Better Network Design Through Measurement and Modeling

by Adam Wierman

Designing and managing networked systems relies on a balance of experimental and analytic tools. Despite the importance of both, recently, experimentation has begun to dominate. Tools like scheduling and queueing theory, which proved invaluable during the development of the internet, are now considered passé. There are many reasons for this shift. The growing complexity of systems makes developing tractable, accurate models increasingly difficult. Further, as we develop a better understanding of system workloads and user behavior, traditional assumptions in models are increasingly being invalidated. Additionally, the traditional metrics studied in scheduling and queueing theory are secondary in importance to measures like power usage, quality of service, and fairness. These changes have resulted in new design paradigms, that no longer fit existing models, e.g. the increasing adoption of multi-channel wireless and multi-core chips to combat power constraints.

“restore balance between experimental and analytic design in networked systems.”

The research in my group seeks to energize analytic performance evaluation, and thus restore balance between experimental and analytic design in networked systems. In order to accomplish this, tools from scheduling and queueing theory need to be “modernized” for today’s computer systems. This research requires working closely with both practitioners, in order to understand design decisions being considered, and theoreticians, in order to develop new mathematical tools.

The Impact of the Lee Center

I arrived at Caltech only 2 years ago. However, in this short time the Lee Center has been instrumental in allowing my research to hit the ground running at Caltech. In fact, the Lee Center even played a role in drawing me to Caltech as a result of the financial freedom it provided and the circle of diverse researchers it brings together. The support from the Lee Center allowed me to quickly build my group and also start new research directions without concern as to when external funding would follow. As a result, in only two years, we have had important research successes in projects that began only after I arrived at Caltech.

Research Successes

Lee center funding was essential for each of these projects. Each began during my initial days at Caltech and benefited greatly from the interactions with other faculty and students in the workshops and seminars organized through the Lee Center.

The science of Green IT: The rapidly increasing power consumption of networked systems is a pressing concern for many reasons, including high operating costs, limited battery lifetimes, and a growing carbon footprint. These concerns arise in devices as small as chips, where power limitations have pushed the industry to adopt multi-core designs, all the way to systems as large as data centers, where energy costs are often 40% of the operating budget. As a result, there is a push both in academia and in industry to develop more energy efficient designs.

For many years, the maxim of system design was “faster is better”, but now that energy efficiency is important, the maxim has changed to “speed costs power”—there is a tradeoff that must be made between “faster” (smaller delay, larger throughput) and “greener” (less energy). Across all levels of computer systems this tradeoff is typically made via speed scaling, i.e., controlling the speed of the system so as to balance energy and delay (e.g. running slower when fewer jobs are waiting). Speed scaling designs are not new and have been applied for many years in wireless devices and chips; however, the fundamental tradeoffs and limitations of speed scaling designs are not understood.

Our work seeks to explore these tradeoffs analytically and has exposed some important new insights:

- *What is the optimal speed scaler?* We have proven that it is impossible for an online speed scaling design to be optimal across all workloads. Over the past decade analytic research has sought to provide near optimal speed scaling algorithms. Our work proposes an algorithm that improves the best known performance guarantee (our algorithm is 2-competitive) and, further, proves that no online algorithm can be better than 2-competitive. Thus, our results show that scheduling for energy and delay is fundamentally harder than scheduling for delay alone (since it is possible to be optimal for mean delay).

- *How sophisticated must a speed scaler be?* We have proven that a simple speed scaling scheme that sleeps when the system is idle and otherwise runs at a constant speed is nearly as good as the optimal speed scaling scheme, which can dynamically adjust speeds at any point of time. However, the optimal scheme provides a different benefit: robustness, e.g., to time-varying workloads.

- *Does speed scaling have any unexpected consequences?* We have proven that speed scaling increases the unfairness of scheduling policies: large jobs are more likely to be in the system when the server speed is slow. However, this unfairness can be countered by paying a small price of increased energy usage, e.g., via increased speeds at low occupancies.

- *How does scheduling interact with speed scaling?* Our results show that in many cases decisions about processing speed and scheduling order can be decoupled with little loss in performance. Thus, optimal speed scaling algorithms can be determined largely independently of the scheduling policy even though it seems that these decisions are highly intertwined.

Non-cooperative cooperative control: Decentralized distributed resource allocation is an increasingly common paradigm across computer networks. Indeed, it is the dominant paradigm in wireless networks, where centralized control is typically impossible, e.g., for access point assignment, power control, and frequency selection problems. The design of distributed protocols for these problems is difficult and typically protocols come with few analytic guarantees. Our work proposes a (non-cooperative) game-theoretic approach to resource allocation that provides general application-independent techniques for developing efficient distributed designs. In particular, our approach designs the decentralized agents as self-interested players in a game, and then engineers the rules of the game in a way that ensures the equilibria of the game (the stable points) are efficient. When taking such an approach, the key engineering decisions are (i) how to design the rules of the game and (ii) how to design the agents that play the game. Our work gives application independent design rules for each of these decisions. Further, we have developed applications of these techniques, including the sensor coverage problem, network coding, power control in wireless networks, and the access point assignment problem.

“For many years, the maxim of system design was ‘faster is better,’ but now that energy efficiency is important, the maxim has changed to ‘speed costs power’”

Tails of scheduling: Traditional scheduling analysis focuses on performance metrics such as mean delay and mean queue length, while providing little insight into the design of scheduling policies that optimize the distribution of delay and queue length. However, modern networked systems seek quality of service (QoS) measures that depend on distributional characteristics not just expected values. To bridge this gap, we are developing analytic tools to study the distributional behavior of scheduling policies in general settings. We have succeeded in analyzing the distributional behavior of delay under a wide array of common policies in very general settings. From this work has emerged some interesting insights. For example, policies that perform well under light-tailed job sizes perform poorly under heavy-tailed job sizes and vice versa. For ten years, it has been widely conjectured that it is impossible to be optimal for the delay distribution in both light-tailed and heavy-tailed settings. We resolved this conjecture by proving that, indeed, no policy that does not know the job size distribution can be optimal in both settings—and further that if a policy is optimal in one setting it must be worst-case in the other setting. Surprisingly, however, we have also shown that if the policy has very little information about the job size distribution—just its mean—then this is already enough to allow the policy to be near-optimal in both regimes.

Each of these projects are in the initial stages and will continue to produce new insights and designs in the coming years. For each project, the Lee Center provided the initial support that allowed the project to develop the first few results, at which point it became possible to attain external funding to support the project. Now, each projects is self-supporting via government and industrial grants. Thus, the initial support from the Lee Center will continue to resonate for years to come.

Adam Wierman is Assistant Professor of Computer Science.

Read more at: http://www.cs.caltech.edu/~adamw

References

[1] A. Wierman and B. Zwart, "Is Tail-Optimal Scheduling Possible?" Under submission.

[2] J. Marden and A. Wierman, "Overcoming the Limitations of Game-Theoretic Distributed Control," Proceedings of the Conference on Decision and Control (CDC), 2009.

[3] H-L. Chen, J. Marden, and A. Wierman, "On The Impact Of Heterogeneity And Back-End Scheduling In Load Balancing Designs," Proceedings of INFOCOM, 2009.

[4] L. Andrew, A. Wierman, and A. Tang, "Power-Aware Speed Scaling in Processor Sharing Systems," Proceedings of INFOCOM, 2009.

[5] J. Marden and A. Wierman, "Distributed Welfare Games with Applications to Sensor Coverage," Proceedings of the Conference on Decision and Control (CDC), 2009.