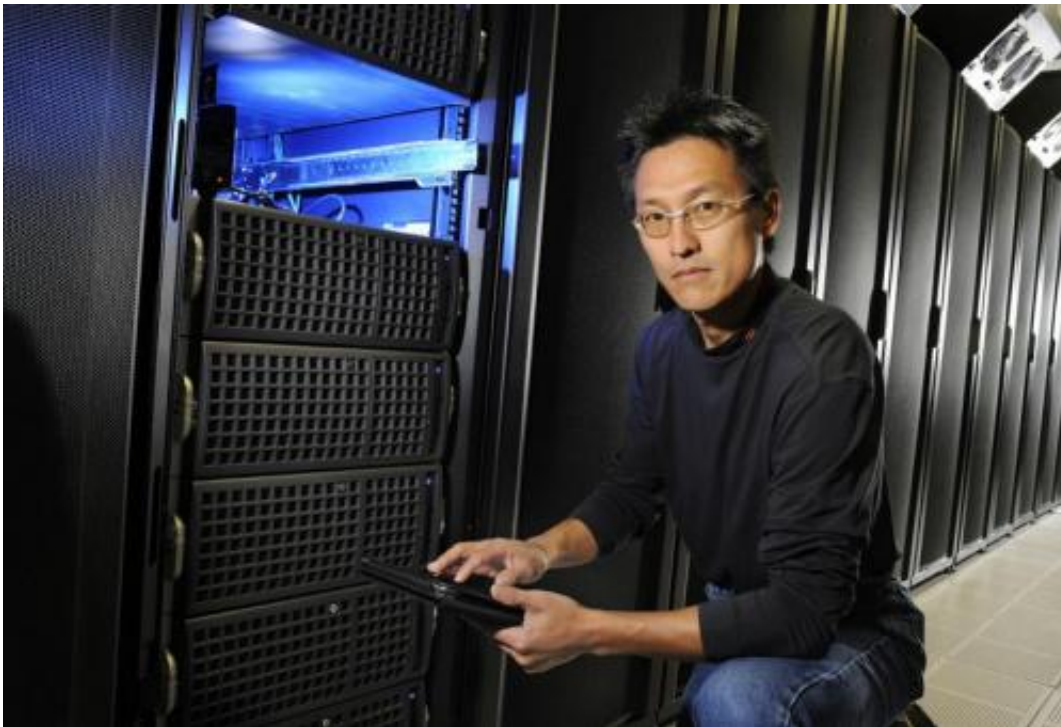


The internet was delivered to the masses; parallel computing is not far behind

August 20 2014



Virginia Tech College of Engineering Professor Wu Feng has garnered support from NSF, Microsoft, the Air Force, and NIH among others, putting together a way to deliver parallel computing to the masses, much like the Internet's worldwide accessibility. Credit: Virginia Tech

During the past few years, Virginia Tech's Wu Feng has built upon a National Science Foundation (NSF) / Microsoft grant from the "Computing in the Cloud" program, and synergistically complemented it

with subsequent collaborative grants, including a \$6 million award from the Air Force on "big computing" for mini-drones and a \$1 million award from NSF and the National Institutes of Health on "big data" for the life sciences.

As he wove together the "parallel computing" aspects from each grant, he was able to tell a much larger, more interconnected story — one of delivering parallel computing to the masses. In doing so, he has worked to apply this democratization of parallel computing to an area of emerging importance—the promise of personalized medicine.

Microsoft took particular notice of Feng's leadership in this cutting-edge research and succinctly worked the supercomputing expert's collaborative ideas into one of its global advertising campaigns, describing Virginia Tech scientists and engineers as "leaders in harnessing supercomputer powers to deliver lifesaving treatments."

This full-page ad ran this summer in the *Washington Post*, *New York Times*, *USA Today*, *Wall Street Journal*, *Bloomberg Businessweek*, *United Hemispheres*, *The Economist*, *Forbes*, *Fortune*, *TIME*, *Popular Mechanics*, and *Golf Digest*, as well as a host of other venues in Philadelphia, Washington, D.C., and Baltimore.

"Delivering personalized medicine to the masses is just one of the grand challenge problems facing society," said Feng, the Elizabeth and James E. Turner Fellow in Virginia Tech's Department of Computer Science. "To accelerate the discovery to such grand challenge problems requires more than the traditional pillars of scientific inquiry, namely theory and experimentation. It requires computing. Computing has become our 'third pillar' of scientific inquiry, complementing theory and experimentation. This third pillar can empower researchers to tackle problems previously viewed as infeasible."

So, if computing faster and more efficiently holds the promise of accelerating discovery and innovation, why can't we simply build faster and faster computers to tackle these grand challenge problems?

"In short, with the rise of 'big data', data is being generated faster than our ability to compute on it," said Feng. "For instance, next-generation sequencers (NGS) double the amount of data generated every eight to nine months while our computational capability doubles only every 24 months, relative to Moore's Law. Clearly, tripling our institutional computational resources every eight months is not a sustainable solution... and clearly not a fiscally responsible one either. This is where parallel computing in the cloud comes in."

As noted by the National Institute of Standards and Technology, cloud computing is "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

"The implicit takeaway here is that the configurable computing resources are hosted and maintained by cloud providers such as Microsoft rather than the institution requiring the computing resources. So, rather than having an institution set-up, maintain, and support an information technology infrastructure that is seldom utilized anywhere near its capacity... and having to triple these resources every eight to nine months to keep up with the data deluge of next-generation sequencing, cloud computing is a viable and more cost effective avenue for accessing necessary computational resources on the fly and then releasing them when not needed," Feng said.

Whether for traditional high-performance computing or [cloud computing](#), Feng is seeking to transform the way that parallel computing

systems and environments are designed and the way that people interact with them.

"My analogy would be the Internet, and how it has transformed the way people interact with information," Feng added. "We need to make a similar transition with parallel computing, whether with the cloud or with traditional high-performance computing such as supercomputers."

The groundwork for Feng's big data research in a "cloud" began in the mid-2000s with a multi-institutional effort to identify missing gene annotations in genomes. This effort combined supercomputers from six U.S. institutions into an ad-hoc cloud and generated 0.5 petabytes of data that could only be stored in Tokyo, Japan.

By combining software Feng developed called mpiBLAST with his ad-hoc 'big data' framework called ParaMEDIC: Parallel Metadata Environment for Distributed I/O and Computing, he and his colleagues successfully addressed an important problem in genomics via supercomputing. They were able to reduce the time it took to identify and store missing gene annotations remotely in Japan from more than three years down to two weeks by simply changing the software infrastructure via parallel computing in their ad-hoc cloud.

This work is now being formalized and extended as part of an NSF/Microsoft Computing in the Cloud grant that seeks to commoditize biocomputing in the cloud.

The advent of NGS, on the heels of the above "missing genes" work, created a larger "big data" problem that resulted in an interdisciplinary grant from NVIDIA Foundation to Compute the Cure (for cancer) and the creation of the Open Genomics Engine, which was presented in a talk at the 2012 Graphic Processing Units (GPU) Technology Conference. Other related "[big data](#)" biocomputing tools that Feng and

his group have created include Burrows-Wheeler Aligner (BWA)-Multicore, SeqInCloud, cuBLASTP, and GPU-RMAP. In turn, this research provided the basis for the recent \$1 million NSF-NIH BIGDATA grant on [parallel computing](#) for next-generation sequencing in the life sciences.

Provided by Virginia Tech

Citation: The internet was delivered to the masses; parallel computing is not far behind (2014, August 20) retrieved 25 April 2024 from <https://phys.org/news/2014-08-internet-masses-parallel.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--