

# Das Computing Grid : zwischen Vision und Realität

Autor(en): **Häberli, Christian**

Objektyp: **Article**

Zeitschrift: **Bulletin.ch : Fachzeitschrift und Verbandsinformationen von Electrosuisse, VSE = revue spécialisée et informations des associations Electrosuisse, AES**

Band (Jahr): **96 (2005)**

Heft 11

PDF erstellt am: **12.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-857814>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

# Das Computing Grid – zwischen Vision und Realität

## Die Hochenergiephysik nutzt das Computing Grid für die Simulation und Analyse von Proton-Proton-Kollisionen

Der Name Computing Grid ist inspiriert durch den für das Stromnetz gebräuchlichen englischen Begriff «Power Grid». Gemäss der Vision des Computing Grid sollten Computerzentren Rechenkapazität in analoger Weise anbieten, wie uns heute die Elektrizitätswerke Strom offerieren. Damit soll Rechenkapazität, deren Umfang die Möglichkeiten des eigenen PCs überschreitet, für alle zugänglich und erschwinglich werden, wie dies heute bei der elektrischen Energie der Fall ist. Statt über einen teuren eigenen Generator im Haus verfügen wir alle über einen Anschluss ans Stromnetz<sup>1)</sup>.

Wie jede Analogie hat auch diejenige zwischen Computing Grid und dem Stromnetz ihre Schwächen. Während im Fall der Stromversorgung nicht benötigte

*Christian Häberli*

Energie in Stauseen gespeichert werden kann, ist die Zeit, die ein Prozessor ohne Rechenauftrag verbringt, unwiederbringlich verloren. Dafür, und im Gegensatz

zur Stromversorgung, können Rechenaufträge in einer Warteschlange gehalten werden, wenn gerade nicht genügend Rechenkapazität vorhanden ist. Schliesslich ist der Bezug von Rechenkapazität ein weitaus komplexerer Vorgang als der Bezug von elektrischer Energie. Während Wechselstrom für den Benutzer im Wesentlichen durch Spannung und Frequenz definiert ist, bedeutet Rechenkapazität, dass ein Programm auf einer Hardware-

Plattform (z.B. Intel Pentium oder AMD Opteron) und auf einem Operating System (z.B. Linux oder Windows) ausgeführt werden muss. Ausserdem muss das Programm Zugang zu den benötigten Input-Daten haben und die Output-Daten zurück an den Auftraggeber transferieren können. Damit nicht genug: das Ausführen von Programmen kann aus verschiedensten Gründen scheitern, etwa wegen eines Fehlers in der Software, wegen beschädigter Hardware oder wegen des Ausfalls von Netzwerkverbindungen. Das heisst, dass dem Auftraggeber der Zugang zu Fehlermeldungen ermöglicht werden muss.

### Die Realität

Wegen der oben erwähnten Schwierigkeiten sind die existierenden Implementierungen von Computing-Grid-Lösungen noch weit von der Vision «Computing Power beziehen wie Strom aus der Steckdose» entfernt. Heute möglich ist das Verbinden von mehreren Computing Clustern (Verbund von einigen wenigen bis einigen hundert PCs an einem Ort) zu

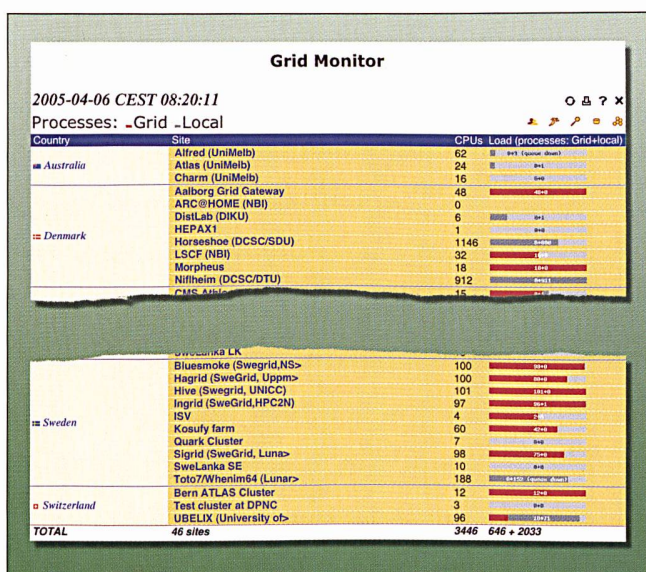


Bild 1 Ausschnitt der im NorduGrid ARC verbundenen Cluster

Auf dem Monitor sind die Anzahl der in den einzelnen Clustern eingebundenen CPU (3446) und die für das Computing Grid aufgewendete Rechenzeit dargestellt. Live: [www.nordugrid.org/monitor](http://www.nordugrid.org/monitor)

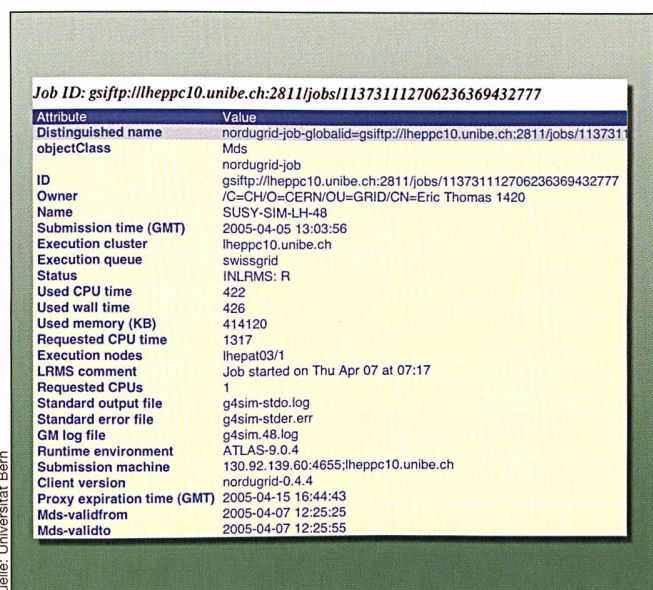
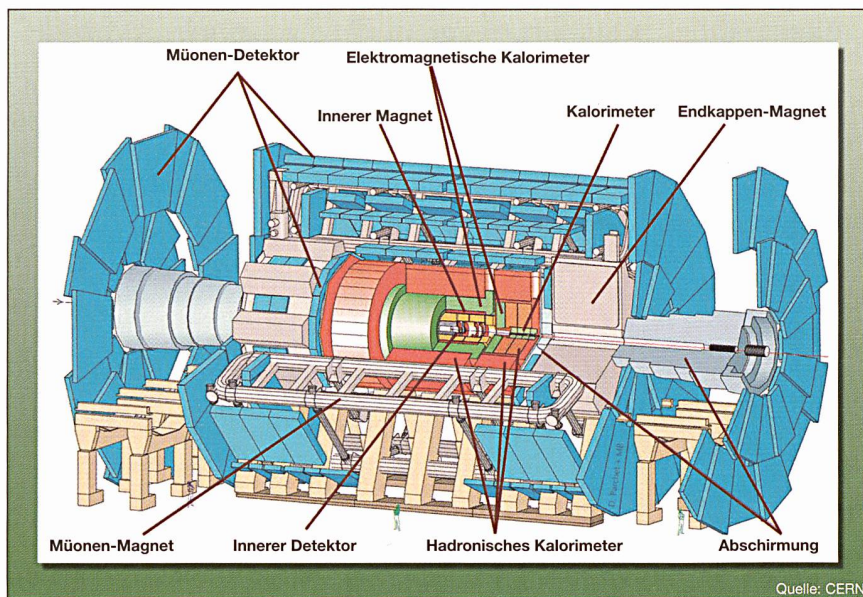


Bild 2 Auszug einer Job-Beschreibung auf dem Computing Grid



**Bild 3 Der ATLAS-Detektor am CERN**  
Der Detektor ist 44 m lang, hat einen Durchmesser von 22 m und ein Gewicht von 7000 t.

### Das Computing Grid in der Hochenergiephysik

Die treibende Kraft hinter den erwähnten Computing-Grid-Lösungen sind die Herausforderungen, welche die Hochenergiephysik an das Computing stellt. Voraussichtlich im Jahr 2007 nimmt am Europäischen Laboratorium für Teilchenphysik (CERN) bei Genf der Large Hadron Collider (LHC) seinen Betrieb auf<sup>6)</sup>. Er wird weltweit der grösste und stärkste Teilchenbeschleuniger sein. Sein Hauptzweck ist es, Proton-Proton-Kollisionen bei einer Energie von je 7 Teraelektronenvolt zu erzeugen. Mittels fünf Detektoren werden die Produkte der Kollisionen vermessen (Bild 3). Dies erlaubt es den Teilchenphysikern, Rückschlüsse auf den Aufbau der Materie zu ziehen. Mit diesem Wissen können auch weiterführende Fragen, vor allem zum Geschehen im frühen Universum kurz nach dem Urknall und damit natürlich auch zum heutigen Universum, beantwortet werden.

Der LHC wird innerhalb eines einzelnen Detektors eine Milliarde Proton-Proton-Kollisionen pro Sekunde produzieren. Eine solch enorme Kollisionsrate ist notwendig, da die gesuchten Prozesse extrem selten auftreten. Von dieser Milliarde Kollisionen werden nach einer strikten Online-Selektion aber nur die 100 Kollisionen mit dem höchsten Entdeckungspotenzial auf Datenträger abgespeichert. Dies entspricht einem Datenfluss von etwa 100 MB/s oder einem Petabyte pro Jahr (1 Petabyte = 1 000 000

einem weltweit verteilten grossen Cluster. Diese Funktionalität wird von verschiedenen Software-Paketen bereitgestellt: so z.B. vom LHC Computing Grid (LCG)<sup>2)</sup> oder vom NorduGrid Advanced Ressource Connector<sup>3)</sup> (NorduGrid ARC, Bild 1). Das LCG wurde hauptsächlich für grosse Cluster in Rechenzentren entwickelt – wie z.B. das CSCS<sup>4)</sup> in Manno – der NorduGrid ARC für kleine und mittelgrosse Cluster an Universitäten. Auf den mittels dieser Software zusammengesetzten, grossen Cluster-Verbänden können parallel voneinander unabhängige Rechenaufträge (Jobs) ausgeführt werden (Bild 2).

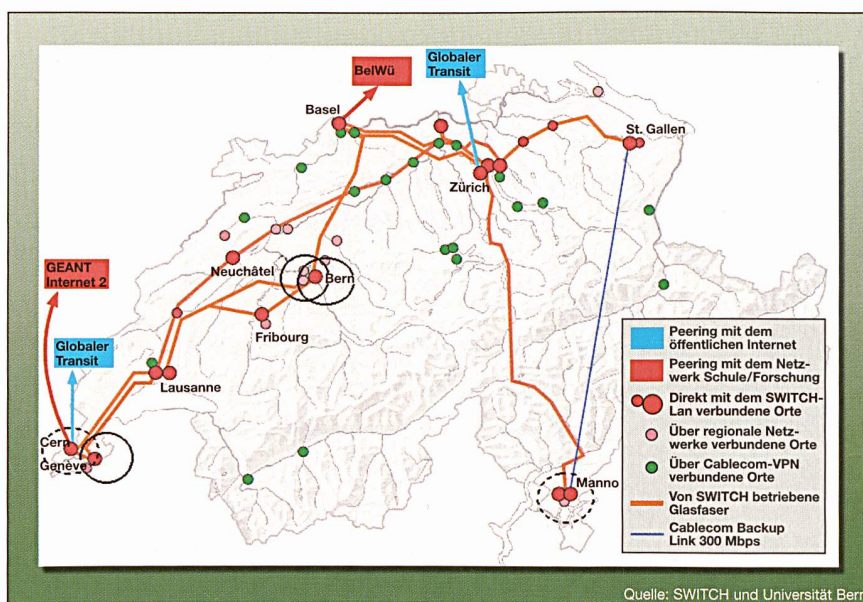
Das Computing Grid verwendet das nachfolgend beschriebene Data Grid für das Management und den Transport der Input- und Output-Daten der Rechenaufträge, die auf dem Computing Grid ausgeführt werden. Die gesamte Kommunikation für das Computing und das Data Grid wird über das Internet abgewickelt.

Zusätzlich sollte mittels Data Grid Tools bereits beim Load Balancing<sup>5)</sup> berücksichtigt werden, ob die für den Rechenauftrag benötigten Input-Daten bereits auf einem Cluster im Computing Grid vorhanden sind. Falls dies nicht der Fall ist, sollte der Job demjenigen Cluster zugewiesen werden, für den die Input-Daten dank einer guten Internet-Verbindung leicht zugänglich sind. Dadurch lassen sich zeitraubende Datentransfers vor dem Ausführen des Rechenauftrags vermeiden.

fachbeiträge

### Das Data Grid

Das Computing Grid ist untrennbar mit dem Data Grid verbunden. Das Data Grid ist die Software-Infrastruktur, die via Internet die Input-Daten eines Jobs zu dem Cluster transportiert, auf dem der Job ausgeführt werden soll. Wenn der Job fertig ausgeführt ist, transferiert die Data-Grid-Infrastruktur die Output-Daten (Resultate und Logs) an einen für den Benutzer zugänglichen Ort (z.B. ein lokaler File Server oder die Harddisk des Desktop-Computers des Benutzers).



**Bild 4 Das Schweizer ATLAS Computing Grid**  
Die ausgezogenen Kreise bezeichnen die derzeit produktiven Cluster, die gestrichelten Kreise bezeichnen den im Aufbau begriffenen Cluster in Manno bzw. die Datenquelle (CERN).

Gigabytes). Es ist weder möglich noch wünschenswert, genügend Computing-Ressourcen an einem Ort zu konzentrieren, um solche Datenmengen analysieren zu können. Ausserdem werden die Detektoren am LHC von grossen internationalen Kollaborationen, bestehend aus Universitäten in aller Welt, gebaut und betrieben. An all diesen Universitäten soll ein Teil der Datenanalyse durchgeführt werden. Die Antwort auf diese Herausforderungen ist das Computing Grid. Die Lösung Grid wird noch attraktiver durch die Tatsache, dass die Rechnungen für die Hochenergiephysik paralleler Natur sind. Eine Proton-Proton-Kollision ist unabhängig von der nächsten. Das heisst, der Computer, der die Rechnung zur Analyse der einen Kollision ausführt, muss nicht mit demjenigen Computer kommunizieren, der die Rechnung für die andere Kollision ausführt.

### Das Schweizer ATLAS Computing Grid

Eines der fünf LHC-Experimente heisst ATLAS<sup>7)</sup>. An diesem Experiment sind die Universitäten Bern und Genf beteiligt. Zur Vorbereitung der Datenanalyse, die im Jahr 2007 beginnen wird, wurde ein ATLAS-Computing-Grid-Prototyp aufgebaut. Der Prototyp basiert auf der NorduGrid-ARC-Software und besteht aus einem Linux Cluster an der Universität Genf und zwei Clustern an der Universität Bern. Zusätzlich wird derzeit ein Cluster am Rechenzentrum der ETH Zürich in Manno bei Lugano aufgebaut. Der Prototyp vereint derzeit insgesamt rund 110 Prozessoren und 2,3 TB Datenspeicher. Die Cluster sind mit dem



Bild 5 Eine simulierte Proton-Proton-Kollision im ATLAS-Detektor

Vor schwarzem Hintergrund: Spuren geladener Teilchen im inneren Detektor. Vor grünem/rotem Hintergrund: Energiedepositionen im elektromagnetischen/hadronischen Kalorimeter

leistungsfähigen SWITCH<sup>8)</sup>-Netzwerk verbunden, das einen zuverlässigen und schnellen Datentransfer ermöglicht (Bild 4).

Das Schweizer ATLAS Computing Grid wird von Forscherinnen und Forschern der Universitäten Genf und Bern hauptsächlich für zwei Zwecke verwendet:

- für die Simulation von Proton-Proton-Kollisionen im ATLAS-Detektor (Bild 5)
- für die Analyse von Test-Daten, die bereits mit einem Segment des ATLAS-Detektors an dem kleineren SPS-Beschleuniger (Super Proton Synchrotron) am CERN aufgezeichnet wurden.

Das Schweizer ATLAS Computing Grid zeichnet sich aus durch eine einfache Bedienung und eine hohe Stabilität – diese Vorzüge sind zu einem grossen Teil auf die Qualitäten der NorduGrid-ARC Software und des Linux Operating System zurückzuführen.

### Die Grenzen des Computing Grid

Das Computing und das Data Grid sind junge Technologien, werden ständig weiter entwickelt und weisen daher Limitationen und fehlende Funktionalität auf. Zusätzlich gibt es, bedingt durch den dezentralen Charakter des Grid, auch prinzipielle Grenzen.

#### Einschränkung «fehlende Interaktivität»

Eine einschneidende Limitation des Computing Grid ist die fehlende Interaktivität. Computer-Benutzer sind sich gewohnt, ihren PC interaktiv zu nutzen, d.h. die Maschine liefert genau dann Rechenkapazität, wenn der Benutzer dies verlangt und der Benutzer kann den Fortschritt der Rechnung mitverfolgen. Im Gegensatz dazu sind die Computing-Grid-Lösungen «Batch»-orientiert, das heisst, dass ein Rechenauftrag an einen Cluster gesendet und dort zunächst einmal in einer Warteschlange gehalten wird, bis die benötigte Ressource im Cluster frei wird. Dann wird der Rechenauftrag ausgeführt. Die Ausführung dauert unter Umständen mehrere Stunden oder Tage, und der Benutzer hat erst ganz am Schluss Zugriff auf die Resultate. Bisher existiert kein überzeugender Ausweg aus dieser unbefriedigenden Situation. In Zukunft werden sich aber wohl so genannte Pull-Modelle etablieren (Bild 6). Die zu lösenden Rechenaufträge werden von einem Master gehalten. Dieser schickt aber anstelle voll definierter Jobs nur so genannte Agenten-Jobs auf das

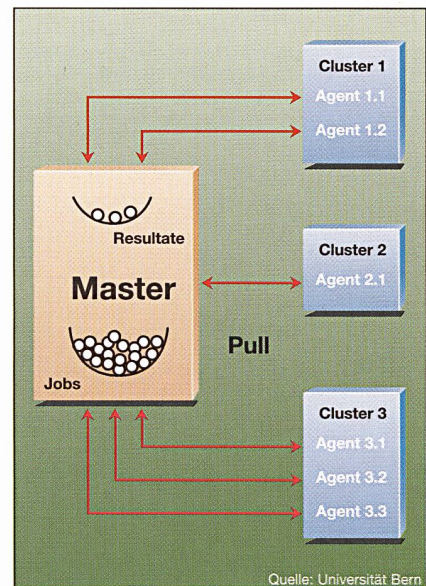


Bild 6 Schematische Darstellung des Pull-Modells für die interaktive Verwendung des Computing Grid

Computing Grid. Sobald ein Agent von einer freien Ressource bearbeitet wird, verlangt er vom Master einen Rechenauftrag, erhält diesen zugeschickt und beginnt mit dessen Ausführung. Nachdem der Job ausgeführt ist, schickt der Agent die Resultate an den Master zurück und verlangt den nächsten Job. Der Master fügt die von den Agenten eintreffenden Resultate fortlaufend zu einem Gesamtergebnis zusammen und zeigt es dem Benutzer an. Dieser kann damit den Fortschritt seiner Rechnungen ständig beobachten. Am Ende der Rechnung hat der Master keine Jobs mehr zu vergeben, und daher verlieren die Agenten ihre Existenzberechtigung. Sie beenden sich selbst und machen die Ressource für nächste Rechenaufträge frei.

#### Einschränkung «Ausführung gekoppelter Probleme»

Eine weitere Limitation besteht in der Schwierigkeit, gekoppelte Probleme auf einer Computing-Grid-Infrastruktur auszuführen. Gekoppelte Probleme zeichnen sich dadurch aus, dass Kommunikation zwischen einzelnen Prozessen notwendig ist. Innerhalb eines einzelnen Computing Clusters ist die Erfüllung dieser Anforderung relativ leicht möglich, wenn die Kommunikation nicht zu oft erfolgt und nicht zu viele Daten ausgetauscht werden müssen. Sobald nun diese Kommunikation zwischen verschiedenen Clustern durchgeführt werden muss, ergeben sich zwei Probleme:

- das Etablieren des Kommunikationskanals: wie wissen die einzelnen Jobs, wo die Partnerjobs ausgeführt werden,

Strecke	Laufzeit [ms]
Innerhalb eines Clusters	0,08
Innerhalb eines Gebäudes	0,12
Uni Bern – CERN	1,5
Uni Bern – Niels Bohr Institut (Kopenhagen)	31
Uni Bern – Fermilab (Chicago)	54
Uni Bern – Uni Melbourne	172

Tabelle Typische Reisezeiten von Internet-Paketen

mit denen sie Daten austauschen müssen?

– die Latenz: es braucht Zeit, bis die Daten via Internet zwischen zwei Jobs ausgetauscht worden sind. Wenn nun Rechenjobs in Bezug auf die Latenzzeit häufig mit andern synchronisiert werden müssen und bis zur erfolgten Synchronisation aufgehalten werden, wird das Ausführen der Rechenaufgabe auf dem Computing Grid ineffizient.

Das erste Problem wird mit der Entwicklung so genannter Message Passing Interfaces (MPI) angegangen. Ein MPI ist eine Software-Infrastruktur, welche die Kommunikation zwischen Jobs auf dem Computing Grid ermöglicht. Anwendbare MPI sind bereits entwickelt, allerdings ist deren Konfiguration vor allem wegen Problemen mit Firewalls schwierig. Aus Sicherheitsgründen öffnen Netzwerkadministratoren die Firewalls für die MPI-Infrastruktur nur ungern.

Das zweite Problem hingegen ist eine grundsätzliche Limitation des Konzepts Computing Grid. Typische Reisezeiten von Internet-Paketen sind in der Tabelle dargestellt.

Als Beispiel sollen voneinander abhängige Rechenaufträge auf dem Computing Grid ausgeführt werden, wobei die Jobs einmal pro Sekunde miteinander 1 kB Daten austauschen müssen. Wenn einige Jobs in Bern und andere Jobs am CERN ausgeführt werden, ist dies überhaupt kein Problem: nach jeder Sekunde Rechenzeit werden weniger als 10 ms für den Informationsaustausch aufgewendet. Wenn aber einige Jobs in Bern und andere in Melbourne ausgeführt werden, so gehen nach jeder Sekunde Rechenzeit mindestens 344 ms (je 172 ms für hin und zurück) für die Kommunikation verloren. Die Ausführung der Rechnung auf dem Computing Grid wird ineffizient. Das heisst, dass das Computing Grid beim Load Balancing die Distanz zwischen einzelnen Clustern berücksichtigen muss.

In unserem Fall hiesse das, dass die Jobs z.B. in Bern, am CERN und in Kopenhagen ausgeführt würden und nicht in Melbourne. Die dafür notwendigen Load Balancing-Algorithmen existieren noch nicht.

### Einschränkung «fehlendes Accounting»

Im Moment gibt es noch keine Lösung dafür, wie auf dem Computing Grid bezogene Rechenleistung abgerechnet werden kann, so wie z.B. das Elektrizitätswerk den bezogenen Strom verrechnet (Accounting). Rechenkapazität gibt es jedoch nicht gratis: Computer müssen gekauft und unterhalten werden, zusätzlich wird eine aufwändige Infrastruktur zu deren Betrieb – vor allem für die Kühlung – benötigt. Daher kann, solange das Accounting-Problem nicht gelöst ist, das Computing Grid keine grosse Verbreitung finden.

### Das Zukunftspotenzial des Computing Grid

Das kurzfristige Potenzial des Computing Grid liegt im Lösen von einfach parallelisierbaren Problemen beispielsweise in der Hochenergiephysik, aber auch in der Biologie und der Pharmazie. Für das Lösen derartiger Probleme ist die notwendige Software-Infrastruktur schon heute vorhanden und funktioniert ohne Probleme.

Für enger gekoppelte Probleme, wie etwa Klimasimulationen, Wetterprognosen oder der Simulation von Vielkörperproblemen kann das Computing Grid längerfristig unter folgenden Voraussetzungen ein wertvolles Werkzeug werden:

- die erfolgreiche Weiterentwicklung der Message Passing Interfaces (MPI) – insbesondere eine bessere Abstimmung auf die gängigen Firewall-Konfigurationen;
- die Entwicklung von Load-Balancing-Algorithmen für gekoppelte Probleme, die insbesondere die Laufzeit von Internet-Paketen zwischen einzelnen Clustern berücksichtigen.

Sobald die Computing-Grid-Software fehlende Funktionalitäten liefern kann, wie das Accounting oder die Interaktivität, hat das Grid das Potenzial, eine weite Verbreitung zu finden. Es wird als wertvolles Werkzeug dienen für alle, die für ihre Arbeit Rechenkapazität benötigen.

### Angaben zum Autor

Dr. **Christian Häberli** hat an der Universität Bern Physik studiert. Während seiner Dissertationszeit am CERN war er im Teilprojekt «Trigger and Data Acquisition» des ATLAS-Experiments tätig. Seit 2004 arbeitet er als Assistent am Laboratorium für Hochenergiephysik der Universität Bern, LHEP. Er ist als Projektleiter verantwortlich für den Aufbau des Schweizer ATLAS Computing Grid.  
christian.haerberli@hep.unibe.ch

<sup>1</sup> Grid Café Website am CERN: <http://gridcafe.web.cern.ch/gridcafe/about.html>

<sup>2</sup> LHC Computing Grid: <http://lcg.web.cern.ch/LCG>

<sup>3</sup> NorduGrid Advanced Resource Connector: [www.nordugrid.org](http://www.nordugrid.org)

<sup>4</sup> Das Centro Svizzero di Calcolo Scientifico ist das Rechenzentrum der ETHZ in Manno bei Lugano. [www.cscs.ch](http://www.cscs.ch)

<sup>5</sup> Load Balancing ist die Beantwortung der Frage, welcher Rechenauftrag auf welchem Cluster innerhalb des Computing Grid ausgeführt werden soll.

<sup>6</sup> Large Hadron Collider: <http://cern.ch/lhc>

<sup>7</sup> ATLAS-Experiment: <http://atlasinfo.cern.ch/atlas>

<sup>8</sup> SWITCH ist der Internet Provider der Schweizer Hochschulen. [www.switch.ch](http://www.switch.ch)

## Computing Grid – entre la vision et la réalité

### La physique de haute énergie utilise le Computing Grid pour la simulation et l'analyse de collisions entre protons

Le nom de Computing Grid est inspiré de l'expression anglaise «Power Grid» couramment utilisée pour désigner le réseau électrique. Selon la vision du Computing Grid, les centres d'ordinateurs devraient proposer de la capacité de calcul de la même manière que les centrales électriques nous fournissent actuellement le courant. Ainsi, une capacité de calcul dépassant les possibilités de l'ordinateur individuel devrait être accessible à tous à un prix abordable, comme c'est actuellement le cas de l'énergie électrique. Au lieu d'une génératrice coûteuse à la maison, nous avons tous un raccordement au réseau électrique.