OpenSense: Open Community Driven Sensing of Environment

Karl Aberer[§] Saket Sathe[§] Dipanjan Chakraborty[‡]
Alcherio Martinoli[§] Guillermo Barrenetxea[§] Boi Faltings[§] Lothar Thiele[†]

[§]EPFL, Switzerland.
{name.surname}@epfl.ch

[‡]IBM Research India,
Academic Guest, EPFL.
cdipanjan@in.ibm.com

[†]ETHZ, Switzerland.
thiele@tik.ee.ethz.ch

ABSTRACT

This paper outlines a vision for community-driven sensing of our environment. At its core, community sensing is a dynamic new form of mobile geosensor network. We believe that community sensing networks, in order to be *widely deployable and sustainable*, need to *follow utilitarian approaches* towards sensing and data management. Current projects exploring community sensing have paid less attention to these underlying fundamental principles. We illustrate this vision through OpenSense – a large project that aims to explore community sensing driven by air pollution monitoring.

Categories and Subject Descriptors

H.1.0 [**Models and Principles**]: General; H.2.0 [**Data Management**]: General

General Terms

Algorithms, Management, Measurement

Keywords

Community sensing, Utility functions, Mobile sensors, Data management, Air pollution, Monitoring

1. INTRODUCTION

Traditional environment sensing principles have been primarily driven by need to optimally reconstruct model phenomenon for consumption by applications and scientists. Air pollution monitoring in urban areas using the community is a rapidly growing area of environmental sensing, and a perfect example of a community driven mobile geosensor network. Data collected from these mobile sensors forms a large geostream corpora, which needs to be mined and analyzed in different dimensions for consumption by applications.

This paper outlines a vision for an *open* community-driven sensing infrastructure (OpenSense) with air quality monitoring as an example. We envision an *open* infrastructure that exploits heterogeneous sensors owned and/or carried by the community for sensing the environment.

IWGS '10, November 2, 2010, San Jose, CA, USA

Community sensing, in order to be sustainable, requires a rigorous and systematic understanding of the model. We believe that in this emerging model, the *environment should be spatially and temporally sampled (and visualized) only at the rate necessary, and not necessarily at the rate to reconstruct the underlying phenomenon.* To realize this, we need to consider factors that converge to model *utility* of data being produced and consumed in this geosensory ecosystem. Furthermore, we need to translate utility models at various layers to utility models used by sensors. To do this, we believe, we need a decentralized approach. This is due to the heterogeneity of service providers, application requirements, sensors, and their decoupled ways of operation in practice.

Utility-optimized approach in community sensing has been considered in a limited context before by Krause *et al.*[13], with a focus on designing near-optimal sensing principles considering privacy and location of privately-owned sensors for reconstructing traffic conditions from GPS data. We argue that *a large variety* of community-sensing based applications could be efficient and sustainable by using utilitarian sensing approaches. These approaches should factor in certain aspects inherent to community sensing – uncontrolled or semi-controlled mobility of sensors, spatial and temporal validity of readings, errors, storage and data management, cost of data production, and privacy requirements of users and applications. The system should be "open", with multiple decoupled infrastructures (sensors or data management) from different service providers, integrated together seamlessly through *data utility models* at every layer of the eco-system. This is unlike wireless/mobile geo-sensor networks [10, 18] where the primary objective is to accurately monitor the environment by specifying the desired mobility/sampling characteristics in order to reconstruct a particular phenomenon.

Air quality monitoring is a topic of extreme importance today. Common air pollutants have direct effect on human health [8] and form a perfect candidate for fine-grained sensing. Since it is costly and infeasible to install sensors everywhere, this can benefit substantially from involving the public into the monitoring task. As such, today there are a variety of projects [1, 3, 7, 11] which use mobile sensors – even mobile phones assisted by sensors – mounted on people for sensing several environmental parameters of urban spaces (toxic gases, diffusion patterns, temperature, humidity, etc.). However, most of the works are primarily focused on, sensor quality improvement, use-case trails, analytics, storage, and visualization of such geostreamed data [6, 12, 17, 20]. Less attention has been focused on the unique characteristics and underlying fundamental principles that should guide large-scale *data production and consumption*, for sustaining a OpenSense-like dynamic and new form of geosensing.

Why Utilitarian Approach?: In essence community-based sensing advocates for efficient *microscopic* [15] monitoring of our environment. However, unlike traditional sensor networks, the community sensing paradigm is different. In traditional sensor networks, producers of the data (i.e., fixed or mobile sensors) are not consumers themselves, but rather are instrumented to create data for consumption by scientists, applications, or community. In community sensing, sensors (producers) and consumers involve people and the devices/entities owned by them; and are tightly coupled. In this case, *sensor* involves devices (e.g., cell phones, thermal watches, thermometers in vehicles, etc.), customized sensing units for measuring dust or toxic gases, and most importantly the "carriers" (cars, buses, people, etc.) – the community. This gives rise to an organic, unstructured sensing paradigm, somewhat analogous to the Web 2.0 model, where the community participates in generating the data. In return, the same community would expect a better value-add than the current macroscopic view of the environment they receive today (e.g., city-wise weather updates).

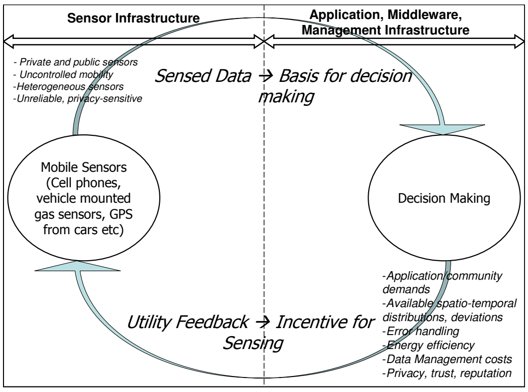

Figure 1: OpenSense cycle of production and consumption of community sensing data.

Consumers of the data (applications or communities) would, in turn, need the producers to produce the data. Since producers and consumers are part of the same cycle, *we believe that making this cycle remain in equilibrium at all times is critical for sustainability and success of this paradigm.* Our vision proposes a way to enable this cycle. In our vision (refer Figure 1), sensed data flowing into the infrastructure forms the basis for decision making – driven by utilitarian approaches that consider several parameters of the *sensing side* and *infrastructure side* (e.g., application demand, data storage cost, transmission costs, error models, privacy, spatio-temporal distribution of available data, deviations, etc.). Data utility computed through this process translates into a feedback for the sensing units, that enables distributed autonomous sensors to determine sensing principles (rules and policies). Decision making is not a global optimization problem, rather is multi-parameter, distributed, and decentralized; with a focus on maintaining the cycle in equilibrium. Next, we illustrate characteristics of this geosensor network through use-cases.

2. PROBLEM ILLUSTRATION

Community sensing faces substantial technical challenges to scale up from isolated, well controlled, small-user-base trials to an open and scalable infrastructure. This infrastructure involves several small-, micro-, or potentially even nano-scale sensors participating in an open "opt-in" model. Let us consider the following use-cases:

Smart Healthcare: Sensors placed on private or public transport moving objects (including fixed mounted sensors at strategic locations) monitor CO, NO_x, etc., in Lausanne, Switzerland. The synthesized data is provided to preventive health researchers to gather a fine-grained model of environmental factors affecting pollution-induced diseases (e.g., asthma, fine particle allergies, etc.). People collaborating in the data collection get discounts and air pollution alerts on their usual travel routes.

Urban Planning: Traffic planning authorities in Delhi, India monitor pollution hotspots of ultrafine particles and determines alternative routing strategies. Nano-scale pollution sensors are attached to cars of volunteers. Furthermore, in-car GPS is used for position approximation. Volunteers get cell phone data charges waived to participate. This information is offered to news media for dissemination and raising public awareness.

Several observations of community-based sensing phenomenon arise from these examples, also revealing some central differences with traditional sensor networks.

1. **Ownership and Participation**: Sensors are private (e.g., phone) or public (e.g., traditional sensing stations) and could be controlled by the owner (e.g., sensors mounted on a car or a public transport vehicle) during the sensing act.
2. **Heterogeneity of Sensor Equipments**: Sensing equipments can be of several types with varying capabilities. They could have different battery capacity, sensor accuracy, or communication methods.
3. **Data Sampling**: Frequent sampling is infeasible as large bulk of data might be wasted. Users also need to invest battery resources (e.g., boards powered by car battery; GPS is energy hungry on cell phones).
4. **Mobility**: Unlike traditional mobile sensor networks, sensors are carried by different entities whose mobility *cannot be controlled* for sensing purposes.
5. **Reliability**: Air quality sensors on moving objects such as buses, cars, or bikes incur mild to major shocks and need recalibration with time. Furthermore, deviations could also be due to the geo-spatial variations (e.g., pollution levels vary drastically within a city [7]).
6. **Trust-worthiness**: Participants in the system (producers, consumers) have incentives to manipulate the data; for instance, to induce bias in environment maps of toxic pollutant distributions.
7. **Privacy:** Participants might be sensitive towards publishing data that can be reverse-engineered to reveal private information (e.g., location).

At its core, traditional sensor network deployments have a *bottom-up* organization with controlled and structured mobility patterns. Also, the sampling rate of the sensor deployment is tuned to the environment. On the contrary, a community-driven geosensory environment has an organic involvement structure, with less control on sensors. Next, we briefly describe the OpenSense project that is geared towards research of sustainable sensing models and management of such geosensor networks.

3. OPENSENSE

OpenSense [4] is an open platform whose major scientific objective is to efficiently and effectively monitor air pollution using wireless and mobile sensors by adopting complex utility driven approaches towards sensing and data management. Optimization goals include: (1) reducing resource consumption (deployment or operational), and thus cost of the infrastructure, while increasing accuracy and value of the information produced, (2) optimizing accuracy of data considering application demands, (3) minimizing transmission, analysis, and storage of measurement data, (4) reducing the latency of real-time information delivery against perturbation factors and uncertainties. For achieving these goals, we will exploit personalized and contextual data consumption patterns of the community and pollution monitoring applications. We aim to study decentralized approaches for management of the geosensory eco-system, using multi-parameter utility-optimized sensing principles.

Deployment Details: In OpenSense, we are deploying *sensing units* on mobile vehicles and stationary monitoring stations around the city of Lausanne, Switzerland. To start with, the project has collaborated with the public transport authority of Lausanne to mount mobile sensing units on public transport buses as well as bus stops (refer Figure 2). These mobile units would enable monitoring of air pollutants, like, Carbon Monoxide (CO), Carbon Dioxide (CO_2), Nitrogen Dioxide (NO_2), etc., at major intersections in the city.

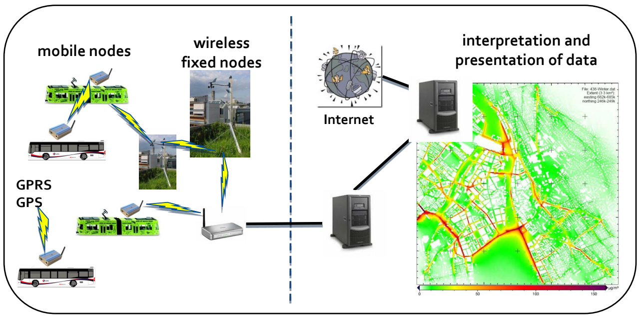

Figure 2: OpenSense Deployment Infrastructure.

3.1 Phenomenon Model Factors

Central to our research is the notion of *data utility*. To realize our vision, our objective is to consider several dimensions of the geosensory eco-system to model the utility of the data being produced and consumed. Secondly, we aim to create a decentralized control system for utility-driven management of the geosensor network ([15] emphasizes need for decentralized spatial computation in geosensor networks). Sustainable community driven air pollution monitoring poses several challenging questions: 1) How to develop sustainable models that encourage users to participate, given the risk of privacy evasion? 2) How to obtain a (quasi)-consistent view of the environmental phenomenon using data contributed by the community? 3) How to unanimously quantify the importance of the data produced by the community for deciding efficient storage and information dissemination strategies using utility models? Next, we critically investigate these questions; we outline the challenges and illustrate factors that need to be considered for achieving sustainable utility-optimized community sensing.

Sensing Model: Unlike traditional sensing, where the primary focus is on optimally sampling the environment, here, sampling policies should be driven by the application-layer requirements and projected utility of data being sampled. This is because price of redundant sampling could be consumption of scarce battery resources or additional transmission and storage cost, and could severely affect the sustainability. Thus, we should have real-time control on the community sensing cycle (refer Figure 1) in order to maintain the data gathering efficiency around the best operational point, possibly under a variety of scenarios characterized by a high degree of uncertainty. This is a challenge for a heterogeneous and private/public owned sensor infrastructure. Furthermore, utility models should incorporate uncontrolled mobility and user behavior. It is a challenge to address varying demands of large user communities and adapt sensing principles. In a way, sensing models form the final outcome of decision making, that are communicated to sensors using utility feedbacks.

Data Management Model: Data storage and management of geosensor networks is costly [6]. Cloud computing infrastructures are adopting economic models since they cannot push arbitrary amounts of data into the management infrastructure [5]. Utility models guiding the OpenSense cycle (Figure 1) need to consider data management aspects (e.g., storage space available, computation capacity, querying/indexing capacities) as factors (parameters) guiding the utility. Furthermore, the data management domain needs to address the following challenges: (i) efficiently applying updates obtained from the continuously streamed data sensed by mobile sensors having unstructured mobility patterns, (ii) as opposed to traditional sensing, users are more likely to be interested only in recent data, thus developing archival techniques that smoothly trade-off representational accuracy with storage space as a function of time are necessary, (iii) all proposed solutions should inherently support uncertain data.

Error Handling Model: It is well-known that sensor hardware often malfunctions, requiring frequent fixes and recalibration. Moreover, unlike traditional sensing, sensors used in community sensing cannot be recalibrated manually. Thus an important challenge is to develop utility functions that tolerate errors while measuring data relevance. We need algorithms for determining uncalibrated sensors and measure the amount of necessary recalibration. These algorithms should be capable of distinguishing between malfunctioning sensors and uncalibrated sensors. There is a need of researching mechanisms for Over-the-air (OTA) transmission of recalibration instructions. In our vision, such control should also be driven by utility (e.g., no need to fix a sensor immediately that is redundant in a certain zone). Moreover, since data produced by the mobile sensors are often imprecise, techniques should be capable of handling imprecision.

Energy Management Model: In traditional sensor networks, one of the prime objective is to conserve energy and increase life-time. Although this objective remains unchanged, sensor hardware shares its energy demands from the same energy sources (e.g., phone battery, car battery), that are also used for other purposes by the user. Thus, utility models should consider adaptive and energy efficient sensing schemes, that take into account communication costs and secondary resource usages on the platform.

Privacy and Reputation Model: There are strong incentives for manipulating sensitive pollution data. This calls for developing robust reputation mechanisms to detect malicious use of sensors to hide pollution. Furthermore, users would demand varying levels of privacy to prevent reverse engineering of their coordinates (e.g., different levels of location cloaking). Hence, utility models should be sensitive to privacy requirements. Reputation management (e.g., signaling or sanctioning reputation schemes [9]) schemes need to be designed based on environmental models of the parameters being monitored.

Application Demand Model: Lastly, applications and community will demand micro-environment updates, potentially over continuous phenomenon or over dynamic geo-spatial events (e.g., pollution hotspots notification). This does not preclude traditional queries for macro updates. Utility of data being produced will critically depend on its importance for answering existing queries in the geosensory ecosystem.

3.2 Management of OpenSense Cycle

To realize our vision of holistic, utility optimized, and sustainable community sensing, we plan to take a step-by-step approach. As a first step, we consider parameters that decide utility of data at every layer of the OpenSense platform. These layers mainly include, (i) sensing environmental pollution parameters, (ii) exchanging data with base stations (communication costs), (iii) efficient storage, (iv) querying data based on application demands, and (v) archiving old/unused data. Notice that the common entity connecting these layers is *data*, which moves from sensors to applications. Moreover, at every layer, important resources are utilized to manage this data. Thus, OpenSense would quantify and track *importance* of data at every layer of the OpenSense eco-system, and measure utility at each layer as a function of local factors and requirements coming from the next layers.

A second aspect of managing the OpenSense cycle is to investigate decentralized decision making and control mechanisms. This is because having a centralized control system with a single optimization goal [13] will not meet demands of heterogeneous producers and consumer applications. As an example, some units may strive to minimize power consumption, while some other units might want to minimize transmission costs. Furthermore, data management costs will vary depending on resource availability. Utility parameters at various layers are used to perform decision making, considering different optimization goals, while accounting for the domain knowledge they build about characteristics of the geosensor network from past data.

Finally, each sensing unit autonomously determines its sampling frequencies after considering the utility of the data it is producing. To quantify the utility of data, we plan to use the *expected utility function*. An expected utility function u maps a set of choices $\mathcal{P}$ to the set of real numbers $\mathbb{R}$, $u : \mathcal{P} \mapsto \mathbb{R}$. This function assigns a real number to every choice $\mathcal{P}$ such that an application's preferences are captured. For e.g., the utility function could be defined over a set of locations, times intervals, or values. Thus, once the utility functions are disseminated in the sensor network, the sensing units can autonomously decide data transmission policies.

4. EXISTING LITERATURE

There are many projects exploring community sensing for efficient air quality monitoring [2, 7, 8, 14, 19]. The Air Project [8] is a public, social experiment in which people are invited to use portable air monitoring devices to explore their neighborhoods and urban environments for pollution and fossil fuel burning hotspots. The Ergo [1] project focuses on delivering nearby air quality readings to mobile users through SMS based updates, using US Government's data sources. The N-Smarts [7] project aims to build a large scale, distributed scientific instrument for characterizing society's relationship to its environment, using environmental sensors embedded in location aware mobile phones. OpenSense targets a much more comprehensive and large-scale monitoring of air quality, with focus on the key scientific questions of sensor heterogeneity, mobility handling, optimal sensing, developing trustworthiness, and fair monitoring eco-systems.

Padhy *et al.* [16] provide energy-aware protocols for adaptive sensing. In their framework the sensors locally – as opposed to based on application requirements – decide their sampling rate based on information theoretic measures. In OpenSense, due to the partially controlled or uncontrolled mobility patterns the sampling obtained is non-uniform and imprecise. Honicky *et al.*[11] advocate a Gaussian Process noise model, a promising approach for handling non-uniform sampling and imprecision produced by location aware sensors. Krause *et al.* [13] propose a complete solution for approximating optimal sensing policies under constraints of sensor availability, context-sensitive value of information, sensor owner preferences about privacy and resource usage. This study is done in the context of road traffic monitoring under centralized settings. OpenSense looks at the general problem of optimal sensing by community-driven sensors. The project considers several challenges including, information dissemination, efficient data management, and energy efficiency in resource utilization.

5. CONCLUSIONS

This paper brings out the generic ecosystem, inherent to the community sensing environments. Though akin to the Web 2.0 model in WWW, important challenges arise for sustaining this paradigm for community-driven geosensor networks. We elaborate our vision of how sensing should be guided using complex utilitarian approaches for sustainability. We elaborate the challenges and discuss our initiatives to address them in OpenSense.

6. ACKNOWLEDGMENTS

The work was supported by NCCR-MICS (http://www.mics.org), a center supported by the Swiss National Science Foundation under grant number 5005-67322, and by the OpenSense project (reference number 839_401) supported by the Nano-Tera initiative (http://www.nano-tera.ch).

7. REFERENCES

[1] Ergo project. http://www.urban-atmospheres.net/Experiments/Ergo.

[2] Urban Atmosphere Project. http://www.urban-atmospheres.net/, 2006.

[3] Cambridge Mobile Urban Sensing. http://www.escience.cam.ac.uk/mobiledata/, 2010.

[4] OpenSense Project. http://www.nano-tera.ch/projects/401.php, 2010.

[5] D. Abadi. Data Management in the Cloud: Limitations and Opportunities. In *IEEE Data Engineering Bulletin*, 2009.

[6] P. Agouris *et al.*. Knowledge Aquisition and Data Storage in Mobile GeoSensor Networks. In *Geosensor Networks*, 2006.

[7] E. Brewer *et al.*. N-Smarts: Networked suite of mobile atmospheric real-time sensors. http://www.cs.berkeley.edu/~honicky/nsmarts/, 2007.

[8] B. Costa *et al.*. Air Project. http://www.pm-air.net/, 2006.

[9] C. Dellarocas. Reputation mechanism design in online trading environments with pure moral hazard. *Info. Sys. Research*, 16(2), 2005.

[10] S. Gandhi *et al.*. Catching Elephants with Mice: Sparse Sampling for Monitoring Sensor Networks. In *Sensys*. ACM, 2007.

[11] R. Honicky *et al.*. Increasing the precision of mobile sensing systems through supersampling. In *UrbanSense*, 2008.

[12] Y. Jung *et al.*. Air Pollution Monitoring System Based On Geosensor Network. In *IGARRS*, 2008.

[13] A. Krause *et al.*. Towards community sensing. In *IPSN*, 2008.

[14] L. Luo *et al.*. Sharing and exploring sensor streams over geocentric interfaces. In *GIS*, 2008.

[15] S. Nittel. A Survey of Geosensor Networks: Advances in Dynamic Environmental Monitoring. In *Sensors*, 2009.

[16] P. Padhy *et al.*. A utility-based adaptive sensing and multi-hop communication protocol for wireless sensor networks. *ACM Transactions On Sensor Networks*, 6(3), 2010.

[17] P. Volgyesi *et al.*. Air Quality Monitoring with SensorMap. In *IPSN*, 2008.

[18] A. Walkowski. Model Based Optimization of Mobile Geosensor Networks . In *Lecture Notes in GeoInformation and Cartography*. The European Information Society, 2008.

[19] W. Willett *et al.*. Common Sense Community: Scaffolding mobile sensing and analysis for novice users. In *Pervasive Computing*, 2010.

[20] J. Yang *et al.*. Integration of wireless sensor networks in environmental monitoring cyber infrastructure. In *WiNet*, 2010.