Chapter XX: Pervasive Sensing and Monitoring for Situational Awareness

Sharad Mehrotra, Nalini Venkatasubramanian
Dept of Computer Science
University of California, Irvine

Mark-Oliver Stehr, Carolyn Talcott
SRI Intl., Menlo Park, USA

1. Introduction

Advances in sensing and multimedia data capture technologies coupled with mechanisms for low power wireless networking have enabled the possibility of creating deeply instrumented cyber-physical spaces. Embedded sensors and data capture devices in such environments enable the possibility of digitally capturing the state of the evolving physical systems and processes which can then be used to gain situational awareness of the activities in the instrumented space. Situational awareness, in a broad sense, refers to a continuum of knowledge that captures the current state of the physical environments being observed, to future projected states of these observed environments. Such awareness is created through processing of data from the sensed environment. Deeply instrumented physical spaces generate sensor data that is used to create digital representations of the physical world, which can then be used for to implement new functionalities or improve existing ones, and to adapt the configuration of the system itself – we refer to such cyberphysical spaces as ***sentient*** spaces. Sentient spaces embody the reflective design principle of "observe-analyze-adapt" wherein a system continuously observes its state in order to adapt its behavior (based on its state). Such adaptations may be at the system level (e.g., adjustment of network parameters to enable more effective information collection, or at the application level to achieve new functionalities or to optimize overall application goals (e.g., automated control of devices based on user behavior to conserve energy). Examples of sentient space applications in the infrastructure security domain include:

- surveillance systems for critical infrastructures such as ports and nuclear facilities or societal spaces such as malls, schools and buildings and
- emergency response systems that provide incident situational awareness during unexpected disasters such as fires, floods.

Sentient spaces offer unprecedented opportunities to bring IT-driven adaptations and control to variety of societal systems in application domains such as energy management, building design, transportation, avionics, agriculture, water management, infrastructure lifelines, etc.

The goal of this chapter is to identify fundamental challenges in building large-scale sentient spaces. Before we discuss challenges and describe emerging technological

advances to address them, we briefly discuss existing work on data streaming systems and sensor networks.

Stream Processing Engines and Sensor Networks: Over the past decade, various stream processing engines (SPEs) like TelegraphCQ[CHA03], STREAMS[ARA04], S3[Ham04], Cayuga[Bre07], Aurora[Car07], Borealis[Ara05], and MedSMan[Liu05] have been proposed in the literature and many related commercial products have been developed (e.g., S4 by Yahoo). Such systems provide on-the-fly techniques to resolving continuous queries and performing analyses on the data streams prior to (or instead of) storing the streaming data into the database. Such approaches are in contrast to traditional database approach wherein streaming data would be first stored into a database and queried/analyzed later. With the exception of Aurora and Borealis, many stream processing systems have focused on providing support for SQL-like queries. Examples include CQL[Ara06] ,MF-CQL[Liu05], TelegraphCQ[Cha03], and TinyDB[Mad04]. These languages extend SQL with window operators, relation-to-stream operators, syntax to specify the sampling period and the life-time of the sensor network, and even syntax to generate output streams based on the query result. In contrast to above SQL style languages, Aurora and Borealis focus on a ``Box-and-Arrow" programming model where one describes queries as a graph of operators with a series of parameters. Service-oriented middlewares(SOM) for pervasive spaces like Gaia[Gaia05], Oxygen[Oxy07], PICO[Kal07],Scooby[Rob04], and Aura[Gar02] take an approach similar to Aurora and Borealis where applications are described as graphs of services. Each device in the pervasive/ubiquitous space advertises its capabilities as services. The main challenges then include how to optimally perform a QoS-based service discovery and composition [Gu03,Kal07], proactive and reactive failure resilience[Gu03], and dynamic swapping of services and service graphs.

SPEs usually execute queries on a centralized server and many mechanisms to scale data stream processing to high data rates given memory and CPU constraints have been devised. These include techniques for load shedding (to dynamically adjust stream rates to those manageable by the stream engine) [Tat03], chain scheduling [Car07], dynamic tuple routing [CHA03,Laz07], load balancing (to distribute stream processing across multiple processors) [Zdonik], and approximate computation (to reduce memory requirements and speed up stream processing computation). Recently, Yahoo's S4 system has explored an actor-based framework to scale stream processing dynamically by exploiting cloud resources [yahoo-paper].

While the work on SPEs has focused on scaling stream processing to high data stream rates, research on wireless sensor networks (WSNs) has focused on in-network processing of sensor information primarily from the goal of minimizing communication in order to maximize battery life of sensor nodes. These include techniques for improved ad-hoc programming of sensor networks via dynamic code upload to each node [Fok05] or providing a database-like view of the sensor network and pushing the execution of the relational operators into the WSN notes [Mad05].

Limitations of Existing Research: Since pervasive sensing and monitoring systems create awareness out of continuous data streams generated at the sensors, many of the techniques for stream processing and sensor networks discussed above are highly relevant to building sentient space applications. While existing work provide effective data processing capabilities over continuous stream of data, it exhibits significant limitations, in our view, to serve as a platform for building sentient spaces. We highlight these challenges below:

Semantic foundations and flexible programming environments: Pervasive applications deal with diverse sensor types that may generate different types of data at different levels of semantic abstraction. Such heterogeneities makes programming pervasive applications very complex especially if applications are required to explicitly deal with failures, disruptions, timeliness properties under diverse networking and system conditions, and missing or partial information. None of the previous approaches provides the level of abstraction desired for programming sentient spaces. All of them still require the application to specify *how* to answer a query. SPEs require applications to specify which streams to connect. WSNs expect applications to specify which sensed data they are interested it. SOMs require applications to specify which services are needed. New programming abstractions for sensor-programming are required that hide application programmers from having to deal with heterogeneity of sensors, low-level details of the specific sensor devices, or to write defensive code to overcome errors and failures. Such a programming environment will empower the application writers to express their higher level application goals which are then translated into lower level sensor specifics programs by the system. Such a framework will also enable effective reasoning about observations, and actions to bring about effective adaptations of both the system behavior and the pervasive application.

(1) ***Scalability:*** To create situational awareness, pervasive spaces are instrumented with large numbers of heterogeneous multimodal sensors that generate voluminous data streams that must be processed in real-time. Techniques to enable accurate and fast processing of relevant data in the presence of communication and computing constraints such as intelligent operator placement, load balancing, etc. must be explored.. While techniques developed in the context of SPEs provide a starting point, a semantically enriched representation of sentient spaces provides new opportunities for optimizations. We will illustrate one such optimization in the form of semantic scheduling of sensors under network constraints.

(2) ***Robustness of sensing:*** The sensing process is inherently unreliable; in addition to sensor and communication errors, pervasive space deployments are unsupervised and often exposed. These changes influence the validity of the information being captured and these uncertainties can propagate to the higher level event processing tasks. Techniques to support robust/trusted situational awareness that will handle small physical perturbations to sensors (e.g. due to wind, tampering), large system failures and network losses must be designed.

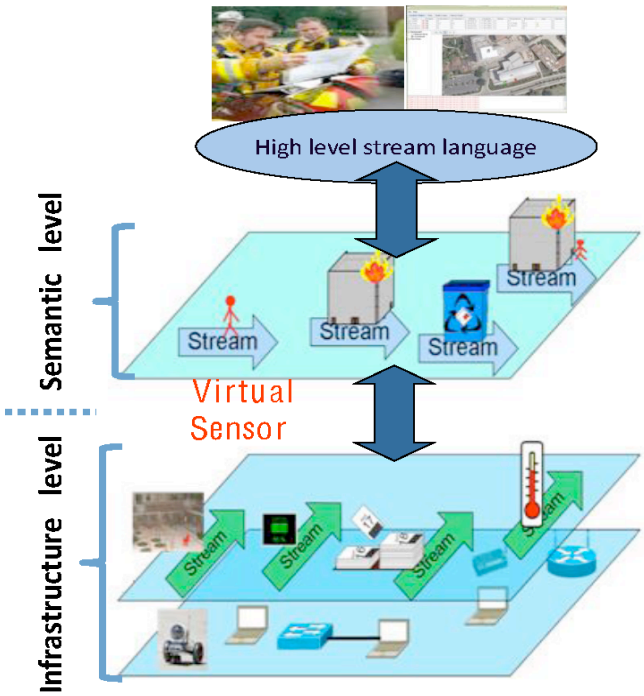
(3) ***Human-centric deployment issues*** – In pervasive spaces that monitor and observe human activities and interactions, additional challenges related to wide-scale deployment further arise. One such concern is that of privacy. While the issue of data privacy has received significant research attention in the context of internet based

applications (wherein web sites store request for individual centric data) and in collecting and disseminating electronic medical records, pervasive systems that continuously capture and process information such as location, activity and interactions using sensing technologies raises many additional challenges by introducing many further inference channels. The chapter will identify the privacy challenges that arise and summarize the progress that has been made in this context.

While later chapters in the book will include details about specific mechanisms to address many of the above discussed challenges (e.g. techniques to detect events from lower-level sensor information, techniques to deal with uncertainty, data mining mechanisms, etc), the thesis of this chapter is that much of the desired functionalities should be incorporated in an adaptive middleware environment. This chapter will discuss design principles in the creation of the sensing and monitoring middleware for pervasive spaces that can address the multifaceted challenges of scalability, robustness and flexibility. It will also discuss the role of formal methods and reasoning in the realization of such a middleware framework. We will discuss technological advances in event processing architectures that can help develop a wide range of situational awareness applications. We will finally discuss our ongoing efforts in developing such a middleware framework - SATWARE built on top of the Responsphere pervasive instrumented space at UC Irvine.

2. Hierarchical Modeling and Reasoning in Cyber-Physical Systems

Figure 1: MultiLevel Information Hierarchy



In a sentient space, a variety of sensing technologies may be used to capture the dynamic state of the evolving real-world which drives applications. Depending upon the applications, sensors may include in-situ sensors, simple mote based temperature/pressure sensors, body-worn physiological sensors, location sensors, video cameras, acoustic sensors, human speech, etc. Such diverse sensors produce widely different types of data which differs from each other in terms of levels of accuracy, reliability, volume of data generated, etc. Furthermore, while some sensors may directly support real-world events (e.g., a motion sensor at the entrance of a room quite accurately may detect someone entering / leaving a room), others may require significant

analysis/processing to convert raw sensor data into a meaningful observation. For instance, in a video surveillance application, raw video needs to be analyzed to generate a phenomena of interest (e.g., tracking a particular entity, detecting anomalous events in a security application). Such differences in the way data can be captured from the sensing device, processed into observations of use to the application, the degree of accuracy and reliability of different sensors, etc. significantly adds to the complexity of building CPS applications. What we desire is an abstract programming framework that hides such complexities from the end-user enabling application writers to focus on the application semantics instead of writing application code that deals with issues of sensor heterogeneity, failures, accuracy, etc. Fundamental to such a programming framework is the underlying modeling of the cyber physical system and its components.

2.1 An Event Oriented Model for CPS

A cyber-physical space can be viewed at multiple levels of abstractions as shown in Figure 1 [Sat]. The first is the *physical layer* consisting of sensors, actuators, networks, and computing and storage nodes. This layer enables information sensed via diverse sensors that monitor the pervasive space to flow to applications that need the information, to monitor the state of the environment and to realize the actions (either automated or through human intervention) taken as a result of monitoring the space.

The second layer is the *semantic abstraction layer* that associates real-world semantics and interpretation to data captured by the sensors. The higher, semantic layers produce semantically meaningful streams that capture occurrences of different events that occurred in the real world being monitored by the pervasive space (e.g., “Shooter on campus in Bldg 315”). An architectural abstraction that we refer to as a *virtual sensor* bridges the gap between the application-level concepts and the raw sensor data using ``operators'' that transform input sensor data streams (e.g. video sensor feeds) to higher level semantic streams that capture application level concepts and entities(e.g. specific people in the room). Virtual sensors are a specific set of transformations that when applied to a set of input streams produce a semantically meaningful output stream that applications can reason with. For example, a virtual sensor used to locate building occupants based on applying a WiFi localization algorithm on a stream of access point signal strengths captured by each user's WiFi Access Point sensor is depicted in the table below.

Name	**Description**
DLinkIndoorCamera	Gets frames from a DLink indoor camera
LinkSysIndoorCamera	Gets frames from a LinkSys indoor camera
WiFiLocalization	Returns location based on sensed WiFi signal
Scan	Gets table from database
ImageBasedMotionDetection	Detects motion based on a stream of frames
Projection	Selects a set of columns from a table
DBLogger	Saves stream contents on a database

Table 1: Example of operators

At the physical level, a pervasive space will be modeled as a set of sensor streams that observe the physical environment. Data from these sensors may be collected continuously, periodically at a specified rate, or triggered by an anomalous event. At the application level, the primary modeling concept will be the notion of an *event*. The derivation of events from raw data via "*virtual sensors*" forms the bridge between the physical observation of a sensor and its interpretation. For example, the event "Bob was in room BH-2001 at 10am" might be derived from data detecting a card with a Bob's ID produced by an RFID reader at the room entrance. We distinguish between primitive events (derived directly from sensor data), and events that are derived by combining or abstracting information from other events.

Key features of the proposed event model are outlined below.
• Associated with events are a set of properties that include entities (people or objects associated with the event), location (where the event took place), and time (when the event occurred). Depending upon the type of event, additional attributes may be associated with events.
• An event may describe change in state, or simply an observation of state.
• Primitive events will have associated derivation and trust information reflecting the uncertainty in the underlying sensor data and analysis procedure. Non-primitive events will have associated evidence that reflects the derivation of the event, and its trustworthiness.
• Events in pervasive spaces are modeled at multiple levels of abstraction supporting a hierarchical information model for pervasive spaces (see Figure 1). For example, "the room is not empty" has less information than "there are three people in the room" which in turn has less information than "Alice, Bob, and Eve are in the room".

Users or applications pose queries using a specification language based on the formal event model. The result of a query could be a single event, or a delimited or continuing stream of events of interest. Several advantages accrue from the use of a multi-layer view of pervasive spaces. Hierarchical information modeling greatly simplifies pervasive application programming by empowering application writers to write applications at appropriate levels of abstraction without worrying about sensors, sensor programming, or sensor scheduling at the physical level. The separation also enables the system to optimize and self tune itself without considering application details.

While the event-based model provides a powerful data abstraction to represent the evolving state of the real-world suitable for building CPS applications, and many concurrent research projects (including ours) are exploring such an approach, many technical challenges/issues will need to be addressed before such a model can be realized. These include: modeling and representing uncertainty with events which are inherent in any event-detection mechanism, techniques to translate errors/inaccuracies/failures in sensor data into corresponding uncertainty in detected events, techniques to store the event oriented representation of the physical world into a database, appropriate extensions to query language to support event-based reasoning, etc.

So far, the pervasive space can be viewed as a hierarchical information model that forms the basis for reasoning about the pervasive space and activities in the space. Such a representation will enable application developers to specify and reason about higher level events that occur in the space independent of the implementation and development environment. In the following section, we present a higher level goal-oriented distributed logic based approach for specification and reasoning about the behavior of components of a cyberphysical space in terms of goals, facts and proofs.

2.2 Reasoning about Cyber-Physical Spaces

In this section, we focus on an event based modeling of sentient spaces and mechanisms to represent and reason with such a model. Specifically, we will describe a logical framework for networked cyberphysical spaces (NCPS) that will provide a high level language for describing robust system behavior (via an executable specification), strategies for execution and reasoning, with clear semantics and clean meta-theoretic properties. The logical framework will serve as a uniform declarative interface to all capabilities of a cyberphysical sentient space. At the same time it provides a semantically well-founded way to represent, manipulate, and share knowledge across the network. In the described framework, logical theories serve as a basis for abstract models that are continuously adapting to new incoming knowledge resulting from local or nonlocal observations.

A Distributed Logical Framework for CyberPhysical Systems: To design a modeling and reasoning framework for distributed networked CPS, various kinds of knowledge need to be expressed including models, facts, goals, and proofs (derivations of goals from facts). For example, facts can represent sensor readings (or virtual sensor readings) at specific locations, and goals can represent queries for information or requests to actors or actuators to perform certain actions. Although there are cases where a goal can be directly satisfied by a single local action, it is typically the case that distributed actions are needed and the more relevant feedback will be conveyed via a feedback loop through the environment. Such indirect feedback can consist of facts (representing observations) from multiple sensors that together can measure the progress toward reaching the original high-level goal. In addition, models may have many different flavors ranging from precise physical models to qualitative commonsense models, and can include approximate and partial models of the real world based on observations. Combinations of different flavors are usually needed. For instance, a virtual sensor may have a precise model of camera locations, but its model of occupancy is approximate and updated as knowledge is gathered and processed.

Apart from a few notable exceptions such as Cyberlogic[Rue03], it is interesting to note that the distributed nature of today's problems is rarely considered in the design of logical frameworks. For cyber-physical systems it is essential, since carrying out proofs may require cooperation across multiple nodes. In many cases, goals and facts cannot be matched locally. Consider an example of gathering certain information from a particular area under observation (e.g., from a sensor network on the ground that is part of a global network including UAVs and satellites). In an interest-driven routing protocol, such as

directed diffusion [Cha03b], a node expresses interest for specific data by sending requests into the network. Data matching the interest is then drawn toward the node from which the interest originates. From a logical point of view, a goal, representing an information request, is injected and disseminated through the network. The goal is a logical formula expressing that the information needs to be of the required kind (content subgoal) and be delivered at the requesting node (delivery subgoal). A fact representing the presence of information at the source will match or satisfy part of the goal --- namely, the content subgoal. Now there is an incentive to route the partially satisfied goal with the requested content toward the interested application, since this will incrementally increase the ***degree of satisfaction*** of the overall goal and eventually complete the distributed proof. In other words, an interest-driven routing and many similar processes can be seen as distributed proofs and optimization strategies that try to bring facts and goals together.

To serve as a formal framework for cyberphysical spaces, the logic must have the capability to express degrees of satisfaction so that both search and optimization become instances of a generalized notion of deduction. As a starting point, we have developed a version of the framework based on first-order logic with equality, real arithmetic, and degrees of confidence [Ste10,Kim10]. Specific application domains are then reflected in the background theory relative to which the reasoning takes place. The domain theory can also influence the search, reasoning, and optimization strategies employed at the higher layers. Due to the resource-constrained nature of many cyber-physical systems, trade-offs between expressiveness and efficiency need to be considered and it is important that the logic be scalable --- i.e., there should be sublanguages and inference systems of adjustable complexity. In the long term, the language needs to go beyond propositional and Horn clause logic, since a functional sublanguage representing cost and utility functions with discrete and continuous parameters and functional parts of the models will be essential. Furthermore, predicates with discrete and continuous parameters are important to support predicate abstraction [Sus97]. To support functional computation as part of reasoning and optimization strategies, the logic should be equipped with operational semantics --- e.g., based on conditional term rewriting similar to that of equational specification languages such as Maude[Cla07], which is key to combining abstract logical models with an efficient notion of execution.

This logical view allows information collection, control, and decision problems to be recast as logical problems that are primarily centered around the duality of two kinds of knowledge: facts and goals. Various classes of distributed algorithms can be declaratively expressed using this duality. Proactive, data-driven, or optimistic algorithms are mostly concerned with the establishment of new facts from existing facts, hoping to satisfy the goal but considering it as a secondary aspect. Reactive, demand-driven, or pessimistic algorithms are primarily goal oriented, meaning that during their execution new subgoals are established based on existing goals, which eventually can be directly established using the facts. It is noteworthy, however, that many interesting practical algorithms (e.g., hybrid routing for sensor nets) are a mixture of different paradigms. Hence, in our logical framework, both facts and goals need to be treated on an equal footing together with corresponding forward and backward inference rules.

Towards a Robust Logic of Degree and Uncertainty: A logical model is an instance from a fixed model class described by a common background theory. In most applications, we are concerned with incomplete information, and the model of the real world is not entirely characterized. Hence, we are almost always concerned with an entire class of models that are consistent with the facts to various degrees. Apart from the natural incompleteness of knowledge due to partial observability, many sources of uncertainty exist in cyber-physical systems, including environmental noise, measurement errors, system perturbations, sensor and actuator delays, and clock drift. Networked systems exhibit further sources of uncertainty caused by delayed, outdated, incomplete, or inconsistent knowledge. Furthermore, uncertainties play a natural role in information fusion and probabilistic algorithms. The consideration of a class of models also allows standard logics to represent certain aspects of uncertainty, but the degree of uncertainty is not explicitly represented. A natural solution would be to use an instance of many-valued logic [Got01] that is sufficiently constrained to be consistent with common probabilistic [Ada98,Fag90], stochastic[Cus00], and quantitative interpretations[Wan09]. To enable expression of priorities between goals or their relative importance (e.g. to differentiate between hard and soft constraints), we furthermore need weighted formulas.

In cyber-physical systems, models, facts, and goals are continuously changing. Therefore, the logical framework must be able to incrementally, and efficiently deal with such changes. Maintaining proofs explicitly at a suitable level of abstraction --- e.g., as partial orders (as opposed to sequential proofs) capturing all dependencies between facts and goals is a first step. Proof maintenance will take advantage of the locality of changes and hence can improve the efficiency of automated deduction and constraint solving/optimization. For instance, an explicit partial-order representation of dependencies enables more sophisticated search and optimization strategies, such as conflict-driven backtracking and logical state composition strategies that do not assume centralized control.

Depending on the nature of changes, proofs can either remain valid, require local adjustments, or become entirely invalid. Clearly, the former case is preferred, which is why we suggest complementing proof maintenance with a notion of proof robustness that, when used as an optimization criterion, allows us to avoid fragile proofs whenever possible. Proofs can be fragile because they are based on rapidly changing or unstable facts or because they lack redundancy. Consider, for instance, the goals of maintaining network connectivity or sensor coverage. Clearly, proofs representing solutions that rely on stable facts about the neighborhood of a node are preferred. Furthermore, in dynamic environments, proofs can be carried out in a robust way that instead of relying on an individual fact, which could become a single point of failure, relies on an abstraction --- e.g., a disjunction of independent facts representing coverage or connectivity via several neighbors that remains invariant under a larger set of network perturbations.

Control and Optimization as Logical Strategies: System control and optimization in NCPS is challenging. Traditional optimization techniques that strive for optimal solutions based on precise models are not suitable for most NCPS, where models have many dimensions of uncertainty, and optimality in the strict sense is neither desirable nor

achievable. What is needed in practice are strategies to find acceptable and robust solutions, sufficient to achieve the goal while taking into account the geometry of the network, limitations of the models, and available resources. Mathematically, the logical framework allows a rich set of conceivable behaviors that need to be constrained to a subset that satisfies the system objectives. Strategies that control and optimize the operation of NCPS will be based on its declarative representation in the logical framework. These strategies will be resource-aware and adaptive. For example, in homogeneous scenarios, our strategies may exploit the parallelism of many nodes so that resource consumption at each node can be low. In heterogeneous cases, they could exploit powerful or energy-rich nodes that perform heavy computations so that low-power nodes can save their resources.

Ideally, logical strategies can exploit the parallelism inherent in search and optimization problems, by allowing nodes to sample the search space independently. Unlike numerical approaches, sampling could be done symbolically, by randomly generating new subgoals that represent entire regions of potential solutions in a finite way. The sampling heuristics may be biased by a nonuniform distribution to express locality and preference for solutions that can be reached more easily or with lower cost. In addition, the cost of reaching a solution may be explicitly quantified and constrained by the system goal. The best stable solution will be shared opportunistically across the nodes and is ultimately used to drive the local actions of the networked distributed CPS. Conflicts may arise, manifesting themselves either as logical inconsistencies or nonacceptable solutions. Local randomized backtracking driven by the conflict itself can be used for resolution, exploiting the dependencies maintained by the underlying logical framework.

Robustness and Composability: There is a natural connection between abstraction, robustness, and composability that is facilitated by a logic based approach to representing cyberphysical spaces. The logical approach to optimization may also enable the composition of (partial) solutions. Rather than aiming at a numerical point solution each node could narrow down the goal to one or multiple solution regions represented by logical formulas. If two nodes have or establish connectivity, the goals could be composed by a logical conjunction resulting in a goal that semantically corresponds to the intersection of solutions acceptable for both nodes. This approach should generalize to entire groups of nodes that merge due to a network topology modification. Composability is thereby enabled by a suitable level of abstraction that avoids over-constrained point solutions; in other words, solutions are robust enough to accommodate at least to some degree the needs of other nodes. The use of an abstract solution region reduces the likelihood of conflicts in the case of composition, but clearly cannot exclude this possibility entirely.

To illustrate how logical inferences are enabled by the distributed logic approach, consider the example of distributed sensing of activity. Assume that there are several robots, each one equipped with only one kind of sensor, either an acoustic sensor or a motion sensor, and some nodes are equipped with a camera. Assume that predicates *Motion(a,t)* and *Noise(a,t)* are true if motion or noise have been detected in an area *a* at time *t* (approximately). Assume furthermore that *Image(I,a,t,t')* means *I* is an image of

area *a* taken in the interval *t,...,t'*, and *Delivered(I,r)* means that the information *I* has been delivered at *r*. Now the following goal is injected at root node *r*:
$Motion(a,t) \vee Noise(a,t) \rightarrow \exists I : Image(I, a, t, t+\Delta t) \wedge Delivered(Extract(Abstract(I)),r)$.

It expresses that an image needs to be taken of a specific area a with maximum delay Δt after motion or noise has been sensed. The image then should be delivered to *r* after abstraction and feature extraction. After the goal is disseminated in the network, each node tries to solve the goal. Let us now assume that a node in area *a* generates a fact *Motion(a,t)* that can be used by another node in that area that is equipped with a camera to simplify the goal to

$$Image(I, a, t, t+\Delta t) \wedge Delivered(Extract(Abstract(I)),r)$$

so that the only way to make progress is to take an image *i t*o satisfy *Image(i, a, t, t+Δt)* leading to the remaining goal *Delivered(Extract(Abstract(i))),r)*.

Let us assume that the abstraction $i' = Abstract(i)$ can be performed immediately after taking the image but feature extraction will be performed at a more powerful node because it is computationally expensive. This node will then simplify *Delivered(Extract(i'),r)* after performing the computation $i'' = Extract(i')$ to *Delivered(i'',r)* which can be incrementally solved by moving *Delivered(i'',r)* closer to *r*, the requesting root node, where it is finally realized by a delivery action. In spite of its simplicity, this example illustrates the combination of logical inference and partial evaluation and their generalization to the distributed setting in which goals and facts can be bound to actions at different locations in the cyber-physical world. In a more complex example, we might easily imagine that *Motion(a,t)* and *Noise(a,t)* cannot be satisfied using the current distribution of nodes so that some nodes will have to move into area *a*. Clearly, this opens a rich trade space of possible solutions, which in our approach will be solved by guiding the distributed logical framework using a more sophisticated strategies.

3. Adaptive Middleware for CyberPhysical Spaces

In this section, we argue for a principled approach to developing a management framework for scalable, dependable CPS. There are several key aspects of instrumented CPSs that merit further exploration. First, the underlying system is *inherently dynamic* – a structured approach to realizing adaptability is essential, especially when the ICPS is long-lived and must operate under unpredictable situations – this requires the *ability to reason about system evolution* to develop adaptations that meet the needs of the CPS application at hand. Second, designing for objectives such as scalability, dependability and security requires an understanding of the end-to-end architecture of the system and the dynamic changes that occur at multiple system levels – e.g. device, network and distributed infrastructure levels.

To represent the end-to-end perspective, we promote a ***cross-layer view of CPS*** environments where we view each device as a vertically layered architecture consisting of application, middleware, network, OS, and hardware layers and the distributed environment and devices connected through a distributed middleware infrastructure to manage information interchange across devices and applications. From our prior

experience in designing a cross-layer approach for timeliness and reliability in resource-constrained mobile embedded systems (e.g., Dynamo [Moh07]) for effective of cross-layer adaptation, xTune [Kim07a, Kin08a], and [Lee08] for protection mechanisms against hardware transient faults), and from other work [Chi07, Cui06, Del05, Kha05, Koz04] it is clear that such distributed cross-layer optimization is required to address end-to-end performance and dependability in dynamic environments such as ICPS.

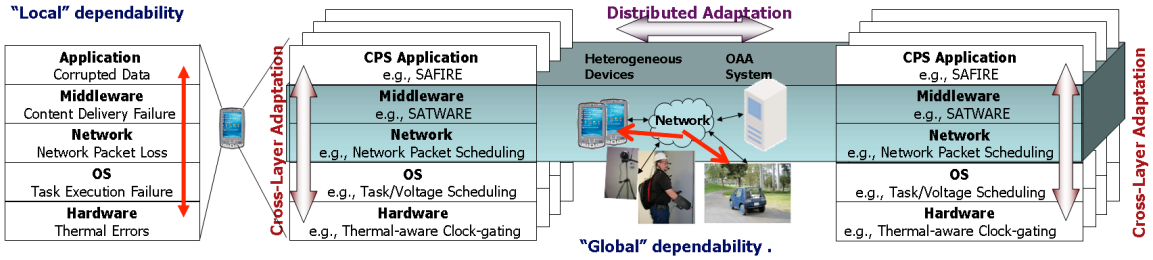


Fig 2. Distributed Cross-layer Adaptation: Local and Global Characteristics

Figure 2 illustrates our conceptual distributed cross-layer view of our adaptive ICPS. At the infrastructure level, a variety of devices are interconnected by a variety of communication channels (e.g., Ethernet, cellular, Wi-Fi) with distributed middleware support (e.g SATWARE [Hor07d]) to operate CPS applications (e.g SAFIRE [SAF], a situational awareness dashboard for firefighters). Adaptation policies to be implemented in the system may be (i)within a layer on a device, i.e independent of other layers, (ii) implemented by a vertical composition of policies on the device across layers *(Cross-Layer Adaptation* in Figure 2), and (iii)realized by a horizontal composition of policies distributed across devices *(Distributed Adaptation* in Figure 2). Clearly, decisions at one layer affect other layers. For example, if the data has high importance with a short expiration time, the middleware layer must adjust the frequency of data dissemination appropriately. Similarly, CPU slowdown to control thermal runaway at the hardware layer may increase deadline misses in OS task scheduling layer; this anomaly bubbles up to the application layer and is manifested as a failure to provide up-to-date data. Furthermore, deadline misses may lead to the delayed delivery of the network packets, which in turn results in a failure for timely delivery of messages.

The cross-layer view of CPS systems has inherent complexities that arise due to dependencies among layers. The first task is to *accurately capture heterogeneity of the CPS system (many sensors with different capabilities) in the cross-layer architecture. This will enable us to* characterize CPS components, abilities, and limitations and determine how application goals map to system-level components and study how behavior can be tuned to make best use of available resources to meet the multifaceted needs of dependability and scalability. Additionally, we must determine *what aspects of each CPS layer need to be observed and what are the end-to-end requirements and how can we relate them to cross-layer parameters.*

From a dependability perspective, both permanent and transient errors need to be modeled and mitigated. For instance, heavy utilization of the device hardware (e.g., for peak performance) can result in excessively high temperatures that may cause hotspots, resulting in thermal errors; to alleviate this, we may trigger task replication or re-execution at the OS layer. The mitigation strategy might cause packet loss due to

buffer overflow, since it requires more processing time. Under such circumstances, the dynamic choice of routing algorithms and their parameters need to consider higher-layer QoS constraints, (partial) knowledge about the network (e.g., sensor density, coverage), heterogeneous devices (with different error sources), and operational context (e.g., prioritizing information flow).

SATWARE - A Middleware Platform for CyberPhysical Spaces: We now present the architecture of the SATware, a distributed multi-level semantics-based middleware for sentient spaces and discuss reconfigurability techniques to enable a scalable and efficient management framework for pervasive spaces.

To capture and enable the cross-layer view of a distributed cyberphysical space, SATWARE consists of ***an efficient CPS Cross-Layer state management service*** that can efficiently capture, represent, process, and store information from the various data producers (e.g., cameras, motes, mesh routers) at desired levels of accuracy and granularity in order to meet the information quality and dependability needs of consumers (e.g., video data for surveillance or link congestion levels for routing) given storage and communication constraints. Central to our approach of designing a state management service is an ***ICPS StateDB*** that stores raw and processed information from different data producers and enables monitoring and management of parameters at different layers of the ICPS system.

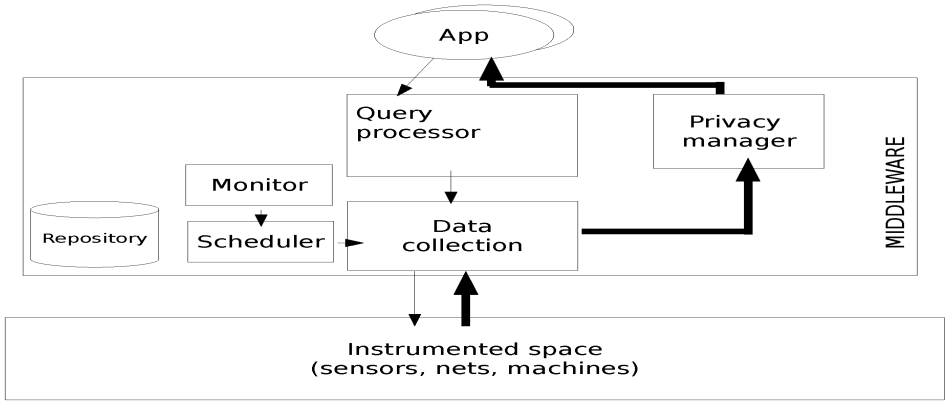


Figure 2: SATWARE System Architecture

Figure 3 depicts the building blocks of the SATware middleware framework - which consists of four key modules: the Query Processor, Data Collection Module,Monitor and Scheduler. The goal of the Query Processor module is enable the multi-level perspective illustrated in Figure 1 where application level semantic concerns are separated from infrastructure level issues. Applications pose continuous queries to the ***Query Processor*** module which in turn selects a set of virtual sensors to provide answers to the continuous queries and forwards this set of virtual sensors to a Data Collection module. The Data

Collection module maps in turn, operators corresponding to these virtual sensors, for execution on physical nodes (machines) in the underlying pervasive computing infrastructure. The resultant streams may be further processed in additional modules prior to being forwarded back to the application. For example, the result streams may pass through a Privacy Module that adapts the query answers to ensure that the output data does not violate privacy constraints (details in [Dan09]).

A monitoring module captures dynamic attributes of the underlying infrastructure (e.g event occurrences, resource availabilities); the monitored information is used to enhance the performance, robustness and scalability of the system. The Scheduler Module combines the events captured by the Monitoring Module with system semantics to direct data collection activities. For example, an increased level of occupancy in a certain region (as captured by motion detectors) can be used to trigger a specific video camera that can capture the activities in that region. Furthermore, based on resource constraints, the scheduler determines the specifics of the sensor data collection plan, e.g. the resolution and frame rate at which the video data must be captured. All modules consult a repository which contains (i) a snapshot of the current infrastructure state containing the location/state of sensors and processing units (ii) virtual sensor definitions and operator implementations available to programmers who can reuse existing virtual sensors or define new ones; and (iii) the semantics of the applications and sentient space.

The above architecture illustrates the building blocks for enabling efficient and robust operation of pervasive spaces; however, suitable techniques must be designed implemented within the various modules (Scheduling and Monitoring, Privacy manager) to achieve the potentially conflicting goals of scalability, robustness and privacy in sentient spaces. In the following sections, we describe challenges that arise in the deployment and management of real-world pervasive spaces, in particular scalability, dependability and privacy and suggest potential techniques to address these issues that can then be incorporated into a middleware framework such as SATWARE.

4. Enabling Scalability in CyberPhysical Spaces

Scalability issues in CPS environments arise at both the infrastructure and information levels. At the infrastructure level, sentient spaces incorporate a plethora of devices/sensors that capture diverse types of information - capturing, processing and storing this data results in network, CPU and storage limitations. CPS applications pose varying data fidelity needs, which in turn can be exploited for intelligent application-driven scheduling of sensor data capture. For example, applications that use a localization service may require accurate position information (e.g. security application), whereas coarser grained information may suffice for other applications (e.g. proximity sensing). The aim would be to exploit these application fidelity tolerances and translate them into scalable, yet useful management mechanisms in the infrastructure. In the remainder of this section, we discuss two key scalability techniques for CPS environments –(i) managing the scale and heterogeneity of the devices and networks in

the *CPS infrastructure* and (ii) enabling *information capture* in a scalable and meaningful manner for the CPS application at hand.

A Scalable, Quality Aware State Management Service: The traditional approach to building distributed state services using (centralized or distributed) database management system (DBMS) technologies has been found deficient when data is produced continuously [Mad02, Ols03]. More recently, several data stream architectures [Aba03, Sri05, Mot03, Bab04a, Bab04b] have explored more intelligent use of data producers explicitly in an attempt to maintain equilibrium between rate of data production and consumption (e.g., via load shedding). There are limitations of this work in the context of ICPS. For example, the architecture and protocols have not considered heterogeneity of data producers that is endemic to ICPS environments. Most techniques (due to limited context for which they were designed) have not significantly exploited "intelligence" at the data producers. A notable exception is "in network" processing in sensor DBs that exploits computing at sensor nodes [Mad02, Mad05, Han07]. However, much of this work has focused on energy optimization to maximize sensor lifetime, which is not the only design criteria in a ICPSs of interest in this proposal.

We envision a quality-aware state management framework that realizes a natural tradeoff between an application's quality and resource needs and thereby supports scalability. Let us view the CPS environment as consisting of data producers (e.g. sensors) scattered in the distributed systems; data consumers (applications that use the captured information in a raw or processed form) that can tolerate a certain degree of degradation in information quality; and a set of servers to store/ process the data that can collectively be viewed as an ICPS StateDB. Data stored at the ICPS StateDB is approximate and can be used instead of current information from sensors to conserve resources for more robust capture when needed. For example, approximate localization based on previous updates consumes less resources than a more accurate assessment (e.g., obtained by fusing input from video, inertial sensors, and access point signal strength analysis).

To support scalable and quality-aware data capture, protocols/algorithms must be in place to synchronize the image of data producers (e.g., current sensor values, available link bandwidths) with the ICPS StateDB where data is maintained. Simplistic algorithms such as *Ad hoc* sampling and summarization will not work because of the high degree of correlation and dynamic variations in data precision (spatial as well as temporal) needs. One promising approach is to use prediction-based techniques wherein producers and storage points agree upon a prediction model and producers communicate only when the sensor value deviates (beyond acceptable tolerances) from the prediction. Implementing a prediction-based producer/consumer view requires answering the following questions: *Who determines the prediction models, the producer or server? How do producers and servers agree on a model? What parameters can be used by the prediction model? How can locality and correlation among sensors be exploited for prediction? How does one dynamically switch models to improve accuracy of prediction? What is the performance overhead due to model maintenance?* While there exist preliminary efforts that aim to answer these questions in a limited context (e.g. camera sensor networks, mote networks), more general solutions for a multisensor space remains a topic of future work.

Enabling Scalability via Semantic Sensor Scheduling: In a multimodal sensing environment, cost-effective capture, delivery and processing of large multimedia data e.g. dynamic video poses a significant challenge. As an example, consider a real-time tracking system which is responsible for monitoring human activity as observed by a large number of camera sensors. When considering systems of relatively large scale, constraints arise at various levels: network bandwidth is required for video delivery, I/O and disk resources are required for writing images, and CPU is consumed for image feature extraction. Assume that, due to resource constraints, only a subset of camera sensors can be probed (i.e. accessed for an image) at any given time unit. The goal of scheduling becomes that of determining the ``best'' subset of sensors to probe under a user-specified objective (e.g., detecting as much motion as possible, maximizing probability of detecting ``suspicious'' events). Under these conditions, one would like to probe a camera only when motion is expected; and conserve resources when there is no activity of interest being captured by the camera, see [Vai09] for more details.

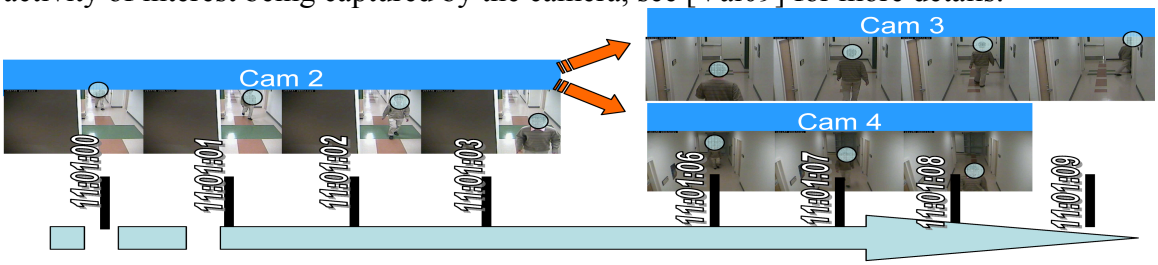


Figure 3: Illustration of a Conditional Correlation of Motion Between Cameras

We advocate a semantic approach toscalable sensor scheduling where semantics learnt from sensor observations is used to guide the scheduling of resources. Our initial work with camera networks suggests a probabilistic model that is based on extracted semantics can predict where events of interest will occur and dedicate resources accordingly. [Vai08,Vai09]. We discuss approaches for semantics based scheduling in the context of motion detection in camera networks since video content is resource intensive; however, the technique generalizes to any sensor data (e.g. audio, motes). Examples of semantics that can be learned over distributed camera sensors include:

- ***A-priori Motion:*** the probability that motion starts ``spontaneously''. In the case of the building, it is likely that the camera at the front door will see more motion than other cameras.
- ***Self Correlated Motion in a Camera Over Time:*** given that a camera observes an event and given the camera's field of view (FOV), one could predict the probability that the event will continue. For instance, a camera focusing on a long corridor will have a person in view for a longer period of time compared to a camera that is focused on an exit door.
- ***Cross-Correlated Motion amongst Cameras:*** a person who exits a FOV of one camera will be captured by another depending upon the trajectory of the individual and the placement of the cameras.

The above semantics learned can be used to predict motion based on which scheduling decisions can be made. ***Formally***, we define a plan, *Plan*, for N cameras to be a binary vector of length N that specifies which cameras will be probed in the next time instant.

Plan = $\{C_i \mid 1 \leq i \leq N\}$, where $C_i \in \{0,1\}$. Assume that the cameras are selected to optimize an application-dependant **benefit function (BF)**. For example, a particular application may want all image frames for which there is motion (all motion events are equally important), while another application may define that two images of two different individuals are more important than two of the same person. Another consideration is the **cost** of a plan, in terms of network resources, referred to as **cost function (CF)**. Different plans may not cost the same in terms of network resources since it may be less expensive to probe the same sensor at the next time instant. In a fully general model, one might also place the number of sensor probes *K* into the cost function. The cost-benefit model described above can be combined with real-time data from the monitoring module to further optimize the scheduling mechanism in real-time. For example, real-time data from the camera sensors can be used to generate accurate predictions of where motion occurs; however, these predictions must arrive early enough for the real-time scheduling process to take action. As the monitoring system is overwhelmed by the sensor data feeds it constantly adjusts its data collection process such that the available resources are assigned to the data sources (cameras) that are most likely to detect an event of interest, based on the cost-benefit functions and the prediction model discussed before. Building on these ideas, we have designed techniques to learn a dynamic probabilistic model of how events that are relevant to the application relate to previously detected events. The goal of this process is to take conditional associations into account and change the sensors deployed in real-time using multi-level feature predictions which can allow us to process more relevant content.

While the scheduling discussion above is motivated by resource constraints, the system may have additional constraints - e.g., in case of an optical zoom camera, the focal length and the field of view of the camera are dynamically adjustable and a particular configuration in many cases competes with another. The scheduling approach discussed above can also be applied to determine the configuration parameters of the sensors at any given time which optimizes the end application goal (e.g., pan/zoom/tilt a camera to where it is expected to maximize the collection of frontal face images). Note that such reconfiguration, may require us to further address quality or reliability tradeoffs. For example, reliable face detection from images requires capture of high resolution of the frontal face image[Yan96]. Activating zoom capabilities will imply temporary loss of "pan" capability which can capture potential events (albeit at lower resolutions) elsewhere in the coverage area of the camera sensor. Ideally, we would like to incur the overhead only when we are confident that the frame contains a face. In prior work, we explore techniques for annotating images with relevant features and predicting when a set of relevant attributes/features are expected to be extracted, camera parameters are set based on this prediction. A promising direction of future research is the possibility of cooperative triggering of multiple[Xu98,Byu03,Sti99,Kru00,Qur09] and heterogeneous sensors in a coverage region with different accuracy, overhead, and coverage profiles to address accuracy/overhead tradeoffs in CPS applications.

5. Dependability in Sentient spaces

Mission-critical pervasive computing applications require the underlying systems to be ***dependable*** , i.e robust to disruptions in the infrastructure that cause failures in sensing,

communications, and computation. Dependability, as defined by the IFIP 10.4 Working Group on Dependable Computing and Fault Tolerance, refers to the trustworthiness of computing systems that allows reliance to be justifiably placed on the services it delivers. Dependability constitutes a variety of non-functional requirements including availability, reliability, maintainability, safety, and integrity. In the context of networked and instrumented cyberphysical spaces, dependability can broadly be classified at two interdependent levels that, combined, can provide a trustworthy platform for building applications.

- *Infrastructure Dependability* – how dependable are the underlying infrastructure components (e.g., sensors, networks, actuators, computing/storage elements, software environments) in the presence of diverse failures that may lead to disruptions, and
- *Information Dependability* – how dependable is the information generated by the underlying infrastructure given errors/uncertainty in sensor readings and data analysis mechanisms.

Consider, for example, a home health monitoring scenario that consists of multiple sensor feeds including RFID-enabled smart pillboxes, video cameras for determining the patient's location, and on-patient body area networks consisting of polar straps for heart rate, oximetry for respiratory conditions, accelerometers to determine position and ambulatory behavior and galvanized skin response sensors.. Infrastructure and information reliability issues arise in this scenario; critical events or a potential emergency situation must be detected from multiple sensors; and actions must be triggered to initiate appropriate medical response.

Dependability is an end-to-end system property – disruptions at any level of the system (hardware, OS, network, software) can hinder application needs – examples include packet drops at the network layer due to congestion, and bit flips in the architectural layer due to soft errors. In this section, we aim to discuss key design principles that can enable reliability at both the infrastructure and information levels including: (1)leveraging the cross-layer approach to capture reliability pitfalls and explores fault tolerance knobs at multiple levels in the system architecture (hardware, OS, network, middleware, application, content); (2) techniques to exploit redundancy of hardware, processing and content to enable infrastructure reliability and (3) a structured use of semantics (of applications and deployment scenarios) in enabling increased information reliability. Our approach offers information reliability while allowing applications to gracefully deal with infrastructure failures. Using the adaptation knobs offered by our cross-layered approach, ICPS applications can achieve high reliability under given resource constraints at all levels of the system. Note that such a cross-layer approach can also enable seamless embedding of human activities and human-in-the-loop decision making into the CPS environment.

Infrastructure Dependability Techniques: Sensors, devices, communication medium, and the application context are subject to constant changes or failures in dynamic environments. For example, devices can be turned on and off, or moved from one place to another; networks may be congested and packets are dropped; the target (e.g., monitored person) may move from one building to another. At the infrastructure-level, the system must incorporate mechanisms to adapt to infrastructure events (e.g., a network

link goes down, device loses connectivity). We view the underlying infrastructure of a cyberphysical space essentially consists of the various sensing and processing nodes and the networking components that implement multiple access networks for sensing and communication.

To support node-level reliability, monitoring mechanisms that obtain node liveness and load status are essential. One direction of research is to explore the design of failure detectors for ICPS devices that distinguish inactive nodes from permanently (and transiently) failed nodes. Such approaches often require periodic exchange of liveness information in a large network that induces unnecessary overhead when there are no/few failures or occupies much needed resources when there are large and significant failures. the design of scalable approaches to detect and manage failures in multi-sensor pervasive spaces requires further investigation. We are currently exploring the use of dynamic clustering techniques to develop scalable protocols for sensors to communicate with each other and the infrastructure: e.g., sensors within a cluster communicate using gossips and cluster heads exchange information using heartbeats.

The ICPS communication scenario is complex, encompassing multiple networks with diverse technologies such wired, WiFi infrastructure, cellular, Zigbee, mobile ad hoc (MANETs), mesh, disruption-tolerant networks (DTNs), and personal area networks (PANs). Some have relatively fixed infrastructures (e.g., cellular networks), whereas others are intermittent (mobile ad hoc networks, Bluetooth) [Coo07]. Lower-level network protocols have been developed to support packet-level routing/scheduling [Meh07], reliability, and timeliness [Jaw08,Mav05,Lu05,Lu97]; at a higher level, group communication protocols support message-level reliability and timeliness [Han04c,Han03,Xin07]. Always Best Connected networks [Gus03,Gru03], and wireless mesh networks [Aky05] aim to support communication over multiple networks; techniques have been developed for network handoff [Moh07,Moh06,Mcn01,Rap91] power management [Jon01] and monitoring [Bai06], and QoS support [Pra05]. Specific combinations of networks (cellular/ad-hoc [Luo03,Cha04], cellular/Wi-Fi [Pan09], Bluetooth/Wi-Fi [Ana09,Bro99]) have been explored. Many of the proposed techniques are at the network and lower layers; such adaptability is difficult to realize with failures and surge demands.

Enabling multinetwork reliability implies (a) monitoring the status of the multiple CPS networks and (b)designing dependable communication strategies that exploit any and all available networks to communicate information. In other words, we aim to expose the requisite network state information to the network adaptation services that ensure communication reliability. Typical approaches to wireless network monitoring include (i) passive monitoring [Bej03, Ciu06, Che06a] where network packets are captured and analyzed in detail (deep packet inspections) and (ii) active monitoring [Kim06, Sai07, Hua07] where probe packets are injected into the network(typically to determine network resource availability). Passive monitoring avoids competition with application and measurement traffic; active monitoring enables faster detection of network errors and consequently supports better fault tolerance in the short term, albeit at higher overheads. In the presence of multinetworks, it is challenging to determine how to employ “active

monitoring" techniques efficiently. Note that connectivity knowledge needed in different parts of the ICPS network varies over time based on node mobility patterns and communication needs. Furthermore, data producers or the communication medium may be intermittently available – leading to robustness challenges. One possible direction is to combine active monitoring with the "quality-aware" approach described earlier to enable application-aware fine tuning of the multi-network monitoring process to ensure reliable detection of network failures. Techniques will need to be developed to efficiently gather connectivity needs from CPS end points (link type, latency, loss-rates, load) and translate this information into network status needs– using this "application information", network state collection can be customized.

Given an reasonably accurate representation of multinetwork state in CPS systems, the next task is to exploit this knowledge to design reliable communication strategies. One approach to enhance communication reliability is by combining the capabilities of multiple access networks that are available in pervasive spaces to form reliable connected networks. When network infrastructure in the ICPS setting is spotty and unreliable, a straightforward approach is to set up a temporary mesh overlay, where mesh routers are brought in and placed around the coverage area in such a way that altogether they form a connected multi-hop network. However, in practice, it is difficult and time-consuming to discover how to place the mesh routers to create overlays that fully and reliably cover a specific area. According to our deployment experiences, the number of mesh routers needed and their placement depend on various factors, such as the size of the area, the obstructions/materials in the area, and the interference sources in the area. When instant network deployment is required, forming connected networks through the direct ad hoc links between adjacent mobile nodes (mobile ad hoc networks) in concert with the mesh networks is a viable option. While the pure ad hoc mode requires a rather dense deployment of the nodes, the autonomously created multinetworks can enable mobile nodes establish indirect connectivity to the outside world via gateway nodes. We envision a middleware driven approach to support adaptive communication techniques in ICPS multinetworks where nodes have the ability to make communication decisions locally, using available knowledge of network status and taking into account tolerance parameters (timing, accuracy, reliability). Such decentralized adaptation will allow us to support reliable communication over diverse network technologies, leveraging the components' network capabilities seamlessly in a quality sensitive manner.

Supporting Information Dependability: Information reliability is a semantic concept – it refers to sensing reliability (as opposed to reliability of individual sensor devices and networks). Given sensing reliability needs, the goal is to design adaptations to the sensing process to enhance confidence in the sensing outcome under dynamic changes building on prior research in enabling dependability in networked sensor systems [Laz09]. Two such techniques are discussed below.

Sensor Fusion to Realize Dependability Requirements: Data captured from sensors may be erroneous due to inherent imprecision of the sensing devices and dynamics of the underlying sensing and communication infrastructure. We wish to exploit knowledge of which technologies work well in specific situations to fuse sensors and sensing

mechanisms to improve information reliability. For example, consider an extensible localization framework that enables seamless fusion of multiple localization technologies that have been developed (e.g., GPS, GSM, Wi-Fi, ultrasound, ultra-wideband (UWB), inertial sensors, and IR) and differ in operational costs, infrastructure requirements, levels of accuracy they can achieve and efforts needed to calibrate and use the technology[Hig01, Che06b, Gas06, Ihl05a,Ihl05b]. For instance, Wi-Fi based localization when coupled with a calibration process using fingerprinting could lead to accuracy levels of about 2 to 3 meters. UWB technology can provide higher accuracy, and can be rapidly deployed; however, it is expensive and also requires appropriate placement of outdoor base units. We are exploring a generic approach whereby diverse sensing technologies can be fused together to meet the diverse needs of the ICPS applications (e.g., different location queries) in a cost-effective manner. Fusion techniques have been used significantly to increase sensing accuracy (e.g. inertial sensing and UWB for improved localization [Scz08, Cle99, Mor77]); this can be futher combined with a query-oriented approach to further tailor sensor fusion needs (e.g. location queries may vary in their location resolution needs). Such a formulation allows us to exploit prior work in the database community on optimal generalized plan generation [Laz07] for evaluation of multi-version predicates (sequences of selection predicates with increasing selectivity and cost) to address the sensor fusion problem. Two interesting approaches include pipelined execution and parallel evaluation of the different sensing technologies. Pipelining is useful to split a technology into a pipeline of stages with increasing efficacies and costs. Parallel evaluation can be used to enhance accuracy. When two or more technologies used in combination, the outputs can be combine to obtain an aggregate estimate providing increased confidence in the result. Applications/queries can use our proposed framework in two complementary fashions. The first approach would be to identify the best answers (with least uncertainty) given a total cost budget. An alternate formulation would be to minimize cost to produce results, if possible, at a given level of quality. Cost metrics will be defined to subsume operational factors.

Sensing Recalibration to Deal with Small Perturbations: Physical tampering of infrastructure components (e.g. light/audio/video sensors) can introduce vulnerabilities that can lead to erroneous information capture. Tampering may be initiated naturally; e.g., earthquake tremors and vibrations may cause a shift in the field of view of a camera. It may also be initiated explicity - online (exploit camera API to pan, tilt, or zoom), physically manipulating the camera or introducing an obstacle in its field of view. Our goal is to detect and alleviate such anomalies. Physical security measures, e.g. tamper-proof installations (camera domes) and access control techniques that require human control for adaptations undermine flexibility and cause increased response times when failures occur. Here again, a semantics-based approach is useful in dealing with undependability of the ICPS physical sensing infrastructure. For instance, one can develop techniques to recalibrate heterogeneous sensing components in the physical infrastructure when possible or at least provide feedback on information validity using the notion of “semantic sensing”[Vai09]. Here sensor readings are translated to a finite set of possible “semantic states”, which represent the observed system’s state. For example, a traffic light transitions between three states of interest: “Red”, “Yellow” and Green; a monitored room may be “empty”, “occupied” or “crowded”. The semantic

characteristics of the monitored system are captured by a temporal state transition model which captures probability of the system being in a particular state at a given time, given knowledge of states at previous times. Given such a stream of past sensor observations, the goal is to determine whether and when the set of detected states deviates significantly from expected behavior. The adaptation component will exploit time sequence system semantics to detect when re-calibration of a sensor's parameters is required and will automatically re-learn a new set of detection parameters for the newly evolved system state.

6. Privacy in Pervasive Spaces

Building pervasive spaces requires collecting and logging information about the state of the environment, its users, and its resources. Such sensor driven information, either in its raw form, or in a suitably aggregated state, is made available to a variety of applications or users that need the information. Humans are often an integral component (and the focus of the observation) in many pervasive systems and applications. Such environments include systems that offer location-based services (based on observing user's location), monitoring and surveillance systems that observe human behavior and interactions in instrumented spaces (such as buildings, malls, critical infrastructures) with the goal of physical security and forensics, and smart-spaces (such as smart buildings, houses, office spaces, hospitals) that exploit sensing infrastructure for customized services (such as personalized medical treatment), and improved efficiencies (e.g., automated patient tracking, drug monitoring). In human-centric pervasive spaces, sensors are used to assess the state of the pervasive environment, resources, and individuals immersed in the space. Such a situational assessment is used to drive automated or semi-automated adaptation to the system (e.g., customization of physical spaces in a smart building example) or to drive a decision-making task (e.g., response to an observation of suspicious activity in a surveillance example).

For the environment to provide utility for its users or due to the very nature of sensors, this collected data may include personal information. For instance, a video camera at the entrance of the building will capture the person as she walks by the field of view of the camera. That fine granularity sensor data capture over time could reveal personalizing information is now well established through multiple studies including the study by Cornell researchers who instrumented and monitored a student residence over a period of 2 weeks gathering both electric usage data from the breaker panel as well as visual data using cameras. Their findings revealed that simple data mining approaches over electric data could be used to decipher common activities such as sleeping/awake, usage of different appliances, etc. Similar studies have been replicated in variety of different contexts using diverse types of sensor data (e.g., using water usage in a building). While the potential of sensor data to reveal private information exists, an important question is whether such information leakage is indeed a concern to individuals. Variety of studies including [Kru09] in the context of location privacy, [Mol10], in the context of smart meters, [Kim09a] in the context of fine grained residential monitoring and more recently [Rai11] have revealed that such information leakage is indeed of significant concern and

the concerns increase as subjects begin to realize the nature of inferences about their behavior that can be made based on sensor data. There is now a growing consensus that indeed privacy concerns are a major deterrent to widespread adoption of emerging sentient technologies. [Cha11].

Privacy technologies in the context of data sharing applications have been extensively studied in the recent literature especially in the context of publishing suitably desensitized medical data for research purposes. The classic problem explores how, given a database of personal records that may contain sensitive data, one can disclose (part of or some properties of) data without revealing any personal information about individuals. A trivial approach to ensuring privacy is to reveal no information or to output random results. However, such a scheme does not offer any utility to data consumers, where utility is a subjective measure that depends upon the exact purpose/need of the data by the consumer. Often, utility of the shared data is measured information theoretically to quantify how much of the information in the original data does the shared data preserve. [Dwo06] has established that perfect privacy is an impossible goal given the utility requiremens. Since perfect privacy is not possible [Dwo06], weaker definitions of privacy that may allow for limited exposure of personal data have been proposed.

Amongst the most well studied such privacy criterion is K-anonymity [Swe02] that ensures in the data outputted, any single individual's record is indistinguishable from that of at least K others. K-anonymity can be achieved through the process of generalizing and/or suppressing individual records to ensure indistinguishability amongst a group of records. K-anonymity, by itself, may not prevent possible inferences about the identity of an individual within a group through de-anonymizing attacks using additional knowledge. Furthermore, it does not prevent inference about the sensitive values associated with an individual in a personal record, if, for example, the entire anonymous group had the same value for the sensitive attribute. Mechanisms such as l-diversity [Kif06a] and t-closeness [Li07] put additional constraints on the anonymity group (e.g., l-most frequent sensitive values within an anonymity group are approximately equi-probable) to reduce effect of such inferences. Such extensions, however, do not prevent adversary from learning an association between an individual and a sensitive value in presence of additional knowledge. Nor do they overcome another limitation of K-anonymity – viz., its adhoc nature as a privacy criterion. Differential privacy introduced in [Dwo06] overcomes these limitations by postulating privacy as a bounded increase in probability about presence or absence of a personal record in the database based on the shared data. Typically, in the context of a query, differential privacy is achieved by adding a least amount of noise to the answer so as to ensure a bound on probability. Mechanisms to achieve differential privacy have been studied for a few class of queries and devising methods to deal with a larger class is an active area of ongoing research.

Much of the above discussed work on privacy in the context of data sharing provides a foundation for developing privacy technologies for pervasive spaces. In a pervasive space, sensors produce data records that may contain personal data (about subjects being monitored in the pervasive space) which is then shared with other entities that need the data to build pervasive functionalities. For instance, in a home medical monitoring

application, data captured by variety of sensors about a patient may be monitored by medical practitioners who may need such data for remote diagnostics. Likewise, in a surveillance application, data captured through cameras and other surveillance sensors may monitor activities/behavior of individuals and the data made available to personnel responsible for security of the space. Unlike the traditional data sharing setting, in the context of pervasive spaces, we need to differentiate between three types of entities: subjects (whose personal information is captured in the sensor data), the environment (which includes the personnel responsible for implementation, deployment, and management of the pervasive space. These include, for instance, human operators and system administrators who may create and manage the database of sensor readings), and finally the data consumers who are the end recepients who need the sensor data to realize the pervasive functionality.

Privacy concerns in pervasive spaces may arise due to lack of trust amongst the different entities in the pervasive space. For instance, subjects may not fully trust the environment (or the humans who operate the environment). In such a case, the privacy concerns have to be addressed at the data collection level. Specifically, techniques need to be designed that guarantee minimally invasive data acquisition just necessary to implement the desired functionality. Desensitizing data at the sensors (e.g., obfuscating faces of individuals in video frames), encrypting sensor data coupled with techniques to compute over encrypted representation, or alternatively employing sensing techniques that are less invasive might be some of the approaches that could be used to alleviate such privacy concerns. We highlight some of the work done along such a direction below [Lan01,Cam02].

Privacy concerns in pervasive spaces might also arise (even if the environment itself is trusted) due to sharing of sensor data with other entities. Such privacy concerns are more similar to the traditional privacy problem in data sharing discussed above. As in data sharing applications, privacy here refers to limiting or preventing disclosure of attributes or information about individuals that is deemed as sensitive and the need is not just to protect data that refers to attributes addressed in privacy policies, but also to establish whether an adversary can infer sensitive knowledge from pieces of information that are by themselves not sensitive.

While the concepts and definitions of privacy developed in the context of database privacy apply, implementing privacy in pervasive applications offers additional challenges. First, unlike traditional setting where privacy has been studied, sensor data typically corresponds to continuous monitoring of a real-world activity which leads to additional inference channels. Consider, for instance, location tracking. One could view the data about a person's location as a sequence of timestamped records each identifying a subjects location at a given time. One could then apply variety of anonymization techniques for instance to hide the association of a person with a particular location. However, such an approach, applied blindly, may place a particular person at Boston or Irvine at a given time, and place the person at either Fairbanks or Los Angeles an hour later. Knowing the geography and distances involved, it might not be too difficult to infer that the subject in question was at Irvine and Los Angeles given the impossibility of

reaching Alaska within an hour from Boston. Traditionally, privacy preserving data sharing techniques have assumed that personal records in the data are largely independent which, while justified for the domains they were designed for, is simply not the case for pervasive applications where sensors capture human activities which exhibit strong spatial and temporal relationships. Such relationships are not just limited to sensor data captured about a single individual. Data captured about one individual may lead to inferences about others. For instance, it might be well known fact that "Alice" and "Bob" usually have lunch together.. Now presence of Alice at the cafeteria (captured through a sensor) might reveal information about Bob's location.. Techniques that can prevent inferences even in presence of real-world constraints/knowledge need to be explored in the context of privacy in pervasive spaces.

Privacy Protection In Untrusted Pervasive Environments: Privacy challenges when pervasive environments are not trusted have been explored in variety of directions. At the network layer combines hop-to-hop routing based on handles with limited public-key cryptography to preserve privacy from eavesdroppers and traffic analyzers. At the architectural level, and in a manner similar to outdoor GPS, solutions such as Cricket[Smi04] and Place Lab [Cur08] protect a user's (private) location by having a user's carry-on device calculate its location based on a series of beacons from the infrastructure rather than having the infrastructure compute the location as in Active Badge[Wan92]. Such mechanisms essentially attempt to use non-invasive sensing wherever possible in order to minimize privacy concerns. Approaches to make sensing non-invasive have also been explored in video surveillance settings [Wic04, Fid04]. One such technique is to strip faces from images (or replace humans in images with blobs) at the video camera (which itself is a trusted device) before transmitting the video to the servers. Assuming that basic events (e.g., a person entering a given region or a room) can be detected at the sensors, [Kla09] has explored mechanisms to detect complex multi-sensor events over encrypted representation that does not reveal information about individuals involved in the event. The problem is motivated by a surveillance setting wherein one is interested in detecting activities deemed to be potentially undesirable behaviors such as a person entering into a restricted area by an individual who does not have credentials, repeated entry and exit of a region which is not normal or expected behavior, meeting between individuals who normally are not expected to meet, etc. The danger is that sensors installed to detect such activities may also empower the environment (or those involved in administering the environment) to spy on and detect other activities of individual which are not undesirable. For instance, an employer who monitors employees to enable for instance locating them in time of need, might very well use the surveillance system for entirely different purpose such as taking account of the amount of time a person takes a break from work. In [Kla09], base level event data is kept encrypted and techniques are developed to detect higher level events that correspond to undesirable behaviors on the encrypted representation of events. This way, the environment is able to detect the undesirable behaviors (which was the advertised/original purpose of the surveillance system) without gaining further knowledge about the subjects through sensor data. The technique developed is inspired

by practical approaches to implement oblivious computation that ensure that the adversary (i.e., the server where event logs are stored) is not able to gain any information about individuals and events they are involved in unless, of course, they are involved in a undesirable behavior.

Privacy Preserving Data Sharing in Pervasive Spaces: If the environment itself can be trusted, the privacy challenge in pervasive spaces becomes that of ensuring that sensor data (that may contain personal information) about the subject should only be revealed to the receiver in accordance with the subjects privacy policies that limit the data that a pervasive space can share with others. Note that in the context of privacy, the issue is *not just about access control, but also inference control*, that is, the information disclosed should not allow the recipient to infer knowledge about properties considered sensitive by the subjects. Consider, for example, the privacy policy of an individual Bob who does not wish Alice to know when (and how often) he visits the smoking-lounge in the office building. Naturally, Bob's policy will disallow Alice to get an update from the sensor at the entry or exit of the smoking-lounge. However, the information about Bob's presence in the corridor is not sensitive in this privacy policy and, thus, could be revealed. If the corridor was the only way to reach the smoking lounge, knowledge of Bob's presence in the corridor would enable Alice to gain information about his visits to the smoking room. Enforcing privacy policies becomes challenging in the context of potential inference. Enforcement requires determination of various inference channels through which an adversary could gain knowledge about the attribute considered sensitive by the policy. One approach to privacy policy enforcement and inference control is to use *privacy as a constraint*, wherein privacy policies (either user-specified or specified by the system on behalf of the user) determine what information can be divulged without violating privacy. This leads to maximal information that is compatible with privacy policies. Another approach is to combine privacy policy based information sharing with the principle of *minimum disclosure* wherein a request for events or data is for a specific purpose, and the goal is to ensure that the least amount of information is divulged while still meeting the information needs of the application.

Privacy policies are specified to limit disclosure of sensitive information to others in the context of data sharing. Multiple privacy and security policy languages have been studied in diverse contexts [Lor02a, Sch03, Lor02b, Mos04, Sch08, Kag03, Lan02]. Policies expressed in such languages control access to (potentially) sensitive data but none of these languages, or their reasoners and enforcement mechanisms, can deal with the combined requirements for privacy in pervasive spaces, including:

- Express and reason about privacy policies that refer to both static information and streams of events.
- Allow users to specify their goals concerning inference control [Joa07]
- Control the ability of the recipient to infer sensitive data by considering adversary knowledge.

Extensions to an existing policy language to address the above mentioned requirements are an important direction for future research. However, the main challenges are not in adapting the syntax of a language, but rather in mechanisms for composing multiple

policies, and in providing a reasoning framework in which policy enforcement for composite policies expressed in such language becomes feasible.

Privacy Policy Enforcement through "Privacy as Constraint": Given that we can specify privacy policies, the next challenge is to develop mechanisms to enforce such policies. Given a sensor data produced that may result in a policy violation, the system will obfuscate the data to the degree such that the data shared does not result in the policy violation. This is, however, a complex task given that the policy violation might not be caused directly by the sensor data under consideration but rather through an inference made using the sensor data. One of the approaches to obfuscate the resulting sensor data is to add controlled amount of noise in the released data to prevent direct or indirect violation of policy. Such an approach is adopted, for instance, in PoolView [Gan08] in the context of participatory sensing. In PoolView, clients (subjects) independently perturb their data using an application-specific noise model. The noise model is robust to reconstruction attacks and yet allows computation of aggregate property of the data by appropriately cancelling the noise. A similar technique is also used in PreSense [Shi10] to support computation of certain aggregation functions. While the approach to controlled noise addition has worked in certain situations, a more general system based on such a technique that may work for diverse sensor data sharing applications has yet to be devised. One can possibly argue that such a general solution that provides privacy guarantees irrespective of the types and nature of inferences possible in the pervasive space will necessarily be very pessimistic. A system built under the worst case assumption of adversarial knowledge will possibly have too low a utility for practical purposes.

An alternate approach to designing privacy preserving data sharing for pervasive spaces is to first explicitly model and represent domain knowledge that capture all the laws of inference in the system. Such domain knowledge may be a result of a combination of developing a formal model of the pervasive space that allows for reasoning and/or rules inducted from past sensor logs generated. Given such a knowledge representation, one can use a policy reasoning algorithm as part of the enforcement mechanism, to filter information being gathered. Given a formal model of information and general inference rules, one can carry out analyses such as whether a given privacy policy can be enforced, and using model-checking and other formal analysis tools, one can detect cases where given set of policies allows information to leak. [Dan09] proposes such a model driven policy enforcement approach. The paper proposes an interactive model wherein the system starts with some generic knowledge (either explicitly specified or gained through mining using logs of the pervasive space). In addition, the knowledge can be refined through interaction with users if a privacy violation occurs. If the outcome of the formal analysis establishes possible violations of privacy policies through inference, different information hiding techniques are used to prevent such information leakage. These techniques use a range of data modification approaches such as coarsening, perturbation, and transformation or clustering etc. Coarsening, for instance, refers to changing the resolution at which an attribute value is revealed. Typically, for categorical attributes, taxonomies or some form of partial-order defined on the domain is used to coarsen the

value. For instance, if the location information is captured at the room level, and "room < floor < building" denotes the partial order, then coarsening may result in Bob's location being revealed only at the building level. The goal in such an approach is minimal modification of the data such that no policy violations occur or number of such violations are limited. An alternate mechanism that could be used is clustering which will group a set of events to make them indistinguishable from each other. For instance, "Bob entered smoking lounge" may be clubbed together with "Alice entered the elevator" and "Tom exited the building". In effect, the observer only knows that one of these 3 events occurred. Notice that such an approach is based on the assumption that it is possible to perform data modifications, associate a measure of disclosure with the modified representation, and choose the least modification that meets the disclosure control requirements. While sometimes this is not difficult – for instance, in the case when events are location updates, in general, this is a significant challenge. Indeed, the space of possible data modifications may be exponential (or perhaps even unbounded), disclosure depends upon the model of adversarial knowledge (inference rules) and might not be easily quantifiable, and identification of the optimal solution is a combinatorial challenge.

Privacy Policy Enforcement through "Minimal Disclosure": So far we have discussed *privacy as a constraint* approach wherein users specify their privacy needs in the form of a policy and the system ensures that the data released to the users (or potential adversaries) meets the privacy requirement of subjects. This approach essentially attempts to maximize information being shared with the user as long as the information does not violate the privacy requirements of the subjects. An alternate approach is that of *minimal disclosure to data sharing*. The latter requires users to specify their goals for the data or purpose for information, and the objective of the system is to provide minimal information to the data user such that the information suffices for the task of the individuals while at the same time minimizing information disclosure. While we are not aware of any such published research, a minimum disclosure approach, we believe would be very useful in the context of pervasive spaces since sensor applications typically are designed to tolerate errors and uncertainty in data. In pervasive space, minimum disclosure approach can be designed to explore a tradeoff between the amount of information disclosure and the probability with which an adversary or untrusted user can determine if an event actually occurred. The notion of information disclosure here is based on the amount of information that the untrusted user gains about the state of the system from the event sequence he observes. Such a setting is very useful for surveillance since the application can negotiate with the underlying system to get additional information if the number of false-positive events is too high for its purpose. For instance, consider that the surveillance system is trying to determine of the number of students currently in the classroom. Assume that we generalize the events to blur the distinction between entry and exit. Such a generalization may lead to wrong estimates suggesting anywhere from 0 to 50 people are in the room. Assume that the class consists of only 10 students and the (potential) estimate of 50 is too high, thus raising an alarm. The surveillance system can then negotiate for a less generalized representation (hence lower privacy) so as to get a more accurate count.

7. Conclusions

In this chapter, we discussed fundamental challenges in building pervasive sensing and monitoring applications. While there have been significant advances in designing, implementing and deploying sensor networks as well as large scale streaming systems, these techniques do not address challenges that arise due to heterogeneity of sensing platforms nor do they exploit the semantic context in which many of these devices and applications are deployed in a principled manner. This chapter argues for a semantic approach to modeling of pervasive spaces and applications where information is modeled/represented at the semantic level - viz., using entities, objects, spaces. We discussed a cross-layer view of pervasive systems and its realization in a middleware framework, SATWARE. We also illustrated how semantics can be exploited to effectively schedule data capture under resource constraints using a cost-benefit model and how lack of reliability at the infrastructure and information levels can be compensated for by intelligent monitoring and use of semantics to enable adaptations for robust application execution. The chapter also makes an argument that deployment issues of privacy can be addressed using the semantic approach which enables the specification of privacy policies in the middleware which can then control disclosure both of the raw and derived sensor data. While we do not elaborate on privacy framework in this chapter, we refer the reader to [Mas09], which develops a utility-based framework in which privacy violations are modeled as negative utility for a target being observed. The framework maximizes the utility of the information being released to an observer given the privacy constraints. Our future research directions will focus on exploiting the foundational frameworks we have developed to address issues related to robustness of pervasive applications, techniques for managing uncertainty in the various phases of event processing and challenges that benefit from user-in-the-loop interactions.

References

[Aba03] D.J. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, and S. Zdonik. Aurora: A new model and architecture for data stream management. VLDB Journal, 12:120-139, 2003.

[Aba05] Daniel J. Abadi and Yanif Ahmad and Magdalena Balazinska and Cetintemel and Mitch Cherniack and Jeong-Hyon Hwang and Wolfgang Lindner and Anurag S. Maskey and Alexander Rasin and Esther Ryvkina and Nesime Tatbul and Ying Xing and Stan Zdonik, The Design of the Borealis Stream Processing Engine, Proceedings of the 2005 CIDR Conference.

[Ada98] E. W. Adams, A Primer of Probability Logic, CSLI Publications, 1998

[Aky05] I. Akyildiz, X. Wang, and W. Wang. Wireless Mesh Networks: A Survey, Computer Networks Journal (Elsevier), March, 2005.

[Ana09] G. Ananthanarayanan, and I. Stoica. Blue-Fi: Enhancing Wi-Fi Performance using Bluetooth Signals, In proceeding of MobiSys, 2009.

[Ara04] A. Arasu and B. Babcock and S. Babu and J. Cieslewicz and M. Datar and K. Ito and R. Motwani and U. Srivastava and J. Widom, STREAM: The Stanford Data Stream Management System, Book chapter, 2004

[Ara06] Arasu, A. and Babu, S. and Widom, J. The CQL continuous query language: Semantic foundations and query execution. The VLDB Journal The International Journal on Very Large Data Bases, 2006

[Bab04a] S. Babu, R. Motwani, K. Munagala, I. Nishizawa, and J. Widom. Adaptive ordering of pipelined stream filters. SIGMOD Conference 2004: 407-418, 2004.

[Bab04b] S. Babu and J. Widom. StreaMon: An adaptive engine for stream query processing. SIGMOD Conference 2004: 931-932, 2004.

[Bai06] M. D. Bailey. A Scalable Hybrid Network Monitoring Architecture for Measuring, Characterizing, and Tracking Internet Threat Dynamics, PhD thesis, University of Michigan, Ann Arbor, MI, USA, 2006.

[Bej03] Y. Bejerano and R Rastogi. Robust monitoring of link delays and faults in IP networks. INFOCOM, 2003.

[Bis07]Joachim Biskup and Jan-Hendrik Lochner. Enforcing confidentiality in relational databases by reducing inference control to access control. In ISC, pages 407–422, 2007.

[Bre07] Lars Brenna, Alan Demers, Johannes Gehrke, Mingsheng Hong, Joel Ossher, Biswanath Panda, Mirek Riedewald, Mohit Thatte, Walker White. "Cayuga: A High-Performance Event Processing Engine" (demo). SIGMOD 2007

[Bro99] J. Broch, D. A. Maltz, and D. B. Johnson, Supporting Hierarchy and Heterogeneous Interfaces in Multi-Hop Wireless Ad Hoc Networks, In Workshop on Mobile Computing held in conjunction with the International Symposium on Parallel Architectures, June 1999.

[Byu03] H. Byun, B. Ko. Robust face detection and tracking for real-life applications, International Journal of Pattern Recognition and Artificial Intelligence, vol. 17, 2003.
[Car07] D. Carney and U. Cetintemel and M. Cherniack and C. Convey and S. Lee and G. Seidman and M. Stonebraker and N. Tatbul and S. Zdonik, Monitoring Streams--A new Class of Data Management Applications, Brown Computer Science, 2007, Technical Report

[Cam02]R. Campbell and et. al. Towards Security and Privacy for Pervasive Computing. In ISSS 2002, 2002

[Cha03] Sirish Chandrasekaran and Owen Cooper and Amol Deshpande and Michael J. Franklin and Joseph M. Hellerstein and Wei Hong and Sailesh Krishnamurthy and Sam Madden and Vijayshankar Raman and Fred Reiss and Mehul Shah, TelegraphCQ: Continuous Dataflow Processing for an Uncertain World, Proceedings of the 2003 CIDR Conference,

[Cha03b]Intanagonwiwat,, Chalermek and Govindan,, Ramesh and Estrin,, Deborah and Heidemann,, John and Silva,, Fabio, Directed diffusion for wireless sensor networking, IEEE/ACM Trans. 2003

[Cha04] R. Chandra, and P. Bahl. MultiNet: Connecting to Multiple IEEE 802.11 Networks Using a Single Wireless Card, In proceeding of INFOCOM, 2004.

[Cha11] Supriyo Chakraborty, Haksoo Choi, Mani B Srivastava, "Demystifying Privacy In Sensory Data: A QoI based approach," The Third International Workshop on Information Quality and Quality of Service for Pervasive Computing , March 2011

[Che06b] B. Chen, L. Tong, and P. Varshney. Channel aware distributed detection in wireless sensor networks, IEEE Signal Processing Mag., vol. 23, no. 4, pp. 16–25, 2006.

[Che06a] K. Chebrolu, and R. Rao. Bandwidth Aggregation for Real-Time Applications in Heterogeneous Wireless Networks, IEEE transactions on mobile computing, volume: 5 issue: 4 page: 388, 2006.

[Chi07] Mung Chiang, Steven H. Low, A. Robert Calderbank, and John C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. In Proceedings of the IEEE, volume 95, pages 255-312, January 2007.

[Ciu06] Ciuff Oletti and M. Polychronakis. Architecture of a Network Monitoring Element. Technical Report TR0033, CoreGRID Project, 2006.

[Cla07] M. Clavel, F. Duran, S. Eker, P. Lincoln, N. M. Oliet, J. Meseguer, and C. Talcott. All about Maude: A high-performance logical framework. Springer, LNCS Vol. 4350, 2007.

[Cle99] R. Clemen. Combining probability distributions from experts in risk analysis, Risk Analysis, vol. 19, pp. 187–203(17), April 1999.

[Coo07] D. J. Cook, and S. K. Das. How Smart Are Our Environments? An Updated Look at the State of the Art Source, in proceeding of Pervasive and Mobile Computing, 2007.

[Cra02a]Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification, 2002.

[Cra02b] Lorrie Cranor, Marc Langheinrich, and Massimo Marchiori. A p3p preference exchange language 1.0 (appel1.0), 2002.

[Cur08] Kevin Curran and Eoghan Furey. 2008. Pinpointing users with location estimation techniques and Wi-Fi hotspot technology. Int. J. Netw. Manag. 18, 5, September 2008.

[Dan09] Daniel Massaguer, Bijit Hore, Mamadou H. Diallo, Sharad Mehrotra, Nalini Venkatasubramania. Middleware for Pervasive Spaces: Balancing Privacy and Utility (MIDDLEWARE 2009, Urbana Champain, USA) [PDF]

[Del05] Franca Delmastro. From pastry to crossroad: Cross-layer ring overlay for ad hoc networks. In PERCOMW '05: Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications Workshops, pages 60-64, Washington, DC, 2005. IEEE Computer Society, 2005.

[Den07] G. Denker, E. Elenius, R. Senanayake, M.-O. Stehr, and D. Wilkins. A Policy Engine For Spectrum Sharing. In F. Jondrall and P. Marshall, editors, 2nd IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN 2007), Dublin, Ireland, 17-20 April, 2007, pages 55–65. IEEE, 2007. http://www.ieee-dyspan.org.

[Dwo06] C. Dwork, Differential Privacy, ICALP, 2006

[Ele07] Daniel Elenius, Grit Denker, Mark-Oliver Stehr, Rukman Senanayake, Carolyn L. Talcott, and David Wilkins. CoRaL - Policy Language and Reasoning Techniques for Spectrum Policies. In POLICY, pages 261–265, 2007.

[Fag90] Ronald Fagin and Joseph Y. Halpern and Nimrod Megiddo, A Logic for Reasoning about Probabilities, Information and Computation, 1990

[Fid04] D. Fidaleo, H. Nguyen, and M. Trivedi, The Networked Sensor Tapestry: A Privacy Enhanced Software Architecture for Interactive Analysis and Processing of Data in Video Sensor Networks, ACM International Workshop on Video Surveillance & Sensor Networks, 2004.

[Fok05] Chien-Liang Fok and Gruia-Catalin Roman and Chenyang Lu, Rapid Development and Flexible Deployment of Adaptive Wireless Sensor Network Applications, ICDCS'05

[Gaia05] Project Gaia OS, 2007

[Gu03] Xiaohui Gu, Klara Nahrstedt, Rong N. Chang and Christopher Ward "QoS Assured Service Composition in Managed Service Overlay Networks", IEEE ICDCS 2003.

[Gan08] R. Ganti, N. Pham, Y. Tsai, and T. Abdelzaher. PoolView: Stream Privacy for Grassroots Participatory Sensing. In ACM SenSys, 2008

[Gas06] M. Gastpar, M. Vetterli, and P.L. Dragotti. Sensing reality and communicating bits: A dangerous liaison, IEEE Signal Processing Mag., vol. 23, no. 4, pp. 70–83, 2006.

[Ged05] Bugra Gedik and Ling Liu. Location-privacy in mobile systems: A personalized anonymization model. In ICDCS, 2005.

[Gru03] M. Gruteser, and D. Grunwald. Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking, in proceeding of MobiSys, 2003.

[Gu03] "Xiaohui Gu and Klara Nahrstedt and Rong N. Chang and Christopher Ward", IEEE Computer Society, 2003

[Gus03] E. Gustafsson, and A. Jonsson. Always Best Connected, IEEE Wireless Communications Magazine, February 2003.

[Gut02] S. Gutierrez-Nolasco and N. Venkat. A reflective middleware framework for communication in dynamic environments. Intl. Symp. on DOA, 2002.

[Ham04] A. Hampapur, L. Brown, J. Connell, M. Lu, H. Merkl, S. Pankanti, A.W. Senior, C.Shu, and Y-L Tian, The IBM Smart Surveillance System, IEEE CVPR, Washington D.C., June 2004

[Han03] Q. Han and N. Venkatasubramanian. Addressing timeliness/accuracy/cost tradeoffs in information collection for dynamic environments. The 24th IEEE International Real-time Systems Symposium (RTSS), Cancun, Mexico, December 2003.

[Han04c] Q. Han, B. Nguyen, S. Irani and N. Venkatasubramanian. Time-Sensitive Computation of Agreegate Functions over Distributed Imprecise Data, In proceeding of 12th IEEE International Workshop on Parallel and Distributed Real-time Systems, April, 2004.

[Han07] L. Paradis and Q. Han. A survey of fault management in wireless sensor networks. Journal of Network and Systems Management (JNSM), Vol. 15, No. 2, June 2007.

[Hig01] J. Hightower and G. Borriello. A survey and taxonomy of location systems for ubiquitous computing, IEEE Computer, Tech. Rep., 2001.[Ho08] T. Ho, and D. S. Lun. Network Coding: An Introduction, Cambridge University Press, Cambridge, U.K., 2008.

[Hor07d] B. Hore, H. Jafarpour, R. Jain, S. Ji, D. Massaguer, S. Mehrotra, N. Venkatasubramanian, and Utz Westermann. Design and implementation of a middleware for sentient spaces. In Proceedings of ISI'07, 2007.
[Hor07e] Bijit Hore, Jehan Wickramasuriya, Sharad Mehrotra, and Nalini Venkatasubramaniam. Privacy preserving event detection in pervasive spaces. UCI-ICS Technical Report, 2007.

[Hua07] F. Huang, Y. Yang, and L. He. A flow-based network monitoring framework for wireless mesh networks. IEEE Wireless Communications 2007.

[Ihl05a] A.T. Ihler. Inference in sensor networks: Graphical models and particle methods, Ph.D. dissertation, MIT, June 2005.

[Ihl05b] A.T. Ihler, J.W. Fisher III, R.L. Moses, and A.S. Willsky. Nonparametric belief propagation for self-localization of sensor networks, IEEE J. Select Areas Commun., vol. 23, no. 4, pp. 809–819, 2005.

[Jaw08] I. Jawhar, Z. Trabelsi, and J. Al-Jaroodi. Towards More Reliable Source Routing in Wireless Networks, International Conference on Networking, Architecture, and Storage, 2008.

[Joa07] Joachim Biskup and Jan-Hendrik Lochner. Enforcing confidentiality in relational databases by reducing inference control to access control. In ISC, pages 407–422, 2007

[Jon01] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen. A Survey of Energy Efficient Network Protocols for Wireless Networks, Wireless Networks. Volume 7, August 2001.

[Kal07] Swaroop Kalasapur and Mohan Kumar and Behrooz Shirazi, "Dynamic Service Composition in Pervasive Computing", IEEE Trans. Parallel Distrib. Syst, 2007

[Kag03] L. Kagal, T. Finin, and A. Joshi. A policy language for a pervasive computing environment, 2003.

[Kha05] S. Khan, M. Sgroi, E.Steinbach, and W. Kellerer. Cross-layer optimization for wireless video streaming - performance and cost. In ICME '05: Proceedings of the IEEE International Conference on Multimedia & Expo, 2005.

[Kif06a] D. Kifer and J. Gehrke, l-Diversity: Privacy Beyond k-Anonymity, ICDE, 2006

[Kim06] K.-H. Kim and K.G. Shin. On accurate measurement of link quality in multi-hop wireless mesh networks. In MobiCom 2006, pp. 38–49, ACM Press, 2006.

[Kim07a] M. Kim, N. Dutt, N. Venkatasubramanian, and C. Talcott. xTune: Online verifiable cross-layer adaptation for distributed real-time embedded systems. In The 28th IEEE Real-Time Systems Symposium (RTSS 2007) Student Forum, 3-6 December 2007, Tucson, AZ, 2007.

[Kim08a] M. Kim, M. Stehr, C. Talcott, N. Dutt, and N. Venkatasubramanian. Constraint refinement for online verifiable cross-layer system adaptation. In DATE ’08: Proceedings of the Design, Automation and Test in Europe Conference and Exposition, 2008.

[Kim09a] Y. Kim, T. Schmid, M.B. Srivastava and Y. Wang, Challenges in resource monitoring for residential spaces, BuildSys, 2009.

[Kla09] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. Exploring privacy concerns about personal sensing. Pervasive Computing, pages 176–183, 2009

[Koz04] Ulas C. Kozat, Iordanis Koutsopoulos, and Leandros Tassiulas. A framework for cross-layer design of energy-efficient communication with QoS provisioning in multi-hop wireless networks. In INFOCOM, 2004.

[Kru00] J. Krumm, S. Harris, B. Meyers, B. Brumitt. Multi-camera multi-person tracking for easyliving, IEEE Workshop on Visual Surveillance, 2000.

[Kru09] J. Krumm, A survey of computational location privacy, Personal Ubiquitous Computing, 2009

[Lan01] M. Langheinrich. Privacy by Design: Principles of Privacy-Aware Ubiquitous Systems. Presented at ACM UbiComp 2001, Atlanta, GA 2001.

[Lan02] Marc Langheinrich. A privacy awareness system for ubiquitous computing environments. In Gaetano Borriello and Lars Erik Holmquist, editors, 4th International Conference on Ubiquitous Computing (Ubicomp 2002), number 2498 in LNCS, pages 237–245. Springer-Verlag, September 2002.

[Laz07] I. Lazaridis, and S. Mehrotra. Optimization of multi-version expensive predicates. In SIGMOD'07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data. New York, NY, USA: ACM, 2007, pp. 797–808, 2007.

[Laz09] I. Lazaridis, Q. Han, S. Mehrotra, and N. Venkatasubramanian. Fault-Tolerant Evaluation of Continuous Queries over Sensor Data, International Journal on Distributed Sensor Networks (IJDSN), Vol. 5, No. 4, 2009.

[Lee08] K. Lee, A. Shrivastava, M. Kim, N. Dutt, and N. Venkatasubramanian. Mitigating the impact of hardware defects on multimedia applications: A cross-layer approach. In MM '08: Proceeding of the 16th ACM international conference on Multimedia, pages 319-328, 2008.

[Li07] N. Li and T. Li, t-Closeness: Privacy Beyond k-Anonymity and l-Diversity, ICDE, 2007

[Liu05] Bin Liu and Amarnath Gupta and Ramesh Jain, MedSMan: A Streaming Data Management System over Live Multimedia, ACM MM'05,

[Lor02a] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification, 2002

[Lor02b] Lorrie Cranor, Marc Langheinrich, and Massimo Marchiori. A p3p preference exchange language 1.0 (appel1.0), 2002

[Lu05] X. Lu, K. Moriyama, I. Luque, M. Kanda, Y. Jiang, R. Takanuki, and K. Mori. Timeliness and Reliability Oriented Autonomous Network-Based Information Services Integration in Multi-Agent Systems, IEICE Transactions on Information and Systems archive, September 2005.

[Lu97] S. Lu, V. Bhargavan, and R. Srikant. Fair Scheduling in Wireless Packet Networks, In Proc. ACM Sigcomm'97, pages 63–74, 1997.

[Luo03] H. Luo, R. Ramjee, P. Sinha, Li. Li, and S. Lu. UCAN: a Unified Cellular and Ad-Hoc Network Architecture, International Conference on Mobile Computing and Networking archive Proceedings of the 9th annual international conference on Mobile computing and networking, 2003.

[Mad02] S. Madden, M. J. Franklin, J. M. Hellerstein, and Wei Hong. TAG: A Tiny Aggregation Service for Ad-Hoc Sensor Networks. OSDI 2002.

[Mad05] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TinyDB: An acquisitional query processing system for sensor networks. ACM Trans. Database Syst. 30(1): 122-173 2005.

[Mav05] C. X. Mavromoustakis, and H. D. Karatza, Bandwidth Clustering for Reliable and Prioritized Network Routing Using Split Agent-Based Method, Fourth International Workshop on Assurance in Distributed Systems and Networks , 2005.

[Mcn01] J. McNair, I. F. Akyildiz, and M. Bender. Handoffs for Real-Time Traffic in Mobile IP version 6 Networks, IEEE Globecom 2001, San Antonio, Texas, 2001.

[Meh07] D. Medhi, and K. Ramasamy. Network Routing: Algorithms, Protocols, and Architectures, Morgan Kaufmann Series in Networking, 2007.

[Moh06] S. Mohanty, and I. F. Akyildiz. A Cross-Layer (Layer 2 + 3) Handoff Management for Next Generation Wireless Systems, IEEE Transactions on Mobile Computing, Vol. 5, No. 10, pp. 1347-1360, 2006.

[Moh07] S. Mohanty, and I. F. Akyildiz. Performance Analysis of Handoff Techniques Based on Mobile IP, TCP-Migrate, and SIP, IEEE Transactions on Mobile Computing, Vol. 6, No. 7, pp. 731-748, July 2007.

[Mol10] A. Molina-Markham et. al., Private memoirs of a smart meter, BuildSys, 2010.

[Mor77] P. Morris. Combining expert judgments: A bayesian approach. Management Science, vol. 23(7), pp. 679–693, 1977.

[Mot03] R. Motwani, J. Widom, A. Arasu, B. Babcock, S. Babu, M.Datar, G. S. Manku, C. Olston, Justin Rosenstein, and R. Varma. Query processing, approximation, and resource management in a data stream management system, CIDR 2003.

[Mos04]T. Moses. extensible access control markup language. Technical report, Oasis, 2004.

[Ols03] C. Olston, J. Jiang, and J. Widom. Adaptive filters for continuous queries over distributed data streams. SIGMOD Conference 2003: 563-574, 2003.

[Oxy07] Project Oxygen, 2007

[Pan09] J. Pang, B. Greenstein, M. Kaminsky, D. McCoy, and S. Seshan. Wifi-Reports: Improving Wireless Network Selection with Collaboration, MobySis2009.

[Pra05] A. Pras , R. Meent, and M. Mandjes. QoS in Hybrid Networks, Lecture Notes in Computer Science, 2005.

[Qur09] F. Qureshi, D. Terzopoulos. Planning Ahead for PTZ Camera Assignment and Handoff, 3rd ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC 2009.

[Rai11] A. Raij, A. Ghosh, S. Kumar, M. Srivastava. privacy Risks Emerging from the Adoption of Innocuous Wearable Sensors in the Mobile Environment. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, CA, May 2011

[Rap91] S. Rappaport, The Multiple-Call Hand-off Problem in High-Capacity Cellular Communications Systems, IEEE Transactions on Vehicular Technology, 1991.

[Rob04] Jon Robinson and Ian Wakeman and Tim Owen, Scooby: middleware for service composition in pervasive computing, MPAC '04: Proceedings of the 2nd workshop on Middleware for pervasive and ad-hoc computing, 2004

[Rue 03] Harald Rue and Natarajan Shankar, Introducing Cyberlogic, National Security Agency's third High Confidence Software and Systems Conference, 2003

[Sai07] F. Sailhan, L. Fallon, K. Quinn, P. Farrell, S. Collins, D. Parker, S. Ghamri Doudane, and Y. Huang. Wireless mesh network monitoring: Design, implementation and experiments. Globecom, 2007

[Scz08] S. Sczyslo, J. Schroeder, S. Galler, and T. Kaiser. Hybrid localization using UWB and inertial sensors, in IEEE International Conference on Ultra-Wideband, 2008. ICUWB 2008, vol. 3, 2008.

[Sch03]M. Schunterand P. Ashley, S. Hada, G. Karjoth, and C. Powers.
Enterprise privacy authorization language (epal 1.1). Technical report, IBM, 2003.

[Shi10] J. Shi, R. Zhang, L. Yunzhong and Y. Zhang, PriSense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems, Infocomm, 2010.

[Smi04] Adam Smith, Hari Balakrishnan, Michel Goraczko, and Nissanka Priyantha. 2004. Tracking moving devices with the cricket location system. In Proceedings of the 2nd international conference on Mobile systems, applications, and services (MobiSys '04). ACM, New York, NY, USA, 190-202. 2004.

[Sri05] U. Srivastava, K. Munagala, and J. Widom. Operator placement for in-network stream query processing. In ACM PODS 2005, 2005.

[Ste08] M. Stehr and C. Talcott. Planning and learning algorithms for routing in disruption-tolerant networks. MILCOM 2008.

[Ste10] Mark-Oliver Stehr and Minyoung Kim and Carolyn L. Talcott, Toward Distributed Declarative Control of Networked Cyber-Physical Systems, Proc. 7th Int. Conf., Ubiquitous Intelligence and Computing (UIC'10)

[Sti99] S. Stillman, R. Tanawongsuwan, I. Essa. A system for tracking and recognizing multiple people with multiple cameras, Proceedings of the Second International Conference on Audio-Vision-based Person Authentication, 1999.

[Sus97] Graf,, Susanne and Sadi, Hassen, Construction of Abstract State Graphs with PVS, CAV '97: Proc. 9th Int. Conf. Computer Aided Verification 1997
[Tat03] Nesime Tatbul and Ugur Cetintemel and Stanley B. Zdonik and Mitch Cherniack and Michael Stonebraker, Load Shedding in a Data Stream Manager", VLDB 2003

[Swe02] L. Sweeney. "Achieving k-anonymity privacy protection using generalization and suppression", International Journal of Uncertainty, 10(5), 571--588."

[Vai08] R. Vaisenberg, S. Mehrotra, and D. Ramanan. Exploiting semantics for scheduling data collection from sensors on real-time to maximize event detection. In Multimedia and Computer Networks (MMCN'09, San Jose, CA).

[Vai09] R. Vaisenberg , S. Ji, B. Hore, S. Mehrotra, and N. Venkatasubramanian. Exploiting semantics for sensor recalibration in event detection systems. In Multimedia and Computer Networks (MMCN'08, San Jose, CA).

[Wan92]Roy Want, Andy Hopper, Veronica Falco, and Jonathan Gibbons. The active badge location system. ACM Trans. Inf. Syst. 10, 1, 91-102. 1992.

[Wan09] Guojun Wang and Hongjun Zhou, Quantitative logic, Inf. Sci., 2009

[Wic04] J. Wickramasuriya, M. Datt, S. Mehrotra and N. Venkatasubramanian: Privacy Protecting Data Collection in Media Spaces. ACM Multimedia. October, 2004.

[Wil07] David Wilkins, Grit Denker, Mark-Oliver Stehr, Daniel Elenius, Rukman Senanayake, and Carolyn Talcott. Policy-based cognitive radios.
IEEE Wireless Communications, 14(4):41–46, 2007. Special Issue on Cognitive Wireless Networks.

[Xin05] B. Xing and N. Venkatasubramanian. Multi-constraint dynamic access selection in always best connected networks. Mobiquitous 2005, 2005.

[Xin07] B. Xing, M. Deshpande, N. Venkatasubramanian and S. Mehrotra. Towards Reliable Application Data Broadcast in Wireless Ad Hoc Networks, IEEE Wireless Communications and Networking Conference (WCNC), 2007.

[Xu98] G. Xu, T. Sugimoto. A Software-Based System for Realtime Face Detection and Tracking Using Pan-Tilt-Zoom Controllable Camera, Proc. Int. Conf. Pattern Recognition, 1998.

[Yan96] J. Yang, and A. Waibel, A real-time face tracker, Proceedings of WACV, vol. 96, pp. 142-147, 1996.

[Yu07] Xingbo Yu, Sharad Mehrotra, and Nalini Venkatasubramaniam.
Untraceability: Towards location privacy preservation in mobile environments. UCI-ICS Technical Report, 2007.