Determining Quality- and Energy-Aware Multiple Contexts in Pervasive Computing Environments

Nirmalya Roy, *Member, IEEE*, Archan Misra, *Senior Member, IEEE*, Sajal K. Das, *Fellow, IEEE*, and Christine Julien, *Senior Member, IEEE*

Abstract—In pervasive computing environments, understanding the *context* of an entity is essential for adapting the application behavior to changing situations. In our view, context is a high-level representation of a user or entity's state and can capture location, activities, social relationships, capabilities, etc. Inherently, however, these high-level context metrics are difficult to capture using uni-modal sensors only and must therefore be inferred using multi-modal sensors. A key challenge in supporting context-aware pervasive computing is how to determine multiple high-level context metrics simultaneously and energy-efficiently using low-level sensor data streams collected from the environment and the entities present therein. A key challenge is addressing the fact that the algorithms that determine different high-level context metrics may *compete* for access to low-level sensors. In this paper, we first highlight the complexities of determining multiple context metrics as compared to a single context and then develop a novel framework and practical implementation for this problem. The proposed framework captures the tradeoff between the accuracy of estimating multiple context metrics and the overhead incurred in acquiring the necessary sensor data streams. In particular, we develop two variants of a heuristic algorithm for multi-context search that compute the optimal set of sensors contributing to the multi-context determination as well as the associated parameters of the sensing tasks (e.g., the frequency of data acquisition). Our goal is to satisfy the application requirements for a specified accuracy at a minimum cost. We compare the performance of our heuristics with a brute-force based approach for multi-context determination. Experimental results with SunSPOT, Shimmer and Smartphone sensors in smart home environments demonstrate the potential impact of the proposed framework.

Index Terms—Context-awareness, energy-efficiency, multi-context recognition, streaming multi-modal sensors.

N. Roy is with the Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD 21250 USA (e-mail: nroy@umbc.edu).

A. Misra is with the School of Information Systems, Singapore Management University, Singapore 188065, Singapore (e-mail: archanm@smu.edu.sg).

S. K. Das is with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: sdas@mst.edu).

C. Julien is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712-0240 USA (e-mail: c.julien@utexas.edu).

I. INTRODUCTION

MANY pervasive computing applications of interest hinge on the ability to extract individual-level *context* from sensor data streams provided by a combination of on-body sensors and infrastructural sensor devices. One illustrative example of such applications centers around remote health and wellness monitoring of individuals, especially the elderly and chronically ill, in smart homes. Commercial solutions already offer remote monitoring within "smart assisted-living homes", using a combination of body-worn medical and non-medical sensors (e.g., SpO_2 monitors and accelerometers) and *in-situ* sensors (e.g., thermal and motion detectors embedded in the home).[1]

Energy remains the most critical resource for both body-worn and in-situ sensors [16], especially as we move towards battery-less infrastructural sensors that operate using energy harvesting. Accordingly, it is vital to develop techniques that can reduce the energy overhead of such remote context monitoring, without sacrificing (or only minimally degrading) the accuracy of the context recognition process. Broadly speaking, *context* represents a single or multiple, sequential or interleaved states of a user. We make two key observations that apply to pervasive applications associated with such relatively sensor-rich smart home environments.

i) The first is that a specific individual context, especially those related to activities of daily living (ADLs), can be inferred (at varying levels of accuracy and energy costs) via *multiple* different combinations of body-worn and infrastructural sensors [17]. For example, we can determine that an individual has been stationary for an unusually long period of time based on (see Fig. 1) either (a) the accelerometer sensor data from the individual's smartphone, (b) passive infrared (PIR) or ultrasonic sensors mounted on ceilings, or (c) force sensitive resistor sensor (FSR) sensors embedded in beds and couches. Each of them is associated with different types of errors—e.g., smartphone-embedded accelerometer data may generate false negatives if the person leaves the phone on the table, whereas PIR sensors may generate false positives due to movement of animals or changing intensity of reflected sunlight. *Multi-modal sensing*, or the fusion of disparate data streams for inferring a single context metric, helps
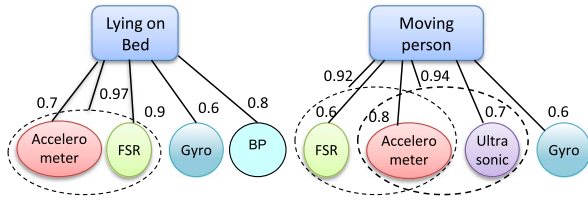
Fig. 1. Impact of choice of different subsets of sensors on $QoINF$ (without considering tolerance range).

mitigate errors that result from relying solely on a single sensing stream [1]. Moreover, in general, *the accuracy of inferred context increases with the use of a larger set of sensors*.

ii) The second is that many applications are interested in *multiple distinct activity contexts simultaneously*. For example, remote wellness applications (see Fig. 1) may be interested in (a) outliers in vital signs (e.g., drop in blood pressure), (b) motion detection and (c) sleep duration monitoring.

In this paper, we tackle a key aspect of a robust application framework for energy-efficient determination of multiple concurrent contexts, each of which may need differing levels of context accuracy. The key idea is to exploit the overlaps and correlations between the multiple distinct combinations of multi-modal sensors that can be used to infer each individual context, and the relation between the inferred context accuracy and *key operating parameters* of each individual sensor, to determine both (i) the best set of sensors that need to be concurrently sensed, and (ii) the optimal parameter setting for each such activated sensor. Here, the terms "best" and "optimal" are used to imply that the choices made should both (a) result in the lowest cumulative energy overhead and (b) satisfy the accuracy requirements associated with each distinct context.

More specifically, in this paper we focus principally on reducing the *wireless communication* energy overheads associated with transferring the generated sensor data from the sensor sources (i.e., the on-body or infrastructural sensors chosen) to a central "server node", where the contexts are computed. This choice is motivated by the observation that communication is often the most energy-intensive component in collecting information and determining context from such embedded sensors. A promising approach to reducing the communication overhead relies on modeling the degree of context inaccuracy that pervasive computing applications can tolerate.

In our previous work [14], we formally defined such a relationship in terms of a metric called *Quality-of-Inference* ($QoINF$). $QoINF$ is defined as the *average error probability in estimating a high-level context state, given the imprecision in the contributing low-level sensor values*, and is related to both (a) the set of sensors used to infer that context state and (b) the imprecision allowed in data acquisition for each sensor in the set. We refer to the latter as the sensor's *tolerance range*: a tolerance range of Δ for a sensor implies that the sensor reports its currently sampled value only when it deviates from the last reported value by at least $\pm\Delta$. A tolerance range based approach allows us to apply an event-driven model of communication by the sensing source; prior work [19] has demonstrated that even a modest increase in tolerance often results in a significant reduction in communication energy overheads. An alternate

approach to reducing the communication energy overheads involves reducing the *sensor sampling rate*; indeed, many prior activity recognition frameworks (e.g., [21] vary the sampling rate (e.g., on accelerometer or gyroscope sensors) to reduce both the sensing and data transfer costs. However, such prior work has looked at an individual context state *in isolation*, and failed to consider the reality that smart home applications often require the determination of multiple contexts simultaneously.

This paper focuses specifically on the additional challenges associated with the problem of simultaneously inferring *multiple (abstract) context* states from an overlapping set of sensors, with the lowest possible energy overhead. To appreciate the additional complexity involved, consider an example, depicted in Fig. 1, from an ambient assisted living environment. In isolation, the single context *moving person* can be determined with 94% accuracy by fusing two sensor streams: one from an accelerometer and another from an ultrasonic sensor. However, when the same environment is asked to simultaneously determine the context *lying on bed* the combination of the accelerometer and a force sensitive resistor sensor (FSR) is best used to help determine this new context, while *moving person* can be determined by combining the accelerometer and FSR, albeit at the reduced accuracy of 92%. As this example demonstrates, solutions to the individual context optimization problem may not be ideal for joint determination of multiple contexts when the underlying sensor streams are shared. Note that Fig. 1 does not explicitly relate the sensors' tolerance ranges to the achieved quality of context recognition. These tolerance ranges add another dimension to the problem. In the example under consideration, however, it is easy to see that the estimation of the *moving person* context will be less accurate in both cases if the accelerometer sensor tolerance range is $\pm40\%$ (indicating that the true reading may be up to 40% higher or lower than the reported value) as opposed to a tolerance range of $\pm10\%$.

To tackle this challenging problem, we shall significantly extend our previous formalization [14] of minimum-cost context information using a generic function that captures the relationship between $QoINF$ and the set of sensors (and their assigned tolerance ranges). Specifically, in this paper we propose a range based sensor selection heuristic algorithm for multiple contexts; this heuristic explores the option of satisfying the $QoINF$ of the additional context metrics without altering the set of activated sensors but simply by tightening the tolerance range of the current set of activated sensors. The major contributions are:

• We formalize the problem of energy-efficient multi-context estimation to incorporate the diversity of distinct sensor streams (both body-worn and infrastructure-based) that are likely to be available in emerging smart home environments.

• We design two low-complexity Lagrangian-based heuristic algorithms to approximate the selection of the best set of sensors and associated tolerance ranges to achieve the specified $QoINF$s for multiple contexts simultaneously at a minimum cost (i.e., the energy expended to acquire context). The first algorithm naïvely grows a set of sensors by incrementally considering additional target contexts. The second algorithm improves on the first's naïvety; it evaluates the increase in cost not only due to the addition of a new sensor to the growing subset but also by altering the sensing parameters (tolerance ranges) of those sensors
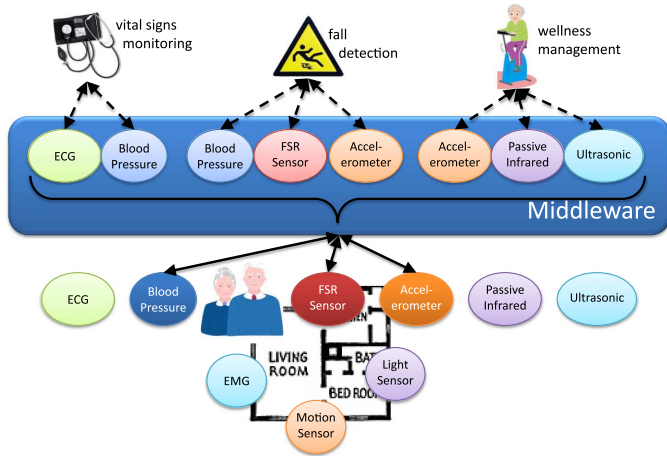
Fig. 2.   Multi-application multi-context inferencing in a sensor-rich pervasive healthcare environment.

already selected. We also present a brute-force algorithm and compare it to our heuristic approaches.
- We develop an empirical method for our $QoINF$ functions. We experimentally evaluate the appropriateness of our $QoINF$ and sensor selection algorithms, using real data traces collected from SunSPOT and Shimmer sensor platforms.

Overall, we demonstrate how a $QoINF$ function can be computed for practical, real-world environments, and thereby develop a systematic framework that intelligently utilizes the diversity of sensor devices to conserve energy, while satisfying the diverse context accuracy needs for multiple concurrently executing context-based applications.

II. Application Scenario

The wide availability of smart healthcare appliances and a variety of standalone and integrated sensor devices is making it progressively easier to ubiquitously and continuously monitor an individual's health-related vital signals and her behavior (i.e., activity). For example, a combination of *body-worn* medical and non-medical sensors (e.g., sensors to monitor blood oxygenation or accelerometers to monitor movements) and *in situ* sensors (e.g., thermal and motion detectors) can be used to determine an individual's *context* in smart homes. Broadly speaking, context here refers to a variety of dynamically changing states, related to either an individual's specific activities (e.g., walking vs. sleeping) or biomedical conditions (e.g., elevated blood pressure, shortness of breath, arrhythmia), or to surrounding environmental conditions (e.g., atmospheric ozone levels, ambient temperature).

As an illustration, consider a remote context monitoring scenario (shown in Fig. 2) in a smart assisted-living environment in which an elderly person resides. The smart home may be equipped with many sensors (light, humidity, PIR, ECG, electromyography, etc.), some of which may be body-worn, while others may be embedded in the environment. A variety of applications and stakeholders (e.g., fall monitoring by a caregiver; wellness activity monitoring by a doctor; vital sign monitoring by a nurse) need to access this low-level sensed information to abstract high level context (both physiological and activity) about the resident. An important observation is that a specific application's context can be satisfied by different combinations

of sensor data types. For example, the fall detection application may utilize data either from multiple video cameras, or from a set of body-worn accelerometer and wall-mounted motion sensors, or from a set of audio sensors, or from some combination of the above. Similarly, the sustained lack of movement inside the home may be determined either from a body-worn accelerometer, from wall-mounted PIR sensors or from room-specific light sensors.

The above example motivates the need for a "matchmaking" software infrastructure that mediates between the context-driven health and wellness applications and the set of available sensors in a way that minimizes the energy overhead, while still ensuring that the applications' needs for high-quality context inferences are met. To enable such a dynamic and automated association between application requirements and the available sensor resources in any environment, we make the following two key contributions:
- First, we suggest the use of an explicit functional model to relate the accuracy of any inferred context to a measure of uncertainty about the true values of the sensor data.
- Then, we develop and evaluate an optimization-based heuristic that uses the model to dynamically select both a set of sensors and the parameters of the sensors to satisfy the context requirements of multiple context-aware applications, while minimizing the energy overhead of sensor data transmission.

III. Context Inference Model

We begin with an overview of our model of the $QoINF$ for context determination in pervasive computing applications, which we developed in our prior work to determine a single context from multiple underlying sensor streams [14]. We extend this original model to support more challenging multi-context recognition based on multiple underlying sensing streams. In both problems, determining a specific context attribute may be viewed as an inference obtained from fusing values acquired through multiple sensor streams.

A. QoINF Model

Given set of sensors, S, let v_i represent the data value reported by sensor $s_i \in S$. Let Λ_i be the range of possible values that s_i could report. Determining a context variable C may be viewed as a mapping function $f_C(.)$ that takes as input the values from a subset of sensors $\theta \subseteq S$ and maps them into the state space of the output context, Λ_C.

$$f_C(\theta) : \prod_{s_i \in \theta} v_i \Rightarrow x : \quad x \in \Lambda_C. \tag{1}$$

Different values of the same context may be inferred with varying accuracy using different subsets of sensors [6]. Therefore, we need to associate an accuracy function with the context inference that expresses the *average accuracy* in estimating the context C using the sensors in θ. We define a particular instance of this accuracy function, $QoINF(.)$ to be one minus the average estimation error incurred by $f_C(\theta)$:

$$QoINF_C(\theta) = 1 - \frac{\sum_{x \in \Lambda_C} err_C(x, \{s_i \in \theta\})}{|\Lambda_C|} \tag{2}$$

where $err_C(x, \{s_i \in \theta\})$ is the probability of error, given accurate readings from the sensor subset θ, when the value for

the context variable C is actually x. Although alternative definitions of accuracy are possible, such definitions are merely alternatives to our definition of $QoINF(.)$; we focus on the use of such functions in expressing the quality of context estimation *in general* and use the definition in (2) as a basis for quality of information founded on sensing errors.

If the state space of C is discrete, then the estimation error is computed as the normalized number of incorrect inferences made by the best possible estimator. If the context space is continuous, the error also depends on the tolerance in the computed output (e.g., it may be acceptable for an inferred location to be inaccurate by ± 3 feet). We do not delve into these technicalities, which relate to the precise specification of the $err(.)$ function. Our focus is how to exploit a *given function $err(.)$* in defining $QoINF$. There exists an extensive body of literature in optimal estimators and the resulting lowest possible error bounds [8], [9].

In our model, the tolerance range is a variable specified by the application and constrains how correct the application's view of the reported sensor value must be. For instance, if the tolerance range is ± 10, and the last reported value was 120, then all the application knows for sure is that all of the subsequent sensor readings must lie in the interval (110, 130). Sensing errors (errors in sampling, calibration etc.) are, on the other hand, an intrinsic part of a sensor. For example, a blood pressure sensor may have an error of ± 2, indicating that a reading of 120 corresponds to a "ground truth" value in the range (118, 122). In general, if the tolerance range is δ and the sensor error is ϵ, applications should typically specify $\delta \geq \epsilon$. Also, given δ and ϵ (which may not always be known) and a last reported value of v, we can say that the "ground truth" of the sensed attribute lies between $(v - \delta - \epsilon, v + \delta + \epsilon)$. In other words, the sensor-specific ϵ contributes an extra uncertainty in addition to the application-specific δ.

B. Uncertainty in the Tolerance Ranges

The context inference model and associated $QoINF$ functions capture complex relationships between the context inference quality and the choice of sensors used to infer that context. Such a model does not, however, completely capture the benefits that can be garnered from using continuous event-driven monitoring. In event-driven monitoring, sensors can be assigned tolerance ranges, allowing them to only report changes that fall outside these ranges.

Previous work has demonstrated a significant reduction in the communication and hence energy overheads due to sensing by providing individual sensors small non-zero tolerance ranges [2], [7].

Let $Q_\theta = \{q_1, q_2, \ldots, q_\theta\}$ be the set of tolerance ranges for the selected sensor subset θ, used to infer a single context metric. We represent the error associated with this choice of sensors by explicitly relating a context's $QoINF$ value to the choice of not only the sensors but also Q_θ:

$$QoINF_C(\theta, Q_\theta)$$
$$= 1 - \sum_{x \in \Lambda_C} err_C(x, \{(s_i, q_i) : s_i \in \theta, q_i \in Q_\theta\}) \quad (3)$$

This computes the average error, summing across all possible values of the context variable C; it implicitly assumes that all context states are equally likely. Again, more sophisticated formulations of $QoINF$ are possible, but the above provides a

sufficiently expressive definition to enable adapting the sensing task to the computed quality of information.

C. Cost Model

One of our primary goals is to reduce the energy consumption due to communication, while ensuring an application-specified minimum $QoINF(.)$ for a given context metric. The cost of acquiring one of the sensor streams needed to update the context variable is a function of both the sensor s_i's tolerance range (q_i) and the (possibly multi-hop) transmission cost from the sensor to the sink. (Note: while the experimental results in this paper all used a 1-hop transfer from each individual sensor source to sink, multi-hop transfer of sensor data has been suggested for both body sensor networks (BSNs) and infrastructural sensors (e.g., based on multi-hop IEEE 802.15.4 links)). Thus, for generalization, we assume that this cost is linear of the number of hops (h_i) in the uplink path from sensor s_i to the sink, given a network in which all links have identical transmission power. Intuitively, the cost and tolerance range should be inversely related: if the tolerance range is large, it is less likely for a value to fall outside the permitted range, thus making communication less frequently necessary. In the absence of any temporal correlation among the sensed samples, we assume the underlying data samples evolve as a random walk, and the *cost* is proportional to h_i/q_i^2 [7]. In this case, the resulting cumulative cost function for the subset θ of selected sensors is given by:

$$COST(\theta, q_\theta) = \kappa \sum_{s_i \in \theta} \frac{h_i}{q_i^2} \quad (4)$$

where κ is a scaling constant and h_i is the hop count.

D. Choice and Characteristics of QoINF Function

While a completely arbitrary $QoINF(.)$ function requires a brute-force search, there are forms of this function that lend themselves to efficient heuristics. Intuitively, as a sensor's tolerance range increases, its data reporting frequency decreases; ultimately, this causes a deterioration in the context inference accuracy. A particularly attractive case occurs when the ith sensor's individual $qoinf(.)$ is represented by an inverse-exponential distribution of the form:

$$qoinf(i) = 1 - \frac{1}{\nu_i} \exp\left(\frac{-1}{\eta_i q_i}\right) \quad (5)$$

where η_i and ν_i are sensitivity constants for sensor s_i. A larger value of ν_i indicates a lower contribution from s_i to infer context C. Consider the estimation of the *Moving person* context in Fig. 1 using the accelerometer with different tolerance ranges. The error rate of the accelerometer based on (5) can be represented as

$$\text{error rate} = 1 - qoinf(i) = \frac{1}{\nu_i} \exp\left(\frac{-1}{\eta_i q_i}\right) \Rightarrow \exp\left(\frac{-1}{q_i}\right)$$

where we assume that $\nu_i = \eta_i = 1$ for simplicity. Clearly, as the tolerance range increases, we observe an increase in the error rate and a decrease in the context inference accuracy ($qoinf$). Moreover, for a selected subset of sensors θ, the resulting $QoINF(.)$ function is:

$$QoINF_C(\theta) = 1 - \prod_{s_i \in \theta} (1 - qoinf_C(i)) \quad (6)$$

This formulation assumes that the estimation errors of different sensors are statistically independent of each other.

It satisfies three key properties of a $QoINF(.)$ function: (i) the value of the function always falls between 0 and 1; (ii) $QoINF_C(.)$ is non-decreasing in θ in the sense that incorporating data from an additional sensor does not lead to a reduced $QoINF$; and (iii) as a sensor's tolerance range increases, the quality of contribution of that sensor decreases towards 0 in the limit. We further validate this choice of $QoINF$ function empirically in Section V.

E. QoINF Function Independence

Our $QoINF$ function is targeted towards scenarios where the same context can be inferred through different modalities of sensing. For example, the context that a person is walking around may be detected by (i) a body worn accelerometer, (ii) a wall mounted motion sensor, (iii) floor-embedded piezo sensors or (iv) a video camera with gait recognition software. In general, given the different sensing modes (the physical property by which the sensing works), we believe that the errors of these sensors will be statistically independent (i.e., uncorrelated). For example, the accelerometer will give errors if it is not mounted properly, the video camera may fail in low lighting, and the motion sensor may fail when multiple people are present in the same room.

Our belief is that activity contexts-such as "the user is walking" or "the user is sleeping" or "the user has been watching TV for >30 minutes" tend to share this independence assumption. On the other hand, specific medical contexts (e.g., "blood pressure is high" or "the body temperature is rising") do not share this independence assumption. Typically, such low level contexts can be measured only by a single type of sensor, and even when measurable by different modes (e.g., heart rate by either Sp02 or blood pressure sensor), they are possibly subject to correlated cases of failure (e.g., both of those may have false readings due to human motion).

Note further that the above $QoINF$ function formulation assumes that at least one of the needed sensors is working. Namely, we look for the context "the user is walking" and declare the user to be walking if *at least one* of the sensors declare her to be walking. This was done for ease of exposition and can be considered a specific embodiment of the proposed model. We plan to investigate other strategies (e.g., majority voting, etc.). These will change the form of the $QoINF$ function and the resulting analytical optimal points, but it would not change the fundamental approach.

IV. JOINT OPTIMIZATION OF MULTIPLE CONTEXTS

Pervasive computing environments entail multiple needs for context information, necessitating simultaneous determination of multiple context metrics from shared underlying sensor streams. As sensor networks become more ubiquitous, they will increasingly be treated as a platform for multi-modal sensing for assessing multiple diverse context metrics simultaneously. This diverges from the current paradigm of designing an individual context recognition model from a specified set of sensors. Recognizing multiple contexts simultaneously from the underlying sensor data streams requires the selection of the best subset of sensors along with their optimal tolerance ranges in a way that simultaneously satisfies the $QoINF$ thresholds for all of the required context metrics with a minimum total cost or energy overhead. However, sharing the same set of sensors across multiple contexts may result in a reduction of an individual context's $QoINF$ accuracy (as compared to the $QoINF$ achievable for that context metric in isolation). To gain a deeper understanding, we formalize the simultaneous determination of multiple contexts as a multi-objective optimization problem. In particular, we propose two sensor selection multi-context search heuristics to choose the best set of sensors and their associated tolerance ranges. Our first heuristic algorithm incrementally adds new sensors to a growing subset to incrementally satisfy additional context metrics at minimal cost. Our second algorithm is a range-based heuristic that can either modify the tolerance ranges of sensors already in the subset or add a new sensor to that subset based on the cost. We also outline a brute-force algorithm for the purpose of evaluating our heuristics.

A. Multi-Context Optimization Problem

We start with a single context determination [14] and extend it for multiple contexts [15]. We first assume we are provided the set of sensors θ and investigate how to determine the optimal tolerance ranges; we then revisit this assumption and investigate how to determine θ in the first place. Given θ, the optimization problem for a *single* context variable is to choose the sensor tolerance ranges $q_1, q_2, \ldots, q_\theta$ that, when used together to infer the value of context variable C, minimize the total cost while ensuring the application-specified $QoINF$, say $QoINF_{min}$. Therefore, the objective is:

$$\mathbf{min}\ COST(\theta, q_\theta)\ \mathbf{subject\ to}\ QoINF_C(\theta) \geq QoINF_{min}$$

The above joint optimization of (θ, q_i) for $1 \leq i \leq \theta$ applies to a single context variable C, which we have investigated in our previous work [14]. Our goal here is to minimize both the total cost (in terms of communication overhead) and the deviation from the application's required quality of inference for *multiple* contexts. The former is to reduce the energy consumption associated with the communication in a distributed sensor network, while the latter ensures that resources are not wasted acquiring context information that is of higher quality than the application requires. Instead, we make best use of shared sensor data streams to satisfy the minimum quality requirements while reducing the communication overhead and energy consumption. Considering the problem of simultaneously determining L separate context variables, $\{C(1), C(2), \ldots, C(L)\}$, we propose the following Lagrangian Optimization problem:

$$\text{minimize} \sum_{s_i \in \theta} \frac{h_i}{q_i^2} + \sum_{l=1}^{L} \lambda_l \left[QoINF_{C(l)}(q_1, q_2, \ldots, q_\theta) \right.$$
$$\left. - QoINF_{min}^l \right] \quad (7)$$

where $QoINF_{min}^l$ denotes the minimum required quality-of-inference value for context l, and λ_l is the Lagrangian multiplier for a specific context l. There are different sensitivity constants ν_{il} and η_{il} for each of the l contexts being inferred because each context has a different sensitivity to the tolerance range of a given sensor s_i. However, each selected sensor s_i has a single selected tolerance range q_i for all contexts it contributes to (i.e., the *sensing* task is shared among the multiple contexts being inferred). A sensor will report a new value only when the new value diverges from the previous one by q_i. In this multi-context determination problem, sharing the same subset of sensors across multiple context types (when possible) can help reduce

the overall cost. For example, in Fig. 1, the two different contexts *Lying on bed* and *Moving person* can rely on the information from an overlapping set of sensor streams, thereby reducing the cost of overall context determination while maintaining the applications' specified $QoINF$ requirements. This also encourages the use of similar $QoINF$ functions across context types (albeit with varying sensitivity values). If the $QoINF_{C(i)}(.)$ functions have the same inverse exponential form, then the collective optimum values of tolerance ranges $\{q_i\}$ may be explicitly computed:

$$\text{minimize} \sum_{s_i \in \theta} \frac{h_i}{q_i^2} + \sum_{l=1}^{L} \lambda_l \left[1 - \prod_{s_i \in \theta} \left[\frac{1}{\nu_{il}} exp \left(\frac{-1}{\eta_{il} q_i} \right) \right] \right. $$
$$\left. - QoINF_{min}^l \right] \quad (8)$$

Lemma 1: If the $QoINF(.)$ functions for the set of l contexts and the set of sensors θ satisfy (5) and (6), then the optimal choices of q_i that minimize the cost function for context l satisfy the following relationship:

$$\frac{\frac{2h_1}{q_1^2}}{\sum_{l=1}^{L} \left[\log(1 - QoINF_{min}^l) + \log(\nu_{1l}) \right]} - $$
$$\frac{\frac{2h_1}{q_1^2}}{\sum_{l=1}^{L-1} \left[\log(1 - QoINF_{min}^l) + \log(\nu_{1l}) \right]} = $$
$$\frac{\frac{2h_1}{q_1^2} + \frac{2h_2}{q_2^2}}{\sum_{l=1}^{L} \left[\log(1 - QoINF_{min}^l) + \log(\nu_{1l}) + \log(\nu_{2l}) \right]} - $$
$$\frac{\frac{2h_1}{q_1^2} + \frac{2h_2}{q_2^2}}{\sum_{l=1}^{L-1} \left[\log(1 - QoINF_{min}^l) + \log(\nu_{1l}) + \log(\nu_{2l}) \right]} = \ldots$$
$$= \frac{\frac{2h_1}{q_1^2} + \frac{2h_2}{q_2^2} + \ldots + \frac{2h_\theta}{q_\theta^2}}{\sum_{l=1}^{L} \left[\log(1 - QoINF_{min}^l) + \sum \log(\nu_{\theta l}) \right]} - $$
$$\frac{\frac{2h_1}{q_1^2} + \frac{2h_2}{q_2^2} + \ldots + \frac{2h_\theta}{q_\theta^2}}{\sum_{l=1}^{L-1} \left[\log(1 - QoINF_{min}^l) + \sum \log(\nu_{\theta l}) \right]} \quad (9)$$

and the optimal value of q_i (the tolerance range assigned to the sensor $s_i \in \theta$) across all contexts is given by:

$$\hat{q}_i = \frac{h_i * \left[\sum_{l \in L} \sum_{s_i \in \theta_l} \frac{1}{h_i * \eta_{il}} \right]}{\sum_{l \in L} \left[\log \left(1 - QoINF_{min}^l \right) + \sum_{s_i \in \theta_l} \log(\nu_{il}) \right]} \quad (10)$$

The minimum cost to achieve the specified inference accuracy using these q_i values is given by:

$$\widehat{COST}(\theta) = \frac{\left(\sum_{l \in L} \left[\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta_l} \log(\nu_{il}) \right] \right)^2}{\sum_{l \in L} \sum_{s_i \in \theta_l} \frac{1}{h_i * \eta_{il}^2}} \quad (11)$$

The proof (in the Appendix) follows by solving the Lagrangian optimization problem for multiple contexts in (8).

In addition to finding the minimum cost for a given subset θ, we also need to determine the best subset of sensors, that minimizes the overall update cost across all the contexts. Clearly, a brute-force approach is to iterate through all possible combinations and compute $\widehat{COST}(.)$ for each of the combinations of sensors, for all contexts.


```
Procedure Brute_Force_Multi_Fusion (input set S, QoINF_{min}^l,
        ∀ l = (1,...,L))
1. Initialize empty set of sensors; θ = φ; COST(θ) = 0;
   MinCost(l) = ∞;
2. Find the power set of S //Assume θ ⊆ P(S)
3. For (i=1; i < |P(S)|; i++){
4.     For (l=1; l <= L; l++) {
5.         Compute the tolerance range q_i
6.         if QoINF_i^l(P(S)) ≥ QoINF_{min}^l ∀l and q_i {
7.             θ = P(S);
8.             Compute the update cost COST(θ) for QoINF_{min}^l
9.                 if (COST(θ) − MinCost(l)) < 0
10.                    MinCost(l) = COST(θ).
11.            } else
12.            break; //move to next power set
13.        } End-For
14. } End-For
15. return {θ, minimum COST(θ)}.
```

Fig. 3. Sensor set selection brute force algorithm for multiple contexts.

B. Brute-Force Sensor Selection for Multiple Contexts

The algorithm shown in Fig. 3 describes a brute-force approach to simultaneously determine multiple contexts. This algorithm iterates over all possible subsets of sensors from the set S, determines for each whether it satisfies the $QoINF$ requirements for all of the contexts, and, if so, what the cost is of using that subset.

After computing this cost for all possible subsets, the brute force approach returns the satisfying subset with lowest cost (if one exists).

While this approach generates an optimal result, its time complexity is impractical. Therefore, we have developed efficient sensor selection heuristic algorithms for this multi-context determination problem.

C. Sensor Selection Heuristic for Multiple Contexts

In this section we propose a heuristic for selecting the set of underlying sensors and associated tolerance ranges for simultaneously determining multiple contexts $\{C(1), C(2), \ldots, C(L)\}$. The heuristic is based on the observation that the additional cost of adding a sensor s_i to an existing subset θ is principally dependent on the term $\log^2(\nu_{il}) * h_i * \eta_{il}^2$. This can be derived from (11) by considering the limiting case of the function $QoINF_{min}^l$. This term can be generalized so that each sensor can have a different sensitivity and hop count factor for each context $C(l)$, where $1 \leq l \leq L$, thus accounting for the possibility that decreasing the quality of a particular sensor stream may have different degrees of impact on the determination of different context metrics. Since a lower value of this term indicates a greater preference for selecting a sensor, our selection heuristic sorts the available sensor set S in ascending order of this term for each context, thus generating L sorted lists, $\Re$. Our approach is to incrementally create a subset of sensors, iteratively considering additional context metrics and adjusting the selected sensor subset to continue to satisfy the growing set of considered context metrics. We first select a single context metric, say $C(1)$, and find the subset of sensors from S such that $C(1)$'s $QoINF$ function can be satisfied with the least cost. We then consider additional context metrics (e.g., $C(2)$ is considered next) and determine whether the subset of sensors selected for determining the previously considered context(s) can also satisfactorily determine each additional context metric. If not, sensors are added to the selected subset

```
Procedure Heuristic_Multi_Fusion (input set S, QoINF^l_min,
        ∀ l = (1,...,L))
1. Initialize empty set of sensors; θ = φ; MinCost(l) = ∞;
2. Sort the sensor set S into a list ℜ_l, ∀ l = (1,...,L)
   in increasing order of sensitivity log²(ν_il) * h_i * η²_il
3. For (l = 1; l <= L; l++)
4.     For (i = 1; i <= |S|; i++) //cardinality of set S
5.         θ = θ + ℜ_l(i); //set-theoretic addition
6.         Compute the tolerance range q_θ
7.         if QoINF^l(θ) ≥ QoINF^l_min for q_θ
8.         Compute the optimal update cost COST̂(θ)
           for QoINF^l_min
9.         if (COST̂(θ) - MinCost(l)) > 0 OR below threshold
10.        break;
11.        else MinCost(l) = COST̂(θ).
12.        For (l = l + 1; l <= L; l++)
13.            ℜ_l = ℜ_l - θ; //set-theoretic substraction
14.        End-For
15.    End-For
16. End-For
17. return {θ, COST̂(θ)}.
```

Fig. 4.   Sensor set selection heuristic algorithm for multiple contexts.

to help determine the newly added context metric given its $QoINF$ function at the least cost. This process continues until all of the contexts have been considered. Fig. 4 shows the pseudocode for this heuristic

D. An Illustrative Example

Consider a set of five sensors $S = \{s_i\}$ for $1 \leq i \leq 5$ used to determine three different contexts $C(l)$ for $1 \leq l \leq 3$. Our goal is to determine the subset of sensors that minimizes the overall update cost. We sort the set of available sensors based on their $[\log^2(\nu_{il}) * h_i * \eta_{il}^2]$ values and generate a sorted list for each context; for our example, this results in the following lists: $\mathfrak{R}_{C(1)} = \{s_1, s_2, s_4, s_5, s_3\}$; $\mathfrak{R}_{C(2)} = \{s_5, s_4, s_3, s_2, s_1\}$ and $\mathfrak{R}_{C(3)} = \{s_3, s_2, s_4, s_1, s_5\}$. For list $\mathfrak{R}_{C(1)}$, assume the optimal choice for the sensor subset θ is $\{s_1, s_2\}$; with minimum cost, these two sensors together can achieve the specified $QoINF$, and including any more sensors increases the update cost. After selecting sensors s_1 and s_2, our algorithm determines what other sensors (if any) need to be added to the subset to satisfy the context $C(2)$ with its required $QoINF$. In each step, we exclude the already chosen sensors from consideration and therefore in this example generate new lists $\mathfrak{R}'_{C(2)} = \{s_5, s_4, s_3\}$ and $\mathfrak{R}'_{C(3)} = \{s_3, s_4, s_5\}$. In both cases, we can already assume the participation of sensors s_1 and s_2, if they can also contribute to the determination of contexts $C(2)$ and $C(3)$. Considering the modified next list $\mathfrak{R}'_{C(2)}$, assume we have an optimal subset $\theta = \{s_1, s_2, s_5\}$ with minimal cost. In other words, adding only sensor s_5 to θ satisfies $C(2)$'s $QoINF$ requirement at a minimal cost. Considering the final list, $\mathfrak{R}'_{C(3)}$, assume we have the optimal subset of sensors, $\theta = \{s_1, s_2, s_5, s_3\}$ with minimal cost. The worst case time complexity of the heuristic is $O(n^2)$.

In contrast, the brute-force search algorithm iterates over the power set of S, which, in this example, contains $2^5 - 1$ elements, across three different contexts $C(1)$, $C(2)$ and $C(3)$. Starting from a singleton sensor subset, we assume that subset $\{s_1, s_2\}$ is good for context $C(1)$ but not satisfactory for contexts $C(2)$ and $C(3)$ simultaneously. After iterating all subsets of three sensors across all contexts, we find the subset $\{s_1, s_2, s_5\}$ simultaneously satisfies $C(1)$ and $C(2)$, but not $C(3)$ with the required conditions. Next iterating over all subsets of four and five sensors across all contexts, we derive an optimal sensor

```
Procedure Range_Heuristic_Multi_Fusion (input set S,
QoINF^l_min, ∀ l = (1,...,L))
1. Initialize empty set of sensors; θ = φ; minCost = ∞;
2. Sort the sensor set S into a list ℜ_l, ∀ l = (1,...,L)
   in increasing order of sensitivity log²(ν_il) * h_i * η²_il
3. FOR (l = 1; l <= L; l++) {
4.     IF (θ == φ) { //initialization for l = 1
5.         θ = ℜ_l.first(); //initialization
6.         {minCost, Q} = optimal update cost COST̂(θ)
           and q_θ for first context {QoINF^1_min}
7.     } END-IF
8.     ℜ_l = ℜ_l - θ; newCostSameTheta = minCost;
       //set-theoretic subtraction
9.     IF (l ≠ 1) { //exclude the first context
10.        {newCostSameTheta, Q_same} = optimal update
           cost COST̂(θ) and q_θ for l contexts
           {QoINF^1_min,..., QoINF^l_min}
11.        IF (newCostSameTheta - minCost < thresh) {
   //only slight increase with use of existing sensors
12.            minCost = newCostSameTheta; Q = Q_same;
13.            break; // move to next context
14.        } END-IF
15.    } END-IF
16.    reduction = true; // to set up the loop for adding
       new sensors
17.    while (ℜ_l.first()! = φ && reduction == true) {
18.        θ_new = θ + ℜ_l.first(); ℜ_l -= ℜ_l.first();
           //set-theoretic addition and subtraction
19.        {newMinCost, Q_new} = optimal update cost
           COST̂(θ_new) and q_θ for l contexts
           {QoINF^1_min,..., QoINF^l_min}
20.        IF (newMinCost < minCost || newMinCost <
           newCostSameTheta) { //adding a new sensor
           reduces cost or is better than using original θ
21.            θ = θ_new; Q = Q_new; Q_same = Q_new;
22.            minCost = newMinCost;
               newCostSameTheta = newMinCost;
23.        } else {
24.            minCost = newCostSameTheta; Q = Q_same;
25.            reduction = false;
               //no point looping over more sensors
26.        } END-IF
27.    } END-WHILE
28. } END-FOR
29. return {θ, Q, COST̂(θ)}.
```

Fig. 5.   Range-based sensor set selection heuristic algorithm for multiple contexts.

subset $\theta = \{s_1, s_2, s_5, s_3\}$ that holds across the three contexts with minimal cost and satisfies the $QoINF$ accuracy. Clearly, the time complexity of the brute force algorithm is exponential in the number of sensors multiplied by the number of distinct context metrics being inferred.

E. Range Based Sensor Selection for Multiple Contexts

Here we propose an enhanced version of the previous heuristic algorithm, shown in Fig. 5, that, for each additional context, tries to compare the total cost from the following two approaches: (a) using the current set of sensors and determining if a modification of the tolerance ranges of this current set is enough to satisfy the $QoINF$ requirement of the additional context metric; or (b) adding an additional sensor to the set of sensors and seeing what tolerance ranges this modified set must have to satisfy all the $QoINF$ requirements of the contexts considered thus far. After computing the costs of each approach, this second heuristic selects the one that both satisfies the $QoINF$ requirements of all of the considered contexts and has the lowest cost. This is in contrast to the approach in the previous algorithm shown in Fig. 4, where the comparison was made only between adding a new sensor and the cost incurred by the current set of sensors (with their tolerance ranges unmodified). In other words, the previous approach did

not explore the option that one could satisfy the $QoINF$ of the additional contexts without altering the set of activated sensors, simply by tightening the tolerance ranges of the current set of selected sensors.

We have thus far assumed that, for a given context, the user is only in one context state at a time, i.e., either the user is in the $Sitting$ state or in the $Walking$ state or in the $Running$ state. There are, however, other scenarios like *Watching TV* and *Speaking on the phone*, which may happen concurrently. Such concurrent context states can also be determined using our model. As shown in our model, the minimum $QoINF$ value and sensitivity factors for these multiple context states will be fundamentally different. For example, consider we have one acoustic sensor for detecting the *Watching TV* context state and one microphone sensor for recognizing the *Speaking on the phone* context state. The operating analytics (tolerance range, etc.) of these two sensors can be computed by our proposed model while still maintaining the underlying objective of sharing sensor data streams to improve the accuracy and minimize the network cost.

F. Time Complexity of the Heuristics

In this section, we discuss the time complexity of our two proposed heuristic algorithms, (a) naive heuristic and (b) range heuristic. We do a step-by-step run time analysis for both algorithms.

1) Naive Heuristic (Figure 4/Algorithm 4): Line 1: Assignment statement runs in constant or one unit of time. Line 2: Runtime of a sorting function such as quick sort is $O(|S|^2)$ where $|S|$ denotes the cardinality of the sensor set S. Lines 3 & 4: Nested for loop has a runtime of $O(L|S|)$ where L represents different number of context states. If L, the number of context states is close to the cardinality of the sensor set S, the runtime becomes $O(|S|^2)$. Lines 5 to 11: All simple operations consisting of additions, subtractions, comparisons, and conditional statements count for one unit of time. Line 12 & 13: For loop with runtime $O(L)$. Lines 14 to 17: Again all the simple operations and return statement count for one unit of time. Given the above analysis, the worst case and best case runtime for the Naive Heuristic is $O(|S|^2)$. In case of best case analysis, the only improvement over the worst case we can introduce is running the sorting function in $O(|S| \log |S|)$ time, though it does not help improve the asymptotic runtime of the Naive Heuristic from $O(|S|^2)$.

2) Range Heuristic (Figure 5/Algorithm 5): Line 1: Assignment statement runs in constant time. Line 2: Runtime of a sorting function such as quick sort is $O(|S|^2)$. Line 3: A single for loop has a runtime of $O(L)$. Lines 4 to 16: All simple operations consisting of additions, subtractions, comparisons, and conditional statements count for one unit of time. Lines 17 & 18: While loop accounts for $O(L)$. Lines 19 to 31: Again all the simple operations and return statement count for one unit of time. Given the above analysis, the worst case runtime for the Range Heuristic is $O(|S|^2)$. In case of best case analysis, the improvement over the worst case we can have is to improve the runtime of the sorting function to $O(|S| \log |S|)$ time, which helps to boost the overall asymptotic runtime of the Range Heuristic from $O(|S|^2)$ to $O(|S| \log |S|)$. We can further improve this by employing a linear time sorting algorithm such as bucket sort; as a consequence, the best case time complexity of the Range Heuristic turns out to be $O(|S|)$.
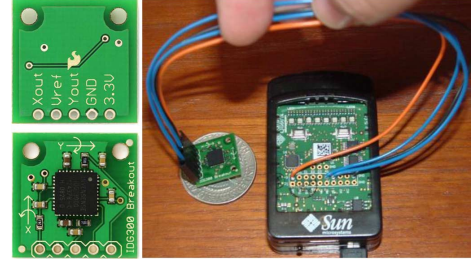

Fig. 6. Single chip dual-axis gyro sensor with SunSPOT.

TABLE I
CALIBRATED ACCELEROMETER (ACCEL.) SAMPLE VALUES

Range ($5^{th} - 95^{th}$ percentile) of Tilt Values (in degree)	Context State
85.21 to 83.33	*Sitting*
68.40 to 33.09	*Walking*
28.00 to -15.60	*Running*

V. EXPERIMENTAL STUDY AND RESULTS

To validate our heuristic approaches and understand the interplay between the tolerance ranges ($\{q_i\}$) and inferencing error, we have performed experiments with SunSPOT[2] and Shimmer sensors.[3] Specifically, we have used a 3-axis accelerometer, a light, and an embedded external gyro sensor on the SunSPOT platform and a 3-axis accelerometer and gyro sensor on the Shimmer platform. Fig. 6 shows the gyro sensor on the left and the wiring of the gyro with the SunSPOT on the right. For our experimental studies, we have utilized readings from these varied sensors to characterize multiple activity contexts such as $Sitting$, $Walking$, and $Running$. We used this setup to evaluate a medical monitoring application that infers a patient's *activity* using different types of sensors.

A. Initial Setup and Methodology

We first describe the capabilities of our experimental platforms and characterize the sensors' values. We used the built-in accelerometer to measure the tilt of the SunSPOT (in degrees) when an individual user was in three different context states: $Sitting$, $Walking$, and $Running$; this is similar to what the commodity sensor Fitbit[4] does. From the collected samples, we computed the 5th and 95th percentile of the tilt readings (normalized X-acceleration values in degrees) corresponding to each context state. Table I shows the resulting ranges in the accelerometer tilt readings observed for each of the three states.There is indeed an observable separation in the ranges of tilt values, i.e., context states can be distinguished with reasonable accuracy even under moderate uncertainty.

We also used the SunSPOT light sensor to measure light intensity for different contexts. Intuitively, low values of light intensity may indicate a *Sleeping* (or inactive) state, while higher values are likely to indicate that the user is *Active* (i.e., $Sitting$, $Walking$, or $Running$). Table II shows the observed ranges for light intensity for these two states. The accuracy of an activity context from the light sensor is much lower, as users may reasonably be *Active* in low light.

---

[2]www.sunspotworld.com

[3]http://www.shimmer-research.com/

[4]http://www.fitbit.com

TABLE II
LIGHT SENSOR SAMPLE VALUES

Average Range of Light Intensity (lumen)	Context State
LightSensor.getValue() = 10 to 50	Turned on $\rightarrow$ *active*
LightSensor.getValue() = 0 to 1	Turned off $\rightarrow$ *sleeping*

TABLE III
GYRO SENSOR SAMPLE VALUES

Average Angular Rate of Rotation (degs/sec)	Context State
X-Rotational Variation (Xout) = 150 to 350	*Walking*
Y-Rotational Variation (Yout) = 150 to 350	
X-Rotational Variation (Xout) = 2 to 8	*Sitting*
Y-Rotational Variation (Yout) = 2 to 8	

Finally, we used the external gyro sensor to measure the variation of the rate of angular rotation when the user is in the $Walking$ and $Sitting$ states. As the gyro is quite sensitive, we use a calibration function to measure the range of average angular rate variation for both states, as shown in Table III.

Given these sensors, we constructed an experimental methodology that relies on event-based data collection emulating an activity monitoring scenario deployed by wellness management professionals to monitor a user's daily activities. We used real collected traces from the SunSPOT wireless sensors. We recruited 5 participants who we instrumented with SunSPOT sensors mounted on their wrists. We logged data onto a laptop through a SunSPOT base station. To study the potential impact of our heuristics, we collected initial traces for the SunSPOT motion, light, and gyro sensors for five participants, who engaged in a *mix* of three activities ($Sitting$, $Walking$, and $Running$) for a period of three days. These participants were all adults with no known serious medical conditions but with differing levels of physical fitness. We then used an emulator to mimic the samples that a sensor would have reported given the trace and an assigned tolerance range and compared the context inferred from the values reported by the emulation against the ground truth. This trace-driven, event-based approach allows us to make meaningful comparisons, as the uncontrollable physiological and environmental variations would otherwise make it impossible to obtain the exact same data stream from two different sessions from the same user.

B. Results for Indistinct Context States

We next investigate various aspects of using the SunSPOT sensors to infer high level context. We ran the experiments on the data collected, $Sitting$, $Walking$, and $Running$ together; without partitioning the data with respect to the corresponding context states. Marginal decrease of $QoINF$ requirements can significantly reduce energy consumption. To quantify the impact of $QoINF$ adaptation on energy consumption, we define the power consumption of SunSPOT sensors using the average power of the eSPOT board (95 mA) and application daughter board (400 mA) in run mode and a wireless $CC2420$ radio transmission power (18 mA). Based on (12), we calculate the average power consumption of each sensor for a given communication frequency.

Power = Power of eSPOT + Power of daughter

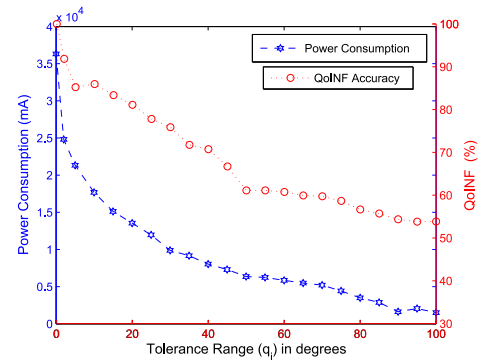$+$(Radio Transmission Power $\times$ Comm. Frequency) (12)
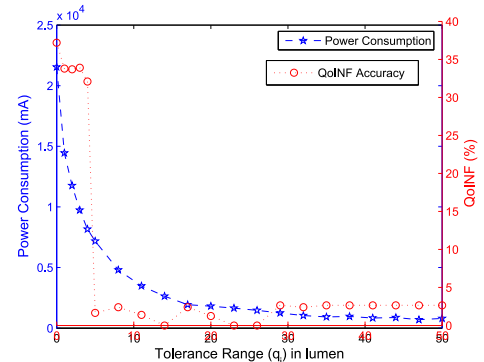

Fig. 7. Motion sensor.
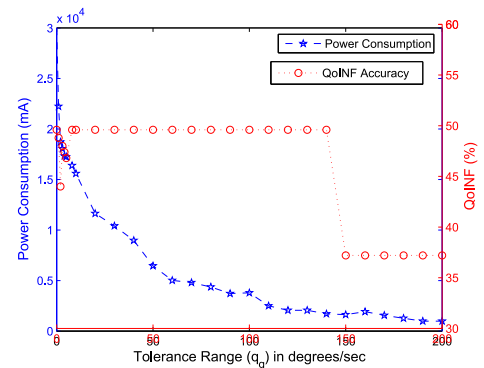

Fig. 8. Light sensor.

Fig. 9. Gyro sensor.

Figs. 7, 9, and 8 show the power consumption and the $QoINF$ accuracy achieved for different values of tolerance range (q_m) for the motion, light, and gyro sensors, respectively. In general, there is a continuous drop in both the power consumption and $QoINF$ as q increases for all three types of sensors. For the motion sensor, a $QoINF$ accuracy of $\approx$81% is achieved for $q = 20$; using this tolerance range reduces the sensor power consumption by $\approx$63% (3.65×10^4 mA $\rightarrow$ 1.35×10^4 mA). This suggests that it is indeed possible to achieve significant savings in energy consumption if one is willing to tolerate a marginal degradation in accuracy. A similar behavior is observed for the light sensor where $q = 4$ incurs a 5% loss in accuracy vs. $\approx$62% reduction in power consumption (2.1×10^4 mA $\rightarrow$ 0.81×10^4 mA). However, as the difference between the light intensity (lumen) ranges for *Active* versus *Sleeping* is only $\approx$10 (Table II), increasing q beyond $\approx$10 leads to a sharp fall in $QoINF$. Similarly, for the gyro sensor, increasing q leads to a decrease in power consumption although the $QoINF$ accuracy remains constant $\approx$49% after it crosses the ranges for $Sitting$ at $\approx$8. For $q = 8$, $QoINF$ accuracy remains $\approx$49%, whereas the
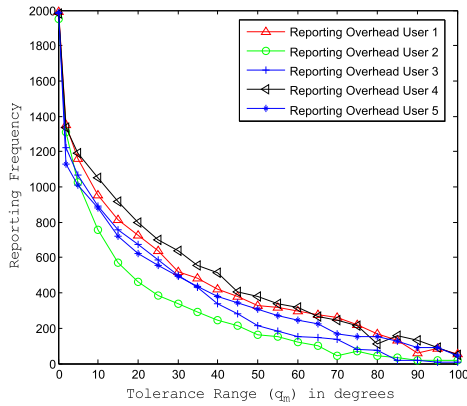
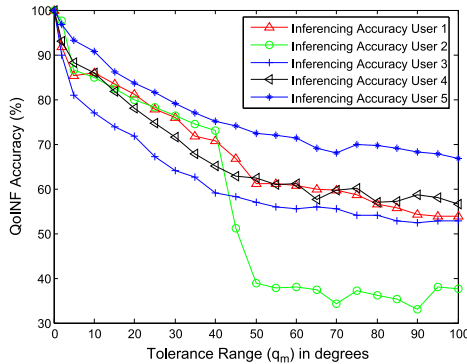Fig. 10. Communication overhead vs. tolerance for motion sensor.

Fig. 11. Accuracy vs. tolerance for motion sensor.

power consumption is reduced to almost half (3.0×10^4 mA $\rightarrow$ 1.6×10^4 mA).

These results demonstrate the applicability of our framework and model for multi-context recognition. Let an application specify $QoINF_{min} \approx 80\%$ and be interested in recognizing all three activity contexts ($Sitting$, $Walking$, and $Running$) simultaneously with an objective of reducing the energy cost. We can conclude that the use of the motion sensor with a tolerance range of $q = 20$ would achieve a $QoINF > 80\%$ with a cost (energy) reduction of $\approx 63\%$.

Personalizing $QoINF$ functions can help maximize the benefits of quality-aware context sensing. We investigated the tradeoff between tolerance ranges and $QoINF$ for the five participants in our experiments. The goal was to study the sensitivity of the tradeoff to individualized activity patterns; we focused on results from only a single sensor (the SunSPOT motion sensor). We replayed the traces through the emulator as before. Figs. 10 and 11 depict the variation, across users, in the communication overhead and inferencing accuracy, as a function of the tolerance range. There are clearly significant differences in the $QoINF$ accuracy across users; in particular, user 2 has a much sharper drop in $QoINF$ once the tolerance range exceeds 40 degrees.

These figures suggest that *personalizing* the $QoINF$ function is important in maximizing the benefits of inference-quality aware sensing. However, even in the absence of personalization, the benefits from quality-aware context inference are significant. For example, if a tolerance range of $q = 20$ is applied to all users, the lower bound of the accuracy achieved is $\approx 71\%$ (for User 3); at the same tolerance range, the worst case (smallest) reduction in the reporting overhead is observed to be $\approx 60\%$ (for User 4).

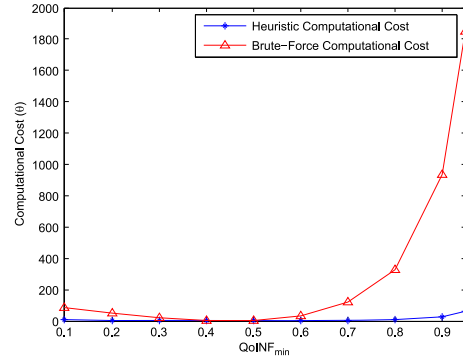| Sensor	η_i	ν_i
Motion	0.0512	1.8474
Light	3.6553	1.2288
Gyro	17.2512	1.8488

Fig. 12. Heuristic and brute-force compute costs.

The heuristic algorithm (Algorithm 4) which approximates the target $QoINF$ accuracy well. Our initial results demonstrate that significant savings can be achieved by relaxing the tolerance of each sensor without compromising the context estimation accuracy. To validate and quantify these benefits, we apply our results to our formal $QoINF$ model. First we fit the $QoINF(.)$ accuracy versus q curves from Figs. 7, 8, and 9 to the inverse exponential model of (5). After obtaining the best parametric fit (using a linear least squares regressor), we use the heuristic to compute the q values for a target $QoINF_{min}$ and then use additional traces to verify if this approach can provide the required accuracy.

We further compare the performance of our heuristic to the brute-force technique. While the latter uses the least-square estimator to compute the "best" inverse-exponential $QoINF$ function, it uses an exhaustive search over the $2^n - 1$ possible combinations, where n is the number of sensors. Table IV shows the estimated coefficients (η_i and ν_i) for each of the sensors for User 1, based on the empirical data. We then use the heuristic and brute-force algorithms to compute the optimal sensor set ($\hat{\theta}$) and associated tolerance ranges $Q(\hat{\theta})$ that minimize the communication overhead for a target $QoINF$.

Fig. 12 compares the computational cost of the heuristic and brute-force methods. Our heuristic always incurs lower cumulative computational cost than the brute-force search. It converges much more quickly than brute-force when the target $QoINF$ value is high or moderately high (≥ 0.6). At higher values of $QoINF$, the heuristic search substantially decreases the computational cost (about 97% for $QoINF = 0.9$).

We plot the minimum cost for the heuristic and brute-force results for different values of $QoINF_{min}$ in Fig. 13; as expected, the brute-force method outperforms the heuristic in all cases. In these experiments, the sensor set selected by the heuristic remains simply $\{motion\}$ until $QoINF_{min} < 0.70$ since the inclusion of other sensors incurs an increase in cost. Beyond that point, the heuristic selects both motion and light sensors. In contrast, in the brute-force approach, at the minimum cost point, the optimal subset of sensors remains $\{motion\}$ until $QoINF_{min} \leq 0.50$. The optimal set becomes $\{motion, light\}$
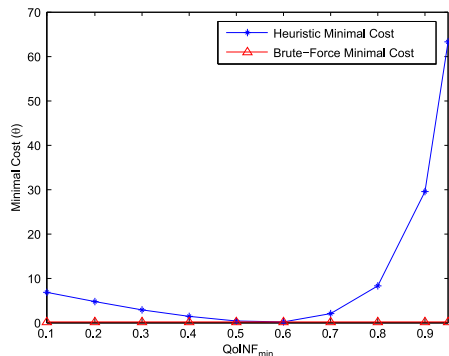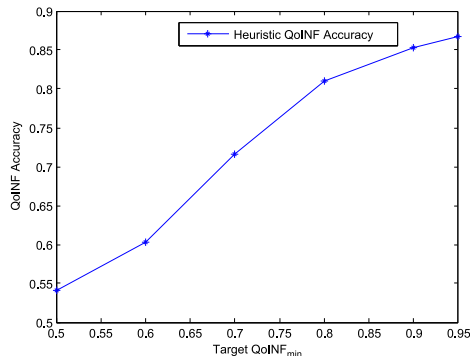
Fig. 13. Heuristic & brute-force minimal costs.

Fig. 14. Heuristic on achieving target $QoINF$.

TABLE V
CALIBRATED SHIMMER ACCELEROMETER SAMPLE VALUES

Range ($5^{th} - 95^{th}$ percentile) of Acceleration Values (in m/sec^2)	Context State
0.60 to 0.64	*Sitting*
0.59 to 0.70	*Walking*
0.46 to 0.76	*Running*

TABLE VI
CALIBRATED SHIMMER GYRO SAMPLE VALUES

Range ($5^{th} - 95^{th}$ percentile) of X-Angular Rotation Values (in $degs/sec$)	Context State
229.98 to 224.98	*Sitting*
246.83 to 221.68	*Walking*
276.15 to 207.10	*Running*

95th percentiles to determine the range for different context states as shown in Table V. Similarly, for the Shimmer gyro sensor, considering the sensitivity factor as 2 mV/deg/s and full-scale range as ($\pm$ 500 deg/s) based on the data sheet,[6] we convert the ADC output from the Shimmer gyro using (ADC $value \times (1000/4096) \times (1/2)$) degrees/sec. The value of the X rotational output has been used to calculate the 5th and 95th percentile as shown in Table VI.

Communication overhead and $QoINF$ accuracy decrease with increasing tolerance ranges for distinct context states. We ran the data traces through the emulator to determine the $QoINF$ accuracy and sensor reporting overheads with respect to the tolerance ranges for different context states. The results are plotted in Figs. 15, 16, and 17 for the SunSPOT accelerometer, in Figs. 18, 19, and 20 for the Shimmer accelerometer; and in Figs. 21, 22, and 23 for the Shimmer gyro.

In general, we do notice that there is a continuous drop in reporting overhead and $QoINF$ accuracy with the increase in tolerance ranges as before. In some cases, due to the specificity in the percentile interval of the sensor values for a specific context and the selected step-size of the tolerance ranges, the results do not follow the continuous drop in reporting overhead and $QoINF$ accuracy with the increase in the tolerance ranges. We have also plotted the 95% confidence interval of $QoINF$ accuracy for $Sitting$, $Walking$, and $Running$ context states respectively for the Shimmer motion and gyro sensors in Figs. 24 and 25. For the Shimmer motion sensor, the mean $QoINF$ values do not vary beyond $\pm9\%$ for a confidence interval of 95%. Similarly for the Shimmer gyro sensor, the mean $QoINF$ values vary within a $\pm7\%$ bound for a confidence interval of 95%. We fit each of these $QoINF$ accuracy curves with respective tolerance ranges to the inverse exponential model of (5) to determine the sensitivity factors of each sensor for different contexts (Table VII). In a few cases, we have to make an approximation (with moderate or large RMS error) in our regression techniques to fit those $QoINF(.)$ accuracy versus q curves and thus determine the sensitivity factors of the sensors.

Our range-based heuristic can achieve application-specified quality and reduce network resource usage substantially. We compare our range-based heuristic (Algorithm 5) with the naïve heuristic (Algorithm 4) and brute-force

when $0.50 < QoINF_{min} < 0.70$, and beyond that the optimal set becomes $\{motion, light, gyro\}$.

More importantly, the heuristic is able to accurately attain the target $QoINF$ at the minimum cost. For a given target $QoINF_{min}$, we first calculate the tolerance range q for all sensors and the minimum cost point, and then determine the optimal subset θ associated with that minimum cost. Using the determined θ and individual q values, we calculate the empirically observed $QoINF$ values, which are plotted in Fig. 14 against the target $QoINF$ values. We observe that the heuristic approach is able to approximate the target objectives well; in particular, the inference accuracy observed by the heuristic is no more than 5% lower than the target $QoINF$.

C. Results for Distinct Context States

In the next set of experiments, we collect data for different context states of the user simultaneously from both the SunSPOT (accelerometer) and Shimmer (accelerometer and gyro) sensor platforms. For the SunSPOT, we follow the same procedure as before but this time we collect different context states separately. We ran the experiments on the partitioned data stored distinctly for individual context states, $Sitting$, $Walking$, and $Running$. The Shimmer was attached on the back of the user's shoe in a vertically downward-facing position. The raw data of X, Y and Z acceleration from the Shimmer ADC have been converted to m/sec^2. We consider the ADC output 4096 as equivalent to the maximum voltage reading of 1200 mV based on the Shimmer accelerometer data sheet.[5] Also as $1g = 9.81$ m/sec^2; equivalent to 800 mV, we convert each ADC output to m/sec^2 using (ADC $value \times (1200/4096) \times (9.81/800)$). We also normalize the X acceleration values and calculate the 5th and

---

[5]www.sparkfun.com/datasheets/Accelerometers/MMA7260Q-Rev1.pdf

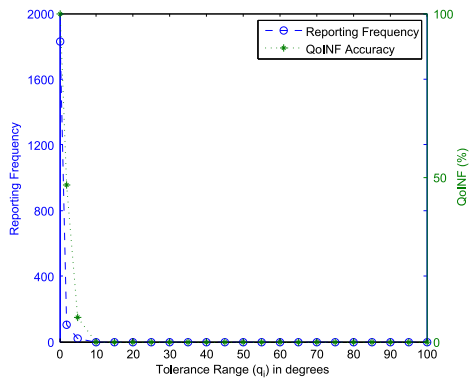[6]www.invensense.com/mems/gyro/documents/PS-IDG-0500-00-06.pdf

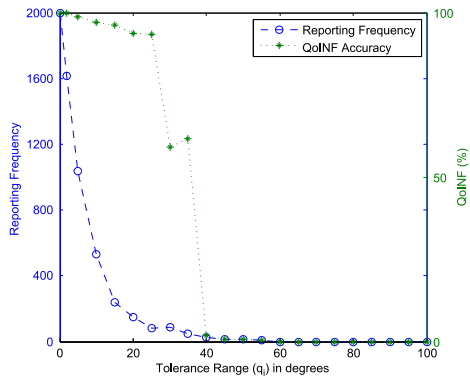Fig. 15. Communication overhead & accuracy vs. tolerance for SunSPOT accel. (Sitting).

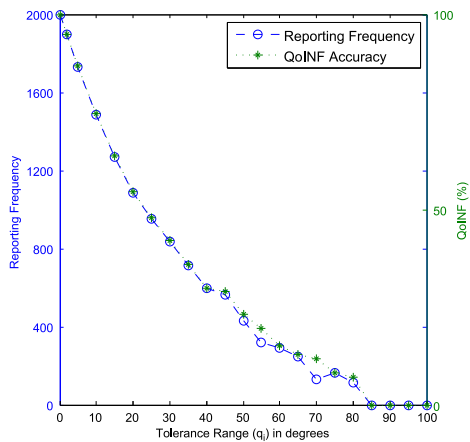Fig. 16. Communication overhead & accuracy vs. tolerance for sunspot accel. (Walking).

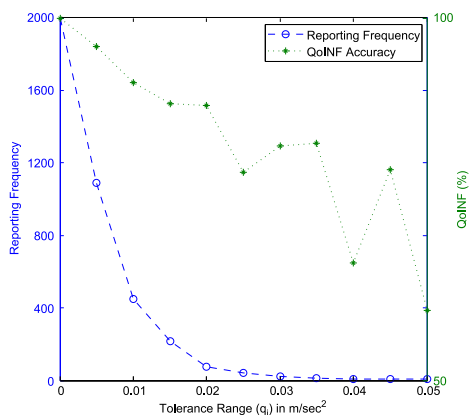Fig. 17. Communication overhead & accuracy vs. tolerance forSunSPOT accel. (Running).

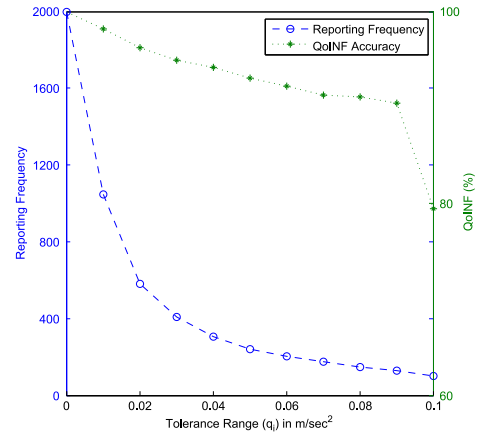Fig. 18. Communication overhead & accuracy vs. tolerance for shimmer accel. (Sitting).

Fig. 19. Communication overhead & accuracy vs. tolerance for shimmer accel. (Walking).
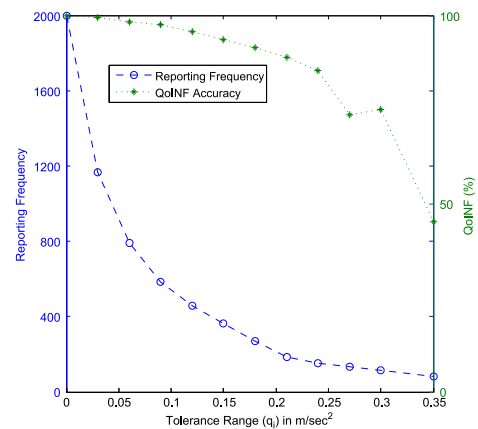

Fig. 20. Communication overhead & accuracy vs. tolerance forshimmer accel. (Running).
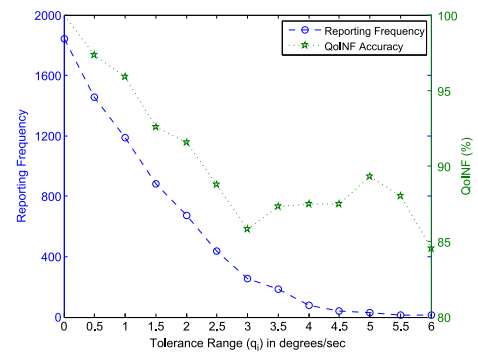

Fig. 21. Communication overhead & accuracy vs. tolerance for shimmer gyro (sitting).
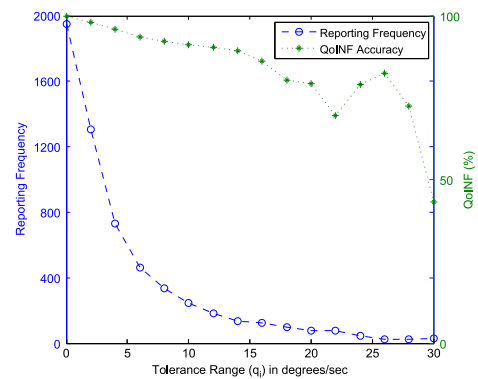

Fig. 22. Communication overhead & accuracy vs. tolerance for shimmer gyro (walking).

TABLE VII
η_i AND ν_i VALUES OBTAINED BY CURVE FITTING

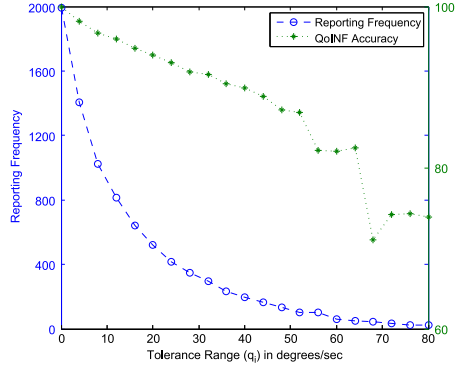| User	Sensor	Sitting		Walking		Running	
		η_i	ν_i	η_i	ν_i	η_i	ν_i
User1	SunSPOT Accel	1.5263	0.0123	0.9952	0.3076	2.4119	0.3074
	Shimmer Accel	0.0998	0.0123	0.5113	0.0125	0.5816	0.0114
	Shimmer Gyro	32.9067	0.0129	1.0007	0.1533	0.7186	0.2343

Fig. 23. Communication overhead & accuracy vs. tolerance for shimmer gyro (running).
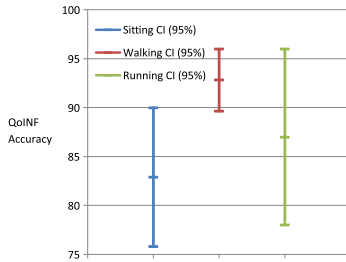

Fig. 24. 95% confidence interval of QoINF accuracy for shimmer accel.
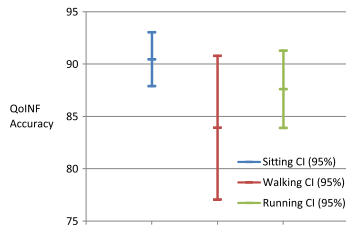

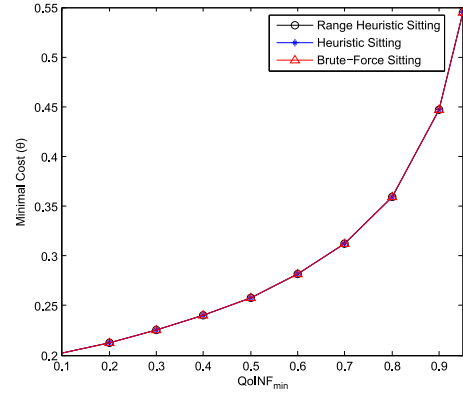Fig. 25. 95% confidence interval of QoINF accuracy for shimmer gyro.

Fig. 26. Range heuristic, heuristic & brute-force minimal cost comparison for *sitting*.
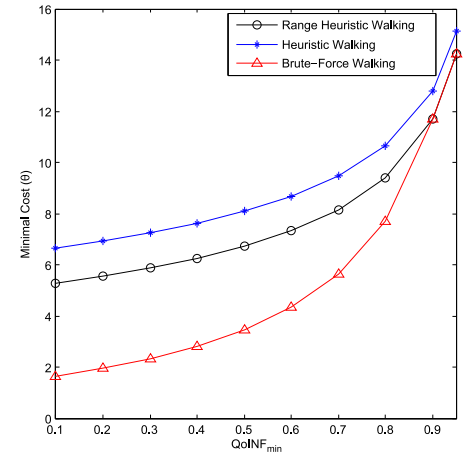
Fig. 27. Range heuristic, heuristic & brute-force minimal cost comparison for *walking*.

search. Based on the derived sensitivity factors, we sort all of the sensors and generate the following sorted lists; $\Re_1 = \{$Shimmer Accel, SunSPOT Accel, Shimmer Gyro$\}$ for context $Sitting$; $\Re_2 = \{$SunSPOT Accel, Shimmer Gyro, Shimmer Accel$\}$ for context $Walking$; and $\Re_3 = \{$Shimmer Gyro, Shimmer Accel, SunSPOT Accel$\}$ for context $Running$. We use each approach to compute the optimal sensor set $(\hat{\theta})$ and associated tolerance ranges $Q(\hat{\theta})$ that minimize the $\widehat{COST}(\theta)$ for a target $QoINF$. We also use the range-heuristic to compute the q values for a target $QoINF_{min}$ and then use those q values to determine the achievable $QoINF$.

Fig. 26 plots the minimal cost associated with the three search methods to determine the optimal subset of sensors and their tolerance ranges for the first context state considered, $Sitting$ in our case. In this example, the range-based heuristic and heuristic perform exactly as the brute force in finding the optimal sensor subset in minimum cost. Fig. 27 compares the performance of these three algorithms for the context $Walking$. The range-based heuristic performs better than the heuristic, and it performs close to brute-force. Similarly, Fig. 28 plots the performance for the $Running$ context, where again the range-based heuristic algorithm out-performs the naïve heuristic. Due to the simple set theoretic addition of sensors from one context to another (without examining the existing sensor set's satisfiability for the new context) in the heuristic algorithm, we observe that first just the Shimmer accelerometer has been selected for the $Sitting$ context; then for $Walking$ both the Shimmer and SunSPOT accelerometers have been selected; and then for $Running$ all three available sensors have been chosen. In the range-based heuristic, only the Shimmer accelerometer is selected for all the contexts at the minimal cost by tightening the tolerance range.

We also evaluate our range based heuristic's ability to attain the application's desired $QoINF$. First we calculate the tolerance ranges for the chosen optimal subset of sensors at minimal cost. Then with those specified tolerance ranges and the determined sensor set, we run our emulation on the already
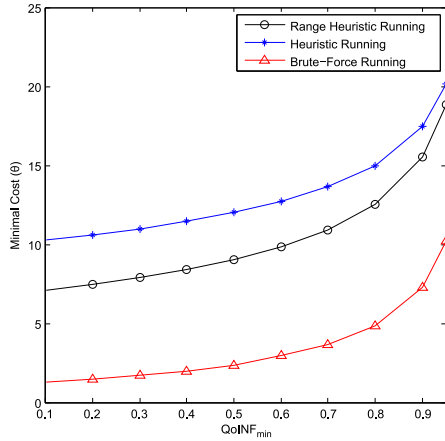
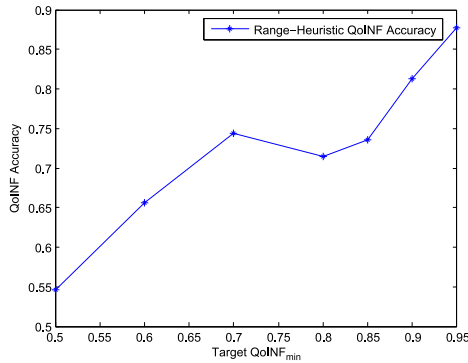Fig. 28. Range heuristic, heuristic & brute-forceminimal cost comparison for *running*.

Fig. 29. Range-based heuristic on achieving target $QoINF$ for *running*.

collected data traces to determine the empirically achieved accuracy of the algorithm. Fig. 29 plots the $QoINF$ achieved by the range-based heuristic algorithm for the context *running*. The range-based heuristic performs well at no more than 10% lower than the target $QoINF$. Nevertheless, we do notice that our range-based heuristic does not perform well in achieving target $QoINF$ accuracy for the other two context states. We believe this is a result of the large approximation in our curve fitting approach. This incurs errors in determining the sensitivity factors and therefore introduces a larger deviation in the q values, which ultimately affects the attainable $QoINF$ accuracy of the range-heuristic algorithm with respect to the target $QoINF$ metric. Adding more sensors to the selection process (as is likely in future pervasive computing scenarios) would be expected to help mitigate this challenge.

D. Performance Evaluation in Smart Home Environments

We evaluated the performance of our Quality-of-Inference (QoINF)-aware model in smart home environments. Our experiments are conducted with real-life smartphone sensor data traces. We have recruited 10 participants and asked them to perform a variety of activities of daily living (ADLs) and instrumental activities of daily living (IADLs) in their own home environment.

App Development and Data Collection. An android application has been developed to collect accelerometer and gyroscope data from an Android based Google Nexus smart phone device for monitoring the activity of a typical user. We included the

TABLE VIII
FEATURE EXTRACTED FROM THE RAW DATA

Feature Name	Definition
Mean	AVG ($\sum x_i$); AVG ($\sum y_i$);AVG ($\sum z_i$)
Mean-Magnitude	AVG ($\sqrt{x_i^2 + y_i^2 + z_i^2}$)
Magnitude-Mean	$\sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}$
Max, Min, Zero-Cross	*max, min, zero-cross*
Variance	VAR ($\sum x_i$); VAR ($\sum y_i$); VAR ($\sum z_i$)
Correlation	$corr(x,y) = \frac{cov(x,y)}{\sigma_x, \sigma_y}$

feature to collect data at different sampling frequencies. To collect the ground truth, we created an option in the app such that the user can label the activity by its semantic name. The user has also the option to label the activity before performing the intended ADLs/IADLs and activating the sensor app for data collection purposes. The user has been asked to keep the smartphone in their pants pocket during the staged experiments.

Activities and Data Processing. We collected samples of ten users performing a variety of low-level and high-level activities. We have collected data for 6 low-level activities first and asked the user to label those as: $\{sitting, standing, walking, running, lying, bending\}$. Similary we have collected data for 6 high-level activities and asked the user to label those as: $\{cleaning, cooking, medication, sweeping, washing\ hands, watering\ plants\}$. We collected samples for time periods between five to sixty minutes based on a specific activity, with sensor data collected respectively at 80 Hz, 70 Hz, 60 Hz, 50 Hz, 40 Hz, 30 Hz, 20 Hz and 10 Hz.

QoINF-Aware Activity Classification. We investigate the behavior of the QoINF-aware activity recognition model in the presence of traditional classification and machine learning algorithms. Our objective in this specific set of experiments is two fold. First, we note the interplay between sampling frequency or sensor sending rate reduction with activity recognition accuracy; second, we quantify the role of sampling frequency reduction, particularly on low-level and high-level activity classification.

We apply feature-based classification techniques such as Multi-layer Perceptron for classifying both the low-level and high-level activities in the presence of a different subset of sensors at different sampling frequencies. The 3-axis accelerometer and the 3-axis gyroscope data streams were broken up into successive $frames$ for different sampling frequency levels. A 30-dimensional feature vector as shown in Table VIII was computed over each frame. The ground-truth annotated training set (aggregated across all 10 participants) was then fed into the classification algorithm (Multi-layer Perceptron) in the presence of various combinations of smartphone sensors, sequence of low-level and high-level activities, and sampling frequency levels. The accuracy of the classifiers was tested using 10-fold cross-validation. Fig. 30 and Fig. 31 plot the average classification accuracy for the low-level and high-level activities at different sending rates; we see that low-level activities such as sitting, standing, walking etc., are classified with higher accuracy even at a lower frequency because they are simple locomotive context states. On the other hand, because the high-level activities such as cooking, medication, sweeping, washing hands, watering plants etc., are composed
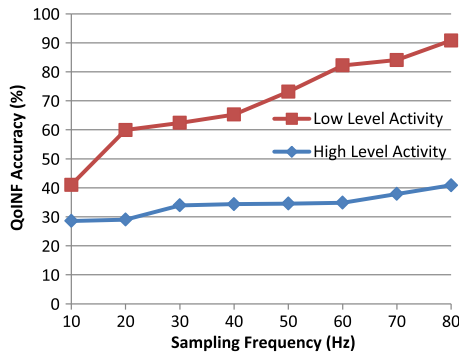
Fig. 30. QoINF (classification) accuracy vs. sending rate for smartphone accelerometer.
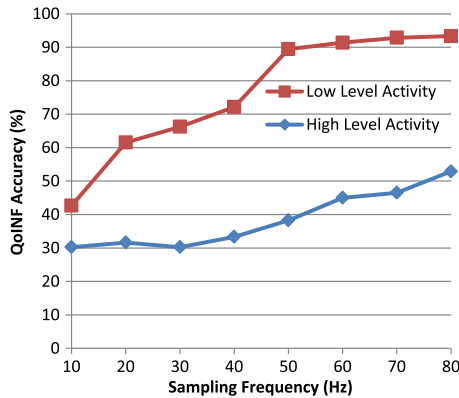

Fig. 31. QoINF (classification) accuracy vs. sending rate for smartphone gyroscope.
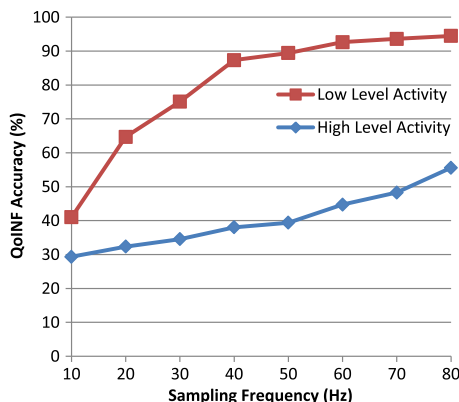

Fig. 32. QoINF (classification) accuracy vs. sending rate for smartphone accelerometer and gyroscope.

of several of low-level activities, the classification accuracy increases monotonically with the increase of the sending rate. It should be noted here that, in general, the high-level activities are not well recognized by only the accelerometer, gyro, or a combination of both. Next we run the experiments for both the accelerometer and gyroscope smartphone sensor and observe the similar behavior. We combine the features from both the accelerometer and gyroscope sensor and use them jointly to understand the effect of sending rate reduction with low- and high-level activity recognition accuracy (Fig. 32). In presence of multiple sensors, low-level activities are classified better at a lower sampling frequency compared to the individual sensor-activity recognition cases.

Looking at Fig. 30 and Fig. 31, we observe that classifying the low-level single state activity (e.g., sitting, standing, walking,

etc.) with moderate accuracy (≈70%) can be achieved with a lower sampling frequency (≈40 Hz) compared to high-level activity classification using the individual accelerometer or gyroscope sensor on the smartphone. Fig. 32 shows that combining the accelerometer with gyroscope helps improve the classification accuracy (≈70% to 90%) of low-level activity while holding the same sampling frequency (≈40 Hz). This corroborates the fact that instead of changing the tolerance range of a preselected sensor (in our case smartphone accelerometer/gyro) adding another sensor may provide more help in improving the classification accuracy of certain context states. Note that for high-level activity recognition we observe only a slight improvement in classification accuracy (≈30% to 40%) with increasing sending rate as shown in Fig. 32.

VI. RELATED WORK

The tradeoff between communication overhead and the quality of reconstructed data was first studied in [13], which envisioned the effect of tolerance ranges on the frequency of sink-initiated fetching vs. source-initiated proactive refreshing. The focus, however, was on snapshot queries and not on *continually satisfying* the $QoINF$ bound of a long-standing subscription. In [5] the authors developed a distributed regression based method for modeling sensor data in wireless network in order to reduce the amount of data transmitted wirelessly while maintaining more complete information about the original data than most aggregation schemes (e.g., averages, maxima, histograms). The idea of exploiting temporal correlation across successive samples of individual sensors for reducing communication overhead for snapshot queries is addressed in [2], which used training data to parameterize a jointly-normal density function. While a precursor to our work, the focus there was on meeting the $QoINF$ requirements for a class of aggregation queries, whereas in this paper we focus on arbitrary relationships between a context variable and the underlying data. A near optimal sensor placement model has been introduced in [9], which maximizes the information gain while minimizes the communication cost regardless of the applications. The authors therein proposed a probabilistic framework of Gaussian Processes not only to model the monitored phenomena, but also to predict communication costs and a polynomial time algorithm for selecting sensor placements at informative and cost-effective locations. Entropy-based sensor selection heuristic algorithms have been proposed in [3], [11], [20]. These works have taken an information theoretic approach for specific application scenarios, where the belief state of the target value is gradually improved by repeatedly selecting the most informative unused sensor until the required accuracy is achieved. The CAPS algorithm [7] is designed for long-running aggregation queries (such as $\{min, max\}$) and computes the optimal set of tolerance ranges for a given set of sensors that minimizes the communication overhead, while guaranteeing the accuracy of the computed response. In contrast, our objective here is to compute *both* the optimal subset of available sensors and their tolerance ranges to achieve the desired accuracy for arbitrary context variables. A probabilistic model to deduce the context prediction accuracy at different context abstraction levels has been proposed in [18]. The optimum processing order of feasible context processing operations that maximizes the expected context processing

accuracy was investigated therein. In [4], the authors developed a dependence mapping between contexts, features and sensors for context prediction and then used it to parameterize sensor usage, sampling and feature generation in order to reduce unnecessary power consumption of the sensors on board of a mobile device.

Additionally, quality of information (QoI) has been studied in the existing literature related to data collection and storage for database systems, with a focus on data consistency, completeness and currency [13]. However, the development of formal models of quality of information (or inference) has not been extensively addressed in the context of sensor-generated data streams. Recent work [22] has suggested the use of specialized QoI models to capture the accuracy of detecting transient events, given a set of sensors. However, this approach does not focus on the optimal joint selection of *sensors and their tolerance ranges*, which is the distinguishing characteristic of our work.

AdaSense [24] proposed a framework that reduces the BSN (Body-area Sensor Networks) sensors' sampling rate while meeting any user-specified accuracy requirements. A lower-power single activity detection sampling strategy and a higher-power multi-activity classification sampling strategy were proposed. To optimize the sampling rate for the sake of saving energy, a genetic programming based algorithm has been employed to select the optimal subset of features that effectively tunes the minimal sampling rate of multi-activity classification under any accuracy requirement. A similar integer programming based quality and energy-aware data acquisition for activity and locomotion recognition framework has been proposed in [23]. While this adaptive sampling [24] and optimization [23] strategies help balance the accuracy and energy efficiency, our approach exploits the sensor's context sensitivity and tolerance ranges to determine an optimal subset that holds across multiple contexts and prevents unnecessarily wasting wireless sensor network resources.

VII. CONCLUSIONS

We have presented a formal framework for energy-efficient recognition of contexts in pervasive computing environments. The key idea is to express the accuracy of context estimation, for arbitrary contextual attributes, through a quality of inference ($QoINF$) function that captures the dependence of estimation accuracy on the selected sensors and their specified tolerance ranges. Such a multi-context model in terms of the uncertainty range of the underlying sensor data streams has not been rigorously investigated in the past and holds promise for reducing the communication overhead in sensor data transmissions. We have described two multi-context search heuristic algorithms to solve the proposed optimization problem dealing with quality-versus-communication cost tradeoff. Experimental traces collected in a laboratory setting demonstrate the significance of our approach and establish that the proposed heuristic is able to provide a close-to-optimal tradeoff between the $QoINF$ value and the communication overhead, at the cost of only modest computational requirements. Experiments using a combination of motion, light and gyroscopic sensors show that our model-based approach provides a $QoINF$ value that is within $\approx 5\%$ of the desired target, but is able to achieve significant reduction (within $\approx 5\%$ of the theoretical minimum) in the communication cost.

APPENDIX

Proof of Lemma 1: The multi-context optimization problem can be stated as:

$$min \sum_{s_i \in \theta} \frac{h_i}{q_i^2}$$

subject to L different constraints, where the lth constraint is:

$$1 - \prod_{s_i \in \theta} \frac{1}{\nu_{il}} exp(-\frac{1}{\eta_{il} q_i}) \geq QoINF_{min}^l \tag{13}$$

Taking the logarithm of each constraint and setting up the Lagrangian, we obtain:

$$\sum_{s_i \in \theta} \frac{h_i}{q_i^2} + \sum_{l \in L} \lambda_l * \left(\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta} \log(\nu_{il}) \right.$$
$$\left. + \sum_{s_i \in \theta} \frac{1}{\eta_{il} q_i} \right) \tag{14}$$

Now taking their derivatives with respect to each q_i, we get:

$$\sum_{s_i \in \theta} \frac{2h_i}{q_i^3} + \sum_{l \in L} \sum_{s_i \in \theta} \frac{\lambda_l}{\eta_{il} q_i^2} = 0 \tag{15}$$

which yields:

$$q_i = \frac{2h_i}{\sum_{l \in L} \frac{\lambda_l}{\eta_{il}}} \tag{16}$$

Further derivatives of the Lagrangian with respect to λ_l yield,

$$\sum_{l \in L} \left(\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta} \log(\nu_{il}) + \sum_{s_i \in \theta} \frac{1}{\eta_{il} q_i} \right) = 0. \tag{17}$$

In order to derive the values of q_i, we solve together the set of L equations ((17)) and the set of $|\theta|$ equations ((15)). Rewriting (15) as follows:

$$\sum_{s_i \in \theta} \frac{2h_i}{q_i^2} + \sum_{l \in L} \lambda_l \sum_{s_i \in \theta} \frac{1}{\eta_{il} q_i} = 0 \tag{18}$$

and replacing $\sum_{s_i \in \theta} \frac{1}{\eta_{il} q_i}$ with the help of (17), we get:

$$\sum_{s_i \in \theta} \frac{2h_i}{q_i^2} - \sum_{l \in L} \lambda_l \left(\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta} \log(\nu_{il}) \right) = 0$$
$$\Rightarrow \sum_{l \in L} \lambda_l = \frac{\sum_{s_i \in \theta} \frac{2h_i}{q_i^2}}{\sum_{l \in L} \left(\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta} \log(\nu_{il}) \right)}$$

For the base case, when $\lambda = 1$,

$$\lambda_1 = \frac{\sum_{s_i \in \theta} \frac{2h_i}{q_i^2}}{\log(1 - QoINF_{min}^1) + \sum_{s_i \in \theta} \log(\nu_{i1})} \tag{19}$$

Similarly for λ_2,

$$\lambda_2 = \sum_{l=1}^{2} \lambda_l - \sum_{l=1}^{1} \lambda_l$$

$$\lambda_2 = \frac{\sum_{s_i \in \theta} \frac{2h_i}{q_i^2}}{\sum_{l=1}^{2} \left(\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta} \log(\nu_{il}) \right)}$$
$$- \frac{\sum_{s_i \in \theta} \frac{2h_i}{q_i^2}}{\log(1 - QoINF_{min}^1) + \sum_{s_i \in \theta} \log(\nu_{i1})}$$

$$\vdots$$

and for λ_L,

$$\lambda_L = \sum_{l=1}^{L} \lambda_l - \sum_{l=1}^{L-1} \lambda_l$$

$$\lambda_L = \frac{\sum_{s_i \in \theta} \frac{2h_i}{q_i^2}}{\sum_{l=1}^{L} \left(\log(1 - QoINF_{min}^l) + \sum_{s_i \in \theta} \log(\nu_{il}) \right)} - \frac{\sum_{s_i \in \theta} \frac{2h_i}{q_i^2}}{\sum_{l=1}^{L-1} \left(\log(1 - QoINF_{min}^1) + \sum_{s_i \in \theta} \log(\nu_{i1}) \right)}$$

ACKNOWLEDGMENT

REFERENCES

[1] O. Amft, "On the need for quality standards in activity recognition using ubiquitous sensors," in *Proc. Workshop How To Do Good Res. Activity Recogn.*, 2010, (in conjunction with Pervasive).

[2] A. Deshpande, C. Guestrin, S. Madden, J. M. Hellerstein, and W. Hong, "Model-based approximate querying in sensor networks," *VLDB J.*, vol. 14, no. 4, pp. 417–443, Nov. 2005.

[3] E. Ertin, J. Fisher, and L. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proc. IPSN*, Apr. 2003, pp. 405–416.

[4] D. Gordon, S. Sigg, Y. Ding, and M. Beigl, "Using prediction to conserve energy in recognition on mobile devices," in *Proc. PerCom Workshops*, Mar. 2011, pp. 364–367.

[5] C. Guestrin, P. Bodik, R. Thibaux, M. Paskin, and S. Madden, "Distributed regression: An efficient framework for modeling sensor network data," in *Proc IPSN*, Apr. 2004, pp. 1–10.

[6] W. Heinzelman, A. L. Murphy, H. S. Carvalho, and M. A. Perillo, "Middleware to support sensor network applications," *IEEE Netw.*, vol. 18, no. 1, pp. 6–14, Feb. 2004.

[7] W. Hu, A. Misra, and R. Shorey, "CAPS: Energy efficient processing of continuous aggregate queries in sensor networks," in *Proc. IEEE PerCom*, Mar. 2006.

[8] N. Katenka, E. Levina, and G. Michailidis, "Local vote decision fusion for target detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 329–338, Jan. 2008.

[9] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication cost," in *Proc. IPSN*, Apr. 2006, pp. 2–10.

[10] A. Krause, A. Smailagic, and D. Siewiorek, "Context-aware mobile computing: Learning context-dependent personal preferences from a wearable sensor array," *IEEE Trans. Mobile Comput.*, vol. 5, no. 2, pp. 113–127, Feb. 2006.

[11] J. Liu, J. Reich, and F. Zhao, "Collaborative in-network processing for target tracking," *EURASIP J. Appl. Signal Process.*, vol. 2003, pp. 378–391, Mar. 2003.

[12] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 4, pp. 462–472, Sep. 2005.

[13] C. Olston and J. Widom, "Offering a precision performance tradeoff for aggregation queries over replicated data," in *Proc. VLDB*, Sep. 2000, pp. 144–155.

[14] N. Roy, A. Misra, S. K. Das, and C. Julien, "Quality-of-Inference (QoINF)-aware context determination in assisted living environments," in *Proc. WiMD*, May 2009, pp. 27–32, (in conjunction with MobiHoc).

[15] N. Roy, A. Misra, C. Julien, S. K. Das, and J. Biswas, "An energy efficient quality adaptive multi-modal sensor framework for context recognition," in *Proc. IEEE PerCom*, Mar. 2011, pp. 63–73.

[16] N. Roy, S. K. Das, and C. Julien, "Resource-optimized quality-assured ambiguous context mediation framework in pervasive environments," *IEEE Trans. Mobile Comput.*, vol. 11, no. 2, pp. 219–229, Feb. 2012.

[17] N. Roy, A. Misra, and D. Cook, "Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments," *J. Ambient Intell. Humanized Comput.*, pp. 1–19, Jun. 2015.

[18] S. Sigg *et al.*, "Investigation of context prediction accuracy for different context abstraction levels," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1047–1059, Jun. 2012.

[19] X. Tang and J. Xu, "Extending network lifetime for precision-constrained data aggregation in wireless sensor networks," in *Proc. IEEE INFOCOM*, Apr. 2006, pp. 1–12.

[20] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. IPSN*, Apr. 2004, pp. 36–45.

[21] Z. Yan, V. Subbaraju, D. Chakraborty, A. Misra, and K. Aberer, "Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach," in *Proc. ISWC*, April 2012, pp. 17–24.

[22] S. Zahedi, M. B. Srivastava, and C. Bisdikian, "A framework for quality of information analysis for detection-oriented sensor network deployments," in *Proc. IEEE MILCOM*, Nov. 2008, pp. 1–7.

[23] N. H. Viet, E. Munthe-Kaas, and T. Plagemann, "Quality and energy aware data acquisition for activity and locomotion recognition," in *Proc. IEEE PERCOM Workshops*, 2013, pp. 493–498.

[24] X. Qi, M. Keally, G. Zhou, Y. Li, and Z. Ren, "AdaSense: Adapting sampling rates for activity recognition in body sensor networks," in *Proc. RTAS*, 2013, pp. 163–172.

Nirmalya Roy is an Assistant Professor in the Department of Information Systems at the University of Maryland, Baltimore County (UMBC). His current research interests include pervasive healthcare, sensor-driven health and green technologies, design and modeling of smart environments, and mobile and pervasive systems. He obtained his Ph.D. and M.S. in computer science and engineering from the University of Texas at Arlington in 2008 and 2004 respectively. He received his Bachelors in computer science and engineering from Jadavpur University, India in 2001.

Archan Misra is an Associate Professor of Information Systems at Singapore Management University (SMU) and a Deputy Director of the joint CMU-SMU Living Analytics Research Center (LARC) with broad research interests in the areas of pervasive computing and mobile systems, including energy-efficient mobile analytics and activity recognition. Archan received his Ph.D. in electrical and computer engineering from the University of Maryland at College Park in May, 2000, and his B.Tech. in electronics and communication engineering from IIT Kharagpur, India in July 1993.

Sajal K. Das is the Chair of Computer Science and Daniel St. Clair Endowed Chair at the Missouri University of Science and Technology, Rolla. His current research interests include theory and practice of wireless and sensor networks, mobile and pervasive computing, cyber-physical systems and smart environments including smart grid and smart healthcare, distributed and cloud computing, security and privacy, biological and social networks, applied graph theory and game theory. He is an IEEE Fellow.

Christine Julien is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin. Her research work focuses on the intersection of software engineering and dynamic, unpredictable networked environments. She earned her D.Sc. in 2004, M.S. in 2003 and B.S. with majors in computer science and biology in 2000 from Washington University in Saint Louis.