Advanced control for data center cooling

Heating, ventilation, and air conditioning (HVAC) systems constitute one fifth of the total energy use in the United States and Europe [1][2], and one half of energy use in buildings [1]. The IT industry in particular is responsible for 2% of carbon emissions worldwide, and up to ½ of power consumption in data centers is due to cooling. Using advances control strategies to reduce the energy consumption for data center cooling is a promising way to reduce this figure.

Despite academic advances in HVAC control theory, industry reaction has been slow. Standard practice still consists of using a single heat pump with PID control logic, together with a single temperature sensor, to control the air conditions in a large thermal zone, often consisting of several rooms. These PID controllers are often poorly-tuned, jeopardizing performance and stability. Furthermore, modern programmable logic controllers (PLCs) are powerful enough to handle much more complex algorithms than PIDs, yet this computational ability is not currently harnessed.

Superior control algorithms for data center cooling should include heightened awareness of the physical state of the data centers, (particularly airflow and pressure), and should be able to incorporate external information, including future weather predictions, data center usage patterns, and varying energy rates from utilities. Some advanced control strategies, such as receding horizon control or stochastic control, can take these factors into account [2][4][5]. Such a controller would use a model of the plant to predict how the control input will affect the future system behavior. The optimal control input would then be obtained by periodically solving an optimization problem, or by referencing a lookup table with a previously-determined optimal policy.

Although recent work has been done to advance receding horizon control strategies in office building HVAC systems [2], these studies have shown only limited success; and implementation is still an issue, and is still limited to academic studies. In these studies, the controller uses a thermodynamic model of the building to predict how changing the temperature setpoints for each heat pump or variable air volume (VAV) box will affect both energy consumption and occupant comfort over the entire day. In this role, the controller, running on a central server, acts in a supervisory role, while the regulatory-level PLCs use PID control logic to maintain the zone temperature at the setpoint given by the MPC.

Data center cooling, however, is a much more promising application of advanced control strategies, as compared to commercial office buildings. One major reason for this is that the system is significantly simpler. For example, one of the major sources of heat in an office building is its occupants; this of course presents of significant problem of predicting occupancy levels in order to provide appropriate levels of cooling. Heat in data centers comes primarily from the servers themselves, which are much more predictable. Another advantage is that data centers have, for the most part, similar architecture: Most have cool air ducted in through an underfloor plenum, which is then directed through the server racks and then collected overhead. Finally, data centers are very energy dense, and the use of advanced control techniques is justified by the high ROI on cooling infrastructure investments. Some companies are already cashing in on this, by selling turn-key data center efficiency packages. These companies, apart from demonstrating that there exists a market for efficient data center cooling, can provide a means for which these control strategies to be implemented: By adding advanced control to their efficiency packages, data center managers would not need to concern themselves with installing or troubleshooting the control system, as it could be included in a more comprehensive packages.

From a research point of view, this problem is an exercise in decentralized control, which

is an interesting and current area. Data centers typically have computer room air conditioners (CRACs) distributed on the periphery of the data center floor. The controllers for these units do of course communicate, but bandwidth is limited (to limited, perhaps, to repeatedly solve a global optimization problem), and the CRACs should be able to function properly (and should guarantee global stability) in the case of a communication failure.

References: **[1]** L. Pérez-Lombard et al, "A review on buildings energy consumption information", Energy and Buildings 40 (2008). **[2]** Y. Ma, et al, "Model Predictive Control for the Operation of Building Cooling Systems", Amer. Contr. Conf., 2010. **[3]** Technology Roadmap, Energy-efficient Buildings: Heating and Cooling Equipment, IEA, 2011. **[4]** F. Oldewurtel, et al, "Increasing Energy Efficiency in Building Climate Control using Weather Forecasts and Model Predictive Control", Clima - RHEVA World Cong., 2010. **[5]** Y. Ma, et al, "A Distributed Predictive Control Approach to Building Temperature Regulation", Amer. Contr. Conf., 2011. **[6]** F. Borrelli, Francesco, et al, Constrained optimal control of linear and hybrid systems, Springer, 2003