de Sensores sem Fio (RSSFs). Na tese, nós investigamos e propusemos soluções que são adequadas para muitos cenários diferentes em RSSFs. Os algoritmos propostos reduzem o número de mensagens necessárias para criar uma estrutura de roteamento, maximizam o número de rotas sobrepostas, selecionam rotas com a maior taxa de agregação, e realizam a transmissão de dados agregados de forma confiável. Além disso, os nossos algoritmos têm sido extensivamente comparado com os que estão disponíveis na literatura e os resultados mostram que eles são possíveis alternativas para realizar a agregação e correlação espaço-temporal de dados durante o processo de roteamento em redes de sensores. Esta tese resultou em quatro artigos em periódicos e 18 trabalhos em conferências.*

Abstract. *This work discusses different approaches for data aggregation and spatio-temporal data correlation in Wireless Sensor Networks (WSNs). In the thesis, we investigate and propose solutions that are suitable for many different scenarios in WSNs. The proposed algorithms reduce the number of messages necessary to set up a routing tree, maximize the number of overlapping routes, select routes with the highest aggregation rate, and perform reliable data aggregation transmission. Moreover, our algorithms have been extensively compared with the ones available in the literature and the results show that they are potential alternatives to perform data aggregation and spatio-temporal data correlation during the routing process in WSNs. This thesis resulted in 4 journal papers and 18 conference papers.*

1. Introduction

Recent developments in the areas of wireless communication and multifunctional sensors with communication and processing capabilities have stimulated the development and use of wireless sensor networks (WSNs) in many different domains such as environmental,

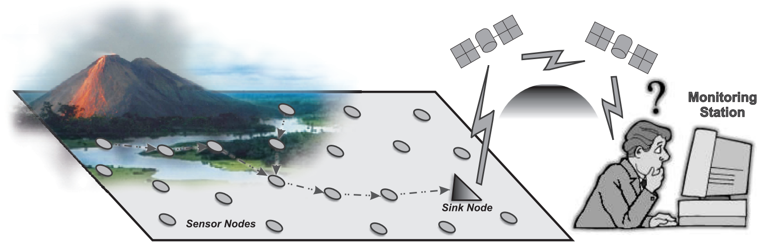

Figura 1. Data aggregation aware routing, a key algorithm for data-driven WSNs.

medical, industrial, military and many others where the human presence is not possible nor desired [Villas et al. 2012a]. A sensor node typically presents a limited sensing capability, but the overall network sensing capability can be increased when nodes are combined with many other nodes forming a WSN. For example, if a gas leak occurs in a room full of gas cylinders and there is only one sensor in this room, it will only be possible to say whether there is a leak or not. On the other hand, if a WSN is used together with appropriate protocols, it is possible not only to detect the leak, but also to indicate where the leak started and how it evolved. A monitoring solution in this way can save lives and assets, and reduce cost insurance [Villas et al. 2011].

Sensor nodes are energy-constrained devices and the energy consumption is generally associated with the amount of gathered data, since communication is often the most expensive activity in terms of energy. For that reason, algorithms and protocols designed for WSNs should consider the energy consumption in their design [Villas et al. 2012a]. Moreover, WSNs are data-driven networks that usually produce a large amount of information that needs to be routed, often in a multihop fashion, toward a sink node, which works as a gateway to a monitoring center (Figure 1). Given this scenario, routing plays an important role in the data gathering process.

For more efficient and effective data gathering with a minimum use of limited resources, sensor nodes should be configured to smartly report data by making local decisions [Villas et al. 2011]. For this, data aggregation[1] and spatio-temporal[2,3] data correlation are effective techniques for saving energy in WSNs. These techniques have been used in WSNs with two purposes: (i) to take advantage of the redundancy and improve data accuracy [Schmid and Schossmaier 2001, Chakrabarty et al. 2002, Liu et al. 2007], and (ii) to reduce the overall data traffic and save energy [Le et al. 2008, Min and Chung 2010, Pham et al. 2010]. Nevertheless, current proposals have a high cost to create routing structures aware of data aggregation and many of them do not deal with node failures and interruptions in communications, which cause data loss and do not guarantee delivery of the sensed data. In addition, these solutions not only introduce delays in data transmissions but also lead to the reception of outdated information by the sink

[1]**Data Aggregation** eliminates inherent redundancy in raw data gathered by sensor nodes and forwards only smaller aggregated information.

[2]**Spatial correlation:** the change pattern of the data sensed by nearby nodes is expected to be the same or similar. Thus, by exploiting the spatial data correlation we can eliminate similar data notifications.

[3]**Temporal correlation:** the change pattern in readings of a sensor node and gathered data is usually similar over a short-time period. Due to the nature of the physical phenomenon, there is a significant temporal correlation among each consecutive observations of a sensor node. Thus, by exploiting the temporal correlation we can eliminate similar data notifications.

node [Deligiannakis and Kotidis 2008, Pham et al. 2010].

In order to overcome these challenges, in this thesis, we went further and proposed solutions that maximize the aggregation rate along the communication route, in a more reliable way, through a fault-tolerant routing mechanism. In addition, the proposed solutions fully exploit both spatial and temporal correlations to perform near real-time data collection in WSNs.

2. Main Contributions

The main contributions of this thesis are the design and development of four different solutions for data aggregation and spatio-temporal data correlation for WSNs, which we refer to as the DAARP, DDAARP, DST, and EAST algorithms, respectively. In summary, we have:

1. Data Aggregation Aware Routing Protocol for WSNs (DAARP): DAARP is a novel reactive data aggregation aware routing protocol for WSNs. The main motivation to design a new data aggregation aware routing protocol is that the proposed solutions in the literature present a high cost to create routing structures aware of data aggregation. The DAARP algorithm builds a routing structure with the shortest paths (in hops) that connect all source nodes to the sink while maximizing data aggregation. Its main contribution is to maximize data aggregation along the communication route, in a more reliable way, through a routing fault-tolerant mechanism. Simulation results reveal that DAARP has some keys aspects required by data aggregation in WSNs such as reduced number of messages for setting up a routing structure, maximized number of overlapping routes, high aggregation rate, and reliable data aggregation and transmission. This algorithm is fully explained and evaluated in Chapter 3 [Villas et al. 2012b].

2. Dynamic Data-Aggregation Aware Routing Protocol for WSNs (DDAARP): DDARP is a novel dynamic data-aggregation aware routing protocol for WSNs, which uses the sink node for processing and configuring routes aware of data aggregation. The main motivation to design a dynamic approach to create dynamic routing structures aware of data aggregation is that the quality of routing structures created by DAARP and most algorithms in the literature depends on the order of event occurrences and, once created, these routes are held fixed during the lifetime of events. The main contribution is that routes created by DDAARP do not depend on the order of event occurrences and are not held fixed during the lifetime of events such as DAARP and others. Simulation results reveal that DDAARP presents a low cost in terms of packet control, improves the quality of the routing structure and maximizes data aggregation along the communication route in a more reliable way, through a routing fault-tolerance mechanism. This algorithm is fully explained and evaluated in Chapter 4 [Villas et al. 2012b].

3. Dynamic and Scalable Tree for WSNs (DST): DST is an efficient data aggregation solution that allows scalable and dynamic routing in WSNs, which builds routing structures with the shortest routes (in Euclidean distance) that connect all source nodes to the sink node maximizing data aggregation and reducing the distance to connect each source node to the sink. Also, the created routing structure does not

depend on the event order. The main motivation to design DST was the lack of solutions in the literature that are scalable, dynamic and present low cost to create routing structures aware of data aggregation. DDAARP presents good results, but it suffers from scalability problems and becomes impractical for large-scale networks. In addition, the sink node needs to have a global knowledge of the network. Simulations results reveal that DST presents a low cost in terms of packet control, maximizes aggregation points and improves the quality of routing structures by offering dynamic routes. This algorithm is fully explained and evaluated in Chapter 5 [Villas et al. 2012b].

4. Efficient Data Collection Aware of Spatio-Temporal Correlation for WSNs (EAST): EAST is an algorithm for energy-aware data forwarding in WSNs that takes full advantage of both spatial and temporal correlation mechanisms to save energy while still maintaining real-time, accurate data report towards the sink node. The main motivation to design EAST is that most of the current spatial and/or temporal correlation algorithms do not consider the energy dissipation during data collection to better choose the representative nodes. Also, these solutions present a high number of control messages and do not exploit efficiently the spatio-temporal correlation nor their dynamicity. The main contribution is an energy-aware spatio-temporal correlation mechanism in which nodes that detect the same event are dynamically grouped in correlated regions and a representative node is selected at each correlation region for observing the phenomenon. The entire region of sensors per event is effectively a set of representative nodes performing the task of data collection and temporal correlation. Simulation results clearly show that by using both spatial and temporal correlation, the event can be sensed with a high accuracy while still saving the residual energy of nodes. This algorithm is fully explained and evaluated in Chapter 6 [Villas et al. 2012b].

3. Final Remarks

This section summarizes this thesis, discusses directions for future research and presents the publications related to this thesis. The objective is to highlight our contributions and to point out some possible directions to proceed with the research. In this context, we first present the thesis conclusions in Section 3.1. In Section 3.2 we present future directions of this work. Finally, in Section 3.3 we present the publications related to this thesis.

3.1. Conclusions

In the current literature of spatial and/or temporal correlation algorithms, most of the proposed studies do not consider the energy dissipation during data collection to better choose the representative nodes. Also, these solutions present a high number of control messages and do not exploit efficiently the spatio-temporal correlation nor their dynamicity. In our solutions, we went further and proposed energy-aware spatio-temporal correlation mechanisms in which nodes that detected the same event are dynamically grouped in correlated regions and a representative node is selected at each correlation region for observing the phenomenon of interest. The entire region of sensors per event is effectively a set of representative nodes performing the task of data collection and spatio-temporal correlation.

The different scenarios in which a WSN can be deployed as well as the broad applicability of WSNs indicate there is no single solution for a problem in WSNs. For this reason, we proposed different solutions for the same problem in WSNs. Each solution is designed to work well in a specific scenario such as static and mobile networks; small, medium, and large scale networks; and sparse and dense networks. Given this observation, the choice of which protocols and algorithms to be used in the routing structure for a WSN depends on both the scenarios and the application requirements.

The solutions developed in this thesis present several advantages: reduce the number of message necessary to set up a routing tree, maximize the number of overlapping routes, select routes with the highest aggregation rate, and perform reliable data aggregation transmission. All these solutions were analyzed and evaluated with other previous state of the art proposals.

3.2. Directions for Future Research

The results obtained in this thesis are very promising. The proposed solutions described above usually take advantage of both data aggregation and spatio-temporal data correlations to improve the routing performance and reduce the energy consumption in the data gathering while preserving both data accuracy and real-time reporting. However, these algorithms may not perform well in mobile networks.

A very promising future work is to extend the proposed solutions to work in mobile wireless sensor networks. It would also be interesting to investigate a special kind of Mobile Ad Hoc Network known as Vehicular Ad Hoc Network (VANET), in which vehicles equipped with wireless and processing capabilities can create a spontaneous network while moving along roads.

3.3. Comments on Publications

As a result of this thesis, we had several publications: four in journals and 12 in conferences of international prestige and great impact in the area of ad hoc networks plus 6 in Brazilian Symposium on Computer Networks and Distributed Systems.

- **Journals**
  - *IEEE Transactions on Computer*
  - *Elsevier Computer Communications*
  - *Elsevier Ad Hoc Networks*
  - *Elsevier Computer Networks*
- **Conferences**
  - 3 articles in *IEEE Symposium on Computers and Communications* - IEEE ISCC
  - 2 articles in *IEEE International Conference on Communications* - IEEE ICC
  - 2 articles in *IEEE Conference on Local Computer Networks* - IEEE LCN
  - 4 articles in *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* - ACM MSWiM
  - 1 article in *ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks* - ACM PE-WASUN
  - 6 articles in *Brazilian Symposium on Computer Networks and Distributed Systems* - SBC SBRC

According to the current Qualis classification, the published papers are classified as:

- Two articles in journals/conferences ranked in Qualis Capes as A1;
- Eight articles in journals/conferences ranked in Qualis Capes as A2;
- Four articles in journals/conferences ranked in Qualis Capes as B1;
- Eight articles in conferences ranked in Qualis Capes as B2.

Referências

[Chakrabarty et al. 2002] Chakrabarty, K., Member, S., Iyengar, S. S., Qi, H., and Cho, E. (2002). Grid coverage for surveillance and target location in distributed sensor networks. *IEEE Transactions on Computers*, 51:1448–1453.

[Deligiannakis and Kotidis 2008] Deligiannakis, A. and Kotidis, Y. (2008). Geosensor networks. In Nittel, S., Labrinidis, A., and Stefanidis, A., editors, *Book chapter: Exploiting Spatio-temporal Correlations for Data Processing in Sensor Networks*, pages 45–65. Springer-Verlag.

[Le et al. 2008] Le, T. D., Pham, N. D., and Choo, H. (2008). Towards a distributed clustering scheme based on spatial correlation in wsns. In *Wireless Communications and Mobile Computing Conference, 2008. IWCMC '08. International*, pages 529 –534.

[Liu et al. 2007] Liu, L., Member, S., and Yu, P. S. (2007). Asap: An adaptive sampling approach to data collection in sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 2007:1766–1783.

[Min and Chung 2010] Min, J.-K. and Chung, C.-W. (2010). Edges: Efficient data gathering in sensor networks using temporal and spatial correlations. *J. Syst. Softw.*, 83:271–282.

[Pham et al. 2010] Pham, N. D., Le, T. D., Park, K., and Choo, H. (2010). Sccs: Spatiotemporal clustering and compressing schemes for efficient data collection applications in wsns. *International Journal of Communication Systems*, 23:1311–1333.

[Schmid and Schossmaier 2001] Schmid, U. and Schossmaier, K. (2001). How to reconcile fault-tolerant interval intersection with the lipschitz condition. *Distrib. Comput.*, 14:101–111.

[Villas et al. 2012a] Villas, L., Boukerche, A., Ramos Filho, H., Oliveira, H., Araujo, R., and Loureiro, A. (2012a). Drina: A lightweight and reliable routing approach for in-network aggregation in wireless sensor networks. *Computers, IEEE Transactions on*, PP(99):1.

[Villas et al. 2012b] Villas, L. A., Boukerche, A., Araujo, R. B., and Loureiro, A. A. (2012b). *Data Aggregation, Spatio-Temporal Correlation and Energy-Aware Solutions to perform Data Collection in Wireless Sensor Networks*. PhD thesis, Federal University of Minas Gerais, http://www.dcc.ufmg.br/pos/cursos/defesas/1197D.PDF.

[Villas et al. 2011] Villas, L. A., Boukerche, A., de Oliveira, H. A., de Araujo, R. B., and Loureiro, A. A. (2011). A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks. *Ad Hoc Networks*.