Networks of Sensors

Onderwater, M.

2016

1

Introduction

Over the past few years, the use of sensors has been growing at an unprecedented rate, and sensors play an increasingly large role in our daily lives. Smartphones, for example, contain a wide variety of sensors. The documentation of Android's API [10] reveals support for sensors measuring acceleration, temperature, relative humidity, air pressure, degrees of rotation, the geomagnetic field, gravity, illumination, and proximity. Apps use these sensors to, e.g., detect shaking (shuffle songs), determine rotation of the screen (prevent recording vertical video), measure a user's level of activity (in a fitness app), and recognize proximity to an ear (switch off the touchscreen during a call).

Another interesting application of sensors is in modern washing machines, in which sensors monitor several aspects of the washing cycle, including water hardness, the amount of dirt and grease, the type of detergent used (liquid/powder), and the weight of the laundry. These observations are then used to control the process, allowing the machine to, e.g., add clean water, change the direction and speed of the spin, and increase the water's temperature. By controlling these parameters, the washing machine can optimize the length of washing cycle, minimize energy usage, and prevent damage to components.

The increasing popularity and relevance of sensors is also evident from Gartner's annual 'Emerging Technologies Hype Cycle' [12]. This graph, shown in Figure 1.1, reflects the level of maturity of a technology during its lifetime (horizontal axis), compared to expectations attached to it by society (vertical axis). The left side of the graph contains innovative, promising technologies that are still fairly unknown and have low expectations. The next stage is character-

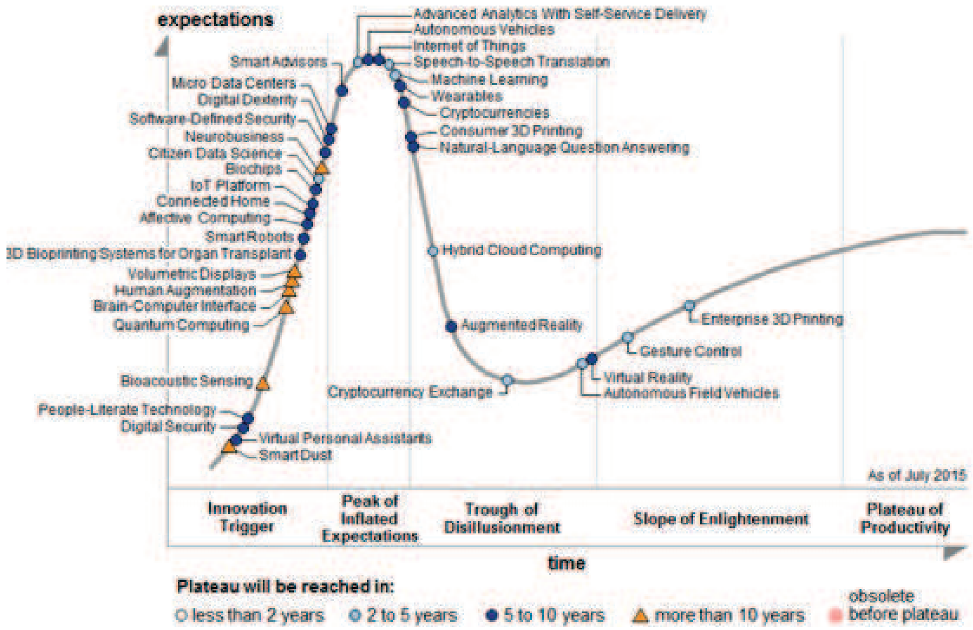

FIGURE 1.1: Gartner's 2015 Hype Cycle of Emerging Technologies [12].

ized by a massive peak of inflated expectations, while the technology is still relatively immature. Typically, this is where a technology reaches the status of 'hype'. Inevitably, this leads to some disillusionment when people realize that expectations about the technology are overrated. Eventually, increased understanding, knowledge, and experience drive the technology towards maturity at a more moderate pace.

Several of the technologies presented in the hype cycle are closely related to sensors. At the top of the hype cycle is 'The Internet of Things' (IoT), a concept referring to a network of physical objects that allow interaction between the physical world and computer-based systems. 'Things' can be, e.g., coffee machines, curtains, smart energy meters, security systems, or watches. Sensors play a central role in IoT, because they are small, energy efficient, have (wireless) connectivity, and still carry sufficient computational resources for basic applications. This makes them ideal to connect 'Things' to 'The Internet'. Moreover, recent technological developments have resulted in cheap and reliable sensors, making them cost-efficient for use in applications, and a driving factor for IoT. Furthermore, note that the term 'IoT Platform' is on the hype cycle as well. This platform facilitates collecting, analyzing, and integration of data from 'Things'. In Chapter 2 we discuss this technology in more detail in a sensor-specific context.

Slightly further along in maturity is 'Wearables', of which the smart watch is probably the most well-known example. The typical smart watch contains several sensors, such as an accelerometer, a thermometer, an altimeter, a barometer, a compass, GPS, and a heart rate sensor. Close integration with smart phones allows wearers to, e.g., play music, get social media notifications, and track physical activity. The Wearables technology is about to enter the third phase on the hype cycle (the 'through of disillusionment'), so it will be interesting to see how it develops in the near future. Other technologies on the hype cycle related to sensor technology are 'Smart Dust' (millimeter-sized devices with embedded sensors), 'Bioaccoustic sensing' (enables the use of, e.g., your arm as a touchpad), and 'Connected Home' (home automation).

The data-driven technologies '[Citizen] Data Science', 'Advanced Analytics [With Self-Service Delivery]'[1], and 'Machine Learning' on the hype cycle are relevant for sensor data as well. Sensors can produce massive amounts of data, because of the potential high frequency of measurements, and the large number of relevant phenomena to monitor. For instance, a typical Boeing aircraft contains 8,000-10,000 sensors that are measuring and reporting every second [39]. Extrapolate this to 5,000 aircraft demonstrates that Boeing receives over fourteen terabyte of sensor data each day. Dealing with such large amounts of data and transforming it to actionable insights is an immense challenge.

1.1 Networks of sensors

The sensors in a smartphone are capable of measuring a large variety of phenomena, both concerning the device itself (e.g., detecting the rotation of the screen), as well as about the immediate environment (e.g., the ambient temperature). Measuring the temperature in, e.g., an office building with a smartphone is, however, unpractical because it requires the phone's owner to continually walk around the building recording the current temperature. Fortunately, with modern day technology it is possible to make small 'mini-computers' to which sensors can be attached, yielding devices smaller than a credit card capable of monitoring of and interaction with the environment. In this thesis we refer to these devices as *sensor nodes*. Typically, besides sensors and basic processing capabilities, a node is also equipped with a small radio that allows it to

[1] Citizen Data Science refers to the fact that improved data science tools allow the average Joe to apply data science, eliminating the need for highly trained data scientists. Advanced Analytics With Self-Service Delivery means that modern, simple to use Business Intelligence tools allow people to analyze data without extensive need of IT support.

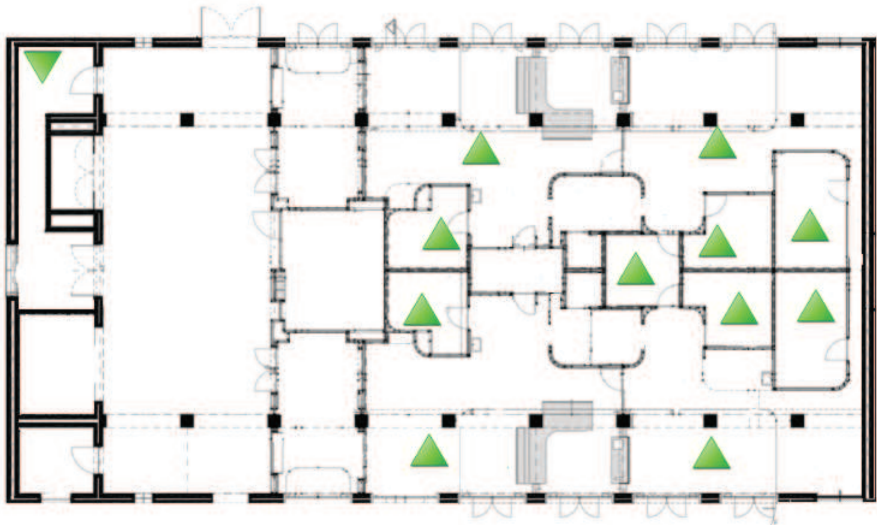

FIGURE 1.2: Floor-plan of a kindergarten, with sensor nodes (triangles) measuring the indoor climate and reporting their measurements to a central node (inverted triangle).

communicate wirelessly. Hence, sensor nodes can report their measurements to interested applications, and thus facilitate remote monitoring and control. Obviously, this is of tremendous practical value when the sensor nodes are spread over a large geographical area. However, the range of the radio of a sensor node is usually small in order to save energy, so communicating across large distances is not possible. As a solution, protocols have been developed that allow the sensor nodes to jointly form a network, so that data can be transmitted over long distances in several smaller steps. These networks are called *wireless sensor networks* (WSNs), and are the topic of the first part of this thesis.

An example deployment of a WSN is in Figure 1.2, showing the floor-plan of a kindergarten. This kindergarten is located in a city in The Netherlands, where directives are in place with respect to the indoor climate. For a kindergarten, a CO_2 level larger than 1,000 parts per million (ppm) is used as an indication of insufficient ventilation. High levels of CO_2 are associated with fatigue, headaches, and reduced concentration [11, 43]. The sensor nodes, marked by triangles, measure CO_2, temperature, humidity, and illumination. The measurements are transmitted wirelessly to a so-called *sink node* (inverted triangle), where the data is collected for further processing. After processing, the administrators of the kindergarten can monitor the current CO_2 level using graphs similar to the one displayed in Figure 1.3. It shows the CO_2 levels

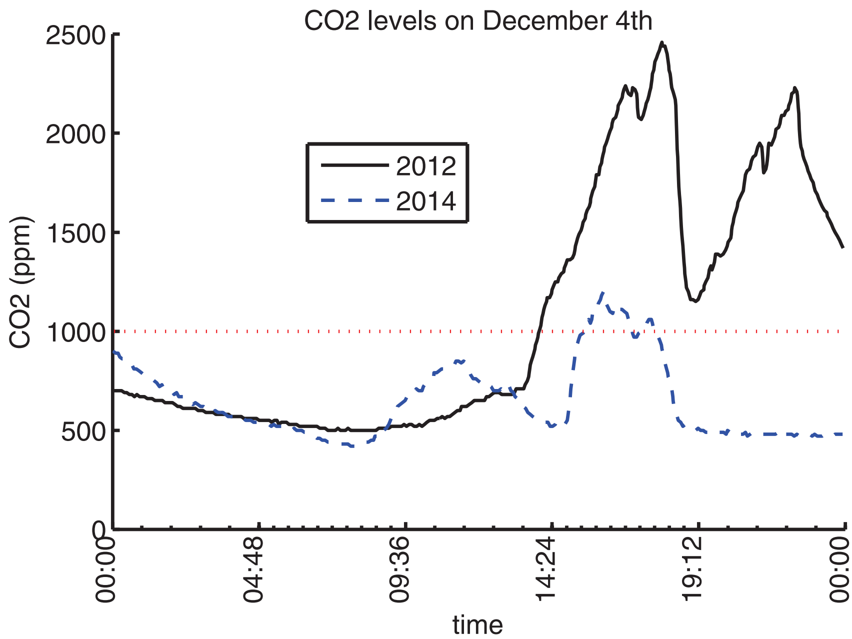

FIGURE 1.3: CO_2 levels in the kindergarten exceeding the recommended level of 1,000 particles per million on December 4th 2012, and a much healthier development on the same day in 2014.

on December 4th 2012 (solid line), on December 4th 2014 (dashed), and the maximum threshold value of 1,000 ppm (dotted). The 2012 line demonstrates that CO_2 levels can exceed the 1,000 ppm threshold, and the 2014 line shows much healthier levels.

1.2 Challenges

Driven by emerging technologies, applications relying on sensor data will become more prevalent in the near future. However, working with sensors involves some distinct challenges that have to be dealt with. For instance, sensor measurements are often 'noisy' and can include a significant level of uncertainty. In the kindergarten in Figure 1.2 several sensor nodes were installed upside down (as compared to the other nodes), because this was more convenient for the power cable. Afterwards we noticed that the temperature measurements at these nodes differed by approximately 1.5° Celsius from the other nodes. The manufacturer of the nodes then confirmed that heat from the sensor node can warm up the temperature sensor if a node is installed upside down. This illustrates how a seemingly small mistake can affect sensor measurements, and how it can contribute to the uncertainty associated with these measurements. Having a network of sensor nodes raises additional challenges, especially when the network is wireless (which is common in practice). The transmission of a packet from one node to another node can easily fail because of, e.g., closed doors, people walking by, or interference on the wireless channel. In this section we discuss the most important challenges in the context of sensors, sensor nodes, and sensor networks. The challenges below are adapted from [18, 110, 132], and appear in no particular order.

- *Limited resources.* Sensor nodes typically have only few resources available, i.e., little storage capacity, a slow CPU, and a battery with limited power. Creating applications on such a device often results in a careful balancing act between satisfying application requirements and managing resources. For instance, measuring with a sensor and transmitting the resulting measurement are two of the most power-consuming operations a node can do. To maximize the life-time of the battery, the frequency of these operations should be minimized, which can be problematic for applications relying on frequent measurements.

- *Dynamic topology.* The topology of a sensor network can be highly dynamic, because sensor nodes can be added, removed, moved, run out of power, or break down. In certain applications, the sensor node might even be mobile instead of stationary. Additionally, interference on the wireless channel might cause a sensor node to be temporarily unavailable.

- *Data redundancy.* The data produced by sensor networks usually contains a large amount of redundancy. For instance, the rooms in the kindergarten in Figure 1.2 are all in the same building and on the same floor, and the temperature measurements by the sensor nodes are expected to be similar. Redundancy can also occur between different types of measurements: in a closed room full of people, both temperature and CO_2 levels are expected to increase, and thus the two types of measurements are correlated. Applications relying on large amounts of sensor data might need to reduce redundancy to remain computationally feasible. At the same time, redundancy in data can also be useful. When a certain level of redundancy is expected but, upon arrival of the measurements, is not observed, then this absence of redundancy might suggest that something happened or that there is a problem.

- *Reliability.* Measurements by sensors can contain random noise, so that it is unclear how accurate a measurement is. This is particularly challenging when monitoring an environment for abnormal events, since an inaccurate measurement and an abnormal measurement can be difficult to distinguish. Also, the wireless channel and the network are susceptible to reliability issues. The channel might be slower than expected, completely unavailable in part of the network, and the topology can change. These issues cause delay or even complete disappearance of measurements, making it challenging for the network to operate reliably.

- *Scalability.* Sensors are relatively cheap and can be used cost-effectively in large quantities. Consequently, sensor networks can potentially be quite large, and the network infrastructure should be scalable to such large sizes.

- *Heterogeneous protocols and data formats.* Sensor nodes can use a wide variety of protocols, access mechanisms, and data formats. This heterogeneity hampers development of applications, since acquiring sensor data requires a significant amount of sensor-specific work.

1.3 Motivation and structure

Sensors and sensor networks are here to stay. Even though IoT is now at the peak of inflated expectations in the hype cycle (see Figure 1.1), sensor-related technology is clearly useful in many applications. The nearing period of improved maturity of IoT suggested by the hype cycle will make the challenges from the previous section increasingly relevant. This raises the need for a deeper understanding of the challenges, for practical methods to deal with them, and for innovative solutions. This is the main motivation for the research in this thesis.

In the following chapters we consider several topics related to the challenges outlined in the previous section. The thesis consists of seven chapters (including this introduction), and is divided in two parts. The first part is about sensor technology, and discusses three different topics in that context. The relation between the three topics is illustrated in Figure 1.4, and explained in more detail in the paragraphs below. In the second part we deal with Markov Decision Processes, a popular framework for taking decisions under uncertainty. The techniques in that part are described in general terms, because the techniques can also be applied in contexts other than sensor networks. The various chap-

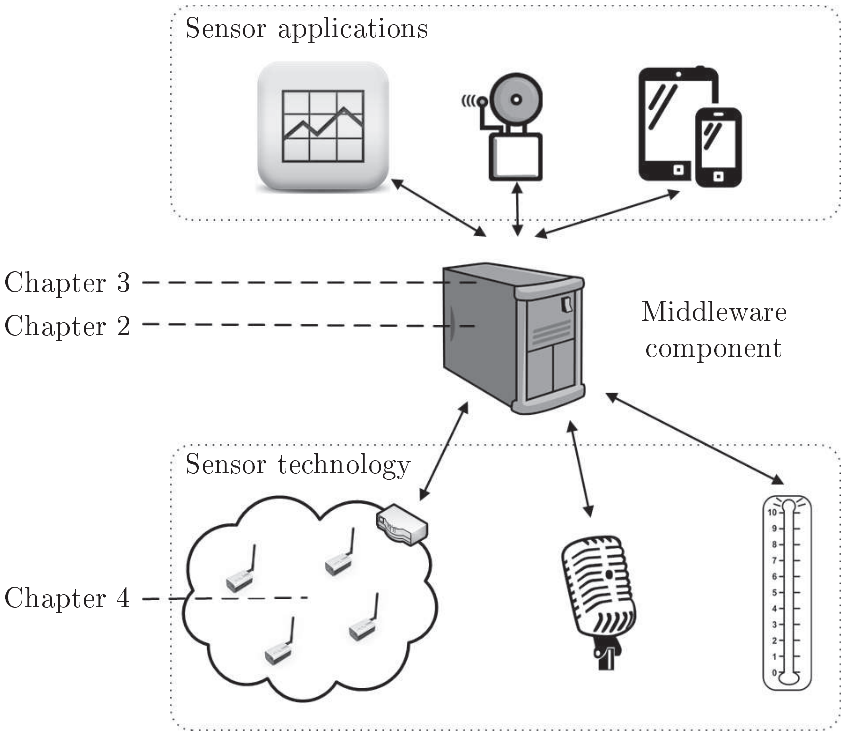

FIGURE 1.4: A middleware component forms a bridge between sensor technology and sensor applications. The figure also illustrates how the three sensor-related chapters in Part I relate: Chapter 2 is about middleware components, Chapter 3 discusses dimensionality reduction techniques (a potential service offered by a middleware component for applications), and Chapter 4 deals with throughput of a sensor network (the left-most example of a sensor technology).

ters in this thesis can be read independently of each other and in arbitrary order, with the exception of Chapter 7, which extends Chapter 6.

We start Part I with a discussion about middleware components for sensor networks in Chapter 2. These components are closely related to the 'IoT Platform' mentioned in the hype cycle in Figure 1.1. The middleware component forms a natural bridge between sensor technology and applications using sensor data. Figure 1.4 illustrates this scenario, with a sensor network and two standalone sensors (for sound and temperature) at the bottom of the figure providing measurements to the middleware component. The top of the figure shows three applications relying on sensor data: a chart, an alarm application, and a mobile app. The middleware component is in the center, and decouples the sensor applications from the sensor technology. In the scientific literature, a wide variety of such middleware components exists and in Chapter 2 we review these components with an often-used categorization. Then we describe that, recently, a new category of middleware components has emerged and we introduce the name 'centralized' for this category. We describe the general architectural form of a centralized middleware component, review four well-known components in the new category, and discuss their relevance for use in sensor networks.

Next, in Chapter 3 we consider *dimensionality reduction techniques*, which aim at removing redundancy from data by finding a short insightful summary. These techniques can be applied to sensor data as well, allowing applications to work with only a small part of the sensor data instead of the full range of measurements. Dimensionality reduction can, e.g., be a service provided by a centralized middleware platform (illustrated in Figure 1.4). The summary resulting from dimensionality reduction is designed to retain the most important part of the information from the original data, but inevitably some information is lost. For certain applications this loss might be unacceptable. For instance, an alarm application watching for abnormal sensor measurements ('outliers') can only use the summarized data if outliers among the sensor data are mapped to outliers among the summarized data by the reduction technique. If the reduction technique does not preserve outliers, the alarm application will miss measurements worthy of an alarm when using the summarized data. In Chapter 3 we discuss three popular dimensionality reduction techniques, and experimentally determine how well they preserve outliers. The experiments identify one of the techniques as best able to preserve outliers, and we discuss the intuitions behind this result.

In Chapter 4 we consider the *saturation throughput*, an important performance indicator of a sensor network (the left-most sensor technology in Figure 1.4). This property reflects at what speed the network is able to process measurements by sensors when a large number of these measurements is offered. We develop a model for analyzing the saturation throughput of the IEEE 802.15.4 MAC protocol, which is the de-facto standard for WSNs and ensures fair access to the channel. To this end, we introduce the concept of a *natural layer*, which reflects the time that a sensor node typically has to wait (as prescribed by the IEEE 802.15.4 MAC protocol) prior to sending a packet. The model is simple and provides insight in how the throughput depends on the protocol parameters and the number of nodes in the network. Validation experiments with simulations demonstrate that the model is highly accurate for a wide range of parameter settings of the MAC protocol, and for both large and small networks.

The first chapter in Part II, Chapter 5, deals with the control of a queueing system with aging state information. The controller of the system has to provide incoming queries with a response. The response can either be a fresh value obtained from a queueing system, or an older value that was cached by the controller. Both choices are imperfect: the first causes a delay because it takes time to generate a fresh response, and the second returns an aged value that is potentially too old for use. Hence, the controller faces a trade-off between response times and data freshness. In practice, a threshold policy is often used to take decisions, where a fresh value is generated when the age of the cached response exceeds a given threshold. Unfortunately, this policy is not always optimal, particularly when the queueing system is heavily loaded, and requesting a fresh response is expensive. In Chapter 5 we demonstrate how such a threshold policy can be improved by taking the load of the system into account. We model the system as a Markov Decision Process, which turns out to be complex. We simplify the model to circumvent these complexities, and then construct a control policy. This policy is demonstrated to have near-optimal performance and achieves lower costs than the threshold policy.

Chapter 6 introduces a novel method for discovery of relative value functions for Markov Decision Processes. This method, which we call Value Function Discovery (VFD), is based on ideas from the Evolutionary Algorithm field. VFD's key feature is that it discovers descriptions of relative value functions that are algebraic in nature. In particular, the descriptions include the model parameters of the MDP. The algebraic expression of the relative value function discovered by VFD can be used in several scenarios, e.g., conversion to a policy (with one-step policy improvement) or control of systems with time-varying parameters. In Chapter 6, we give a detailed description of VFD and illustrate its application on an example MDP. For this MDP we let VFD discover an algebraic description of a relative value function that closely resembles the relative value function corresponding to the optimal policy. The discovered relative value function is then used to obtain a policy, which we compare numerically to the optimal policy of the MDP. The resulting policy has excellent performance on a wide range of model parameters.

We continue work on VFD in Chapter 7, where we demonstrate how additional information about the structure of an MDP can be included in VFD. For this we use the same MDP as in Chapter 6, and include prior knowledge that the optimal policy is of threshold type. We let VFD learn an expression for this threshold in terms of the model parameters, and numerically inspect its performance. We demonstrate that this alternative use of VFD also yields near-optimal policies, illustrating that VFD is not restricted to learning relative value functions and can be applied more generally.

1.4 Publications

This thesis is based on the following publications:

- M. Onderwater. An overview of centralised middleware components for sensor networks. To appear in *International Journal of Ad Hoc and Ubiquitous Computing*, 2015.
- M. Onderwater. Outlier preservation by dimensionality reduction techniques. *International Journal of Data Analysis Techniques and Strategies*, 7(3):231–252, 2015.
- M. Onderwater, G. J. Hoekstra, and R. D. van der Mei. Throughput modeling of the IEEE MAC for sensor networks. *Under review*, 2015.
- M. Onderwater, S. Bhulai, and R. D. van der Mei. On the control of a queueing system with aging state information. *Stochastic Models*, 31(4): 588–617, 2015.
- M. Onderwater, S. Bhulai, and R. D. van der Mei. Value Function Discovery in Markov Decision Processes with Evolutionary Algorithms. To appear in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2015.
- M. Onderwater, S. Bhulai, and R. D. van der Mei. Learning optimal policies in Markov Decision Processes with Value Function Discovery. *Performance Evaluation Review*, 43(2):7–9, 2015.
- M. Onderwater, S. Bhulai, and R. D. van der Mei. Discovery of structured optimal policies in Markov Decision Processes. *Under review*, 2015.
- M. Mitici, M. Onderwater, M. de Graaf, J. van Ommeren, N. van Dijk, J. Goseling, and R. J. Boucherie. Optimal query assignment for wireless sensor networks. *International Journal of Electronics and Communications*, 69(8):1102 – 1112, 2015.