Towards Discovering Data Center Genome Using Sensor Nets

Jie Liu
Microsoft Research
One Microsoft Way
Redmond, WA
liuj@microsoft.com

Bodhi Priyantha
Microsoft Research
One Microsoft Way
Redmond, WA
bodhip@microsoft.com

Feng Zhao
Microsoft Research
One Microsoft Way
Redmond, WA
zhao@microsoft.com

Chieh-Jan Mike Liang
Computer Science
Johns Hopkins Univ.
Baltimore, MD
cliang4@cs.jhu.edu

Qiang Wang
Control Science & Engr.
Harbin Inst. of Tech.
Harbin, China
wangqiang@hit.edu.cn

Sean James
Microsoft Corp.
One Microsoft Way
Redmond, WA
seanja@microsoft.com

Abstract

The IT industry is the fastest growing sector in US energy consumption. Improving data center energy efficiency is a pressing issue with significant economic and environmental consequences. Heat distribution is a key operational parameter that affects data center cooling and energy consumption. However, typical data centers lack effective fine-grained sensing systems to monitor heat distribution at a large scale. In this paper, we motivate the use of sensor networks as a dense instrumentation technology to understand and control cooling in data centers. We present Microsoft Research Genomote sensors designed for data center monitoring, and the RACNet for reliable data acquisition. We describe lessons learned from early pilot deployments, and discuss architectural and technical challenges in developing data center sensor networks.

Categories and Subject Descriptors

C.3 [**Special-Purpose and Application-Based Systems**]: Real-time and embedded systems

General Terms

Experimentation

Keywords

Sensor Network, Data Center

1 Introduction

Computer servers and network devices are at the core of IT infrastructure. As enterprise and Internet computing services scale up, these devices are consolidated into data centers to take advantage of the economy of scale. Data centers provide centralized cooling, power, and networking services. The IT industry is the fastest growing sector in US energy consumption. According to U.S. Environmental Protection Agency [2], the United States data centers consumed 61 billion kWh electricity in 2006, which was enough to power up 5.8 million average U.S. households. The amount is expected to double in the next five years under the current trend. Data center energy saving is a pressing issue with great economical and social impact.

In a typical data center, roughly half of the energy is used by IT equipment, while the other half is used in power distribution and cooling [1]. Since the servers and network devices are huge investment for an enterprise, and their reliable operation is key to the success of the businesses, data center operations are usually quite conservative: power is over provisioned [3]; and devices are over cooled. Heat distribution in data centers has complex dynamics, related to many factors, such as room sizes, ceiling heights, rack layout, air vent locations, server types, and workload distribution. However, facility operators lack sufficient visibility into how heat is generated, distributed, and exchanged in data centers. Providing this visibility to data center operators in real time can reduce over cooling, encourage innovation in rack layout design, increase operation effectiveness, and ultimately save the energy used by the facility.

In this paper, we motivate the use of wireless sensor networks as a key technology in data center operation monitoring and control. They can provide non-intrusive and fine-grained data collection with a relatively low cost. At the same time, the large-scale and dense deployments also challenge existing sensor network and system technologies in many ways: for example, how to power the nodes, how to effectively use wireless bandwidth, and how to reliably collect data in real time. We share our experiences and thoughts on building data center monitoring sensor networks, and analyze future technical challenges.

The remainder of the paper is organized as follows. In section 2, we describe the data center cooling management challenges and the advantage of using wireless sensors in such environments. In section 3, we present *Genomotes*, a wireless sensor platform specifically designed for data center monitoring, and *RACNet*, a system for collecting data in

HotEmNets'08 June 2-3, 2008, Charlottesville, Virginia, USA

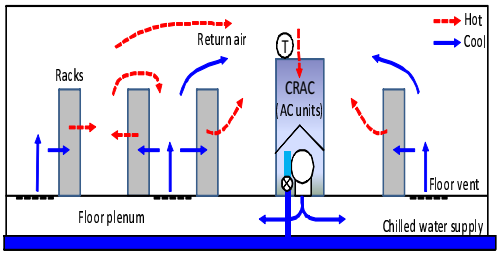

Figure 1. An illustration of a data center server colocation [10].

large-scale and dense sensor networks. We share our experiences from a few early trials and discuss the technical challenges for RACNet in section 4.

2 Sensors on Racks

2.1 Data Center Cooling Basics

Data centers are built in many ways. Figure 1 is an example of the cross section of a data center room (called server colocation, or *colo* for short). Racks are installed on a raised floor in aisles. Cool air is blown by computer room air conditioning (CRAC) systems to the sub-floor. Some floor tiles are perforated as vents to make cool air available to servers. The aisles with these vents are called *cold aisles*. Typically, servers in the racks draw cool air from the front, and blow hot exhaust air to the back – hot aisles. To effectively use the cool air, servers are arranged face to face alongside the cold aisles. As illustrated in the figure, cool air and hot air are eventually mixed near the ceiling and is drawn into the CRAC. Inside the CRAC, the mixed air exchanges heat with chilled water. Usually, there is a temperature sensor at the intake of the CRAC, and the chilled water valve opening is controlled to regulate that temperature to a setpoint.

Heat distribution in data centers can be analyzed in two ways: through computational fluid dynamic (CFD) simulation, or through environmental sensing. With good thermodynamic and air-flow models, CFD simulations can be quite accurate, and is useful to perform "what if" analysis. However, the thermo-properties of materials in data centers can be very diverse. Rack contents and rack layout also change over time. It is hard to keep CFD models up to date as building those models is time consuming and expensive. On the other hand, in-situ temperature and humidity data collected from sensors can provide direct visibility to the heat distribution and can be used to adjust air conditioning in real time [10].

Most servers built in recent years have multiple on-board temperature sensors, and in some cases, these data can be accessed by high-level software and collected to a central place. However, since these sensors are rationally placed near thermo-sensitive devices such as CPU and disks to prevent them from overheating, their readings are very sensitive to server workload, thus cannot be directly used to control air conditioning operations. In comparison, data collected from sensors attached to the racks are less noisy and tangible for CRAC control.

We use a concrete experience to motivate the usefulness of sensors on racks. As a pilot deployment, we installed

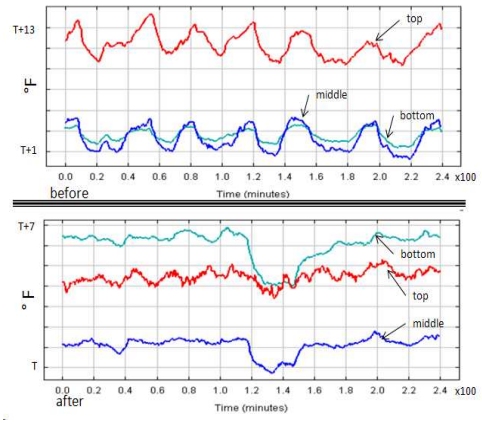

Figure 2. Temperature measurements from three sensors before and after air blockers are installed.

sensors on several server racks: three in the front of each rack, to monitor the intake air temperature, and three in the back of each rack, to monitor the exhaust air temperature. One day, a data center operation manager wanted to test an idea of isolating cold aisles from hot aisles by putting air blockers at the end of the aisles. After the installation, some servers started to send overheat alarms. Surprisingly, all the overheated servers were near the bottom of the racks, which should be the coolest spots. Naturally, when there were overheat alarms, the operation engineers increased the CRAC fan speed to provide more cool air circulation. Unfortunately, this action made the situation worse. Even more servers near the bottom gave out alarms.

Our sensor data revealed the cause. Figure 2 shows the temperature readings before and after the installation of the air blockers. We can see that after installing the air blockers, almost all temperature data were lower than before. The temperature at the top of the rack dropped almost 5^oC. However, the bottom of the rack, rather than being the coolest area, becomes the hottest area. This can be explained by the Bernoulli's principle, which states that an increase in fluid speed causes a decrease in pressure. After installing the air blockers, less air was diffused through the gaps between racks, which increased the speed of the air flow, especially near the bottom of the racks. The low pressure pockets created by the high speed airflow drew warm air from the back of the rack through the gaps between the bottom servers and the floor. After sealing the bottom of the rack better, and *reducing* air flow speed, the problems went away.

This example shows the complex air dynamics in data centers, and the value of having dense sensors for troubleshooting and supporting operational decision making.

2.2 Using Wireless Sensors

Although data centers are highly engineered environments, filled with network infrastructure, there are several advantages in using *wireless* environmental sensors.

- Non-intrusive: Deploying wireless sensors does not require changes to existing data center infrastructure. Data center facility management is a huge challenge. Misconfiguration of network routers can cause signif-

icant performance degradation. Having thousands of sensors over wired TCP/IP network is a big system management risk. In addition, the use of wireless sensors does not require the installation of any software on servers. Since these sensors are completely "out of band", adopting them does not require the buy-in from users who run application on servers.

- Adaptive to changes: In data centers, environmental sensors can be useful in two ways: general monitoring and on-demand troubleshooting. Temperature and humidity change slowly over time and diffuse slowly over space. General monitoring only needs sparse sensors and slow sampling. However, when significant changes are made to a colo, such as adding new racks, decommissioning old servers, or tuning up major facility equipment, dense and frequent monitoring is critical. Wireless sensors can be quickly deployed and relocated, which gives data center operators a flexible instrument for on-demand monitoring.
- Low cost: Typically, the cost of a rack in a data center, including the power and networking cables and non-IT components inside it, is between 5,000 to 10,000 dollars. The servers in a rack can worth hundreds of thousand dollars. Imposing an additional couple of hundred dollars for sensors is acceptable, especially considering the amount of cooling energy they can help save.

In particular, we focus on IEEE 802.15.4 enabled wireless sensors, since from a hardware perspective they are low power and can be built at low cost. The amount of data collected over the sensors is also relatively small.

3 DC Genome System

The Data Center Genome project at Microsoft Research aims to understand how energy is consumed as a function of server hardware, application performance, network load, heat distribution, and many other factors at the data center level, and to improve data center computing efficiency by dynamic resource provisioning and control. At the core of the system is a Reliable ACquisition Network (RACNet) that provides fine-grain sensing of the physical facility. We start with collecting environmental data such as temperature and humidity. The system can be used to collect other data as well.

Figure 3 shows the architecture of DC Genome System. Wireless sensors are deployed around the racks. Data are collected to gateway stations, where they are temporarily staged for further processing. The gateways provide SNMP interfaces and a set of web-service interfaces for other systems to access the data for visualization, analysis and archiving. Due to the long latency introduced by accessing archival data warehouse, recent data are visualized directly from the staging database.

In the rest of this section, we describe some key components in this architecture.

3.1 Genomote

Genomotes are sensors designed specifically for the DC Genome project. We use a combination of wired and wireless links among the sensors to reduce the number of nodes

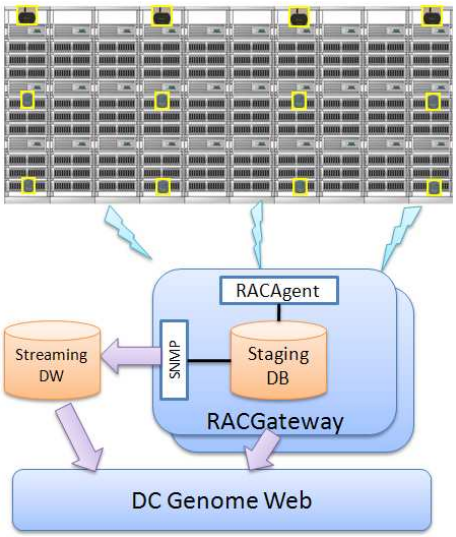

Figure 3. The architecture of RACNet.

that compete for the same radio medium.

There are two kinds of Genomotes, *masters* and *slaves*, as shown in Figure 4. Their capabilities are similar to Telos motes [11]. A master is a wireless node, with a MSP430 microcontroller, a CC2420 radio, 1MB of flash memory, and a temperature/humidity sensor (SHT11). The master node also has a RS232 port that allows it to connect to a slave node. A slave node has a MSP430 microcontroller, a temperature sensor (ADT7301), but no radio capability or flash. It has two RS232 ports, one upstream, connecting to a master node, and one downstream, optionally connecting to another slave node. In this way, one master node and multiple slave nodes can form a daisy chain.

The daisy chain formation is ideal for installing on server racks. When monitoring heat distribution, one usually needs to compare the temperatures at various heights of the rack. The chain naturally drapes down from the top of the rack, and the density can easily be adjusted. The daisy chain design reduces the number of wireless nodes in the network, but still allows individual racks to be relocated without tangling wires. Another benefit of the design is to reduce cost. Humidity, measured in terms of dew point, albeit being a important parameter to monitor, does not vary much within a colo. The cost of a slave node can be made much lower without a humidity sensor.

The sensors are powered by USB ports of servers in racks, since USB ports are ubiquitously available. The chain is designed in a way that one USB connection can power the entire chain of up to 8 nodes. To make these sensors non-intrusive to the servers, the data wires in the USB connection can be disabled by a switch on the motes. So, the host servers will not pop up dialog boxes for driver installation.

A master node also has a re-chargeable battery. In the rare cases when the server providing the USB power is shut down, a fully charged battery (1100mAh at 3.6V) can power up a 8-mote chain for more than 8 hours, so that the server will come back on or a data center operator can find another

Figure 4. The Genomotes hardware platform.

USB power for the sensors.

Each sensor has a unique electronic ID and a unique barcode sticker on the case. They are mapped to the location, in terms of the host rack and slot number (representing height), at the installation time. The location information is then used in data retrieval and visualization.

3.2 Data Staging and Visualization

Sensor data are collected by RACAgents that run on dedicated gateway servers located in each colo. The agents retrieve sensor data and insert them into a staging database. Data across multiple RACAgents are synchronized and checked for consistency. We have built a back-end database and corresponding web interfaces for data staging and visualization. The back-end database is a relational database implemented in Microsoft SQL Server for flexible query processing. We also implemented a standard SNMP interface for retrieving sensor data. In particular, the interface is used by a streaming data warehouse for archiving purposes. The data warehouse is specifically chosen to archive time series with small disk spaces by doing signal compression.

We also pull other data sources available in data centers, such as CRAC valve opening, CRAC return air temperature, device power consumption, and device load. The web interface provides various ways to visualize and export these data. For example, users can directly plot the data streams across sensor types in a browser to study the trends and correlations. Users can also visualize spatio-temporal sensor data through animated temperature contour maps. Figure 5 shows the front and back temperature contour, each interpolated from 12 sensing points, overlaid on top of a row of 10 racks and their contents. From the contours, we can clearly see how gaps in the racks affect the heat distribution.

The query interface on the website allows users to export the data, synchronized over time, to external tools, such as Excel and Matlab for further analysis.

4 Technical Challenges

We have deployed several pilot sensor networks in various data centers to understand the application constraints for RACNets, and discovered some key technical challenges in this application that have not been addressed in previous sensor network research. In RACNet,

- Power is not a dominant concern. To provide long-term fine-grained monitoring with low-maintenance, sensors have access to line power from nearby servers.
- Network is dense and large-scale. A RACNet consists of thousands of sensors in a data center to perform fine-grained monitoring. Every sensor generates 10 byte per sampling period. This implies several hundred kbps burst traffic per sensor type. In addition, due to the physical layout of the data center and the deployment density, dozens of sensors may be within one-hop communication range from each other.

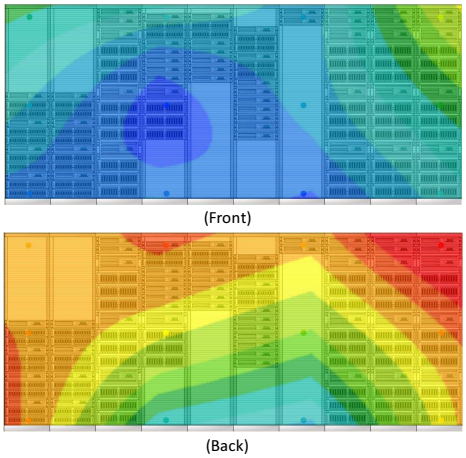

Figure 5. The temperature contour map of the front and the back of a row of servers.

- The sensor network is expected to be a production system. Unlike many scientific applications where the sensors are only deployed for a couple of days or weeks, RACNets are expected to be part of daily data center operation. They are operated by people with little knowledge of wireless sensors. It is important that they can monitor their own health and require zero attention from operators.

These characteristics distinguish RACNet from existing sensor network applications. For example, in most military, environmental or habitat monitoring deployments, sensors are completely un-tethered [9, 12, 5]. Designers are forced to make trade-offs between power consumption and application performance. Duty cycling is a common approach in reducing overall power consumption of the network. Usually, the price to pay is reduced data yield and long data collection latency.

To overcome network density, one might choose to limit the radio power to encourage spatial spectrum reuse. However, our experiments show that with Genomote or TelosB-class motes, it takes about an additional 25ms to forward one 128-byte packet (maximum size allowed in IEEE 802.15.4 protocol) one more hop. Thus, increasing the number of hops introduces extra delay to data collection latency.

RACNets are also different from Tenets [4], where dumb sensors are controlled by a large number of powerful microservers. Given the cost and availability of rack space, it is hard to deploy a large number of microservers. We rely on a multi-hop network among the motes to reliably relay data. To do this, the network needs to be resilient and be able to accommodate failures and disruptions.

The key goal of RACNets design is to provide reliable, low latency data collection over large-scale and dense sensor networks. We address the following research issues:

- **Multi-channel multi-hop networks.** At the networking

layer, the relatively low data rate of IEEE 802.15.4 radios has difficulty in supporting dense network and burst traffic. However, modern radio chips, such as TI CC2420 on Genomotes, provide channel diversity. Since transmissions happening on distant channels do not collide with each other, channel diversity can improve spatial multiplexing [13, 8, 7]. Building multi-hop data collection network over multiple channels is challenging. Metallic objects in data centers, such as racks and panels, can alter the radio propagation pattern. We cannot rely on node locations for radio scheduling. Instead, we expect the number of nodes on each channel to dynamically adapt to link quality changes. At the same time, collection trees on each channel need to be balanced and stable to maximize overall throughput. Given the size of the network and low sink to sensor ratio, efficient signaling between nodes is crucial to preserve bandwidth for data communication.

- **End-to-end reliable data collection with low latency.** An accurate heat distribution map for troubleshooting requires the complete data set from sensors. Our trial deployments using off-the-shelf TinyOS 1.x Oscilloscope and TinyOS 2.0 MultihopOscilloscopeLQI give about 60% and 80% data yield, respectively, which cannot meet application requirements. Therefore, at the application level, we need reliable and low-latency data collection protocols. Unlike in Flush [6], our large-scale and dense network topologies inherently have bandwidth problems, especially in the case where sensors are constantly generating data. Although the external flash memory on Genomotes is capable of temporarily caching data to tolerate communication glitches, we want to avoid communication collisions as much as possible due to the low bandwidth constraints.
- **Simplification to network management.** One of the design goals of RACNet is ease of management. For example, software on all sensors has to be the same, with no extra configurations for node IDs, channels, or communication power levels at deployment time. In addition, administrators should be able to gather network health information through the base stations, and disseminate network-wide commands and code updates. The management protocol should be efficient in a large-scale setting. In addition, its traffic should not affect the data transmission between sensors and sinks.

5 Conclusion

RACNets will be some of the largest sensor networks deployed in in-door environments with clear application requirements and distinct system characteristics. We believe that lessens learned from this project, such as multi-channel reliable data collection protocols, network management techniques, and deployment experiences, will be broadly applicable to other sensor systems, such as building management, museum assistance, and factory floor sensing.

Collecting data to understand heat distributions is a first step toward improving data centers energy efficiency. A long term goal of DC Genome is to close the loop between large-scale sensing and distributed actuation to dynamically change the environmental conditions and resource allocation in data centers. These feedback control systems are particularly challenging, since they involve both physical and computational processes with very different time scales.

6 Acknowledgment

The authors would like to thank Microsoft Data Center Services team, especially Mike Manos, Amaya Souares, Jeff O'Reilly, and Kelly Roark, for their support and collaboration. Michael Sosebee, Rasyamond Raihan, and Martin Cruz also made engineering contributions to the project.

7 References

[1] BELADY, C. L. In the data center, power and cooling costs more than the it equipment it supports. *ElectronicsCooling magazine 3*, 1 (February 2007).

[2] EPA Report on Server and Data Center Energy Efficiency. U.S. Environmental Protection Agency, ENERGY STAR Program, 2007.

[3] FAN, X., WEBER, W.-D., AND BARROSO, L. A. Power Provisioning for a Warehouse-sized Computer. In *ISCA* (2007).

[4] GNAWALI, O., JANG, K.-Y., PAEK, J., VIEIRA, M., GOVINDAN, R., GREENSTEIN, B., JOKI, A., ESTRIN, D., AND KOHLER, E. The tenet architecture for tiered sensor networks. In *ACM SenSys* (2006).

[5] HARTUNG, C., HOLBROOK, S., HAN, R., AND SEIELSTAD, C. FireWxNet: A multi-tiered portable wireless system for monitoring weather conditions in wildland fire environments. In *Mobisys* (2006).

[6] KIM, S., FONSECA, R., DUTTA, P., TAVAKOLI, A., CULLER, D. E., LEVIS, P., SHENKER, S., AND STOICA, I. Flush: A reliable bulk transport protocol for multihop wireless network. In *5th ACM Conference on Embedded Networked Sensor Systems (SenSys '07)* (Sydney, Australia, Novermber 2007).

[7] LE, H. K., HENRIKSSON, D., AND ABDELZAHER, T. A practical multi-channel medium access control protocol for wireless sensor networks. In *IPSN '08* (2008).

[8] LIANG, C.-J., MUSALOIU-E., R., AND TERZIS, A. Typhoon: A reliable data dissemination protocol for wireless sensor networks. In *European conference on Wireless Sensor Networks 2008* (Bologna, Italy, January 2008).

[9] MAINWARING, A., POLASTRE, J., SZEWCZYK, R., CULLER, D., AND ANDERSON, J. Wireless sensor networks for habitat monitoring. In *First ACM Workshop on Wireless Sensor Networks and Applications* (Atlanta, GA, USA, September 2002).

[10] PATEL, C. D., BASH, C. E., SHARMA, R., BEITELMAL, M., AND FRIEDRICH, R. Smart cooling of data centers. In *Proceedings of International Electronic Packaging Technical Conference and Exhibition* (Maui, Hawaii, June 2003).

[11] POLASTRE, J., SZEWCZYK, R., AND CULLER, D. Telos: Enabling ultra-low power wireless research. In *2005 Internal Conference on Information Processing in Sensor Networks, SPOTS track* (Los Angeles, California, April 2005).

[12] WERNER-ALLEN, G., LORINCZ, K., JOHNSON, J., LEES, J., AND WELSH, M. Fidelity and yield in a volcano monitoring sensor network. In *OSDI* (2006).

[13] WU, Y., STANKOVIC, J., HE, T., AND LIN, S. Realistic and efficient multi-channel communications in dense sensor networks. *INFOCOM 2008. 27th IEEE International Conference on Computer Communications. Proceedings* (April 2008).