Privacy and Security in Environmental Monitoring Systems: Issues and Solutions

Sabrina De Capitani di Vimercati, Angelo Genovese, Giovanni Livraga,
Vincenzo Piuri, Fabio Scotti

Dipartimento di Informatica
Università degli Studi di Milano
Crema, Italy
email: *firstname.lastname*@unimi.it

Abstract

There is today an increasing interest in environmental monitoring for a variety of specific applications, with great impact especially on natural resource management and preservation, economy, and people's life and health. Typical uses encompass, for example, Earth observation, meteorology, natural resource monitoring, agricultural and forest monitoring, pollution control, natural disaster observation and prediction, and critical infrastructure monitoring. While on one hand these systems play an important role in our society, on the other hand their adoption can raise a number of security and privacy concerns, which can represent an obstacle for the development of future environmental applications. In this chapter, we identify the main security and privacy issues characterizing the environmental data as well as the environmental monitoring infrastructures. We then provide an overview of possible countermeasures for diminishing the effects of these security and privacy issues.

1 Introduction

Environmental monitoring systems allow the study of physical phenomena and the design of prediction and reaction mechanisms to dangerous situations. In its general form, a monitoring system is composed by a certain number of sensors designed to measure different physical quantities, one or more processing nodes, and a communication network. The sensors provide in output analogical signals, which are conditioned and converted into the digital domain. The digital signals are then transmitted to the computing devices, which perform the aggregation of the obtained data to understand the measured phenomenon.

In our modern society, these systems are becoming more and more important for keeping under control the state of the environment. In fact, they have a fundamental role for detecting new environmental issues and for providing evidences that can help in prioritizing the environmental policies. Such systems are also useful to better understand the relationship between environment, economical activities, and daily life and health of people. For instance, weather affects agriculture prosperity and forest well-being, while environment pollution affects human health and reduces the quality of water, land, cultivations, and forest. There is then great interest in monitoring the environment to associate possible effects with observed phenomena and predict critical or dangerous situations. For instance, today we know that there is a direct link between the exposure to PM10 and PM2,5 and different pathologies of vascular systems. Besides, natural resource management and preservation can also greatly benefit from using monitoring systems to observe the status and its evolution so as to initiate conservation actions when needed. Similarly, natural disaster detection, observation and, eventually, prediction can be based on monitoring the geographical areas of interest. Another sector in which these systems are becoming highly significant is the monitoring of critical infrastructure, in particular encompassing railways, highways, gas pipelines, and electric energy distribution networks.

In the last years, the environmental monitoring systems have been subject to fundamental changes due to the rapid advancements of the technology as well as the development of a global information infrastructure such as Internet that allows an easy and rapid diffusion of the information worldwide. As an example, the advances in spectral and spatial resolutions, new satellite technologies, and the progress in communication technologies have improved the level of detail of satellite Earth observations, thus making available high resolution spatial and spectral data. Although such technological developments have the positive effect of expanding the application fields where environmental data can be successfully used, there is also a negative effect related to the increase of possible misuses of environmental data and systems. As a matter of fact, seemingly innocuous environmental information can lead to privacy concerns. For instance, ambient environmental monitoring data could be used to identify small geographic areas. Property owners identified in the vicinity of a hazardous waste site or other pollution sources could experience decreased property values or increased insurance costs.

In this chapter, we aim at providing a comprehensive analysis of the main security and privacy issues that can arise when collecting, processing, and sharing environmental data. The main contribution of this chapter is the analysis of these security and privacy issues, which involve both the infrastructure of the environmental monitoring systems as well as the data collected and disseminated, along with possible countermeasures for mitigating them. The remainder of the chapter is organized as follows. Section 2 discusses the different kinds of systems and architectures used for environmental monitoring. Section 3 resents what kinds of environmental data are typically collected and analyzed. Section 4 illustrates the main security and privacy issues related to the collection, processing, and sharing of environmental data. Section 5 discusses how such security and privacy risks can be counteracted by adopting suitable protection techniques. Finally, Section 6 concludes the chapter.

2 System Architectures

Environmental monitoring systems have evolved from a simple computer with sensors to composite structures which include specialized subcomponents addressing particular data collection issues. These systems are typically classified by considering the system architecture, the geographical extension of the monitored phenomenon, or the number of functions performed by the system.

Based on the system architecture, environmental monitoring systems can be classified in *centralized*, *distributed*, and *remote sensing systems* [1]. Centralized systems are composed by a single processor or controller, a limited number of sensors and a simple output presentation interface (e.g., a single value on a display). Data are collected by sensors and transmitted to the processing unit which performs data analysis and feature extraction required by the application, and stores all relevant information as specified by the application itself. They may have small dimensions and be easily transported. Examples of centralized environmental monitoring systems are radiation detectors, gas detectors, and laboratory equipment. Centralized systems include also monitoring systems based on a single observation point or systems that use robotic architectures to perform the monitoring of hostile or remote environments [2].

Distributed systems are composed by a high number of sensing nodes and can exploit distributed computing and storing abilities. A sensing node contains a limited number of sensors, a processing unit, and a network communication channel. Sensing nodes collect data, may perform some local processing, and route data and information towards some processing nodes in the distributed structure. Some nodes have interfaces to deliver results of their elaborations and storage devices to save acquired sensor data and processed information. Sensing nodes are deployed in a fixed position or may be mobile on board of robots to explore the environment [3]. Some intelligence may be distributed in sensing and processing nodes to provide local abilities for data processing to extract knowledge as nearer to the sensors as possible, reducing the transmitted data or taking earlier local actions [4]. Sensing nodes can have self-configuration capabilities to adapt their operation to the environment and allow for easier deployment, especially when the environmental conditions are harsh or humans cannot reach the monitored place. Mechanisms are also introduced for automatic network configuration if nodes are added or removed [4], for determining if node measurements are not necessary and thus save energy, or for allowing the nodes to move when a more suitable position is found [5]. Self-calibration techniques are used to set the operating parameters [6]. The distributed structures may help in limiting costs and impact on the environment (e.g., small and inexpensive sensors, shorter and cheaper sensor connections, small low-cost processing units for real-time operation, and possibly wireless transmission for limited interconnection costs).

In the most simple network topology, a central node processes data (Fig. 1), even though continuous data transmission from sensing nodes leads to higher energy consumption, adjacent nodes may measure redundant or highly correlated data, and scalability may be limited due to computational and bandwidth issues.

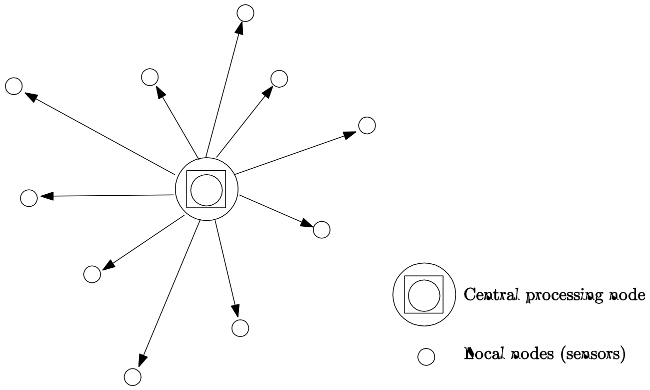

Figure 1. Sensor network with a central processing node

To overcome these problems, hierarchical sensor networks have been used, usually composed by three-levels: *local nodes* (sensors), *intermediate nodes* (local aggregation centers, gateways, or base stations), and a *central processing* node. Some nodes may coordinate some sensors (cluster) by performing synchronization and data fusion [4] (Fig. 2). Computation is distributed in the hierarchical structure to create abstract views of the environment at different abstraction levels and compact the information by extracting the relevant knowledge as locally as possible. Local processing should be performed carefully to avoid possible erroneous interpretation of the corresponding data at higher levels. Appropriate data aggregation techniques must be adopted to achieve a global understanding of the measured phenomena, while avoiding data loss and redundant transmissions [7].

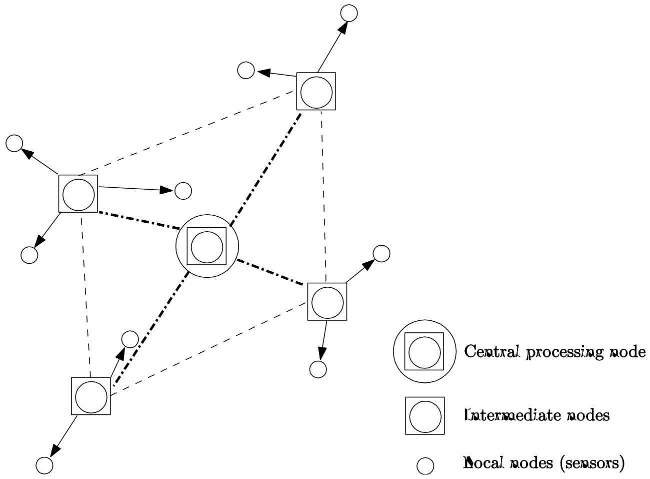

Figure 2 Hierarchical sensor network

Communications are a critical aspect in sensor networks. They can be wired as in the conventional architectures or wireless (Wireless Sensor Networks, WSN). The use of cables to power sensors and transmit the data can create difficulties. Low-power communication protocols and wireless interconnections are often used [8]. In these architectures, geographical position of nodes may not be known a-priori: GPS or GIS systems are used to trace the positions of the data collected from sensors.

Sensing can be performed by using sensors for the specific quantities to be measured and placed locally in the point in which the measure has to be taken. In some environments, direct local sensing may be difficult or even impossible due to costs or environmental/operating conditions. To overcome this problem, for some quantities indirect measures can be taken by observing the point of interest from some distance. Visual Sensor Networks (VSN) are an example of this approach: their nodes are equipped with image-capturing devices and use image-based monitoring techniques. However, they require more complex devices, a greater memory usage, a higher bandwidth, and also nodes with more power consumption. Hierarchical sensor network architectures, composed by heterogeneous nodes, can be used to reduce the costs and the computational load [9].

Remote sensing systems are based on signals and images acquired by sensors installed on artificial satellites or aircrafts and are used for vast geographical phenomena. These systems can capture several types of quantities at a significant distance, for example by aircrafts or artificial satellites. Such systems can be passive or active. In the first case, the sensors only detect quantities naturally produced by the object (e.g., the radiations of the reflected sunlight emitted by the objects). Many passive sensors can be used according to the chosen wavelength and signal dimension (e.g., radiometers, multispectral and hyperspectral imaging). Active systems, instead, send a signal to the object to be monitored and measure the reflected pulse (e.g., RADAR, LIDAR, laser altimeters). Remote sensing techniques can be merged with terrestrial sensor networks to integrate local data with large-scale observations to enhance the observation quality [10].

Environmental monitoring systems can be also classified, according to their geographical extension, in *large-scale*, *regional*, or *localized* monitoring systems [11]. Large-scale environmental monitoring systems are deployed when there is the need to cover a vast geographical area, such as several countries, or even the whole earth globe. They are typically based on distributed networks or remote sensing, and are used, for example, for monitoring seismic activity [12, 13, 14], geophysical monitoring [15], earth pollution [16, 17], global water quality [18], wildfire [19], meteorological data [20, 21, 22, 23], artic ice and snow [20, 24], deserts sand storms [25], or their combinations [26, 27, 28, 29].

Regional monitoring systems typically cover areas such as cities, forests, or a region. They are used, for example, for monitoring water quality [30], air quality [31, 32], meteorological information [33, 34, 35, 36], regional oceanographic processes [33, 36], or wildfires [37, 38, 39, 40, 41].

Localized systems are used for monitoring very localized points, for example, lakes, volcanos, indoor environments, or buildings. Several practical cases are available, e.g., for the quality of the water in lakes, rivers, or small bays [42, 43, 44], the state of glaciers [45], underwater currents [46], air quality in small environments [47, 48, 49], and urban pollution (noise [50], radiation [51]). Localized systems are also used for disaster prevention, e.g., for active volcanos [52], landslides [53], and critical infrastructures [54, 55].

More complex measurement systems, called *heterogeneous sensor networks* [11], have been created by integrating combinations of sub-systems of the above types, with different scales and functions, especially when applications use systems already deployed in the environment of interest or when quantities must be measured in an heterogeneous setting. Some examples of this kind are the UK Climate Change Network [21, 23] for land and aquatic places in the UK, the Global Earth Observation System (GEOSS) [28] for different environmental processes all over the world, the ORION project [56] for oceans. Heterogeneous systems may combine information from local sensor networks with satellite information, for example, linking local sensor networks on a planetary scale [10] or aggregating local imaging data with satellite imaging techniques [57].

Environmental monitoring systems are also characterized by the type of functionalities performed [11]. In *mono-function systems* the measured quantities are directed to provide knowledge for a single application, as in monitoring volcanos [52] or buildings [55]. In *multiple-function systems* data are collected (possibly in subsets of different types from different locations) and used by different applications and even for different global purposes, thus integrating various monitoring systems into a single infrastructure (e.g., [58] supports environmental monitoring, border control, and surveillance applications, while [26, 27] deals with climate and resource monitoring, topography, and disaster prevention).

3 Environmental Data

Before describing the security and privacy issues that characterize an environmental monitoring system, it is fundamental to clarify what kinds of environmental data can be typically collected and possibly released to the public.

Different data types are used in environmental monitoring systems, depending on the applicative context. The used sensors can, in fact, measure data related to different physical quantities: movement, speed, acceleration, force, pressure, humidity, radiation, luminosity, chemical concentration, audio, video, and so on. Usually, the acquired data consist in monodimensional or multidimensional signals (images/frame sequences). The data used by large-scale environmental monitoring systems are inherent to the physical quantities chosen to measure a single phenomenon, and the capture and aggregation of the data are done at a high frequency to perform a continuous monitoring of the phenomenon.

In most cases, the geographical positions of the measuring nodes are fixed, known a-priori and released publicly. For instance, the system described in [34] was composed of 192 measurement stations with fixed and known positions, and performed a continuous monitoring of air temperature, humidity, precipitations, solar radiations, wind speed and direction, and atmospheric pressure. The system described in [12, 14] was composed by more than 150 measurement stations with fixed and known positions, and measured data from seismographs. The system proposed in [16, 17] used different UV radiation detectors to perform a continuous monitoring of radiations. In the case of regional or localized environmental monitoring networks with multiple functions, nodes may not have fixed or known a-priori positions, are equipped with GPS devices, use wireless transmission techniques, and are powered using batteries. For this reason, the data transmission frequency is often smaller than the one used in large-scale environmental monitoring systems. For instance, the system described in [33, 36] performed the continuous monitoring of the waves along the coasts of Louisiana and the Mexican gulf, measuring the wave height, their period, the direction of propagation, the water level, and the direction and speed of the currents. Different kinds of nodes with wireless transmission capabilities can be used. For instance, a volcano monitoring system is described in [52], and uses nodes with infrasound sensors and GPS devices. An experimental visual sensor network for fire monitoring is proposed in [37, 40].

At high level, the lifecycle of environmental data can be divided in three macro-steps: *collection*, *storage*, and *publication*. Data are collected from the environment and stored at the sensor and/or processing nodes. The format of the stored data depends on the specific purpose for which such data have been collected. Authorized parties can access the environmental data for analysis or other purposes. The environmental data (or a subset of them) can then be made publicly or semipublicly available. The publication of the data is typically in the form of *macrodata* (i.e., tables reporting aggregated information about an environmental phenomenon) or *microdata* (i.e., records reporting data related to specific physical measurements) [59].

In the remainder of this chapter, we illustrate some security and privacy risks that may arise in the different steps of data lifecycle. To fix ideas and make the following discussion clear, we refer our examples to a scenario characterized by a localized network in the city of San Francisco, which is under the control of the local municipality. The system is distributed and the sensor nodes are organized according to a centralized configuration. The collected data are stored at processing node $\mathcal{PN}$. *Alice* is an adversary that tries to violate the monitoring system and to discover sensitive information. We also consider a fictitious factory $\mathcal{A}$, which improperly releases pollutants and production rejects in the environment.

4 Security and Privacy Issues in Environmental Monitoring

Environmental monitoring systems and the data they collect can be vulnerable to security and privacy risks [60]. In particular, security risks are related to the threats that can undermine *confidentiality*, *integrity*, and *availability* of both the data and the monitoring systems in their entirety (e.g., system architecture and communication infrastructure). Conversely, privacy risks are related to those threats that can allow an adversary to use the environmental data for *inferring sensitive information*, which is not intended for disclosure and should be kept private. Security and privacy risks are not independent: they are often correlated, and an adversary can exploit a security violation for breaching data privacy. As an example, suppose that *Alice* successfully violates the physical security of processing node $\mathcal{PN}$, causing a security violation that can allow her to access private information related to the pollutant levels in the air of San Francisco. This security violation can allow *Alice* to infer pathologies of the citizens of a given area of the city, violating therefore their privacy.

In this section, we present and illustrate through examples the main security and privacy risks that can arise in the context of environmental monitoring. Note that, in the following discussion, we neither consider the classical security problems related to failures of systems and applications due to errors, nor the reliability and dependability aspects characterizing the system, as our goal is to focus on the less-known security and privacy issues.

4.1 Security risks

Broadly speaking, in our environmental monitoring scenario security risks are related to all threats that can: *i)* damage the infrastructure of the monitoring system; *ii)* violate communication channels connecting different components of the monitoring system; *iii)* allow unauthorized parties to intrude into the monitoring system for malicious purposes. We now describe in details these threats.

Damages to the system infrastructure. Any attack performed with the aim of physically damaging the monitoring system can put at risk the confidentiality, integrity, and availability of the collected environmental data. For instance, suppose that the local municipality of San Francisco wants to build a new playground for children and, to determine the safest location, it analyzes the collected environmental data to discard polluted areas of the city. Suppose also that *Alice* maliciously damages the sensor nodes close to factory $\mathcal{A}$, to hide evidences of the pollutants and production rejects release. Clearly, this compromises the collection of the environmental data, since these sensor nodes become not available (data availability violation). An analysis of the partial environmental data available to the local municipality can erroneously identify an area close to factory $\mathcal{A}$ as the safest area where building the new playground. If this were to happen, children would be exposed to pollutants and production rejects. The same risks apply when all sensor nodes are working properly but the processing node gets attacked and becomes unavailable: in this case, the analysis of the environmental data would not be based on the latest measurements of the sensor nodes, and the results might be compromised. Note that these attacks can impact any of the three steps of the environmental data lifecycle, as similar problems arise when an adversary succeeds in compromising the nodes collecting data (collection step), the database where environmental data are stored (storage step), or the systems where they are published (publication step).

Violation of the communication channels. All communication channels connecting the different components of a sensor network can represent a possible target for an adversary. In particular, the adversary might be a *passive adversary*, that is, she could be only interested in monitoring the communication channels to observe information that she would not be able to access, or an *active adversary*, that is, she could attempt to delete or modify data transmitted on such channels. These two scenarios configure two "classical" security attacks, which can intuitively violate the confidentiality and integrity of the data. Besides such attacks, an adversary can also be interested in monitoring the *accesses* performed on the data by the authorized parties, to discover some sensitive information about them. For instance, the fact that an authorized party accesses data related to the concentration of particulates discloses the fact that the party is interested in discovering the polluted areas. If the party is a building constructor, this may imply that the party is interested in building a new apartment complex, and therefore the adversary can speculate on the costs of the lands. Effective protection of data access also requires the protection of *access patterns*: an adversary should not be able to see whether two accesses performed by two different parties aimed at the same data. For instance, *Alice* should not be able to see if two competitors are interested in performing similar analysis on the environmental data. If so, *Alice* would be able to sell this knowledge to one of the two competitors. Note how the latter two attack scenarios configure two examples of a security violation causing a privacy breach.

Unauthorized access. Environmental data should be available only to users and parties authorized by the data owner. Clearly, restrictions on accesses to environmental data only apply when such data are not publicly released. Unauthorized accesses can possibly involve the database where environmental data are stored after their collection and analysis, or the sensor nodes. The storage server can be a local server, under the control of the data owner, or an external, third-party storage server. In the first case, the server can be considered trusted (i.e., data can be safely stored) and access control should only be enforced against users requesting access to the stored data. In the second case, the external storage server is not considered trusted, and therefore access restrictions should also take into account the fact that the server itself should not be able to access the stored data. An adversary intruding into sensor nodes can be interested in accessing raw data to update them, or to inject false data so that tampered data are sent to the processing node. For instance, *Alice* can be interested in manipulating the measurements performed by the sensor nodes close to factory $\mathcal{A}$ to reduce the concentration of a specific harmful substance. An adversary intruding into the storage servers is clearly interested in accessing environmental data after their collection, normalization, and analysis. Note that collected data can also be stored together with other datasets and, as a consequence, the adversary can discover correlations and dependencies among these different datasets. In all these cases, both data confidentiality and integrity are at risk.

4.2 Privacy risks

Privacy risks are related to all threats that can allow an adversary to infer sensitive information from the collected environmental data. Such inferences can be *direct*, that is, caused by observations in the data collection (e.g., an adversary observing production rejects can discover confidential details of the productive processes of a company), or indirect (e.g., studies on the presence of polluting substances in geographical areas or workplaces can be correlated with studies on the relationship between correlating pollutants and diseases, revealing possible illnesses of individuals living in those areas). Inferred sensitive information can involve individuals, the environmental area on which data have been collected, and also areas close to or correlated with it. As an example, the knowledge that some geographical areas are polluted with harmful substances can also affect individuals who live in other areas if they own properties in the polluted areas. In fact, due to such knowledge, the value of their properties could decrease. Privacy risks can occur when environmental data are made publicly available (publication step) or when they are (properly or improperly) accessed, and can be a consequence of data correlations and associations, observations of data evolutions, unusual data, or the knowledge of users' locations.

- *Data correlation and association.* A possible means through which sensitive information can be inferred is represented by the natural correlations existing among different phenomena. To illustrate, consider a life and sickness insurance company in San Francisco. Suppose that a third-party organization releases a study illustrating the relationship existing between pollutants and rare diseases. Suppose also that the insurance company accesses this study. By analyzing environmental data collected by the local municipality, and comparing them with the study, the insurance company can decide to increase the risk associated with citizens living in polluted areas of San Francisco and re-compute their insurance policies. In addition to correlation, also the association of environmental data with other information coming from different sources can be exploited for inferring sensitive information. For instance, suppose that *Alice* can access a collection of data recording the medical histories of a community of patients. *Alice* might then link such data with airborne pollution studies (by exploiting city and county zones that are used to identify population exposed to specific airborne pollutants), and violate patients' privacy.

- *Data evolutions.* To obtain more meaningful data, sensor nodes can perform several measurements of quantities of interest over time. For instance, a measuring station can continuously record the noise level in a given area of a city. While a high number of samples allows for better analysis of a given phenomenon, such repeated measurements can open the door to possible inference channels leaking sensitive information. For instance, suppose that *Alice* wants to discover the timetable of the freight trains traversing the railroad in San Francisco, which is kept secret by the local train company. Suppose also that the environmental monitoring of the local municipality includes the measurements of the noise pollution in the city. Having access to the measurements collected close to the railway, *Alice* can notice peaks in the noise levels and correlate this information with the public timetables of passenger trains, thus re-constructing the freight trains timetable.
- *Unusual data.* Intuitively, if the measurements obtained from an environmental monitoring system deviate from what is expected or considered as usual, a high risk of sensitive information inference can arise. To illustrate, suppose that the results of the environmental monitoring of the San Francisco city area show a high level of radioactivity. If the neighbor cities do not show such a high level of radioactivity, then these values can be considered surprising, and may witness the existence of a neighbor location storing radioactive material (e.g., nuclear weapons, or rejects of nuclear power plants). Otherwise, if the same level of radioactivity is observed also in other cities, the radioactivity in San Francisco can be due to some peculiarities of the soil.
- *Users' locations.* Mobile phones and smartphones are portable computers that more and more users have and carry with them all times. In the near future, we can imagine that our phones will be equipped with sensors and applications specifically targeted to the environmental monitoring, leading to a *pervasive* environmental monitoring where the sensing will be directly performed by users who will collect data related to the locations they visit. Since users move around the space, measurements have to be tagged with the location in which they have been captured. An adversary able to track the movements of a given user can violate her privacy discovering her frequent addresses (e.g., home and workplace), usual movements (e.g., from home to work), habits, and, accordingly, infer sensitive information about her. For instance, suppose that *Alice* gains access to the set of location-tagged environmental measurements

performed by her colleague *Bob* with his smartphone. *Alice* can notice that *Bob* visits every day a clinic for cardiovascular diseases, discovering that *Bob*, or one of his relatives or close friends, suffers form a heart problem.

5 Countermeasures

We now describe possible countermeasures that can be adopted to avoid or mitigate the security and privacy risks described in the previous section. In the remainder of this chapter, we will refer our examples to the environmental data in Table 1, reporting a possible example of a collection of noise and PM10 values measured in the area of San Francisco. Each row reports the GPS coordinates of the node that performed the measurement, personal information (name, date of birth, and ZIP code) of the owner of the area in which the sensor node is placed, and the noise and PM10 values measured by the node, expressed in dB and µg/m^3, respectively.

	Owner personal data				
Sensor Position	**Name**	**DoB**	**ZIP**	**PM10**	**Noise**
37.739404,-122.483128	Arnold	21/06/1980	94210	60	40
37.748313,-122.583017	Bob	12/06/1980	94211	60	42
37.737222,-122.451906	Carol	07/06/1980	94152	42	60
37.746131,-122.442895	David	26/06/1980	94112	30	51
37.735048,-122.533784	Emma	01/07/1970	95113	50	38
37.744957,-122.534673	Fred	10/07/1970	95141	20	40
37.733864,-122.625562	George	05/07/1970	95217	35	43
37.742772,-122.416451	Hillary	12/07/1970	95235	38	61

Table 1 – An example of a collection of environmental data

5.1 Counteracting security risks

The security risks related to the system architecture can be prevented by the hardening of the physical security of the whole system architecture and by adopting intrusion detection systems [61]. Fault-tolerance solutions can also be helpful when an adversary turns out to be successful and some parts of the system report damages. For instance, a simple solution for ensuring the availability of the data stored in the processing node consists in replicating the data on several machines, possibly located in different sites. The classical attacks on the communication channels can be prevented by encrypting the traffic, though lightweight solutions appear to be suitable for an environmental monitoring scenario, where data measurements are typically performed by sensor nodes with limited computational capabilities [62]. More challenging are the problems of ensuring appropriate protection against non-classical attacks that analyze data access and access patterns (see Section 4.1), and of enforcing access restrictions under the assumption that the set of authorized users can dynamically change and might not be known a priori. In the remainder of this section, we illustrate possible strategies that can be adopted for addressing these two issues.

5.1.1 Protecting environmental data access patterns

The problem of protecting data access and access patterns from external observers and the storage server itself has been mainly studied in the database field [63]. A possible solution to the problem of ensuring that an adversary cannot infer any sensitive information from the observations of accesses to data is to change the physical location (blocks of the hard disk) where data are stored at each access. The technique in [63] goes in this direction, enabling authorized parties to access the stored data while guaranteeing: *i) content* confidentiality (i.e., data privacy is maintained); *ii) access* confidentiality (i.e., the fact that an access aims at a specific data item is protected); and *iii) pattern* confidentiality (i.e., the fact that two different accesses aim at the same data items is protected) from any observer, including the storage server itself. The technique is originally proposed in scenarios of data outsourcing, but it nicely fits a scenario in which a collection of environmental data needs to be stored and maintained private, and each access to certain information is performed by a request issued by a *trusted client*, directly interacting with the storage server.

Adopting this proposal, content, access and pattern confidentiality are guaranteed by organizing data in an ad-hoc data structure, called *shuffle index*. Such a shuffle index assumes data to be organized in an unchained B+-tree, and encrypts data at the node level, so that real (plaintext) values are protected from the (possibly untrusted) storage server. In the B+-tree, data are indexed over a candidate key defined for the data collection, and actual data items are stored in the leaves of the tree according to their index values. Accesses to the data items stored in the tree are based on the value of the associated indexes. Note that, to avoid improper leakages of information to the storage server, the B+-tree does not include any link from a leaf to the next one. The rationale behind this is that such links would expose the order relationship among index values in different nodes.

Data encryption ensures content confidentiality while access and pattern confidentiality are safeguarded by the client by means of: *i)* hiding the real (target) request within cover (fake) requests; *ii)* caching target searches recently performed by users; and *iii)* shuffling, at each request, the content among blocks stored at the server. These three strategies work as follows.

- Cover searches hide a request in a set of fake ones, thus introducing confusion on the requested target. Cover searches are executed in parallel to the target search, and the number of cover searches can be customized to tune the offered protection level.
- Cache avoids the client to search in the B+-tree for the same target in two close queries. The client maintains a local copy of the nodes forming a path in the B+-tree reaching a target value. The size of the cache determines the number of last target searches that are maintained in the cache itself.
- Shuffling implies modifying the data structure at every access, shuffling content among its blocks. The shuffling operation destroys the one-to-one correspondence otherwise existing between a block and the node of the B+-tree stored in it. In this way, repeated accesses to the same node might actually refer to searches for different data items, while different accesses to different nodes might refer to searches for the same data item.

5.1.2 Enforcing access restrictions on environmental data

To prevent unauthorized access to the system, an access control mechanism is needed. A peculiarity of the environmental monitoring scenario is that the set of users authorized to access collected environmental data is typically very dynamic and may not be known a priori. For instance, consider the monitoring of air pollutants in the area of San Francisco. The collected and analyzed data could be accessed for analysis by the local municipality, but also by young researchers of local universities, which may have collaborations with other universities and be therefore part of a dynamic research group. According to this observation, the identity of the users accessing the data may not always be known in advance, and traditional identity-based access control techniques [64] might not be applicable. To overcome this problem, attribute-based access control might represent a viable solution [65]. In this case, rather than considering users' identities, the authorizations stating who can access what data are defined by taking into consideration properties (e.g., age, nationality, occupation) of the authorized parties. For instance, suppose that the local municipality of San Francisco aims at giving access to the collected environmental data only to U.S. citizens. To this aim, the access control policy might grant access to users showing that they hold a U.S. passport, regardless of their identity. Attribute-based access control has been introduced as a means for enforcing this kind of access restrictions in open environments. It is based on the assumption that typically each interacting party (e.g., a client and a server) has a portfolio of *credentials* and *declarations*, either issued and certified by trusted authorities, or self-declared by the party herself [65]. More precisely, a credential includes a list (possibly empty) of certified attributes of the form ⟨`attribute name, attribute value`⟩ representing the subject's attributes (e.g., name and surname contained in an electronic passport), the issuer's public key, the subject's public key, a validity period, and a digital signature. Declarations are pairs of the form ⟨`attribute name, attribute value`⟩ specifying the party's attributes (e.g., the professional status communicated by a user during a registration process) and are produced by the party itself, without any certification from a legal authority. A common assumption underlying attribute-based access control systems is that the set of credentials and declarations that can be released by a party is stored in a profile associated with the party itself.

Attribute-based authorizations involve a *subject*, an *object*, and a set of *actions* to which the authorization refers. A *subject* can be defined as a Boolean formula over declarations and/or credentials. Analogously, an *object* can be defined as a Boolean formula of *predicates* specifying given conditions on the metadata associated with objects. An authorization therefore states that all subjects whit a profile that satisfies the conditions in the *subject* field can perform *actions* on the objects whose metadata satisfy the conditions in the *object* field [65]. An authorization might also contain other elements imposing further conditions on the authorization, such as the purpose of access, or generic conditions that must be satisfied by the access request. For instance, consider the environmental data in Table 1. To read (action) a specific set of PM10 measurements in San Francisco area (object) collected from a certain set of ZIP codes (condition to be satisfied by the object profile), an authorization can require the proof of majority age and a U.S. nationality (conditions to be satisfied by the subject's profile).

When an access request is submitted to the storage server (service provider), it is evaluated with respect to the authorizations applicable to it. An access request is allowed if the conditions for the required access are satisfied; it is denied if none of the specified conditions that might grant the requested access can be fulfilled. However, it may happen that the currently available information is insufficient to determine whether the access request should be granted or denied: in such cases, additional information is needed and the requester receives an undefined response with a list of requests that she must fulfill to gain the access.

5.2 Counteracting privacy risks

To protect environmental data from inferences it is necessary to adopt techniques limiting the analysis that an adversary can perform on them, and obfuscating correlations, associations, and dependencies among them. As previously mentioned (see Section 4.2), these kinds of inferences can arise whenever environmental data are properly or improperly accessed (i.e., when they are stored or outsourced), or when they are made publicly available. In the first case, environmental data privacy can be protected by adopting privacy-enhancing solutions devised for data storage and outsourcing (e.g., encryption and fragmentation). In the latter case, solutions investigated in the context of privacy-preserving data publishing can be adopted. In the remainder of this section, we discuss some of these possible solutions, and briefly overview how location privacy can be ensured in the context of environmental monitoring.

5.2.1 Encrypting stored and outsourced environmental data

Properly storing and maintaining a collection of environmental data that can include, for example, raw data, analysis results, and evidences of correlations among environmental factors is not a trivial problem due to the possible inferences that can arise when accessing such data. Ensuring an appropriate degree of data privacy is of paramount importance, especially when also the storage server is not trusted for accessing the data. Clearly, storing environmental data in *encrypted form* can represent an intuitive solution to guarantee protection against inferences. In fact, an encrypted data collection will be accessible for analysis only to authorized users, that is, those who are provided by the data owner with a decryption key.

Ensuring proper access to encrypted data is however a challenging problem, since different users are typically authorized to access different portions of the stored data. To ensure that all authorized parties can access *all and only* the data for which they have the appropriate authorization, data encryption can be combined with access control, leading to a peculiar kind of encryption usually referred to as *selective encryption* [66,67]. Adopting selective encryption, the keys with which data items are encrypted are regulated by the authorizations holding on the data, and different data items are encrypted with different keys, mapping an *authorization policy* into an equivalent *encryption policy*. As a consequence, an authorization to access a data item translates into the knowledge of the key with which the data item is encrypted (for efficiency reasons, selective encryption is typically assumed to use symmetric encryption). An intuitive solution for enforcing selective encryption consists in encrypting each data item with a different key, and providing each user with a set of keys including all those used to encrypt the data items she can access. Such a naïve solution is however not viable in practice due to the unacceptable key management burden left to users: each user would be required to manage as many keys as the number of data items she is authorized to access. This issue can be conveniently overcome by adopting *key derivation* methods. Basically, a key derivation method allows the computation of an encryption key starting from another key and some public information [66]. Adopting a key derivation technique, each user in the system is provided with a unique key. The set of keys in the system is then built in such a way that, starting from her own key and according to a *key derivation structure*, each user can compute all and only the keys needed for decrypting the resources she can access.

Among the possible key derivation strategies, *token-based* key derivation [66] results particularly appealing for storing or outsourcing (environmental) data. In fact, this solution minimizes the amount of re-encrypting and re-keying required to enforce changes and updates to the authorization policy. Broadly speaking, token-based key derivation works as follows. Given a key k_i in the set of keys of the system, identified by public label l_i, a different key k_j can be derived from k_i and l_j through a so-called *token* $d_{i,j}$, computed as $k_j \oplus h(k_i, l_j)$, where $\oplus$ is the bitwise xor operator and h is a cryptographic function (e.g., a secure hash function). Note that the key derivation can be iteratively applied via a chain of tokens and, since tokens are public pieces of information, all tokens defined in the system are stored in a public catalog. For instance, given three different keys k_i, k_j, and k_h, and two tokens $d_{i,j}$ and $d_{j,h}$, a user who knows (or can derive) key k_i can first use $d_{i,j}$ to derive k_j and, from k_j and $d_{j,h}$, she can then derive k_h. The effect of providing a user with a set $K = \{k_1, \ldots, k_n\}$ of keys is therefore conveniently obtained by providing the user with a single key $k_i \in K$ and publishing a set of tokens allowing the (direct or indirect) derivation of all keys $k_j \in K$, $i \neq j$. In this way, the user can derive all the n encryption keys, while having to worry about a single one.

To implement updates in the authorization policy regulating access to the stored data (i.e., insertion/deletion of a user or data item, and grant/revoke of a permission), a subset of the keys and of the tokens defined in the system must be updated, and some data items must be accordingly re-encrypted. To limit computational burden, the solution in [66] proposes a two-layer encryption strategy, called *over-encryption*. Adopting over-encryption, policy updates can be performed on encrypted resources themselves, without need of decrypting them: in this way, the storage server itself can directly manage policy updates.

5.2.2 Fragmenting stored or outsourced environmental data

When encryption results too heavy or when encrypting the whole data is an overdue, alternative solutions can be adopted. As a matter of fact, if what is sensitive is the data association, instead of specific data values, solutions based on the vertical fragmentation of the data can be adopted. The intuition is very simple: when the joint visibility of some pieces of information is sensitive, such pieces of information are split in different portions not joinable. Fragmentation can be adopted by itself, or coupled with encryption. For instance, suppose that the collected environmental data include information about the concentration of a pollutant in an area, the area, and the owner of the properties within the area. Suppose also that the data holder wants to protect the identities of the owners of polluted properties. Such collection of environmental data can be easily split in two fragments: one fragment includes the concentration of the pollutant and the corresponding area (with the information about the properties' s owner possibly encrypted) and the other fragment includes the information about the owners.

Data fragmentation has been deeply studied in the context of data outsourcing and publication, to vertically fragment the set of attributes composing the schema of a relation to be outsourced or published in such a way to satisfy all confidentiality constraints defined by the data holder. Confidentiality constraints are subsets of attributes composing the original schema. Depending on the number of attributes involved, confidentiality constraints can be classified as: *i) singleton constraints*, stating that the values of the attribute involved in the constraint are sensitive and cannot be released (e.g., the SSN of patients hospitalized for a given respiratory disease due to PM10 exposure are sensitive *per se* and should be kept private); and *ii) association constraints*, stating that the association among the values of the attributes in the constraint is sensitive and cannot be released (e.g., the association between the name and the respiratory illness of a patient can be considered sensitive and should be protected from disclosure). Several fragmentation techniques have been proposed in the literature, and these techniques can be classified based on how they fragment the original relation schema, and whether they adopt encryption.

The first strategy [68] couples fragmentation with encryption, and is based on the assumption that fragments can be stored on two non-communicating servers. When some confidentiality constraints cannot be solved by fragmentation, at least one attribute appearing in such constraints is encrypted. This technique strictly relies on the absence of communications between the servers storing the fragments. However, since collusions among servers can restore the original relation schema compromising the protection of sensitive data, alternative techniques have been proposed for enforcing confidentiality constraints.

The technique in [69] enforces confidentiality constraints coupling fragmentation with encryption, while removing the assumption of absence of communication among the storage servers. This technique satisfies singleton constraints by encrypting the values of the involved attributes. Association constraints are satisfied adopting either fragmentation (i.e., storing the involved attributes in different fragments) or encryption (i.e., encrypting at least one of the involved attributes). However, this technique favors fragmentation over encryption: if a confidentiality constraint can be satisfied via encryption or fragmentation, such a constraint will be enforced with fragmentation. To ensure that no sensitive association can be reconstructed, each attribute must appear in the clear in at most one fragment. This makes the different fragments not joinable and, therefore, all fragments might also be stored on a single storage server. Also, to guarantee the possibility for authorized users to run queries against the data collection, at the physical level each fragment stores all attributes of the original relation schema, either in the clear or encrypted, so that no confidentiality constraint is violated. For instance, consider the environmental data reported in Table 1, and suppose that there are seven confidentiality constraints (c_0,…,c_6) as reported in Figure 3.

- c_0={SensorPosition}
- c_1={Name, DoB}
- c_2={Name, ZIP}
- c_3={Name, PM10}
- c_4={Name, Noise}
- c_5={DoB, ZIP, PM10}
- c_6={DoB, ZIP, Noise}

Figure 3 – An example of confidentiality constraints

Intuitively, these confidentiality constraints state that: *i)* the list of the sensor GPS positions is considered sensitive (c_0); *ii)* the association of the land owners' names with any other information in the relation is considered sensitive (c_1,…,c_4); and *iii)* attributes DoB and ZIP can be exploited to infer the identity of the land owners and, therefore, their associations with the collected noise and PM10 values are considered sensitive (c_5 and c_6). Table 2 represents a possible fragmentation of Table 1 satisfying all the defined confidentiality constraints.

Name	Enc_T
Arnold	Gfg5656d!
Bob	Dfgh45rer
Carol	Fg9324gd
David	Hd72pjc"L
Emma	543rfet4[f
Fred	2q34rxa1q
George	Jkr8478'q
Hillary	0932hjdfk

DoB	ZIP	Enc_T
21/06/1980	94210	Jhfdshjew
12/06/1980	94211	Hde832a8
07/06/1980	94152	Jw92[oq\
26/06/1980	94112	He82n1-x
01/07/1970	95113	Nhw92d3
10/07/1970	95141	9832ie9f
05/07/1970	95217	Hj282nf2
12/07/1970	95235	83jdpvjw

PM10	Noise	Enc_T
60	40	Jr8kds32j-
60	42	Jhu2982nd
42	60	Njef9832m
30	51	Ne983mvs
50	38	][NJ9,PDH
20	40	Jd0wKL34
35	43	/.USHSD8
38	61	[/'jdipw8m

Table 2 – An example of fragmentation (multiple fragments)

Attribute `Enc_T` contains the encrypted version of all attributes appearing in the original relation but not appearing in the clear in the fragment. Note that attribute `SensorPosition` is the only attribute not appearing in the clear in any fragment since it is the only attribute involved in a singleton. Therefore, attribute `Enc_T` of the first fragment on the left-hand-side in Table 2 includes in encrypted form the set `{SensorPosition, DoB, ZIP, PM10, Noise}` of attributes. Similarly, attribute `Enc_T` of the second and third fragment in Table 2 includes the sets `{SensorPosition, Name }` and `{SensorPosition, Name, DoB, ZIP}` of attributes, respectively.

Favoring fragmentation over encryption, the technique in [69] aims at limiting the overhead conveyed by encryption. There are however situations calling from a complete departure from encryption. The technique in [70] avoids the use of encryption, and relies solely on fragmentation for satisfying confidentiality constraints. The assumption is that the data owner is willing to store a limited portion of the data, whenever needed for enforcing confidentiality constraints. In this context, confidentiality constraints are satisfied by storing (at least) one attribute for each constraint at the data owner side. This fragmentation technique builds a pair of fragments, one stored on the data owner, and the other one at the external storage server. Assuming that the storage capacity of the data owner is however limited, each attribute of the original schema should appear in only one fragment to avoid the replication of attributes already stored at the server side also at the data owner side. To illustrate, consider the environmental dataset in Table 1, and the set of confidentiality constraints in Figure 3.

Table 3 reports a possible fragmentation where attributes `SensorPosition`, `Name` and `ZIP` are stored at the data owner side, while attributes `DoB`, `PM10` and `Noise` are stored externally. Note that, differently from the fragmentation in Table 2, no attribute is encrypted (i.e., all attributes belonging to the original schema appear in the clear in exactly one fragment).

T_Id	SensorPosition	Name	ZIP
1	37.739404,-122.483128	Arnold	94210
2	37.748313,-122.583017	Bob	94211
3	37.737222,-122.451906	Carol	94152
4	37.746131,-122.442895	David	94112
5	37.735048,-122.533784	Emma	95113
6	37.744957,-122.534673	Fred	95141

T_Id	DoB	PM10	Noise
1	21/06/1980	60	40
2	12/06/1980	60	42
3	07/06/1980	42	60
4	26/06/1980	30	51
5	01/07/1970	50	38
6	10/07/1970	20	40

7	37.733864,-122.625562	George	95217
8	37.742772,-122.416451	Hillary	95235

7	05/07/1970	35	43
8	12/07/1970	38	61

Table 3 – An example of fragmentation (no encryption, two fragments)

Adopting this technique, the execution of queries involving attributes stored in the two fragments requires that the two fragments must have a common key attribute, so to guarantee a lossless join property (attribute `T_Id` in the fragments in Table 3).

To increase the utility of fragmented data, the fragmentation process can also take into consideration *visibility constraints*, expressing views over data that the fragmentation should satisfy. Visibility constraints permit the expression of different needs of visibility, such as visibility over the *values* of a single attribute, visibility over the *association* among the values of the attributes, or *alternative* visibility over different attributes [71]. Furthermore, fragments can be complemented with a sanitized release of the sensitive associations broken by fragmentation. Such a release takes the form of *loose associations*, defined in a way to guarantee a specified degree of privacy. A loose association reveals some information on the association broken by fragmentation by hiding tuples participating in the associations in groups, and providing information on the associations only at the group level (in contrast to the tuple level) [71, 72].

5.2.3 Protecting published environmental data

When environmental data are publicly released, the possible countermeasures for their protection depend on the format of the data themselves (see Section 3). In the following, we illustrate how it is possible to publish environmental data while ensuring appropriate privacy protection, both in the cases of macrodata and microdata.

Publishing environmental macrodata. If environmental data are published through macrodata tables, they are released as aggregate values and do not contain information specifically related to single individuals or single environmental measurements. However, sensitive information can still be leaked. For instance, consider a macrodata table reporting the concentration of a pollutant during the day and night for each county of a given region. The cells of the macrodata table that contain a high value can be considered sensitive since they indicate that the persons living in the high polluted counties may have a high probability of suffering from specific illnesses. The content of these cells needs therefore to be somehow protected.

A macrodata table can be protected before or after tabulation. In the first case, the objective is to apply some protection techniques to the collected data (e.g., data swapping, sampling, noise addition) so that the computed aggregate values can be considered safe. In the latter case, the protection techniques typically operate in two steps since they first discover sensitive cells, that is, cells that can be easily associated with a specific respondent, and then protect them [59]. We now describe how sensitive cells can be discovered and protected.

- *Detecting sensitive cells.* Sensitive cells can be identified according to different strategies [72]. An intuitive strategy is the so-called *threshold rule*, according to which a cell is sensitive if the number of respondents who contribute to the value stored in the cell is less than a given threshold. The *(n,k)-rule* states that a cell is sensitive if less than *n* respondents contribute to more than *k%* of the total cell value. Other examples of techniques are the *p-percent rule* and the *pq-rule*. According to the *p-percent rule*, a cell is sensitive if the total value of the cell minus the largest reported value v_1 minus the second largest reported value v_2 is less than $(p/100)\cdot v_1$ (the reported value of some respondents can be estimated too accurately). The *pq-rule* is similar to the *p-percent rule*, but takes into consideration the value *q* representing how accurately a respondent can estimate another respondent's sensitive value ($p < q < 100$).
- *Protecting sensitive cells.* Once detected, sensitive cells can be protected by applying several techniques: *cell suppression*, *rounding*, *roll up categories*, *sampling*, *controlled tabular adjustment function* (*CTA*), and *confidential edit* are possible examples of protection techniques. In particular, cell suppression consists in protecting a cell by removing its value (*primary suppression*). However, if some partial (marginal) totals of the table are revealed or publicly known, it might still be possible to re-determine the value of a suppressed cell, or restrict the uncertainty about it. To counteract this risk, additional cells can be suppressed (*secondary suppression*). The rounding technique modifies the original value of a sensitive cell by rounding it up or down to a near multiple of a chosen base number. The roll up categories technique modifies the original macrodata table so that a less detailed (i.e., of smaller size) table is released. Sampling implies that, rather than through a census, the macrodata table is obtained through a sample survey. The CTA technique consists in replacing the value of a sensitive cell with a different value, not considered sensitive with respect to the rule chosen to detect sensitive cells. In a subsequent step, linear programming techniques are used to selectively adjust the values of the non-sensitive cells. The rational behind confidential edit is to compute the macrodata table on a dataset being slightly modified with respect to the original collection. In particular, a sample of the original records are selected and matched (i.e., a set of records with the same values on a specific set of

attributes) in other geographical regions, and the attributes of the matching records are then swapped.

Publishing environmental microdata. Microdata tables contain specific information related to single entities (called respondents). To illustrate, consider the environmental data reported in Table 1, and suppose that the local municipality of San Francisco decides to publicly release the PM10 values in the area. Table 4 illustrates a microdata table that the municipality can prepare from the collected data and can then publicly release. Intuitively, the publication of a microdata table increases the privacy risks and extreme attention has to be devoted for ensuring that no sensitive information is improperly leaked due to the release of such a table. In particular, in our example, the municipality must protect the fact that a given individual lives in an area with a high concentration of PM10 since an adversary may infer that the individuals living in such areas have a high probability of suffering from respiratory diseases.

	Owner personal data			
SensorPosition	**Name**	**DoB**	**ZIP**	**PM10**
37.739404,-122.483128	Arnold	21/06/1980	94210	60
37.748313,-122.583017	Bob	12/06/1980	94211	60
37.737222,-122.451906	Carol	07/06/1980	94152	42
37.746131,-122.442895	David	26/06/1980	94112	30
37.735048,-122.533784	Emma	01/07/1970	95113	50
37.744957,-122.534673	Fred	10/07/1970	95141	20
37.733864,-122.625562	George	05/07/1970	95217	35
37.742772,-122.416451	Hillary	12/07/1970	95235	38

Table 4 – An example of an environmental microdata table

Before publishing an environmental microdata table, all explicit identifiers have to be removed (or encrypted). For instance, Table 5 is a de-identified version of Table 4. In this table, the name of the land owners and the GPS position of the sensing devices (which would univocally identify the associated owner) have been removed by replacing them with value ***.

	Owner personal data			
SensorPosition	**Name**	**DoB**	**ZIP**	**PM10**
***	***	21/06/1980	94210	60
***	***	12/06/1980	94211	60
***	***	07/06/1980	94152	42
***	***	26/06/1980	94112	30
***	***	01/07/1970	95113	50

***	***	10/07/1970	95141	20
***	***	05/07/1970	95217	35
***	***	12/07/1970	95235	38

Table 5 – An example of a de-identified environmental microdata table

Name	DoB	Address	ZIP	City	Job
…	…	…	…	…	…
Arnold Doe	21/06/1980	1201, Main Street	94210	San Francisco	Dentist
…	…	…	…	…	…

Table 6 – An example of a public voter list

However, a de-identified table does not provide any guarantee of anonymity: in fact, besides identifiers, there can exist other attributes such as race, ZIP code, or gender (usually referred to as *quasi-identifiers*) that might be linked to publicly available information to re-identify respondents. For instance, consider the public voter list reported in Table 6 and the de-identified microdata in Table 5 where there is only one land owner born on 21/06/1980 and living in the 94210 area. If this combination is unique in the external world as well, it identifies the first tuple of the microdata in Table 5 as pertaining to Adam Doe, 1201 Main Street, San Francisco 94210, thus revealing that Adam is the owner of an area where the level of PM10 is 60 µg/m^3.

Effective protection of data privacy can be achieved adopting techniques that, for example, generalize the data while preserving data truthfulness: *k*-anonymity is the pioneering technique in this direction [73]. *k*-Anonymity enforces the well-known protection requirement, typically applied by statistical agencies, demanding that any released information should be indistinguishably related to no less than a certain number of respondents. This general requirement is reformulated in the context of *k*-anonymity as follows: *Each release of data must be such that every combination of values of quasi-identifiers can be indistinctly matched to at least k respondents*. Since, typically, each respondent is assumed to be represented by at most one tuple in the released table and vice-versa (i.e. each tuple includes information related to one respondent only), a microdata table satisfies the *k*-anonymity requirement if and only if: i) each tuple in the released table cannot be related to less than *k* individuals in the population; and ii) each individual in the population cannot be related to less than *k* tuples in the table. Taking a safe approach, a microdata table is said to be *k*-anonymous if each combination of values of the quasi-identifier in the table appears with at least *k* occurrences. In this way, each respondent cannot be associated with less than *k* tuples in the table, and each tuple cannot be related to less than *k* respondents in the population, guaranteeing the satisfaction of the *k*-anonymity requirement. To guarantee data truthfulness, *k*-anonymity is typically achieved by applying *generalization* and *suppression* over quasi-identifying attributes. Generalization substitutes the original values with more general values. For instance, the date of birth can be generalized by removing the day, or the day and the month of birth. Suppression consists in removing information from the microdata table. As an example, suppose that the quasi-identifier for Table 5 is composed by attributes DoB and ZIP. Table 7 represents a possible 2-anonymous version of the environmental data in Table 5. The 2-anonymous version has been produced generalizing the date of birth of the land owners (releasing only the month and year) and the ZIP code (releasing only the first three digits of the code). It is easy to see that comparing the 2-anonymous table with the voter list in Table 6, and adversary cannot determine which one between the first two tuples is related to Adam Doe, since both of them share the same combination of attributes DoB and ZIP. More precisely, each combination of values for attributes DoB and ZIP appear in the table with (at least) two different.

	Owner personal data			
SensorPosition	**Name**	**DoB**	**ZIP**	**PM10**
***	***	**/06/1980	942**	60
***	***	**/06/1980	942**	60
***	***	**/06/1980	941**	42
***	***	**/06/1980	941**	30
***	***	**/07/1970	951**	50
***	***	**/07/1970	951**	20
***	***	**/07/1970	952**	35
***	***	**/07/1970	952**	38

Table 7 – An example of a 2-anonymous microdata table

k-Anonymity has been designed for counteracting *identity disclosure*, that is, it represents an effective solution for protecting the identities of the respondents of a microdata table. The original definition of *k*-anonymity has been extended to counteract also the risk that sensitive information is leaked releasing a microdata table (*attribute disclosure*). As an example, ℓ-diversity [74] and *t*-closeness [75] are two well-known extensions of *k*-anonymity, which slightly modify the *k*-anonymity requirement to ensure that neither identities nor sensitive information related to a respondent can be leaked when releasing a microdata table. The basic idea behind these approaches is that of extending the *k*-anonymity requirement considering not only quasi-identifiers, but also sensitive attribute values when computing a privacy-preserving microdata table. To illustrate, consider the 2-anonymous microdata in Table 7. Although an adversary cannot precisely identify the tuple of Adam Doe between the first two in the table, they both share the same value for the PM10 measurement. As a consequence, the adversary is still able to discover that Adam Doe is the owner of a high polluted area. Table 8 illustrates a 3-diverse version of the microdata in Table 5, obtained by generalizing the date of birth to the year of birth, and the ZIP code by releasing only the first two digits. In this case, the tuple of Adam Doe can be one of the first four tuples of the table but since these tuples assume three (hence the 3-diversity) different values for the PM10 concentration, the adversary cannot determine which is the concentration associated with Adam Doe's area.

k-Anonymity, ℓ-diversity and t-closeness represent have recently been modified and/or extended to suit particular releasing scenarios, characterized by particular assumptions, constraints and privacy requirements, such as multiple table releases [76], [77], data republication [78], non-predefined or dynamic quasi-identifiers [79], customizable privacy protection [80].

	Owner personal data			
SensorPosition	**Name**	**DoB**	**ZIP**	**PM10**
***	***	**/**/1980	94***	60
***	***	**/**/1980	94***	60
***	***	**/**/1980	94***	42
***	***	**/**/1980	94***	30
***	***	**/**/1970	95***	50
***	***	**/**/1970	95***	20
***	***	**/**/1970	95***	35
***	***	**/**/1970	95***	38

Table 8 – An example of a 3-diverse microdata table

5.2.4 Protecting privacy of location information in environmental data

The problem of protecting users' positions and movements has recently gained an increasing interest due to the proliferation of both mobile devices equipped with location capabilities and location-based services [81]. This has lead to the definition of different techniques for protecting location information, which can be nicely adapted to the scenario of pervasive environmental monitoring. In the remainder of this section, we survey three different classes of works that can be adopted in this scenario for protecting users' privacy.

The first class of works aims at protecting the privacy of anonymous users communicating with a location-based service provider whenever their real identities are not relevant for the service provision [81]. Goal of these techniques is to avoid the possibility to *re-identify* users observing their position. Since in traditional location-based services users communicate with the service provider posing queries associated with their position, the intuition is that of ensuring that a same location be shared by at least a certain number of different users. These techniques guarantee indistinguishability of users by typically enforcing the requirement of *k*-anonymity [72], specifically tailored to fit the location-based scenario. In our environmental context, instead of issuing queries to a service provider, users communicate some environmental measurements: this translates to the requirement that a same sensed location should be shared by at least a certain number of different sensing users.

The second class of works aims at obfuscating the real position of the users in scenarios in which users are not anonymized and must provide their real identity to the service provider. The idea is that of *degrading the accuracy* of the location measurement. An intuitive strategy might consist in hiding the real position of a user with a set of other *n* fake positions, characterized by the same probability [82]. A different strategy is based on the adoption of some *obfuscation operators*, with the goal of balancing the accuracy of the position and the privacy requirements of the users. For instance, the technique in [83] quantifies privacy with respect to the accuracy of the location measurement, since the more accurate the measurement, the less the privacy. The defined obfuscation operators change the radius, or the center, of the original location measurement, and are used to degrade the accuracy of the location measurement in such a way that, for each user, her privacy preferences are satisfied.

The third class of works focuses on *path privacy*, and aims at releasing a path shared by multiple users so to make them indistinguishable [84]. For instance, these solutions are based on a dynamic grouping of users [85], and protect path privacy enforcing a modified version of *k*-anonymity that requires that all *k* users associated with a specific location remain grouped together as time passes. A different solution is instead based on the release of fake (i.e., simulated) locations [86]. This technique adopts probabilistic models of driving behaviors, applied for creating realistic driving trips, and GPS noise to decrease the precision of the starting point of a trip, and is therefore more suitable for scenarios in which environmental sensing devices are placed on vehicles.

6 Conclusions

In this chapter, we provided an overview of the systems and architectures used for environmental monitoring. We also presented an overview of the main security and privacy issues in environmental monitoring systems, and discussed possible countermeasures for mitigating such issues. Our work can help in better understanding the security and privacy issues that characterized the environmental monitoring systems, and in designing novel environmental systems and applications that guarantee a privacy-aware collection, management, and dissemination of environmental data.

Acknowledgements

This work was supported in part by the Italian Ministry of Research within the PRIN 2008 project "PEPPER" (2008SY2PH4), and by the Università degli Studi di Milano within the "UNIMI per il Futuro – 5 per Mille" project "PREVIOUS".

7 References

[1] F. Amigoni, A. Brandolini, V. Caglioti, V. D. Lecce, A. Guerriero, M. Lazzaroni, F. Lombardi, R. Ottoboni, E. Pasero, V. Piuri, O. Scotti and D. Somenzi, "Agencies for perception in environmental monitoring," *IEEE Transactions on Instrumentation and Measurement,* vol. 55, no. 4, p. 1038–1050, 2006.

[2] M. Dunbabin and L. Marques, "Robots for environmental monitoring: Significant advancements and applications," *Robotics Automation Magazine, IEEE,* vol. 19, no. 1, pp. 24-39, March 2012.

[3] A. Rodic, D. Katie and G. Mester, "Ambient intelligent robot-sensor networks for environmental surveillance and remote sensing," 2009.

[4] M. Tubaishat and S. Madria, "Sensor networks: an overview," *IEEE Potentials,* vol. 22, no. 2, pp. 20-33, 2003.

[5] T. Wong, T. Tsuchiya and T. Kikuno, "A self-organizing technique for sensor placement in wireless micro-sensor networks," 2004.

[6] H. Leung, S. Chandana and S. Wei, "Distributed sensing based on intelligent sensor networks," *Circuits and Systems Magazine, IEEE,* vol. 8, no. 2, pp. 38-52, 2008.

[7] C. Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities and challenges," *Proc. of the IEEE,* vol. 91, no. 8, 2003.

[8] "ZigBee Allianz," [Online]. Available: http://www.zigbee.org.

[9] P. Kulkarni, D. Ganesan, P. Shenoy and Q. Lu, "SensEye: A multitier camera sensor network," in *Proc. of Multimedia 2005*, Singapore, 2005.

[10] D. Aksoy and A. Aksoy, "Satellite-linked sensor networks for planetary scale monitoring," in *Proc. of VTC 2004*, Los Angeles, CA, USA, 2004.

[11] J. K. Hart and K. Martinez, "Environmental sensor networks: A revolution in the earth system science?," *Earth Science Reviews,* vol. 78, no. 3-4, pp. 177-191,

2006.

[12] R. Butler, T. Lay, K. Creager, P. Earl, K. Fischer, J. Gaherty, L. Gabi, B. Leith, J. Park, M. Ritzwoller, J. Tromp and L. Wen, "The Global Seismographic Network surpasses its design goal," *EOS,* vol. 85, no. 23, pp. 225-229, 2004.

[13] "NOAA Center for Tsunami Research, DART (Deep-ocean Assessment and Reporting of Tsunamis)," [Online]. Available: http://nctr.pmel.noaa.gov/Dart.

[14] "Global Seismographic Network," [Online]. Available: http://www.iris.edu/hq/programs/gsn.

[15] "Hawaii Institute of Geophysics and Planetology," [Online]. Available: http://www.higp.hawaii.edu/.

[16] G. Bernhard, C. Booth and J. Ehramjian, "Real-time UV and column ozone from multi-channel UV radiometers deployed in the national science foundation's UV monitoring network," *Ultraviolet Ground - and Space-Based Measurements, Models, and Effects III: Proceedings of SPIE,* vol. 5156, pp. 167-178, 2003.

[17] "NSF Polar Programs UV Monitoring Network," [Online]. Available: http://uv.biospherical.com/.

[18] "GEMSTAT global environment monitoring system," [Online]. Available: http://www.gemstat.org/.

[19] J. Vogelmann, J. Kost, B. Tolk, S. Howard, K. Short, X. Chen, C. Huang, K. Pabst and M. Rollins, "Monitoring landscape change for LANDFIRE using multi-temporal satellite imagery and ancillary data," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* vol. 4, no. 2, p. 252–264, June 2011.

[20] G. Schaefer and R. Paetzold, "SNOTEL (SNOwpack TELemetry) And SCAN (Soil Climate Analysis Network)," *Automated Weather Stations for Applications in Agriculture and Water Resources Management: Current Use and Future Perspectives,* March 2000.

[21] "UK climate change network," [Online]. Available: http://www.ecn.ac.uk.

[22] J. Kimball, L. Jones, K. Zhang, F. Heinsch, K. McDonald and W. Oechel, "A satellite approach to estimate land CO2 atmosphere exchange for boreal and arctic biomes using MODIS and AMSR-E," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 47, no. 2, pp. 569-587, February 2009.

[23] A. Lane, "The UK environmental change network database: An integrated information resource for long-term monitoring and research," *Journal of Environmental Management,* vol. 51, no. 1, pp. 87-105, 1997.

[24] S. Ngheim and P. Clemete-Colon, "Arctic sea ice mapping with satellite radars," *Aerospace and Electronic Systems Magazine, IEEE,* vol. 24, no. 11, pp. 41-44, November 2009.

[25] J. Qu, X. Hao, M. Kafatos and L. Wang, "Asian dust storm monitoring combining terra and aqua MODIS SRB measurements," *Geoscience and Remote Sensing Letters, IEEE,* vol. 3, no. 4, pp. 484-486, October 2006.

[26] M. Shimada, T. Tadono and A. Rosenqvist, "Advanced land observing satellite (ALOS) and monitoring global environmental change," *Proceedings of the IEEE,* vol. 98, no. 5, pp. 780-799, May 2010.

[27] A. Rosenqvist, M. Shimada, N. Ito and M.Watanabe, "ALOS PALSAR: A pathfinder mission for global-scale monitoring of the environment," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 45, no. 11, pp. 3307-3316, November 2007.

[28] "National Oceanic and Atmospheric Administration NOA, United States Department of Commerce, Global Earth Observation System," [Online]. Available: http://www.noaa.gov/eos.html.

[29] "United states environmental protection agency, national environmental monitoring initiative," [Online]. Available: http://www.epa.gov/cludygxb/html/choices.htm.

[30] "King county natural resources and parks," [Online]. Available: http://www.kingcounty.gov/environment/dnrp.aspx.

[31] M. Carotta, G. Martinelli, L. Crema, C. Malagu, M. Merli, G. Ghiotti and E. Traversa, "Nanostructured thick-film gas sensors for atmospheric pollutant

monitoring: quantitative analysis on field tests," *Sensors and Actuators B,* vol. 76, pp. 336-342, 2001.

[32] G. Andria, G. Cavone, V. D. Lecce and A. Lanzolla, "Model characterization in measurements of environmental pollutants via data correlation of sensor outputs," *Instrumentation and Measurement, IEEE Transactions on,* vol. 54, no. 3, pp. 1061-1066, June 2005.

[33] "WAVCIS Wave-Current-Surge Information System for Coastal Louisiana," [Online]. Available: http://www.wavcis.lsu.edu.

[34] G. Hoogenboom, "The Georgia automated environmental monitoring network," *Southeastern Climate Review,* vol. 4, no. 0, pp. 12-18, 1993.

[35] "Chesapeake Bay Observatory System," [Online]. Available: http://www.cbos.org.

[36] G. Stone, X. Zhang, J. Li and A. Sheremet, "Coastal observing systems: key to the future of coastal dynamics investigations," *GCAGS/GCSSEPM Transactions,* vol. 53, pp. 783-799, 2003.

[37] A. Genovese, R. Donida Labati, V. Piuri and F. Scotti, "Wildfire smoke detection using computational intelligence techniques," in *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA 2011)*, Ottawa, Canada, 2011.

[38] A. Genovese, R. Donida Labati, V. Piuri and F. Scotti, "Virtual environment for synthetic smoke clouds generation," in *IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measure ment Systems (VECIMS 2011)*, Ottawa, Canada.

[39] Z. Liu and A. Kim, "Review of recent developments in fire detection technologies," *Journal of Fire Protection Engineering,* vol. 13, no. 2, pp. 129-149, May 2003.

[40] Q. Li, Q. Hao and K. Zhang, "Smart wireless video sensor network for fire alarm," in *Proc. of WiCOM 2010*, Chengdu, China, 2010.

[41] B. Son, Y.-s. Her and J.-G. Kim, "A design and implementation of forest-fires

surveillance system based on wireless sensor networks for south korea mountains," *IJCSNS International Journal of Computer Science and Network Security,* vol. 6, no. 9B, September 2006.

[42] J. Tschmelak, G. Proll, J. Riedt, J. Kaiser, P. Kraemmer, L. Bárzaga, J. Wilkinson, P. Hua, J. P. Hole, R. Nudd, M. Jackson, R. Abuknesha, D. Barceló, S. Rodriguez-Mozaz, M. L. d. Alda, F. Sacher, J. Stien, J. Slobodník, P. Oswald, H. Kozmenko, E. Korenková, Z. Krascsenits, G. Gauglitz and L. Tóthová, "Automated Water Analyser Computer Supported System (AWACSS) Part I: project objectives, basic technology immunoassay development, software design and networking," *Biosensors and Bioelectronics,* vol. 20, no. 8, pp. 1499-1508, 2005.

[43] T. Bendikov, J. Kim and T. Harmon, "Development and environmental applications of a nitrate selective microsensor based on doped polypyrrole films," in *204th Meeting of the Electrochemical Society*, 2003.

[44] C. Alippi, R. Camplani, C. Galperti and M. Roveri, "A robust, adaptive solar-powered WSN framework for aquatic environmental monitoring," *Sensors Journal, IEEE,* vol. 11, no. 1, pp. 45-55, January 2011.

[45] K. Martinez, J. Hart and R. Ong, "Environmental sensor networks," *Computer,* vol. 37, no. 8, pp. 50-56, 2004.

[46] G. Acar and A. Adams, "ACMENet: an underwater acoustic sensor network protocol for real-time environmental monitoring in coastal areas," *Radar, Sonar and Navigation, IEE Proceedings,* vol. 153, no. 4, pp. 365-380, August 2006.

[47] K. Persaud, "smart gas sensor for monitoring environmental changes in closed systems: Results from the MIR space station," *Sensors and Actuators B,* Vols. 2-3, no. 55, pp. 118-126, 1999.

[48] A. Kumar, I. Singh and S. Sud, "Energy efficient and low-cost indoor environment monitoring system based on the IEEE 1451 standard," *Sensors Journal, IEEE,* vol. 11, no. 10, p. 2598–2610, October 2011.

[49] J. Guevara, F. Barrero, E. Vargas, J. Becerra and S. Toral, "Environmental wireless sensor network for road traffic applications," *Intelligent Transport*

Systems, IET, vol. 6, no. 2, pp. 177-186, June 2012.

[50] S. Santini and A. Vitaletti, "Wireless sensor networks for environmental noise monitoring," *GI/ITG KuVS Fachgespraech Drahtlose Sensornetze,* pp. 98-101, July 2007.

[51] L. Ioriatti, M. Martinelli, F. Viani, M. Benedetti and A. Massa, "Realtime distributed monitoring of electromagnetic pollution in urban environments," in *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, 2009.

[52] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees and M. Welsh, "Monitoring volcanic eruptions with a wireless sensor network," in *Proc. of EWSN 2005*, Istanbul, Turkey, 2005.

[53] M. V. Ramesh, "Real-time wireless sensor network for landslide detection," in *Proceedings of the 2009 Third International Conference on Sensor Technologies and Applications*, Washington, DC, USA, 2009.

[54] L. Buttyan, D. Gessner, A. Hessler and P. Langendoerfer, "Application of wireless sensor networks in critical infrastructure protection: challenges and design options," *Wireless Communications, IEEE,* vol. 17, no. 5, pp. 44-49, October 2010.

[55] T. Harms, S. Sedigh and F. Bastianini, "Structural health monitoring of bridges using wireless sensor networks," *Instrumentation Measurement Magazine, IEEE,* vol. 13, no. 6, pp. 14-18, December 2010.

[56] "ORION Project," [Online]. Available: http://orion.lookingtosea.ucsd.edu/.

[57] E. Bradley, M. Toomey, C. Still and D. Roberts, "Multi-scale sensor fusion with an online application: Integrating GOES, MODIS, and webcam imagery for environmental monitoring," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* vol. 3, no. 4, pp. 497-506, December 2010.

[58] P. Ferraro, M. Bauersachs, J. Burns and G. Bataller, "A system for the measurement of the Amazon," *Aerospace and Electronic Systems Magazine, IEEE,* vol. 22, no. 8, pp. 9-19, August 2007.

[59] V. Ciriani, S. De Capitani di Vimercati, S. Foresti and P. Samarati, "Microdata Protection," in *Secure Data Management in Decentralized Systems*, T. Jajodia and S. Yu, Eds., Springer-Verlag, 2007.

[60] S. De Capitani di Vimercati, G. Livraga, V. Piuri and F. Scotti, "Privacy and Security in Environmental Monitoring Systems, " in *Proc. of ESTEL 2012*, Rome, Italy, 2012.

[61] W. Stallings, Network Security Essentials: Applications and Standards, 4th ed., Upper Saddle River, NJ: Prentice Hall Press, 2010.

[62] C. Castelluccia, A. C.-F. Chan, E. Mykletun and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks," *ACM TOSN,* vol. vol. 5, no. 3, pp. 1-36, 2009.

[63] S. De Capitani di Vimercati, S. Foresti, S. Paraboschi, G. Pelosi and P. Samarati, "Efficient and Private Access to Outsourced Data," in *Proc. of ICDCS 2011*, Minneapolis, MN, USA, 2011.

[64] S. De Capitani di Vimercati and P. Samarati, "Access Control in Federated Systems," in *Proc. of NSPW*, Lake Arrowhead, CA, USA, 1996.

[65] C. Ardagna, M. Cremonini, S. De Capitani di Vimercati and P. Samarati, "A Privacy-Aware Access Control System," *Jounal of Computer Security,* vol. 16, no. 4, 2008.

[66] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi and P. Samarati, "Encryption Policies for Regulating Access to Outsourced Data," ACM TODS, vol. vol. 35, no. 2, pp. 1-46, 2010.

[67] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi and P. Samarati, "A Data Outsourcing Architecture Combining Cryptography and Access Control," in *Proc. of CSAW 2007*, Fairfax, VA, USA, 2007.

[68] G. Aggarwal, M. Bawa, P. Ganesan, H. Garcia-molina, K. Kenthapadi, R. Motwani, U. Srivastava, D. Thomas and Y. Xu, "Two can keep a secret: A distributed architecture for secure database services," in *Proc. of CIDR 2005*, Asilomar, CA, USA, 2005.

[69] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi and P. Samarati, "Combining Fragmentation and Encryption to Protect Privacy in Data Storage," *ACM TISSEC,* vol. vol. 13, no. 9, 2010.

[70] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi and P. Samarat, "Keep a Few: Outsourcing Data while Maintaining Confidentiality," in *Proc. of ESORICS 2009*, Saint-Malo, France, 2009.

[71] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi and P. Samarati, "Fragments and Loose Associations: Respecting Privacy in Data Publishing," *PVLDB,* vol. vol. 3, no. 1, pp. 1370-1381, 2010.

[72] Federal Committee on Statistical Methodology, Statistical Policy Working Paper 22, 2nd ed., Washington, DC, 2005.

[73] P. Samarati, "Protecting Respondents' Identities in Microdata Release," *IEEE TKDE,* vol. vol. 13, no. 6, pp. 1010-1027, 2001.

[74] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM TKDD,* vol. vol. 1, no. 1, pp. 3-52, 2007.

[75] N. Li, T. Li and S. Venkatasubramanian, "t-Closeness: Privacy beyond k-anonymity and l-diversity.," in *Proc. of ICDE 2007*, Istanbul, Turkey, 2007.

[76] K. Wang and B. Fung, "Anonymizing sequential releases," in *Proc. of KDD 2006*, Philadelphia, PA, USA , 2006.

[77] M. Nergiz, C. Clifton and A. Nergiz, "Multirelational k-anonymity.," in *Proc. of ICDE 2007*, Istanbul, Turkey, 2007.

[78] X. Xiao and Y. Tao, "m-invariance: Towards privacy preserving re-publication of dynamic datasets," in *Proc. of SIGMOD 2007*, Beijing, China, 2007.

[79] M. Terrovitis, N. Mamoulis and P. Kalnis, "Privacy-preserving anonymization of set-valued data.," *PVLDB,* vol. vol. 1, no. 1, pp. 115-125, 2008.

[80] Frikken and Y. Zhang, "Yet another privacy metric for publishing micro-data," in *Proc. of WPES 2008*, Alexandria, VA, USA, 2008.

[81] C. Bettini, S. Jajodia, P. Samarati and X. S. Wang, Eds., vol. LNCS 5599, Springer, 2009.

[82] M. Duckham and L. Kulik, *A formal model of obfuscation and negotiation for location privacy,* Munich, Germany, 2005.

[83] C. Ardagna, M. Cremonini, S. De Capitani di Vimercati and P. Samarati, "An Obfuscation-based Approach for Protecting Location Privacy," *IEEE TDSC,* vol. vol. 8, no. 1, pp. 13-27, 2011.

[84] C.-Y. Chow and M. F. Mokbel, "Trajectory privacy in location-based services and data publication," *SIGKDD Explorations Newsletter,* vol. vol. 13, no. 11, pp. 19-29, 2011.

[85] C.-Y. a. M. M. Chow, "Enabling private continuous queries for revealed user locations," in *Proc. of SSTD 2007*, Boston, MA, USA, 2007.

[86] J. Krumm, "Realistic Driving Trips For Location Privacy," in *Proc. of Pervasive 2009*, Nara, Japan, 2009.