

A Data Analysis Driven Streaming Framework for Body Sensor Area Networks

Ming Li
California State University, Fresno
2576 E. San Ramon
Fresno, CA 93619, USA
+1 559-278-4792
mingli@csufresno.edu

Yu Cao
University of Massachusetts Lowell
198 Riverside St.,
Lowell, MA 01854, USA
+1 978-934-3620
ycao@cs.uml.edu

B. Prabhakaran
The University of Texas at Dallas
800 W. Campbell Road; MS EC31
Richardson, TX 75080, USA
+1 972 883 4680
bprabhakaran@utdallas.edu

ABSTRACT

Streaming and analysis of time series body sensor data have been well investigated recently for various applications, especially on health data monitoring. However, existing strategies work independently from each other. Obviously, lacking appropriate information sharing, feedback, and interaction mechanisms, these strategies, even in combination, do not provide an efficient and effective solution for real time body sensor data collection, transmission, and analysis. In this work, we propose a data analysis driven framework with feedback for efficient streaming of body sensor data. The core idea of this framework is based on a data analysis algorithm specific variance threshold that identifies the data reliability requirement. Then, a reliability index is generated and sent from the data aggregator to sensors as a feedback to guide the streaming protocol. At the sensor side, a data importance ranking and grouping strategy is designed so that samples that affect data analysis most significantly are given higher priority for transmission.

Categories and Subject Descriptors

C.2.2: Network Protocols – protocol architecture.

General Terms

Algorithms, Performance, Reliability.

Keywords

Framework, data analysis, streaming, time series data, body sensor networks.

1. INTRODUCTION

Recently, there has been increasing investigation on a new type of network architecture, generally known as body sensor networks (BSNs [1]) or wireless body area networks (WBANs [2]), attributed to the exciting advances in design of lightweight, small size, ultra-low power, and intelligent monitoring wearable sensors. Compared with existing technologies such as WLANs, BSNs enable the wireless communications inside or around the

vicinity of human body, therefore further extending the desirable concept of pervasive wireless computing into a completely new level. Key applications that may benefit from this new technique include remote health monitoring, sports training, and entertainments.

Basically, these applications exhibit unique characteristics from two interesting aspects:

- *Data aspect*: almost all sensors sample time series data at certain frequencies. For some sensors such as ECG, the sample rate can be quite high. In addition, these data have to be analyzed online at both the data aggregator for early detection and the physician's server for comprehensive diagnosis. In some specific applications such as assessment of rehabilitation, synchronous data from multiple sensors might have to be analyzed for feature extraction and pattern matching.
- *Network aspect*: due to the nature of energy conservation for portability and availability, the bandwidth achieved in body sensor motes are significantly limited compared to other wireless technologies such as WiFi and Bluetooth. This bandwidth is further affected by factors such as body movements. When sensors with various capabilities are present, providing reliable and real-time sensor data collection requires additional consideration.

To address these issues, quality of service (QoS) scheduling [3, 4] for streaming body sensor data have been well investigated. In a similar manner, efficient data analyzing and aggregating techniques have been developed. However, most of these strategies work independent of each other without any proper feedback. For instance, most streaming strategies treat all sensor data samples equally and attempt to improve throughput and reduce delay without basic understanding of the exact data reliability requirements for further analysis. By doing so, the streaming strategies may not provide an efficient and effective solution to support the success of data analysis algorithms.

We argue that the major challenge of streaming time series body sensor data is to satisfy data reliability requirements as specified by the data analysis algorithms. To address this issue, we propose *DADS: Data Analysis Driven Streaming* for body sensor data. This framework consists of two major components at the sensor side: (i) a data importance based grouping scheme to minimize the effect of data loss on the accuracy of data analysis algorithms; (ii) a hybrid transmission protocol that satisfies the basic data reliability requirement (specified by analysis algorithm. At the data aggregator side, a data reliability analysis algorithm that establishes the relationship between data reliability index (in term

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

BODYNETS 2013, September 30-October 02, Boston, United States

Copyright © 2013 ICST 978-1-936968-89-3

DOI 10.4108/icst.bodynets.2013.253914

of the average normalized variance of a received data group) and the accuracy of data analysis algorithms is proposed. This algorithm generates a feedback of a sensor-DA calculated reliability index to guide the data transmission protocol.

To the best of our knowledge, this is one of the first works that provide a framework for efficient and reliable transmission of time series body sensor data. Compared with pure networking solutions that attempt to maximize data reliability by retransmissions without understanding the exact data analysis requirement, the proposed framework saves more bandwidth by meeting the data variance threshold. With this additional bandwidth saving, the framework has a better potential of supporting streaming requirements of multiple sensors, making it more scalable. In addition, the framework allows easy integration with medium access control (MAC) layer strategies to further improve the network performance.

2. RELATED WORKS

Quality of service (QoS) scheduling for streaming body sensor data has been well investigated. Zhou and Lu [5] proposed BodyQoS, a virtual MAC for quality of service scheduling in BSNs. The approach basically measures the effective bandwidth and adaptively allocates remaining resources to meet the QoS requirements of applications. A desirable feature of BodyQoS is that it does not require the modification of the underlying MAC layer implementation. Fulford-Jones and Malan [6] proposed CodeBlue, an ad hoc infrastructure for emergent medical care. In this project, several types of body sensors (e.g., pulse oximeter, ECG/EKG sensor) are individually connected to Zigbee enabled radio transmitters, which communicate with access points directly. Due to the ad hoc architecture and the capability of self-organizing, CodeBlue yields scalability for network expanding and flexibility to connect various wireless devices. Jiang and Cao [7] proposed CareNet, an integrated wireless environment used for remote health care systems. CareNet offers features such as high reliability and performance, scalability, security and integration with web based portal systems. High reliability is achieved by using two-tier architecture. Younis and Akkaya [8] proposed Distributed Queuing Body Area Network, a MAC protocol aiming at providing better QoS. It uses a cross-layer fuzzy rule based scheduling algorithm to optimize MAC layer performance in terms of QoS and energy efficiency. In addition, there have been considerable standardization efforts during recent years. An IEEE 802.15.6 [9] task force is finalizing a wireless body area network (WBAN) standard.

Data analytics for WBANs has been very active recently. Existing data analytics algorithms can be organized into two categories: the traditional threshold-based or light-weight machine learning-based approaches [10, 11], which are usually executed at the data aggregator side in a real-time fashion, and the emerging sophisticated pattern-matching algorithms [12, 13], which run at the remote server in a near real-time or offline fashion. Our proposed data analysis driven stream framework is mainly focusing on the first category but can be extended to coordinate with the analytics algorithms in the second category in the future.

3. SYSTEM MODEL

In body sensor network based applications, the data reliability requirements can be dynamic. For example, in health care monitoring, sensors needed for measuring the detailed human

condition are associated with different events (e.g., human activity and gestures). During the walking activity, the most important measurement is walking speed. In this case, accelerometers are enough. However, to measure the hand motion (e.g., move hand from kneel to head), electromyography and gyroscopes have to be employed in addition to accelerometers. Basically, this dynamic demand for multiple and synchronous sensor data collection needs to be determined either by continuous data analysis or sporadic human expert interruption. For these sensors, high priority should be given with satisfactory reliability for streaming.

As illustrated in Figure 1, in the proposed data analysis driven streaming framework, one or more data aggregators (usually a smartphone running iOS or Android, illustrated at the right side of Figure 1 with font in Bold) run one or more data analysis algorithms for various purposes such as fall detection, walking speed estimation, and heart function diagnosis. For each such algorithm, there is a corresponding data analyzer that performs (i) online event detection to determine if a new human movement has started or if there is a need to sample multiple sensors simultaneously. These sensors will be identified as high priority sensors; (ii) online or offline reliability threshold determination for the specific algorithm. A reliability index (value between 0 and 1) based on the threshold can be calculated to indicate how reliable the received data samples are. Then, the (sensor id, priority, reliability index) tuple is broadcasted to the sensors, which in turn operates based on the information through a lightweight data pre-processor. The functions in the data pre-processor (illustrated at the left side of Figure 1 with font in Bold) include adjusting sample rate based on its priority and monitoring critical activities from a preliminary screening of data samples (thereby setting the sensor as high priority if necessary). Each sensor determines its own actions in a distributed manner. In addition, data pre-processor also ranks the importance of each data sample, groups samples of the same importance, queues packets according to importance, transmits data groups through either reliable or unreliable channels, and performs QoS scheduling to make sure the reliability requirements of requested sensor data are satisfied. Note that there can be more than one data aggregators with each communicating a subset of body sensors in a practical scenario. However, since all communications share the same channel, the issue for channel contention is similar to the case with only one aggregator.

3.1 Rationale of the Proposed Approach

As we have argued earlier, the major reason for the inefficiency and ineffectiveness of a pure network measurement (such as delivery ratio) based streaming approach is the lack of understanding of data analysis requirement. From the networking aspect, all data are equally important. Therefore, a well performing protocol should try to deliver as many data reliably as possible. However, for data analysis, not all data samples are required (although desired) in order to have accurate event detection/recognition. If certain samples are lost due to link error or network congestion, interpolation can be applied to obtain an estimate. Usually, the bigger the gap, the more inaccurate the estimate will be. As long as the difference between the estimate and the original values do not differ significantly, it is sufficient to claim that the data reliability for analysis algorithms has been achieved. With this understanding, we propose to identify the importance of data samples, group them for transmission with a specific reliability requirement, and then queue them for minimum effect on the accuracy of data analysis algorithms.

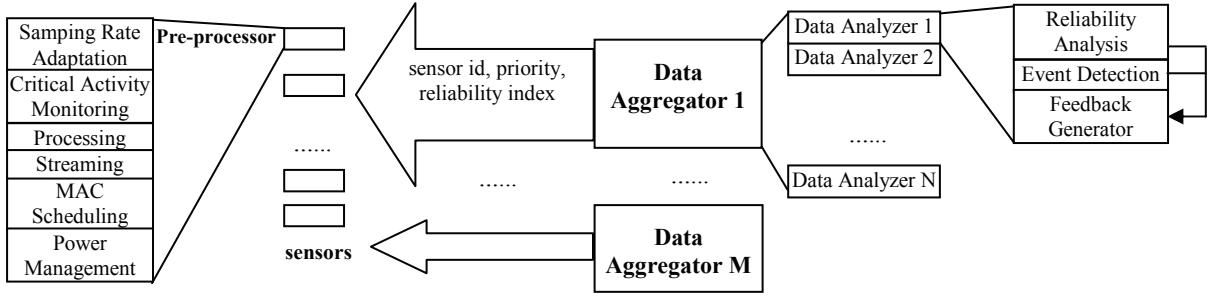


Figure 1. System model for reliable body sensor data streaming

3.2 Data Reliability Modeling

In DADS, how to measure data reliability is critical for design and evaluation of the proposed strategies. For a given group of data samples, we define the *data reliability index (RI)* as the ratio of the average normalized variance (ANV) threshold over averaged normalized variance of the received data samples. We also define the variance of a sample i as the different between the actual value and the estimated value from higher importance data samples, i.e., $\text{variance}(i) = d_i - d_{i,est}$. Then we have that

$$RI = \frac{\text{average normalized variance threshold}}{\sqrt{\sum_i \left(1 - \frac{d_{i,est}}{d_i}\right)^2 / N}} \quad (1)$$

Where average normalized variance threshold (ANV threshold) is determined by the data analysis algorithms. Obviously, if $RI < 1$, then the actual variance of a received group is greater than the threshold on average. Otherwise, the data reliability is satisfactory. Based on the RI , we can generate feedback information from the data aggregator to guide the transmission protocol at the sensor side.

4. DESCRIPTION OF THE FRAMEWORK

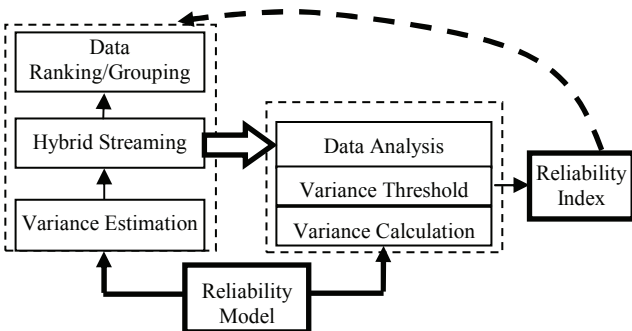


Figure 2. Components of the proposed framework

Figure 2 depicts the components of the proposed framework. At the center of the framework is the reliability model (introduced in section 3.3) and the reliability index based feedback. Since the sensors have the original samples, ANV has to be pre-computed at the sensor side and sent to the receiver (DA) side, which will have to recover the values of missing samples through estimation. In addition, since data are sent through packets containing one or more data sample groups, data are lost in groups and ANV calculation needs to be performed for each such group. These variance values are sent to the data aggregator through a special

packet over reliable channel to ensure its receipt. At the DA side, ANV threshold and computed actual ANV can be used to obtain reliability index in (1). At the sensor side, data ranking and grouping are performed to prepare transmission of sample groups with reliable or unreliable channels. Such decision is also adjusted based on the received feedback message.

4.1 Data Importance Ranking and Grouping

Basically, when sensor data samples are delayed (or lost) during transmission, some prediction technique may be used to estimate the missing data samples based on the ones already received. When such prediction-based recovery of missing sensor data samples is allowed to be used in data analytics algorithm, it is necessary to identify samples that (i) help predict other samples with more accurate approximation, i.e., have differential priorities for data samples, or (ii) cannot be predicted with good accuracy if lost. Obviously, further understanding of the relationships among data sample, especially adjacent samples, needs to be obtained for this purpose. Differentiation of samples with various importance level is also critical. Afterwards, it is easy to group samples with the same importance level and design a priority queue such that maximum data reliability can be achieved by sending more important groups over reliable channel.

4.2 Hybrid Minimum Cost Transmission

To achieve high efficiency, we propose a hybrid transmission scheme that maintains a tradeoff between the data reliability and transmission cost. Here, the data reliability index (RI) is determined by the requirement of the specific data analysis algorithm and the transmission cost ($COST$) can be defined as the total number of packet transmission attempts during a fixed time interval. The problem statement as follows: *given a specific network environment such as loss ratio l , how to select minimum number of samples, m , to transmit over reliable channel such that the data reliability index $RI \geq 1$* . Obviously, with less samples transmitted reliably, lower transmission cost is required to meet the data reliability requirement. The bandwidth saving will help support other sensors and applications in the same BSN. Therefore, we name the proposed scheme *hybrid minimum cost transmission*.

Algorithm 1 described the proposed hybrid minimum cost transmission scheme. Once m is determined, the following transmission order is followed: (i) a special packet containing pre-computed ANVs is created and sent through reliable channel; (ii) top m highest importance groups are sent over reliable channel. A group id is attached for appropriate receipt at the DA; (iii) remaining importance groups are sent over unreliable channel. A group id is attached for appropriate receipt at the DA.

Algorithm 1: Finding m highest importance groups for reliable transmission

```
estimate loss ratio  $l$ 
for each group  $i$  from 1 to  $n$ 
    calculate the expected ANV  $v(i)$  assuming all groups from
     $n+1$  to  $i+1$  that are sent with unreliable channel.
    if  $v(i) < v_{\text{threshold}}$ 
         $i++$ 
return  $i-1$ 
```

4.3 Variance Calculation

Three steps are involved in variance calculation in the framework:

Step 1: pre-calculation of ANVs for each data importance group. This value is the summation of the individual variance of all samples if a packet is lost.

Step 2: determination of variance threshold for each algorithm. This should be done by offline experiment to generate a variance value corresponding to a minimum satisfactory data analysis.

Step 3: reliability index calculation. At the DA, once a data sample set is received, it will check if all groups are received successfully. Then, based on the received or loss status of each specific group and the corresponding variance value inserted in the special packet of the same group, an actual ANV v_{received} can be obtained. Then, the actual reliability index is calculated as $RI = v_{\text{received}} / v_{\text{threshold}}$.

4.4 Feedback Generation

The DA calculated RI can be used to generate feedback messages and enable self-adaptation and reliability in BSN data streaming. A feedback tuple (sensor id, priority, RI) is generated when

- One or more sensor data collection is requested by a newly launched data analyzer. In this case, the RI field is set to 0;
- One or more sensor data collection is canceled by a newly terminated data analyzer. In this case, the RI field is set to MAX (such as 1000);
- $RI < 1$, i.e., not reliable.

Upon the receipt of a broadcasted feedback message, sensors should adjust their operation as follows:

- If $RI=0$ and the sensor id matches, the receiver sensor is identified as high priority and the minimum cost transmission scheme (Algorithm 1) is used.
- If $RI=MAX$ and the sensor id matches, the receiver sensor is identified as low priority and the basic hybrid transmission.
- If $RI < 1$ and the sensor id matches, the receiver sensor starts to increase the number of highest importance groups transmitted over TCP by 1.
- If $RI < 1$ and the sensor id does not match, the receiver sensor who is in low priority starts to further reduce traffic.

5. SUMMARY

Existing network protocols have not incorporated the data analysis requirements, thereby not achieving efficient and effective operation, especially when considering multiple sensor data collection scenarios. In this paper, we have designed a data analysis driven framework with feedback for this purpose. We described the components of this framework in details and

outlined the solutions for various issues. In the future, we will focus on implementing key algorithms such as identification of data importance and evaluation of the performance of the framework for further improvement.

6. REFERENCES

- [1] G.-Z. Yang, *Body Sensor Networks*. New York, USA: Springer Science+Business Media LLC, 2006.
- [2] C. Cordeiro and M. Patel, "Body area networking standardization: present and future directions," presented at the Proceedings of the ICST 2nd international conference on Body area networks, Florence, Italy, 2007.
- [3] B. Otal, L. Alonso, and C. Verikoukis, "Novel QoS scheduling and energy-saving MAC protocol for body sensor networks optimization," presented at the Proceedings of the ICST 3rd international conference on Body area networks, Tempe, Arizona, 2008.
- [4] G. Zhou, C.-Y. Wan, M. D. Yarvis, and J. A. Stankovic, "Aggregator-centric QoS for body sensor networks," presented at the Proceedings of the 6th international conference on Information processing in sensor networks, Cambridge, Massachusetts, USA, 2007.
- [5] G. Zhou, J. Lu, C. Wan, M. Yarvis, and J. Stankovic, "Bodyqos: Adaptive and radio-agnostic qos for body sensor networks," in *IEEE INFOCOM*, Phoenix, AZ, 2008, pp. 565-573.
- [6] T. Fulford-Jones, D. Malan, M. Welsh, and S. Moulton, "CodeBlue: An ad hoc sensor network infrastructure for emergency medical care," in *Proceedings of International Workshop on Body Sensor Networks*, 2004.
- [7] S. Jiang, Y. Cao, S. Iyengar, P. Kuryloski, R. Jafari, Y. Xue, R. Bajcsy, and S. Wicker, "CareNet: an integrated wireless sensor networking environment for remote healthcare," presented at the Proceedings of the ICST 3rd international conference on Body area networks, Tempe, Arizona, 2008.
- [8] M. Younis, Akkaya, K., Eltoweissy, M. and Wadaa A., "On handling QoS traffic in wireless sensor networks," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, Big Island, HI, USA, 2004.
- [9] K. Kyung Sup, M. A. Ameen, K. Daehan, L. Cheolhyo, and L. Hyungsoo, "A study on proposed IEEE 802.15 WBAN MAC protocols," in *Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on*, 2009, pp. 834-840.
- [10] J. Chen, K. Karric, D. Chang, J. Luk, and R. Bajcsy, "Wearable Sensors for Reliable Fall Detection," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2005, pp. 3551-3554.
- [11] L. Qiang, J. A. Stankovic, M. A. Hanson, A. T. Barth, J. Lach, and Z. Gang, "Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information," in *Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on*, 2009, pp. 138-143.
- [12] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," presented at the Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data, San Diego, California, 2011.
- [13] J. C. Castillo, D. Carneiro, J. Serrano-Cuerda, P. Novais, A. Fernández-Caballero, and J. Neves, "A multi-modal approach for activity classification and fall detection," *International Journal of Systems Science*, pp. 1-15, 2013.