Syncing data center computers at the speed of light

December 16 2016, by Bill Steele

When several computers need to work together, as when calculating in parallel on several parts of a single large problem or managing a large database, they need to keep in step. These situations are particularly common in large data centers that make up "the cloud" where institutions farm out their computing needs. So a Cornell computer scientist has come up with a new system, "Datacenter Time Protocol" (DTP), in which signals sent at the speed of light over fiber-optic cables between the computers enable them to stay in sync to within a few nanoseconds.

"Time is important," said Hakim Weatherspoon, professor of computer science. "The closer you can synchronize clocks, the higher performance your systems can achieve." For the corporate or individual user this can mean cloud computing efficiency, faster search engines and speedier apps on a mobile phone.

Weatherspoon was on sabbatical during fall 2015 at Microsoft Research, working with Microsoft researchers Hitesh Ballani and Paolo Costa in their Rack-Scale Computing group. They have continued their collaboration since Weatherspoon's return to Cornell.

"Rack-scale computing" refers to the latest step forward in cloud centers: Beyond linking arrays of individual servers, the systems link racks that hold several computers each. A "switch" on each rack keeps the computers on the rack working together and connects the rack to the rest of the center.

Weatherspoon and doctoral students Ki Suh Lee, Han Wang and Vishal Shrivaslav presented the research in a paper, "Globally Synchronized Time via Datacenter Networks," at the ACM Special Interest Group on Data Communication conference in Florianópolis, Brazil, in August.

The central processing unit in a computer generates a "clock signal" – a steady beat that tells the rest of the chip when to move forward a step in its program, like a drummer keeping a marching band together. Even if all the CPUs in a data center are the same model from the same manufacturer, differences in temperature, power supply or load can cause some of them to run at different rates. The goal is for each computer to recognize the "offset" between itself and another to which it is linked, and adjust the timing of its messages to the other accordingly.

DTP works behind the scenes at the "physical level" of a network. On Ethernet networks, data ordinarily travel in "packets" – strings of ones and zeros marked with a beginning and an end and containing a destination address and a payload of information. DTP sends computers information about their time difference by directly modifying the pulses of light that travel through fiber-optic cables, with no effect on the higher level packets.

Upgrading a data center to use the system would not be expensive, the researchers said, although the hardware that connects computers to the network would have to be modified to read and write the DTP signals. Weatherspoon is hoping for the addition of a DTP standard to the IEEE standards for Ethernet networks.

In tests at a Microsoft data center, DTP kept computers in sync to within 25.6 nanoseconds. "We scheduled every packet from every machine in a rack-scale computer system, resulting in reduced power, cost and size of the network fabric," said Weatherspoon. "They are very interested in this technology."