Making Sense of Sensor Networks

data-flow services

Abstract

Sensor networks are novel data routing and processing structures that allow for the emergence of data-centric applications, such as the controlled irrigation of agricultural fields and the continuous monitoring of an elderly person's health. Numerous challenges exist including sensor coverage, power consumption, interoperability, security, integrated display and user bandwidth. In this white paper, we concentrate on the challenges presented in trying to make sense of the collected data. We first introduce the conventional techniques emphasizing the prominent ones: neural networks, Bayesian networks, Kalman filters and Dempster-Shafer methods. We then briefly present our solution, and the steps involved in its development.

1.0 Introduction

Sensor networks are composed of multiple interconnected and distributed sensors that collect information on areas or objects of interest. Sensor nodes (SNODEs) make up each sensor network and consist of three major components: (i) parameter, event and object **sensing**, (ii) data **processing**, and (iii) data **communications**.

Sensor networks became a feasible reality in the mid 1990s, when computing and communications capacities became economical enough at their higher ends of the spectrum. At the outset of the technology, military applications were abundant due to their immediate need for scalable and robust surveillance systems. As with most other technologies developed in the military, SNETs easily migrated into commercial application development earlier this decade (see Figure 1 for the SNET chronology). The new entry coincided with the sudden realization and urgent need for personal and communal security (e.g. anti-civilian actions and threats), and the corresponding organizational restructuring to bring about solutions to these pressing concerns (e.g. Homeland Security Department in the US).

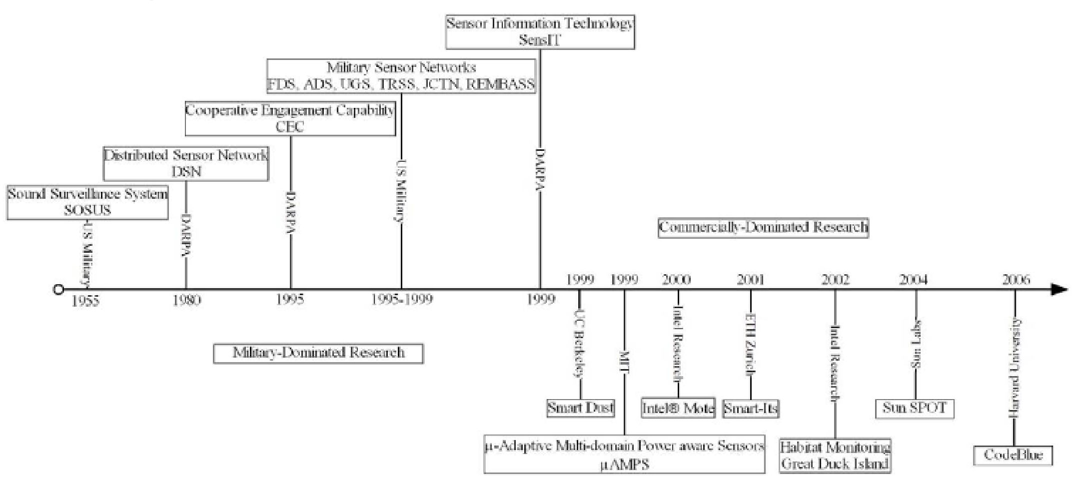

Figure 1. SNET chronology, © 2006-2009 Larus Technologies

There are numerous taxonomies that differentiate and classify SNETs; however, we have identified the five main areas that indiscriminately delineate one SNET from another:

1. Node services: pertaining to the properties of the SNODE. Examples are sensing unit, processing unit, communications unit, power unit, localization, mobility and physical size.
2. Network services: pertaining to the properties of the network. Examples are self-organization, self-discovery, network topology, security and network protocols.
3. Data-flow services: pertaining to the properties of how the data is handled throughout the SNET. Examples are fusion, diffusion, aggregation, dissemination, classification, in-network processing, area monitoring and target tracking.
4. Control-flow services: pertaining to the properties of how the data is controlled throughout the SNET. Examples are storing, tasking and querying.
5. Environment services: pertaining to the properties of the environment that the SNET resides in. Examples are deployment, landscape and survivability.

In this white paper, we will concentrate on the data-flow services, which define part of the client's subscription to the sensor network. We can imagine each of these services as ones that could be subscribed to by a client. For example, a person residing in Ottawa (Canada) would subscribe to the area monitoring data-flow service provided by the Rideau Canal ice formation sensor network, in order to receive a clear indication of the current skating conditions.

2.0 The Problem: Fusion and Diffusion

Fusion has been tackled from many different angles, and for various assorted applications. The underlying goal is the analysis of a refined state estimate for the process under observation. Its reciprocal, diffusion, describes the distribution of entities from a common source. Its underlying goal is the synthesis of a refined command dissemination for the process under control. Note that an entity can be a datum, a packet or even a signal.
In order to complicate matters further, data can be retrieved from multiple, and often dissimilar, sources, and commands can be sent to multiple, and often independent, sources. For that matter, we call the former process data source fusion (DSF) and the latter control source diffusion (CSD).

According to a market survey [1], the world market for non-military sensors will grow at an average annual rate of 4.5% between 1998 and 2008, culminating in a US $50.3 billion market capitalization in 2008. In a more recent [2], and yet unpublished survey, it was found that amongst the world wireless markets that exceeded $5 billion in the most recent surveyed year, three specific ones had cumulative annual growth rates (CAGR) of over 15%. They are mobile entertainment, machine-to-machine (M2M) communications and location-based services. Examples of other surveyed markets include world telecommunications equipment, world mobile services, satellite communications, smart cards and fixed wireless broadband.

Observing a few more market trends:

1. Sensors are getting smaller in size and variable in nature
2. Computing power is getting bigger and is being embedded
3. Communications bandwidth is getting higher, while transceivers are getting smaller

We can notice that the amount and variance of data is becoming quite overwhelming, and we will need to come up with improved methods to deal with this data overload. Furthermore, we will need to extract useful features and properties from the assorted data, without compromising its real-world and real- time nature. For example, attributes such as time, location, error margin and reliability for a particular data source must be maintained for the ad-hoc network.

Sensor networks exhibit two innovative characteristics that are typically not found in the today's networks: **in-network** processing and **data-centric applications**. The former describes a method of processing the data near where it is generated, while the latter describes a technique of controlling the data with respect to its physical properties. The core idea behind in-network processing is to filter erroneous data or previously known information and only propagate anomalies or novel information to the backend servers; this reduces SNET bandwidth and better utilizes the available communications channels. As for data-centric applications, the core idea is that the data being acquired is more important than the location it is coming from; however, knowing the confidence and uncertainty of the data sources becomes paramount. Compared to conventional processing (e.g. client/server model, peer-to-peer model) and address-centric applications (e.g. TCP/IP-based), these new characteristics, themselves byproducts of the network structure and the overall application, need new ways of leveraging their benefits: the old methods simply do not work.

The problem, then, is to allow the decision makers to have access to the right information, at the right time, in order to make the right decision. And the challenge is to integrate this framework onto a distributed and resource-constrained sensor network, while leveraging its unique characteristics.

3.0 Technology Background

There exists numerous methods to deal with multiple data of the same type; however, few methods exist to deal with multiple data of different types. This multi- type multi-source (MTMS) data fusion is quite complicated indeed, typically requiring a complex mathematical algorithm that is computationally expensive. The prevalent single-type multi-source data fusion techniques include Neural Network learning, Bayesian learning, Kalman filtering and Dempster-Shafer evidential reasoning.

Not all of the aforementioned techniques can be applied to heterogeneous (multi-type) data fusion; however, we can describe how they can, on their own or through variants, be applied. Neural networks, for example, can be extended to integrate heterogeneous data, as shown by the authors in [3]. This model allows us to extract very simple and accurate models from the search space in question, even with much of the data missing. Its combination of neuro-fuzzy and genetic algorithm parameters maximizes the chances of discovering good models suitable for describing heterogeneous, incomplete, imprecise and time-dependent data.

Bayesian learning can also be extended to correlate heterogeneous data. Work has been started on learning from single-type sources, with promising results; hence, an extension to MTMS data fusion is the next logical step. The authors in [4] describe a hierarchical Bayesian belief network, called a Hierarchical Temporal Memory (HTM), which attempts to discover causes in the world and infer causes of novel input. HTMs have two optional functions, mainly making predictions and directing behavior. They are typically thought of as special Bayesian networks, but with some significant additions to handle time, self-training, and the discovery of causes.

The remaining two techniques provide methods of refining state estimates. Kalman filtering has been previously used as a data fusion technique; however, to utilize such a technique for sensor networks, it cannot be centralized. To that end, the authors in [5] recently described a distributed Kalman filter (DKF) consisting of two separate dynamic consensus problems. The idea is to divide the central KF into micro-Kalman filters, which are collectively capable of providing an estimate of the state of the process that is identical to the estimate obtained by a central Kalman filter.

Finally, Dempster-Shafer (D-S) [6][7] theory extended Bayesian beliefs to allow for the explicit representation of uncertainty. In certain situations, a classification algorithm cannot classify a target or cannot exhaustively list all the classes it belongs to; hence, degrees of belief are collected from previous predictions to merge multiple pieces of information. There are solutions to overcome the problems found in D-S classifications, such as non-conflicting outputs resulting in counterintuitive decisions [8]; and the method is commonly used when a set of alternatives exists that may not have been previously classified as a possible state. Four major problems exist with the prevalent techniques. Firstly, they are not natively meant to handle multi-type data. Secondly, they are not meant to be physically and computationally distributed. Thirdly, they do not inherently resolve both fusion and diffusion. Fourthly, they do not inherently fuse real-time data. According to the Joint Director of Labs (JDL) Data Fusion Group, data fusion is defined as [9]:

> [..] a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates for observed entities, and to achieve complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and by evaluation of the need for additional sources, or modification of the process itself, to achieve improved results.

The JDL model has been taken as a lingua franca for data fusion problems. It has been revised twice, once in March 1999 [10] and another in December 2004 [11]. Other fusion models exist, including the Data Fusion Information Group (DFIG) model [12], DDF model [13], the Omnibus model [14] and the perceptual reasoning model [15]. We will not attempt to redefine the term; however, it is important to outline the difference between a data fusion model and a data fusion technique. The former is a successively refined process by which low-level data is presented to the intended user, whilst the latter is a well-defined method of correlating data, at any level, for the purpose of state refinement. Finally, let us note that fusion is a data-flow service provided by a sensor network. There are other services that could be provided by such a network, such as target tracking, classification/identification, aggregation, dissemination, position/state estimation, and so on. We will not be presenting the details of these services in this white paper.

4.0 The Larus' Solution

Larus Technologies is working on sensor networking solutions, for the real-time association of multi-data and multi-control sources, to extract information about the time and location of environment events, and to integrate the resulting intelligence products into immersive models. Shown in Figure 2 is the overview of the proposed end-to-end solution.

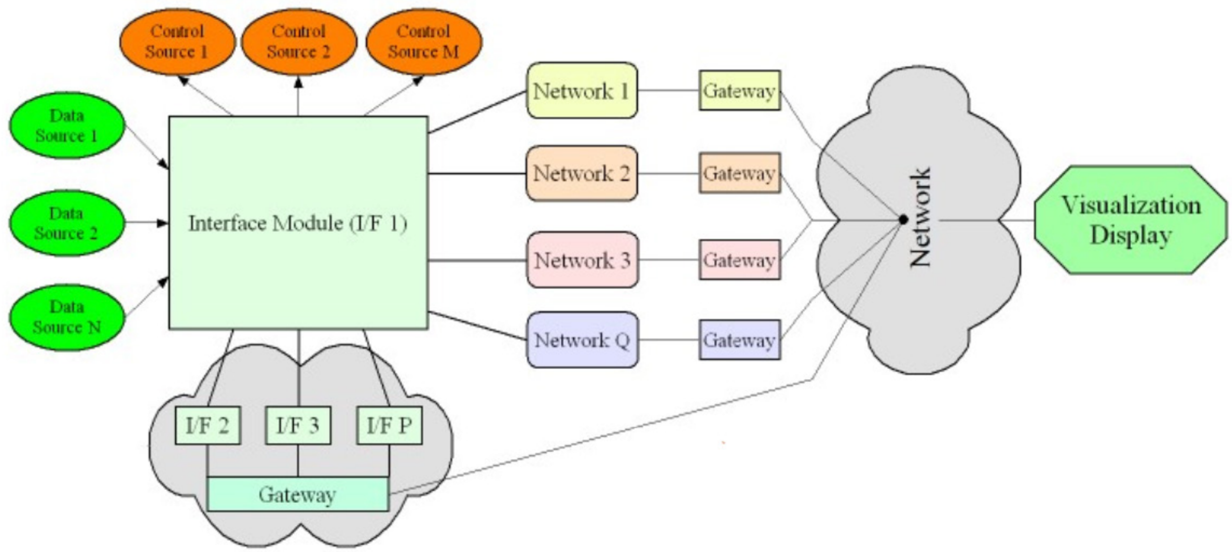

Figure 2. End-to-end system solution, © 2006-2009 Larus Technologies

Larus Technologies has developed a method to improve sensor coverage and consistency of new information at each SNODE, resulting in enhanced overall decision accuracy at each node. Our solution allows for the real-time monitoring of an environment by integrating low intelligence nodes together in order to provide a high intelligence collective.

When planning the deployment of SNETs, it is imperative to evaluate various configurations that will maximize the productivity of the overall network. Larus Technologies has developed a virtual prototyping tool to assist in optimizing the SNET configuration.

Larus Technologies has developed a novel architecture that paves the way for the realization of intelligent systems. The architecture is directly mapped onto the SNET, and performs pattern learning, storage, recall and prediction. The advantages of the proposed predictive modeling approach, compared to conventional approaches, include faster and more accurate event predictions, more realistic heterogeneous sensor fusion estimates, and a more concrete grounding in state-of-the-art neurocomputing research. The association method and its computational architecture are capable of dealing with MTMS data, within a real- time sensor network framework, and are also symmetrical in their ability to fuse data (DSF) and diffuse commands (CSD). This provides a methodology that does not suffer from the aforementioned four major problems that impede the utilization of the prevalent techniques in these types of applications.

Currently, SNET information presents the environment to the operator in a raw format with minimal visual cues. Larus Technologies has developed an immersive virtualized reality model of the SNET environment. The virtualized reality model is an interactive representation of the real world; however, it is constantly updated to reflect the real-time activity as perceived by the deployed sensors.

Larus Technologies helps you make sense of sensor networks.

5.0 Sensor Network Applications

A sensor network is exactly that: a network of sensors. The devices themselves possess a constrained supply of resources, be it in the form of energy, memory, computational power and communication bandwidth. The SNODEs are also small and inexpensive to manufacture, so as to allow for high distribution and turnover rates. The intention is that a large number of sensor nodes working together, in a coordinated manner, form a network that can be represented as a single data source to higher-level processing levels. For example, augmenting the reception of hundreds of individual temperature measurements in a biodome, one would also receive a fused information stream describing the dynamic weather patterns in the environment, and possible scenarios of how to mitigate dangerous crop weather conditions. In summary, sensor networks provide flexibility, fault-tolerance, high sensing fidelity, low cost and rapid deployment.

Henceforth, sensor networks can be applied to a myriad of areas: security (e.g. threat tracking), health (e.g. vital sign monitoring), environment (e.g. natural habitat analysis), home (e.g. motion detection), manufacturing (e.g. assembly line fault-detection), entertainment (e.g. virtual gaming) and the digital lifestyle (e.g. parking spot tracking). Concentrating on security-related applications, we can identify sample end-user applications such as strategic area surveillance, path prediction, target detection, classification and tracking, integrated views and state estimation.

6.0 Conclusion

Without the right fusion method, one will not be able to properly and effectively make sense of the collected data gathered from their sensor network. This problem directly and negatively affects the subsequent data flow tasks, such as downstream processing and dissemination, by weakening the quality of the resulting intelligence products. It also limits the potential actionable intelligence and adversely affects both the reaction time and efficiency of a decision maker. **Larus Technologies has developed solutions to make sense of the overwhelming data collected by the sensor network by allowing one to seamlessly handle the collected data for a more uniform and integrated operating environment.**

References

[1] Norbert Schroder, “Sensor Markets 2008: Worldwide Analyses and Forecasts for the Sensor Markets until 2008”, Intechno Consulting, 1999.

[2] D.R. Senik and Associates Inc., “ICTC Wireless Technology Roadmap”, November, 2006.

[3] J. Valdes, “Similarity-Based Heterogeneous Neurons in the Context of General Observational Models,” Neural Network World, vol. 12, no. 5, pp. 499-508, November 2002.

[4] Hawkins, J. and George, D. Hierarchical Temporal Memory. Concepts, Theory, and Terminology. Numenta Inc, 2006.

[5] Reza Olfati-Saber, “Distributed Kalman Filtering and Sensor Fusion in Sensor Networks”, 2005.

[6] A.P. Dempster, “A Generalisation of Bayesian Inference,” J. Royal Statistical Soc., vol. 30, pp. 205-247, 1968.

[7] G. Shafer, A Mathematical Theory of Evidence. Princeton Univ. Press, 1976.

[8] N.-H. Yu and Y. Yin, “Multiple Level Parallel Decision Fusion Model with Distributed Sensors Based on Dempster-Shafer Evidence Theory,” Proc. Second Int’l Conf. Machine Learning and Cybernetics, Nov. 2003.

[9] JDL Data Fusion Group

[10] Alan Steinberg, Christopher Bowman and Franklin White, “Revisions to the JDL Data Fusion Model,” March 1999.

[11] James Llinas, Christopher Bowman, Galina Rogova, Alan Steinberg, Ed Waltz and Franklin White, “Revisiting the JDL Data Fusion Model II,” December 2004.

[12] E. Blasch, I. Kadar, J. Salerno, M. Kokar, S. Das, G. Powell, D. Corkill and E. Ruspini, “Issues and Challenges in Situation Assessment (Level 2 Fusion),” Journal of Advanced Information Fusion, vol. 1, no. 2, pp. 122-139, December 2006.

[13] B. Dasarathy, “Decision Fusion Strategies in Multisensor Environments “, IEEE Transactions on Systems, Man, and Cybernetics, pp 1140-1154, vol. 21, 1991.

[14] M. Bedworth and J. Obrien, “The Omnibus Model: A New model of data fusion?”, AES Magazine, April 2000.

[15] Kadar, I, “Perceptual Reasoning in Adaptive Fusion Processing,” SPIE, 2002.

About Larus

Through our culture of innovation and research, Larus Technologies has developed the next generation of embedded technology for developers of mission-critical C4ISR Systems and Security Systems.

With a solid foundation pioneering high level information fusion (HLIF) for the ever-changing defense and security industries, Larus is perfectly positioned to help Original Equipment Manufacturers (OEMs) make a world of difference. Working at the higher levels of the US Department of Defense's Joint Director of Laboratories (JDL) information fusion model, our technology not only delivers more knowledge, its adaptive learning algorithms deliver more accurate and more predictive information—faster.

170 Laurier Ave West, Suite 310
Ottawa, Ontario K1P 5V5
Canada

Phone +1 (613) 244-8916
Fax +1 (613) 244-9941

www.larus.com