# List of Amendments in Addressing Prof. Giusoppe Anastasi’s Comments

## Chapter 2

* A number of examples are presented in this chapter to illustrate the proposed event definition / evaluation language, e.g. GEM and JECTOR. Most of the examples are not explained, at least not satisfactorily. While I understand that this chapter is mainly intended at providing a review of the previous works, nevertheless the lack of appropriate explanations makes examples difficult to understand. Hence, they become pointless (and could be even removed):

More explanations have been added to these sections. In particular, for the section on event definition, yeast is studied because of its distinct feature in expressing temporal relations while those works in active database provide a generic model in expressing complicated relations among events.

For the section on event evaluation, works like EPS and SIENA are related but they have not addressed the issues facing WSN. GEM is more related because it provides a more generic rule-based framework. It inspired the design of the event detection framework for PSWare.

For the section on event operators, I’ve added a short discussion in the beginning of the section. The purpose for this part is to summarize the existing event operators / functions so that we can have an idea on what types of operators / functions PSWare should support.

* Similar remarks apply to examples in Figure 2.10 and Figure 2.11.

Figure 2.10 is for a related work named q-digest. More explanation has been added to explain the figure and its relevance. In essence, q-digest tries to balance the trade-off between accuracy and storage by summarizing some part of the data. Such trade-off is actually an important concern in WSN when resource constraint is considered.

Figure 2.11 is another related work similar to q-digest. More explanation has been added. This work is actually an improvement over q-digest in the sense that it not only tries to balance the trade-off between accuracy and storage but also tries to balance the weights among different nodes in the network. As a result, this work is also highly relevant if we consider all sensor nodes in the network as even.

* Page 108: it should be explicitly stated that the message cost is related to energy consumption. Otherwise, the objective of reducing energy consumption is not well justified.

I have added a remark for clarifying that we only consider message cost in our energy consumption.

* Page 108: how is the message cost defined?

A short description has been added: message cost is the number of hops for the event to be delivered from the event source to its destination. Such destination could be an event fusion point (fusion point will be discussed in the later sections) for detecting higher level composite events or the sink node where the events will be delivered.

* Page 112-114: this part is unclear to me. Formulae derived in this section are not adequately justified. This makes understanding almost impossible.

The corresponding paragraphs have been modified to show how the formulae are derived. In particular, cost1 is derived based on law of cosine while cost2 is derived based on law of cosine.

## Chapter 4

* Many formulae are derived in this chapter but, generally, they are not adequately justified in the text. This, in addition to a non-perfect presentation, makes understanding very difficult.

The chapter has been updated to include more description for the derived formulae. In particular, the most critical formula is in the chapter is the one for calculating the average distance between a fusion point and any sensor nodes randomly distributed in a particular area. This is actually a calculus problem for computing the mean distance from vertex to interior of plane figures and the solution to this problem involves some non-trivial calculus techniques. Instead of replicating the very detailed steps for calculation, a reference has been added so that the reader can know a bit more about the problem’s background.

## Chapter 7

* Page 164-165: energy cost are derived without any discussion

To be more precise, message cost should be used instead of energy cost. Message cost is also used in the problem formulation. As a result, the corresponding paragraphs have been modified to address message cost.

* Apparently D is used with two different meanings, when deriving the energy cost (section 7.1.1) and when analyzing the detection delay (section 7.1.2), respectively. If so, please use different identifiers for clarity

Another identifier, K, has replaced D in Section 7.1.2 to represent the average distance between events and sink.

* In Section 7.1, the energy cost of TED and SPF is derived. I agree that the considered metric provides an indication of the cost to pay, also in terms of energy consumption, when using the two different approaches. However, in my opinion, the term “energy cost” is not completely correct as the latter also includes other components that are not considered here (e.g., energy consumed by sensor nodes when idle). More correctly, the cost metric is referred to as “message cost” in Section 7.2 (Simulation). It would be better to use the same definition in both section, for clarity, and I would prefer the second option (i.e., message cost).

Energy cost is a broader term than message cost and in Section 7.1, it is more accurate to use message cost. This is also the case in the problem formulation. Therefore, Section 7.1 has been modified to use message cost consistently.

* Figure 7.2-7.4 are intended to show the influence of event distance on message cost. However, the results are dispersed in three different figures; this does not allow a fair comparison. The results should be re-organized by showing the distance on the X-axis. Actually, the impact of the event distance on message cost is better emphasized in Figure 7.5-7.7, which are originally intended to show the influence of the event size.

There was indeed some mislabeling in the figures. The problems have been corrected so that for each comparison, the X-axis shows the variable being discussed.

* Similar remarks apply to Figure 7.5-7.7. They do not allow appreciating the impact of event size on message cost. As above, they should be re-organized by showing the event size on the X-axis.

Some figures in this part have also been mislabeled. Similar to Figures 7.2-7.4, they are re-organized for clarity.

* How is event size defined? Is it the number of primitive events that concur to generate a composite event?

Event size is defined as the number of nodes participated in detecting a particular event. For composite events, it includes all the nodes participated in detecting their sub-events.

* In Section 7.2 many important simulation details are omitted. What about the accuracy of the presented results (e.g., confidence intervals or std deviation)? How many simulation runs for each experiment? How long is each run?

The details have been added, for each simulation, we run 10,000 times and for each set of the data we obtain, the standard deviation less than 0.5.

* In the same section, the simulation analysis only considers the message cost. The other metrics defined in Section 7.1, i.e., detection delay, is not considered at all. Is there a reason for that? In my opinion, the current analysis is not complete. I strongly encourage the candidate additional simulation experiments and investigate the impact of the considered factors also on the event detection delay.

New simulation has been performed to study the impact on delay. For each figure on message cost, a new simulation has been performed and studied to study the corresponding impact on delay. In summary, the centralized TED will introduce the most delay as expected. SPT will usually has smaller delay than distributed TED but when the event probability is low in the network, distributed TED can still outperform SPT.

* Section 7.3: is the “message cost” considered in this section the “energy cost” defined in Section 7.1? Or is it a different metric?

Section 7.1 has been modified to also use message cost so that the metric is consistent in the whole thesis. The “message cost” in Section 7.3 refers to the same cost metric.

* In Section 7.3 the experimental setup used for different application domains, as well as the experimental results shown, are not discussed in a satisfactory way. This is a very important part of the thesis and, hence, it would deserve more discussion.

This section has been modified to include more discussion. For the car park experiment, more description is added for the experimental setup. For the results, there’s a spike in each figure because of the rush hour and sudden surge of data. For the ITS experiment, details are added to illustrate how the test bed works and which components are used for the experiment. For the indoor monitoring experiment, more details have been added to describe the results - we tried to emulate the events such as fire alarm by putting some sensors on the heater and hence the spike in each figure.