Workshops der
Wissenschaftlichen Konferenz
Kommunikation in Verteilten Systemen 2009
(WowKiVS 2009)

Identifying the Challenges in Reducing Latency in GSN using Predictors

Andreas Benzing, Klaus Herrmann, Boris Koldehofe and Kurt Rothermel

8 pages

Identifying the Challenges in Reducing Latency in GSN using Predictors

Andreas Benzing, Klaus Herrmann, Boris Koldehofe and Kurt Rothermel

$\langle firstname.lastname \rangle$@ipvs.uni-stuttgart.de
Institut für Parallele und Verteilte Systeme
Abteilung Verteilte Systeme
Universität Stuttgart, Germany

Abstract: Simulations based on real-time data continuously gathered from sensor networks all over the world have received growing attention due to the increasing availability of measured data. Furthermore, predictive techniques have been employed in the realm of such networks to reduce communication for energy-efficiency. However, research has focused on the high amounts of data transferred rather than latency requirements posed by the applications. We propose using predictors to supply data with low latency as required for accurate simulations. This paper investigates requirements for a successful combination of these concepts and discusses challenges that arise.

Keywords: Global Sensor Networks, Wireless Sensor Networks, Predictors

1 Introduction

Simulation technology is applied in a broad range of applications to reduce cost or because real experiments require too much resources to be built at all. Several applications also employ such technology to estimate the future state of the physical world. Examples include weather forecasting, predicting natural disasters such as earthquakes and volcanic eruptions, and minimizing the number and size of warehouses in supply chain management by estimating the time of arrival of deliveries for just-in-time assembly.

Up to now, data for these systems has been acquired using isolated sets of wired sensors which are all read out directly by the application. However, deploying sensors and data links separately for each application involves high cost and effort. Proprietary solutions also prohibit re-usage of existing sensor networks or require expensive adaptation to multiple special interfaces. Finally, this approach also requires transferring all measurements to the application for further processing. It does not allow for data processing and reduction close to the sensor nodes.

In this paper we present preliminary research in our effort building a Global Sensor Grid Middleware (GSGM). It abstracts away from proprietary formats and enables access to preexisting sensors and the growing number of wireless sensor networks in the field. In particular, we focus on how GSGM can benefit from the concept of predictors, which has been initially proposed to reduce communication and therefore energy consumption in wireless sensor networks [GI01], in order to allow for low latencies.

Each predictor consists of two parts which independently estimate the next measurement from previously observed value patterns. The actual sensor reading is compared to the prediction at
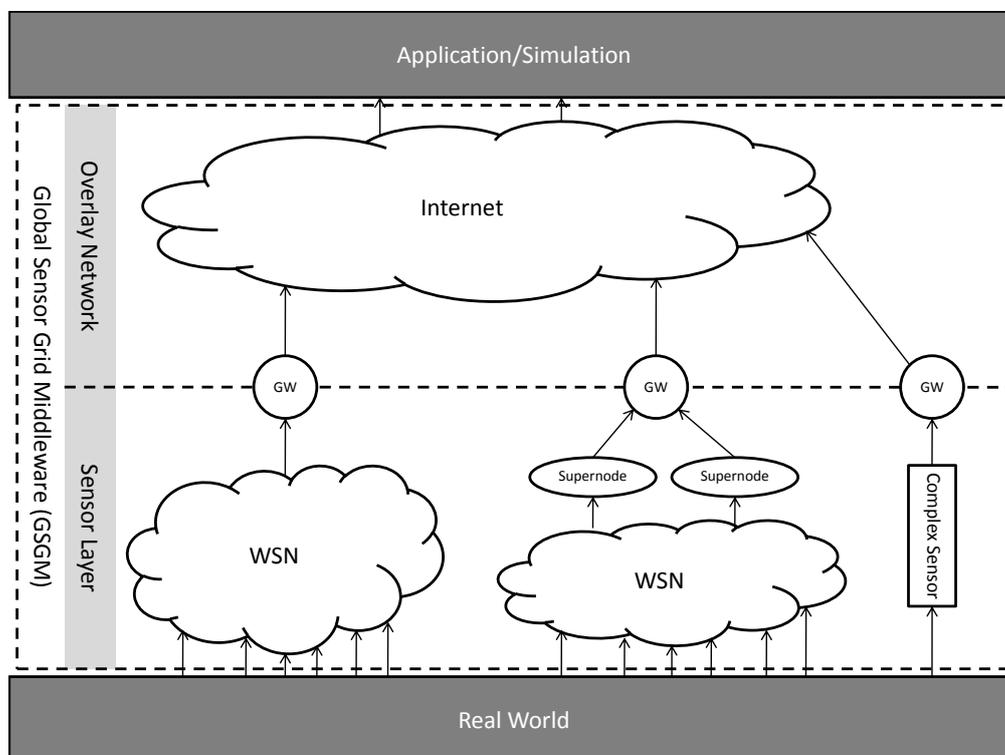
Figure 1: The main components of the Global Sensor Grid Middleware

the source and if it deviates more than a certain threshold an update is sent to the corresponding sink keeping the two parts in sync. Depending on the threshold, the number of update messages can be greatly reduced which directly results in reduced energy consumption. On the other hand, the predictor at the sink can always instantly provide a current value at the expense of inaccurate data. To exploit this capability for GSGM, predictors have to be established throughout the whole network and provide the application with estimations in real-time.

The rest of this paper is structured as follows. In Section 2, we introduce GSGM and present the concept of predictors and a classification. Section 3 gives an overview about related work. Research challenges are presented in Section 4 and finally, Section 5 concludes.

2 The Concept of Predictors in GSGM

GSGM aims to integrate pre-existing and upcoming sensor data sources all over the world to provide a single interface to arbitrary sensor information. Applications like real-world simulations can then monitor the current state of the physical world by posing continuous queries for any type of sensor information from any area to this interface. To distribute the load on the sensor data sources, the application nodes will be integrated into the Global Sensor Grid according to the peer-to-peer paradigm. In addition, applications can specify their required quality of service and data parameters like latency or precision to which the system then adapts accordingly.

2.1 System Model

Figure 1 shows the main components of the GSGM. The whole system is divided into two major parts. The top part contains the overlay network consisting of resource-rich nodes equipped with high-bandwidth internet connections. The lower part contains the sensor layer where homogeneous, hierarchical and other (wireless) sensor networks (WSN) as well as single complex sensors are located. These devices have severe restrictions on computational power and energy resources that depend on the actual hardware. These two layers are connected by gateways (GW) that simultaneously act as nodes in the overlay network and as data sinks for the WSN.

Applications can pose their queries to any node in the overlay network which is initially formed only by the gateways. Nodes that establish continuous queries are integrated into the overlay network to release the gateways from computational and network load. The query is split up for and routed to the corresponding gateways which deploy it inside the targeted WSNs, thereby building a distributed query tree. This previously distributed query tree is used to route the emerging streams of sensor data back to the application.

2.2 Classification of Predictors

To illustrate the application of predictors in this system, we first identify classes of predictors and give an introduction to their respective mode of operation. The main distinguishing feature is the type of information which is queried, namely sensor values about the environment and object locations. Sensor value predictors can be further categorized depending on the type of correlations on which they are based. We therefore identify three classes of predictors for sensor values: **temporal**, **spatial**, and **spatio-temporal**. To support queries for object positions efficiently in GSGM, **location predictors** form a fourth class. The following paragraphs point out characteristic properties of each class and present clarifying examples.

Pure **temporal predictors** only involve a single sensor and estimate future measurements based on previous observations of this sensor. The prediction function that calculates the expected next measurement on both parts of the predictor can therefore be generated directly on the sensor nodes using simple adaptive algorithms like Least-Mean-Square. It can also be built up using greater computational effort on the gateways and then be pushed to the respective node.

In contrast, **spatial predictors** are used to reduce the number of sensor nodes that have to be queried by exploiting spatial correlation between neighboring sensors. If, for example, 45 out of 50 sensors in a narrow area report a temperature between 16°C and 18°C, it is very likely the missing five measurements are in the same range. Generating a predictor of this class requires readings from a whole area to build a model of the territory observed. Since this data is not available on a single node, no prediction of this type can be run inside a WSN and therefore cannot contribute to reducing data. However, using this model, spatial predictors can instantly provide values for uncovered areas by evaluating the measurements of surrounding sensor nodes and still provide the desired low latency.

To capture streams, e.g. an airflow taking warm air from one region to another, value patterns have to be monitored for spatio-temporal correlations. Given the route of the stream and the evolution of measurements over time in the originating region, value predictions for the destination area of the stream can be derived and pushed towards the respective sensors. These **spatio-**

temporal predictors require a large amount of data from past measurements of a number of sensors and also high computational effort to identify relevant streams.

Unlike physical measurements, object locations are represented by complex positional information. Since predicting such data is costly, current approaches only estimate time windows in which an actual sensor needs to be queried to keep track of an object. If speed and direction can be measured, single sensors suffice to generate a **location predictor**, but in most cases this task also requires a more global view of the object's recent locations. In addition, gathering this information from multiple sensor nodes allows to detect more complex movement patterns than linear continuation.

3 Related Work

Current work on predictors focuses on reducing communication inside WSNs to preserve energy resources as much as possible. Since this can be achieved most efficiently with predictions on single sensor nodes, much work has been done on temporal predictors. Symmetric approaches [SR06], where predictors on data source and sink perform basically the same computations, have been proposed. PRESTO [DGL+05], which also considers latency-reducing properties, follows an asymmetric strategy where predicting functions are generated on the predictor at the data sink and pushed towards sensor nodes.

To model spatial correlations, BBQ [DGM+04] uses time-varying multivariate Gaussians. Using this probabilistic model over the observed area, only those sensors are selectively queried which most improve reliance until the target confidence, required by the application, is met. A similar idea has also been proposed in [LCLC04].

In [GI01] each measuring point is treated like a pixel of an image. Time series of measurements can then be considered as a video and prediction techniques used in current video compression algorithms can be applied to model spatio-temporal correlations.

For object tracking sensor networks, location predictors [XWL04] have been proposed to minimize the number of nodes that need to be in active sensing state. More recent approaches [GQP06, TLH08] propose exploiting regular movement patterns, which, in case of simulations, could be supplied by the application.

Although PRESTO also considers the problem of high latencies for queries in WSNs, to our knowledge no further research has been conducted to optimize predictors for latency or use this approach outside WSNs. By applying the idea of predictors from WSNs to all parts of the GSGM and optimizing it for latency using the available resources in the overlay network, we aim to provide applications with real-time sensor information.

4 Research Challenges

Exploiting the potential of predictors for reducing latency by combining the different classes poses several challenges concerning the structure of predictors and the GSGM. This section introduces the main challenges that arise from integrating predictors in the distributed stream processing overlay network of GSGM. The placement problem of predictors is discussed in detail and requirements for the interaction with applications are presented.

4.1 Stream Processing with Predictors

In the stream processing overlay network of GSGM window operators materialize the stream for further processing. Thereby, the window size is given as the number of samples or as a time frame. While a time frame is not affected by predictors, the number of samples the predictor gets does not contain any information about the number actually taken at the sensor any more. Since the sampling of sensor data is only used due to digital processing and packet based networking, we will convert sample-based windows to time frames. The required information about the sensor sampling rate, which presents an upper bound for predictor and sensor data accuracy, will be implicitly encoded in the predictor model.

As per description, predictors in WSN provide a good mechanism to efficiently monitor continuous observations. However, GSGM also aims to support discrete sensor events, which do not follow a continuous characteristic. Although some applications query for the occurrence rate of such events, which is again continuous, a new class of predictors has to be developed to support event prediction. Our approach is to evaluate future values obtained from temporal predictors to detect the triggering of filter operators, which cause such events, in advance. However, for events from sensors, like the ones that are generated on registration of RFID-enabled pallets, new approaches have to be found.

4.2 Placement of Predictors in the Overlay Network

Since a predictor's sink can instantly reply to queries, it is desirable to move it as close to the application as possible. If its source exists, it should be placed directly on the sensors to reduce the amount of data transmitted. However, a variety of reasons complicate this approach.

First of all, as the classification of predictors shows, all but temporal predictors require more than the local view. Transmitting all the required data to the predictor sink at the application and then synchronize it with its corresponding source on the sensor incurs a high overhead, even if only local correlations are considered.

Due to the vast differences of hard- and software between WSNs and the overlay network, we will not initially focus on predictors that act across the gateways despite the availability of IP-enabled wireless sensor nodes [DAW$^+$08]. This also allows the usage of other established energy-preserving techniques – also not predictor-based ones – that have been especially developed for WSNs. Furthermore, we can easily integrate arbitrary WSNs into the GSGM on a well-defined interface.

While the placement inside a WSN is implicitly given by the type of predictor used, all types can be arbitrarily combined and placed in the resource-rich overlay network. We propose a common interface to predictors to easily enable their application on every single link between stream processing in-network operators in the overlay network of GSGM. This way, different classes of predictors can also be directly connected and a single predictor can be re-used to provide multiple adjacent operators and predictors as well as applications with measurement data.

Instead of single sensor readings, many applications will require aggregated values from a certain geographic area. These should be computed as close to the data sources as possible to minimize bandwidth usage. To maximize the profit of re-using data streams for multiple queries,

operators that compute these partial results have to be placed in the network in a way that allows both early filtering of information and most possible re-usage of data. With the integration of temporal predictors, this problem becomes even more complex because of their two-part nature. As already mentioned in earlier, the optimal positions of the aggregation operator and the predictor are contrary, which at first sight contradicts the placement strategies commonly used for operators in distributed stream processing systems. This problem is solved in GSGM by splitting the predictor if necessary, thereby building a multi-hop temporal predictor. To prevent an increase of latency by the intermediate stopover, the predictor updates will be forwarded unchanged parallel to the recreation of the complete continuous data stream.

4.3 Interacting with the Application

A general problem in GSGM is the discovery of relevant sensor data sources. First of all, this concerns addressing schemas for the application interface, but partial results of existing queries also have to be found inside the stream processing overlay network. Since different applications can specify different bounds on data quality in terms of latency and accuracy, the challenge is no longer only to find the data, but to also provide it with the required quality. To reduce costly adaption in the overlay network by transferring predictors from one node to another, we focus on dynamic predictor models whose accuracy can be adjusted quickly at runtime. Where possible, application nodes in the overlay network are placed in order of decreasing QoS requirements along the data streams to avoid meticulous data overhead.

Another challenge in stream processing systems is, due to the high and possibly bursty data rates, load shedding in temporary overload situations. Since at least temporal predictors directly reduce the amount of data transmitted and other predictors are intended to cope with a reduced input stream, we can provide an intelligent load shedding mechanism, which still supplies the relevant information queried but reduces the accuracy provided. However, if the overload is caused by shortage in computational or memory resources on a node rather than link congestion, the predictors have to switch to a simpler model rather than increased thresholds. In order to pick the right model according to the situation, adaptive model selection [BSB07] will be employed.

Without additional information from the applications using the GSGM, predictors need very conservative thresholds for triggering an update from the data source. To use their full potential, the application can specify information about the maximum tolerated error bounds and staleness. If the application can handle a higher latency, GSGM can provide higher accuracy by requesting an explicit model update for example. Vice versa, if high uncertainty of values is tolerated, the actual prediction might already be sufficient. The challenge in GSGM is to find metrics and algorithms to trade staleness for accuracy and provide an appropriate interface for applications to specify the required quality of service and data. Approaches from the domain of approximate stream processing will be evaluated for this purpose.

Finally, especially for spatial and location predictors, we expect that applications can provide information about spatial correlations and object movement patterns respectively. This information will be integrated into the predictors to enhance the prediction model and therefore data quality. However, if different applications provide conflicting data about correlations or movement patterns, a reliable conflict solution has to be found. Work on such reasoning problems has already been done in other areas whose applicability for GSGM has to be checked.

5 Conclusion

We have provided the motivation for and described the usage of predictors in the Global Sensor Grid Middleware to provide applications with low latency sensor data. Existing predictive techniques have been classified and their integration in the GSGM has been proposed. Basic requirements for predictors, applications and the GSGM for a successful implementation of this approach have been identified. Resulting research challenges have been detected that are subject to future work.

The application of the concept of predictors in the resource-rich overlay network offers new possibilities for the combination of different classes of and more complex predictors. Universal deployment of predictors from sensors to applications is therefore expected to significantly improve data acquisition for close to real-time purposes.

We will implement the presented concepts within GSGM to build a Global Sensor Grid for simulation applications as part of the Stuttgart Research Centre and Cluster of Excellence "Simulation Technology" (SimTech). To investigate requirements on quality of service and data and to evaluate the potential of predictors for the Global Sensor Grid under real conditions, cooperation partners inside the SimTech Cluster will provide their applications and the respective quality requirements.

Bibliography

[BSB07] Y.-A. L. Borgne, S. Santini, G. Bontempi. Adaptive model selection for time series prediction in wireless sensor networks. *Signal Processing* 87(12):3010 – 3020, 2007. Special Section: Information Processing and Data Management in Wireless Sensor Networks.
doi:DOI: 10.1016/j.sigpro.2007.05.015

[DAW+08] M. Durvy, J. Abeillé, P. Wetterwald, C. O'Flynn, B. Leverett, E. Gnoske, M. Vidales, G. Mulligan, N. Tsiftes, N. Finne, A. Dunkels. Making Sensor Networks IPv6 Ready. In *Proceedings of the Sixth ACM Conference on Networked Embedded Sensor Systems (ACM SenSys 2008), poster session*. Raleigh, North Carolina, USA, Nov 2008.
http://www.sics.se/~adam/durvy08making.pdf

[DGL+05] P. Desnoyers, D. Ganesan, H. Li, M. Li, P. Shenoy. PRESTO: a predictive storage architecture for sensor networks. In *HOTOS'05: Proceedings of the 10th conference on Hot Topics in Operating Systems*. Pp. 23–23. USENIX Association, Berkeley, CA, USA, 2005.
http://portal.acm.org/citation.cfm?id=1251146

[DGM+04] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, W. Hong. Model-driven data acquisition in sensor networks. In *VLDB '04: Proceedings of the Thirtieth international conference on Very large data bases*. Pp. 588–599. VLDB Endowment, 2004.
http://portal.acm.org/citation.cfm?id=1316741

[GI01] S. Goel, T. Imielinski. Prediction-based monitoring in sensor networks: taking lessons from MPEG. *SIGCOMM Comput. Commun. Rev.* 31(5):82–98, 2001.
doi:10.1145/1037107.1037117

[GQP06] O. Garcia, A. Quintero, S. Pierre. Profile-based energy minimisation strategy for Object Tracking Wireless Sensor Networks. In *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. WiMob'2006*. Pp. 372–379. June 2006.
doi:10.1109/WIMOB.2006.1696370

[LCLC04] K.-Y. Lam, R. Cheng, B. Liang, J. Chau. Sensor Node Selection for Execution of Continuous Probabilistic Queries in Wireless Sensor Networks. In *VSSN '04: Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*. Pp. 63–71. ACM, New York, NY, USA, 2004.
doi:10.1145/1026799.1026811

[SR06] S. Santini, K. Römer. An adaptive strategy for quality-based data reduction in wireless sensor networks. In *Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS'06)*. Pp. 29–36. 2006.
http://www.vs.inf.ethz.ch/publ/papers/santinis_inss2006.pdf

[TLH08] V. Tseng, K. Lin, M.-H. Hsieh. Energy Efficient Object Tracking in Sensor Networks by Mining Temporal Moving Patterns. In *IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing. SUTC '08*. Pp. 170–176. June 2008.
doi:10.1109/SUTC.2008.27

[XWL04] Y. Xu, J. Winter, W.-C. Lee. Dual prediction-based reporting for object tracking sensor networks. In *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services. MOBIQUITOUS 2004*. Pp. 154–163. August 2004.
doi:10.1109/MOBIQ.2004.1331722