Information Fusion in Wireless Sensor Networks*

Eduardo Freire Nakamura[1,2], **Antonio A. F. Loureiro**[2]

[1]Department of Computer Science – Federal University of Minas Gerais (UFMG)
DCC/ICEx – Av. Antonio Carlos, 6627 – 31270-010 – Belo Horizonte – MG – Brazil

[2]Research and Technological Development Center (FUCAPI)
CESF – Av. Danilo Areosa, 381 – 69075-351 – Manaus – AM – Brazil

eduardo.nakamura@fucapi.br, loureiro@dcc.ufmg.br

Abstract. *In wireless sensor networks (WSNs), energy consumption and data quality are two important issues due to limited energy resources and the need for accurate data. In this scenario, information fusion techniques (also reffered to as data fusion or data aggregation) have been used to distributedly process sensor data to save energy – by reducing the overall communication load – and increase accuracy – by removing noisy measurements of correlated data. The thesis herein summarized presents a study about how information fusion techniques can be used in WSNs to improve data accuracy and reduce the energy consumption. The main contributions include: (1) an information-fusion framework for WSNs; (2) a novel algorithm that applies information fusion to detect when a routing tree needs to be rebuilt; (3) a novel routing strategy, based on role assingment, which maximizes the gains of an information-fusion application; and (4) a critical survey about information fusion in WSNs.*

1. Introduction

Wireless Sensor Networks (WSNs) are composed of a large number of nodes with sensing capability [Akyildiz et al. 2002]. The applicability of such networks includes several areas such as environmental, medical, industrial, and military applications. Usually, WSNs have strong constraints regarding energy resources and computational capacity. In addition, these networks demand self-organizing features to autonomously adapt themselves to eventual changes resulting from external interventions, reaction to a detected event, or requests performed by an external entity.

In general, WSNs are deployed in environments where sensors can be exposed to conditions that might interfere with the sensor readings or even destroy the sensor nodes. As a result, sensor measurements may be more imprecise than expected, and the sensing coverage may be reduced. A natural solution to overcome failures and imprecise measurements is to use redundant nodes that cooperate with each other to monitor the environment. However, redundancy poses scalability problems caused by potential packet collisions and transmissions of redundant data. To overcome such a problem, information fusion is frequently used. Briefly, information fusion comprises theories, algorithms, and tools used to process several sources of information generating an output that is, in some sense, better than the individual sources. The proper meaning of "better" depends on the application. For WSNs, "better" has at least two meanings: cheaper and more accurate [Nakamura et al. 2007b].

*Thesis available at: `http://www.dcc.ufmg.br/pos/cursos/defesas/46D.PDF`.

The thesis, herein summarized, aims at providing a general discussion for information fusion in WSNs, allowing us to identify open issues, understand the requirements and the implications regarding information fusion and the resource-constrained WSNs. Then, we choose specific issues to provide new solutions based on information fusion. The main contributions can be summarized as:

1. **A survey about information fusion in WSNs.** This is not the thesis central contribution. However, the survey provides an ample view of information fusion in the domain of WSNs. As a result, the survey allows us to identify open issues and opportunities to use information fusion in WSNs.
2. **Diffuse: an information-fusion framework.** By assessing the solutions identified in the survey, we propose a framework, called Diffuse, to apply information fusion in WSNs. This framework encompasses the main functions and activities of a general fusion process, and a specific API that implements some useful algorithms for WSNs. The Diffuse framework has an ample applicability and should be seen as a tool for helping the designer to reason about what types of information fusion and what methods should be used, furthermore, how these methods should be used to accomplish an information-fusion task or application.
3. **Using information fusion to achieve reliability in data routing.** This contribution consists in specifying a data routing protocol that applies information-fusion techniques to achieve reliability in an environment with failures. In this case, information fusion plays a supporting role in the data routing task, illustrating how to use information fusion in a different application domain. This contribution is designed by using the Diffuse framework.
4. **A role assignment and routing strategy for event detection.** Solutions of information fusion for event detection usually evaluate the detection efficiency, but communication aspects are often put aside. This contribution comprises a data routing strategy that specifies the communication behavior during the detection and notification phases. Hence, it aims at providing an example of how we should design internal tasks in WSNs having in mind an information-fusion application.

The rest of this paper is organized as follows. Section 2 discusses the contributions directly resulted from the thesis research. In Section 3, we summarize a few additional contributions resulted from the carried out research and the interaction with other members of our research group. Section 4 presents our conclusions, while Section 5 discusses some future directions.

2. Thesis Key Contributions

In this section, we list the thesis key contributions in the order they appear in the original thesis document [Nakamura 2007]. For the sake of space, we do not provide extended details about each contribuition. Please check the thesis document [Nakamura 2007] for the details about the algorithms and implementations related to each contribution.

2.1. State-of-the-Art

Despite not being the thesis central contribution, this comprehensive survey about information fusion in WSNs is worth to be mentioned. The survey provides an ample view of information fusion in the domain of wireless sensor networks. It shows the methods and architectures that have been proposed and their corresponding benefits and limitations. Every method and architecture is contextualized by making references to the available literature about WSNs. As a result, the survey allows us to identify open issues and opportunities to use information fusion in WSNs. Additional contributions resulted from the identified state-of-the-art include a survey [Nakamura et al. 2007b], published in the *ACM Computing Surveys*, and two invited chapters [Nakamura et al. 2005b, Boukerche et al. 2008a].

2.2. The Diffuse Framework

By assessing the architectures, models, and methods identified in the survey, we propose the Diffuse framework [Nakamura et al. 2007a] to guide the use of apply information-fusion techniques in WSNs. This framework encompasses the main functions and activities of a general fusion process, and an API that implements useful algorithms for WSNs. The Diffuse framework has an ample applicability and should be seen as a tool for helping the designer to reason about what types of information fusion, what methods should be used, and how these methods should be used to accomplish an information-fusion task or application. Although its applicability is ample, as a proof of concept we show how the Diffuse framework can be used to achieve reliability in tree-based routing protocols (see next contribution). However, in the thesis document [Nakamura 2007], we also illustrate how we can use the framework in other scenarios, such as how to adapt the routing tree to improve data aggregation and avoid low-energy areas (nodes).

The elements of the architecture in Figure 1 are described in the following. The **Information Fusion API** includes the basic fusion methods demanded by sensor networks, which are used by the other framework components. The API is organized in three sub-components: *Aggregation*, *Estimation*, and *Inference* functions. In the **Signal Processor**, input signals measured by sensor nodes may embed some noise. The Signal Processor is responsible for choosing the proper information fusion methods to filter the noise and/or predict the signal behavior. Sensors provide measurements or signals that describe a measurable property of the sensed entity. However, to use such data, we usually translate signals into features that better describe the monitored entity. Thus, the **Feature Processor** translates signals into features. As an example of an input feature, the energy map may be used to detect the need for a topology rebuilding aiming at energy balance. The **State Estimator** applies inference methods to determine the state of a monitored entity based on the features describing the sensed entity. Inference methods commonly adopted are the Bayesian Inference, Neural Networks, and Dempster-Shafer Evidential Reasoning. The **Decision Maker** takes as input the state of the monitored entity, and decides which action should be taken in response to the identified state.

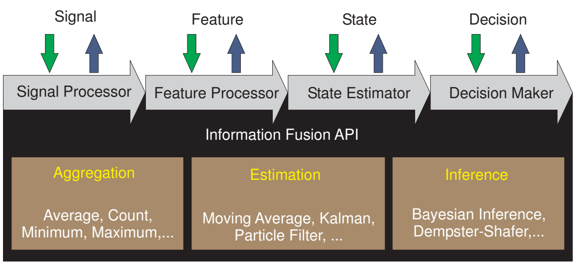
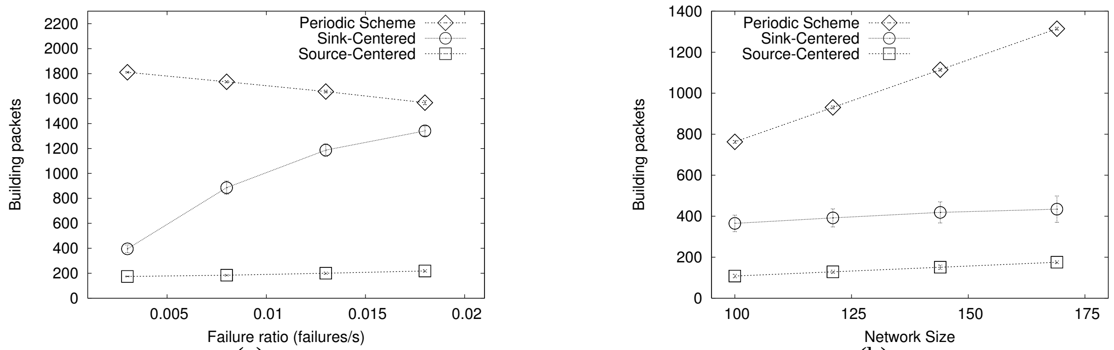


Figure 1. Diffuse architecture.

2.3. The Topology Building Algorithm

This contribution consists in specifying a data routing protocol [Nakamura et al. 2005d, Nakamura et al. 2005a, Nakamura et al. 2005c] that applies information-fusion techniques to achieve reliability in an environment with failures. In this case, information fusion plays a supporting role in the data routing task, illustrating how to use information fusion in a different application domain. This contribution is designed by using the Diffuse framework. In this solution, we estimate the data traffic (by using a moving average filter), extract features (assessing decays in the traffic level) that are translated into failure probabilities, then we use the Dempster-Shafer inference to decide whether or not we should rebuild the routing tree. The idea is to rebuild the routing tree only when necessary and save resources, in contrast to the proactive strategy that periodically rebuilds the routing tree even when no failures occur. The solution is implemented in two versions: *Sink-Centered*, in which the sink node detects failures and rebuilds the whole routing tree, and *Source-Centered*, in which the source node closest to the failure detects such an event and partially rebuilds the routing tree.

Figure 2 shows the performance of our solution compared to the proactive rebuilding strategy (*Periodic Scheme*). As we can see, when the number of failures increases (Figure 2(a)), the Sink-Centered scheme gets closer to the Periodic scheme. The reason is that the Sink-Centered scheme rebuilds the whole routing tree, and so does the Periodic scheme. However, since the Source-Centered scheme only partially rebuilds the routing tree, is performance is superior in the same situation. However, when the number of failures is constant and the network size increases (Figure 2(b)), both Sink- and Source-Centered schemes are always superior to the Periodic scheme. For networks with 169 nodes, the traffic overhead generated by the Source-Centered Diffuse is nearly 85% smaller than the traffic generated by the periodic rebuilding.



(a) (b)

Figure 2. Performance of the Diffuse-based solutions: (a) under different failure ratios; (b) under different network sizes.

2.4. The InFRA Algorithm

Solutions of information fusion for event detection, usually evaluate the detection efficiency. However, communication aspects are often put aside. This contribution comprises a data routing strategy that specifies the communication behavior during the detection and notification phases. Such a strategy must consider the fusion requirements imposed by the application to guarantee the desired quality of service (QoS). Hence, this contribution aims to provide an example of how we should design internal tasks in WSNs having in mind an information-fusion application. The proposed solution, called Information-Fusion-based Role Assignment (InFRA) [Nakamura et al. 2006a, Nakamura et al. 2006b, Nakamura et al. 2008], reactively assigns roles to build a routing tree that maximizes data aggregation [Nakamura et al. 2004]. This is essentially the Steiner Tree problem [Hwang et al. 1992]. Thus, the InFRA algorithm is actually a reactive and distributed heuristic for the Steiner Tree [Nakamura et al. 2006a]. In a nutshell, the InFRA algorithm builds a routing tree that finds the shortest paths that maximizes route overlapping, and whenever two routes overlap, the data from these routes are aggregated into a single data (information fusion), therefore, saving energy. Theoretical results demonstrated in the thesis document [Nakamura 2007] show that this heuristic has $O(1)$-approximation ratio when the network diameter remains constant and, in large-scale networks it has a k-approximation ratio, where k is the number of simultaneous events.

Figure 3 compares the InFRA algorithm with the reactive CNS (routing tree centered at the nearest source) and the reactive REFAT (a simple shortest-path tree). This figure shows that the InFRA is more scalable than the other algorithms. The reason is that the InFRA reduces the data transmissions by increasing data aggregation (Figure 3(a)). However, the most important result is that, as the network size increases, the InFRA spends less energy to process the data packets generated by source nodes (Figure 3(b)). Particularly, when the network size is 1024, even though InFRA has greater overhead, it spends nearly 70% of the energy used by REFAT.

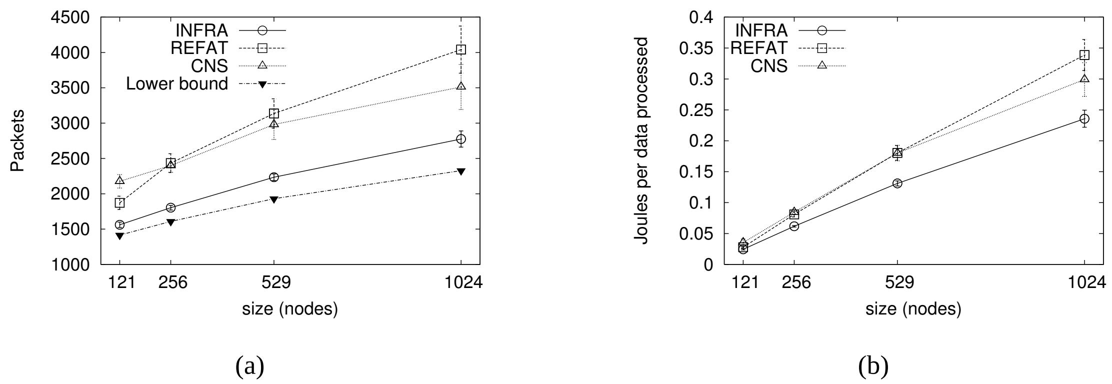


Figure 3. Performance of the InFRA algorithm: (a) packet transmissions; (b) energy efficiency.

3. Additional Contributions

During the thesis research, we have identified the possibility to colaborate with other researches by adding an information-fusion perspective. The results of such collaborations are summarized in the following.

3.1. Algorithms for Estimating Nodes' Location

Node location algorithms (or localization algorithms) for WSNs are usually distributed algorithms that compute nodes' location based on a set of reference nodes (nodes aware of their locations) [Boukerche et al. 2007]. As a consequence of errors in the location of such references nodes (e.g., due to GPS limitations) and limitations of distance estimation methods, localization algorithms present non-negligible errors [Oliveira et al. 2005].

Thus, information-fusion methods for value estimation are natural candidates for improving localization algorithms. In this context, we have applied information-fusion methods, such as the Least-Squares [Nakamura et al. 2007b], to compute a node location based on several references [Oliveira et al. 2007, Boukerche et al. 2008b, Boukerche et al. 2008d, Boukerche et al. 2008e, Boukerche et al. 2008c]. In particular, we have contributed for locatization in vehicular networks by illustrating how we can apply information fusion in different fases of the location estimation process [Boukerche et al. 2008e].

3.2. Data Reduction Algorithms

The data explosion problem caused by several redundant nodes leads to energy waste, and information fusion has been adopted as a leading solution to reduce the effects of such a problem [Nakamura et al. 2007b]. We have adapted data-stream techniques, namely *Data Sketching* [Aquino et al. 2007b] and *Data Sampling* [Aquino et al. 2007c], to process a stream of sensor data to reduce the data volume and preserve a desired data quality in WSNs [Aquino et al. 2007a].

3.3. Data Routing Algorithms

In the data routing segment, we have also contributed to the Multi algorithm [Figueiredo et al. 2004] by applying information fusion methods, namely the moving average filter [Nakamura et al. 2007b], to improve the event-detection estimation model [Figueiredo et al. 2007] that is used to decide which routing strategy, proactive or reactive, should be chosen to enroute data from source nodes towards the sink node.

4. Conclusions

In the thesis here summarized, we have provided an ample perspective of information fusion in WSNs by surveying the available state-of-the-art. This survey allows us to understand how information fusion has been used in WSN, and how it can still be used to address open issues of WSNs. Based on the discussion the survey presents, we have proposed a general framework, called Diffuse, to assist in the design of information fusion tasks in WSNs. Thus, we do understand that the Diffuse framework is a key contribution of this thesis. Another contribution of this thesis is the mechanism we propose to make tree-based routing algorithms more reliable by reconstructing the routing topology only when critical failures occur. In some cases, the traffic generated by this mechanism, which results from the use of the Diffuse framework, is 85% smaller than the traffic generated by the Periodic Rebuilding, which is one of the current solutions adopted to rebuild the routing topology. The final key contribution is the InFRA algorithm. The InFRA algorithm saves energy by searching for the shortest paths that maximizes data aggregation, and by starting the assignment process only when an event is detected, therefore, saving energy during the periods of inactivity. In some cases, the InFRA algorithm uses only 70% of the energy spent by a reactive shortest-path tree.

5. Outlook

Regarding the Diffuse framework, we believe it gives us the opportunity for several other applications. For instance, we could use Diffuse to apply information fusion methods and reduce the errors of the current methods for estimating the location of the nodes in a WSN. Possibly, the field of WSNs more suitable for using the Diffuse framework is actually the design of applications, especially in event-based scenarios. The reason is that filters, feature maps, and inference methods are natural candidates for improving data accuracy and making decisions and inferences about the environment being perceived. A very promising extension of this work is the integrated use of Diffuse to provide a reliable and energy-efficient routing algorithm for a network executing an information-fusion application that also uses the Diffuse framework to detect events and make inferences about the environment. For instance, an environmental application to monitor and track animals in process of extinction.

References

[Akyildiz et al. 2002] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cyirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks*, 38(4):393–422.

[Aquino et al. 2007a] Aquino, A. L. L., Figueiredo, C., Nakamura, E. F., Buriol, L., Loureiro, A., Fernandes, A., and Coelho, C. (2007a). On the use data reduction algorithms for real-time wireless sensor networks. In *IEEE ISCC'07*, pages 583–588, Aveiro, Portugal.

[Aquino et al. 2007b] Aquino, A. L. L., Figueiredo, C. M. S., Nakamura, E. F., Buriol, L., Loureiro, A. A. F., Fernandes, A. O., and Coelho, C. J. N. (2007b). Data stream based algorithms for wireless sensor networks. In *IEEE AINA'07*, pages 869–876, Niagara Falls, Canada.

[Aquino et al. 2007c] Aquino, A. L. L., Figueiredo, C. M. S., Nakamura, E. F., Buriol, L., Loureiro, A. A. F., Fernandes, A. O., and Coelho, C. J. N. (2007c). A sampling data stream algorithm for wireless sensor networks. In *IEEE ICC'07*, pages 3207–3212, Glasgow, UK.

[Boukerche et al. 2008a] Boukerche, A., Nakamura, E. F., and Loureiro, A. A. F. (2008a). Algorithms for mobile and sensor networks: Present and future. In Boukerche, A., editor, *Algorithms and Protocols for Wireless Sensor Networks*. John Wiley & Sons, Inc. (Accepted).

[Boukerche et al. 2007] Boukerche, A., Oliveira, H. A. B. F., Nakamura, E. F., and Loureiro, A. A. F. (2007). Localization systems for wireless sensor networks. *IEEE Wireless Communications*, 14(6):6–12.

[Boukerche et al. 2008b] Boukerche, A., Oliveira, H. A. B. F., Nakamura, E. F., and Loureiro, A. A. F. (2008b). An efficient directed localization recursion protocol for wireless sensor networks. *IEEE Transactions on Computers*. (Accepted).

[Boukerche et al. 2008c] Boukerche, A., Oliveira, H. A. B. F., Nakamura, E. F., and Loureiro, A. A. F. (2008c). An enchanced lightweight algorithm for time-space localization in wireless sensor networks. *Performance Evaluation*. (Accepted).

[Boukerche et al. 2008d] Boukerche, A., Oliveira, H. A. B. F., Nakamura, E. F., and Loureiro, A. A. F. (2008d). Secure localization algorithms and protocols for wireless sensor networks. *IEEE Communications Magazine*. (Accepted).

[Boukerche et al. 2008e] Boukerche, A., Oliveira, H. A. B. F., Nakamura, E. F., and Loureiro, A. A. F. (2008e). Vehicular ad-hoc networks – a new challenge for localization. *Computer Communications*. (Accepted).

[Figueiredo et al. 2004] Figueiredo, C. M. S., Nakamura, E. F., and Loureiro, A. A. F. (2004). A hybrid adaptive dissemination protocol for wireless sensor networks. In *ALGOSENSORS'04*, volume 3121 of *Lecture Notes in Computer Science*, pages 171–186, Turku, Finland.

[Figueiredo et al. 2007] Figueiredo, C. M. S., Nakamura, E. F., and Loureiro, A. A. F. (2007). An event-detection estimation model for hybrid adaptive routing in wireless sensor networks. In *IEEE ICC'07*, pages 3887–3894, Glasgow, UK.

[Hwang et al. 1992] Hwang, F., Richards, D., and Winter, P. (1992). *The Steiner Problem*. Annals of Discrete Mathematics. North-Holland, Amsterdam, Holland.

[Nakamura 2007] Nakamura, E. F. (2007). *Information Fusion in Wireless Sensor Networks*. PhD thesis, Federal University of Minas Gerais, Belo Horizonte, Brazil. Available at: `http://www.dcc.ufmg.br/pos/cursos/defesas/46D.PDF`.

[Nakamura et al. 2004] Nakamura, E. F., Figueiredo, C. M. S., and Loureiro, A. A. F. (2004). Disseminação adaptativa de dados em redes de sensores sem fio auto-organizáveis. In *SBRC'04*, pages 29–42, Gramado, Brazil.

[Nakamura et al. 2005a] Nakamura, E. F., Figueiredo, C. M. S., and Loureiro, A. A. F. (2005a). Detecção de falhas em redes de sensores sem fio baseada na medição do tráfego e em técnicas de fusão de dados. In *SBRC'05*, pages 579–592, Fortaleza, Brazil.

[Nakamura et al. 2005b] Nakamura, E. F., Figueiredo, C. M. S., and Loureiro, A. A. F. (2005b). Information fusion algorithms for wireless sensor networks. In Boukerche, A., editor, *Handbook of Algorithms for Wireless Networks and Mobile Computing*. Chapman & Hall/CRC.

[Nakamura et al. 2005c] Nakamura, E. F., Figueiredo, C. M. S., and Loureiro, A. A. F. (2005c). Information fusion for data dissemination in self-organizing wireless sensor networks. In *IARIA ICN'05*, volume 3420 of *Lecture Notes in Computer Science*, pages 585–593, Reunion Island.

[Nakamura et al. 2007a] Nakamura, E. F., Figueiredo, C. M. S., Nakamura, F. G., and Loureiro, A. A. F. (2007a). Diffuse: A topology building engine for wireless sensor networks. *Signal Processing*, 87(12):2991–3009.

[Nakamura et al. 2007b] Nakamura, E. F., Loureiro, A. A. F., and Frery, A. C. (2007b). Information fusion for wireless sensor networks: Methods, models, and classifications. *ACM Computing Surveys*, 39(3):A9/1–A9/55.

[Nakamura et al. 2005d] Nakamura, E. F., Nakamura, F. G., Figueiredo, C. M. S., and Loureiro, A. A. F. (2005d). Using information fusion to assist data dissemination in wireless sensor networks. *Telecommunication Systems*, 30(1-3):237–254.

[Nakamura et al. 2008] Nakamura, E. F., Oliveira, H. A. B. F., and Loureiro, A. A. F. (2008). A reactive role assignment for data routing in event-based wireless sensor networks. *Computer Networks*. (Accepted).

[Nakamura et al. 2006a] Nakamura, E. F., Oliveira, H. A. B. F., Pontello, L. F., and Loureiro, A. A. F. (2006a). Atribuição dinâmica de papéis para agregação de dados em redes de sensores sem fio. In *SBRC'06*, pages 779–794, Curitiba, Brazil.

[Nakamura et al. 2006b] Nakamura, E. F., Oliveira, H. A. B. F., Pontello, L. F., and Loureiro, A. A. F. (2006b). On demand role assignment for event-detection in sensor networks. In *IEEE ISCC'06*, pages 941–947, Sardenha, Italy.

[Oliveira et al. 2005] Oliveira, H. A. B. F., Nakamura, E. F., Loureiro, A. A. F., and Boukerche, A. (2005). Error analysis of localization systems in sensor networks. In *ACM GIS'05*, pages 71–78, Bremen, Germany.

[Oliveira et al. 2007] Oliveira, H. A. B. F., Nakamura, E. F., Loureiro, A. A. F., and Boukerche, A. (2007). Towards an integrated solution for node localization and data routing in sensor networks. In *IEEE ISCC'07*, pages 449–454, Aveiro, Portugal.