



[Features](#) - September 4, 2008

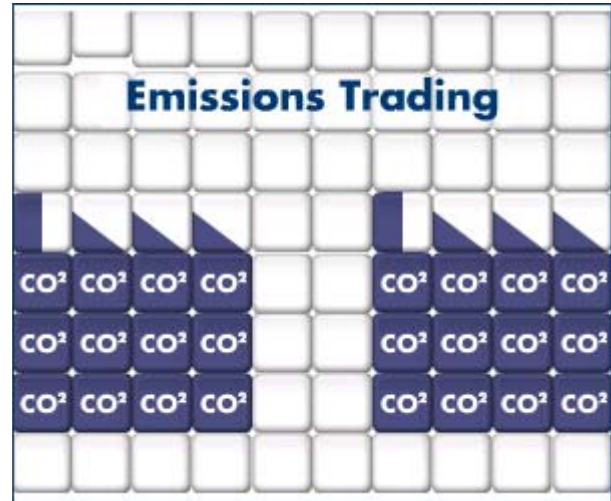
How the Large Hadron Collider Might Change the Web The LHC Computing Grid may teach the Internet how to quietly handle reams of information

By Mark Anderson

When the [Large Hadron Collider \(LHC\)](#) begins smashing protons together this fall inside its 17-mile- (27-kilometer-) circumference underground particle racetrack near Geneva, Switzerland, it will usher in a new era not only of physics but also of computing.

Before the year is out, the LHC is projected to begin pumping out a tsunami of raw data equivalent to one DVD (five gigabytes) every five seconds. Its annual output of 15 petabytes (15 million gigabytes) will soon dwarf that of any other scientific experiment in history.

The challenge is making that data accessible to a scientist anywhere in the world at the execution of a few commands on her laptop. The solution is a global computer network called the [LHC Computing Grid](#), and with any luck, it may be giving us a glimpse of the Internet of the future.



Once the LHC reaches full capacity sometime next year, it will be churning out snapshots of particle collisions by the hundreds every second, captured in four subterranean detectors standing from one and a half to eight stories tall.* It is the grid's job to find the extremely rare events—a bit of missing energy here, a pattern of particles there—that could solve lingering mysteries such as the origin of mass or the nature of dark matter.

A generation earlier, research fellow Tim Berners-Lee of the European Organization for Nuclear Research (CERN) set out to create a global ["pool of information"](#) to meet a similar challenge. Then, as now, hundreds of collaborators across the planet were all trying to stay on top of rapidly evolving data from CERN experiments. Berners-Lee's solution became the World Wide Web.

But the fire hose of data that is the LHC requires special treatment. "If I look at the LHC and what it's doing for the future," said David Bader, executive director of high performance computing at the Georgia Institute of Technology, "the one thing that [the Web](#) hasn't been able to do is manage a phenomenal wealth of data." Bandwidth alone is a major bottleneck. Bader said that for researchers running supercomputer simulations, it's cheaper to write the data to terabyte hard drives and ship them from one supercomputer center to another via FedEx than it is to transfer the gigantic data sets over the net.

The LHC Computing Grid handles data in stages, referred to as tiers. "Tier 0," located at CERN, is a massively parallel computer network composed of 100,000 of today's fastest CPUs that stores and manages the raw data ([1s and 0s](#)) from the experiments. It ships portions of data over dedicated 10-gigabit-per-second fiber-optic lines to 11 "Tier 1" sites across North America, Asia and Europe. Brookhaven National Laboratory in Upton, N.Y., for example, receives data from the ALICE experiment, which collides lead ions.

From those sites the data is parceled out for easier access among 140 Tier 2 computer networks based at universities, government labs and even private companies around the globe. Tier 2 is where scientists will actually access data and perform the kinds of hands-on numerical analysis needed to translate the raw 1s and

Os into energies and trajectories of particles.

The crucial element that will make the data accessible, said project leader Ian Bird of CERN's information technology (IT) department in Geneva, is a type of software known as "middleware". The information a user wants may be spread among petabytes of data on different servers and stored in different formats. An open-source middleware platform called [Globus](#) is designed to gather that information seamlessly as though it's sitting in a folder on one's own desktop PC.

The trial by fire that LHC programmers will be putting Globus through—and the modifications that emerge as a result—may be the first practical outgrowth of the LHC grid. If project scientists can tame massive, worldwide fields of networked data and computing cycles in particle physics, their solutions could well apply across the Internet—in much the same way that Berners-Lee's specialized HTML invention morphed into the very backbone of modern technological society.

Bader imagines future middleware allowing home computers to provide instant weather forecasts by accessing information from nearby environmental sensors. Or it might help sift through a life's accumulation of personal medical records or years of home video footage looking for dimly remembered events.

Ironically, CERN's next great contribution to the Internet could be all but transparent to the end user. In a perfect world, Globus or its successors would simply make everything on a given grid straightforwardly and transparently accessible from any computer. "If Globus is a success," Bader said, "then you won't hear about it."

**Correction (9/3/08): This article originally stated that the LHC will produce millions of snapshots of particle collisions; "millions" refers to the number of collisions, only a fraction of which will be recorded.*

Further Reading

[Internet Eavesdropping: A Brave New World of Wiretapping](#)

[Industry Roundtable: Experts Discuss Improving Online Security](#)

[Data Fusion: The Ups and Downs of All-Encompassing Digital Profiles](#)

[Cryptography: How to Keep Your Secrets Safe](#)

[International Report: What Impact Is Technology Having on Privacy around the World?](#)

[How I Stole Someone's Identity](#)

[Pedophile-Proof Chat Rooms?](#)

[Planning to E-Vote? Read This First](#)