

Kapitel 2

BOINC ist Public-Resource Computing

BOINC (Berkeley Open Infrastructure for Network Computing) is a software system that makes it easy for scientists to create and operate Public-Resource Computing projects. It supports diverse applications, including those with large storage or communication requirements. PC owners can participate in multiple BOINC projects, and can specify how their resources are allocated among these projects.

David P. Anderson [4]

Das Public-Resource Computing (PRC) ist eine vielversprechende Möglichkeit, die im vorherigen Abschn. 1.3.2 erwähnten versteckten Rechenleistungen für wissenschaftliche Anwendungsgebiete nutzbar zu machen. Verschiedene Umsetzungen prägen das Bild von PRC-Projekten, folgende Möglichkeiten stehen als Grundgedanke bei der Verwendung von BOINC zur Verfügung:

Volunteer-Computing Durch das Volunteer-Computing, was soviel bedeutet wie *freiwilliges Rechnen*, melden sich Benutzer an ein oder mehreren Projekten an und spenden sozusagen ihre Rechnerressourcen den jeweiligen wissenschaftlichen Projekten und rechnen für diese.

Desktop-Grids Ein Desktop-Grid kann wie das Volunteer-Computing ausgeprägt sein. Allerdings handelt es sich bei solch einem Grid um ein eigenständiges Netzwerk, z. B. ein Firmennetzwerk oder das Campus-Netzwerk eines Institutes (Universität, Volkshochschule, Bibliothek). In diesem Fall wird auf den Rechnern der Firma oder dem Institut zusätzliche Arbeit verteilt. Sinnvoll erscheint das in dem Zusammenhang, wenn man bedenkt, dass die meisten Rechner sowieso viel Zeit im Wartezustand verbleiben und eher selten durchgehend auf Volllast arbeiten. Die Sekretärin oder der Pförtner müssen im Normalfall keine größeren Applikationen ausführen und daher können ihre Rechner zusätzlich für andere Aufgaben zusätzlich genutzt werden.

2.1 Anforderungen

PRC-Projekte fallen in die Kategorie der Grid-Computing-Installationen, allerdings muss ein solches Projekt zusätzliche Eigenschaften erfüllen [30]:

Granularität Das Internet ist ein sehr heterogenes Netz. Bandbreiten schwanken zwischen einigen GBit/s (nationale Backbones) bis hin zu 56 kBit/s (analoger Telefonanschluss). Die Datenpakete der aktuellen BOINC-Projekte haben Größen zwischen 340 kByte (SETI@home) und einigen MBytes (Rosetta@home).

Dies stellt natürlich ein Problem dar: Was machen Anwender mit einer eher niedrigen Übertragungsgeschwindigkeit, denn heutige DSL-Geschwindigkeiten sind noch immer nicht überall verfügbar. Es kann dann schon einmal vorkommen, dass das Herunterladen eines Arbeitspaketes länger als die Berechnung selber dauert, und dies sollte möglichst vermieden werden, weil die Anwender dadurch sicherlich frustriert sind. Ein Mechanismus für die Prüfung bzw. Messung der Netzleitung, um daraus zu ermitteln, welche Größe am geeignetsten für den Anwender ist, gibt es leider nicht. Diese Verantwortung tragen bisher noch die Administratoren eines BOINC-Projektes.

Diversität Die teilnehmenden Rechner sind sehr unterschiedlich. Um Windows-PCs, Macintosh-Rechner und diverse Unix-Maschinen nicht auszuschließen, müssen die BOINC-Projektbetreiber für jede dieser Plattformen die passende Client-Software bereitstellen. In der Regel werden mindestens eine Windows- und eine Unix-Variante für das Bearbeiten von Arbeitspaketen zur Verfügung gestellt. In vielen BOINC-Projekten fehlt teilweise einfach die Hardware, um eine Portierung vorzunehmen und zu testen.

Robustheit Die Client-Rechner unterliegen nicht der Kontrolle eines BOINC-Projektes. Man muss damit rechnen, dass übernommene Arbeitspakete nicht bearbeitet werden oder zurückgelieferte Berechnungsergebnisse falsch sind. Selbst Sabotage kann nicht ausgeschlossen werden. Der BOINC-Server erstellt deshalb für ein Arbeitspaket mehrere Aufteilungen und versendet diese an verschiedene Clients. Zur Überprüfung erhalten Clients außerdem regelmäßig Aufgaben, deren Berechnungsergebnisse bekannt sind. Zur Robustheit gehört auch, dass die Client-Software ein vorzeitiges Abschalten des Rechners „überlebt“ und nach dem Einschalten dort weitermacht, wo sie unterbrochen wurde. Dies wird durch das Erstellen von sog. Checkpoint-Dateien ermöglicht.

Attraktivität Die Berechnungssoftware stört den normalen Rechenbetrieb nicht. Auf Unix-Maschinen läuft sie mit niedriger Priorität im Hintergrund, für Windows-Rechner ist eine Bildschirmschoner-Variante erhältlich. Unterstützer müssen allerdings sowohl Zeit (Installation) als auch Geld (Online- und Stromkosten) investieren. Dazu sind die Leute umso eher bereit, je attraktiver das BOINC-Projekt erscheint. Für viele ist auch die wissenschaftliche Fragestellung interessant und beflügelt die Anwender, keine Kosten und Mühen zu scheuen, wenn die BOINC-Teilnehmer Potential an der Fragestellung erkennen. Noch attraktiver wird es, wenn die einzelnen Betreiber der BOINC-Projekte versprechen, dass die Teilnehmer mit den entsprechenden passenden Berechnungsergebnissen in den internationalen Forschungsberichten erwähnt werden.

2.2 Was ist umsetzbar?

Abbildung 2.1 zeigt die prinzipiellen Möglichkeiten, wie Arbeitsaufträge abgearbeitet werden können; dies ist auch davon abhängig, wie die einzelnen Arbeitspakete mit der Außenwelt kommunizieren. Die einzelnen Abarbeitungsmöglichkeiten

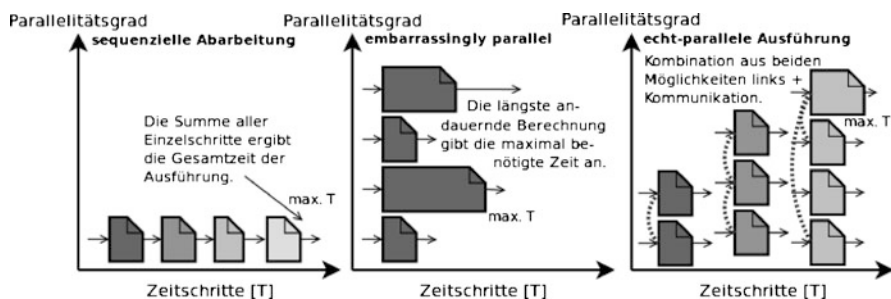


Abb. 2.1 Auf einem Rechencluster lassen sich die unterschiedlichsten Probleme lösen, abhängig vom Problem muss keine oder muss zwangsläufig eine Kommunikation zwischen den Rechenknoten stattfinden. Diese drei Konstellationen sind möglich: (1) sequenzielle Abarbeitung, (2) embarrassingly parallel und (3) echt-parallele Ausführung [32]

werden im Grad der Parallelität über die Zeit dargestellt und beschrieben. Je mehr Arbeitsaufträge übereinander angeordnet sind, desto mehr Arbeit kann in einem Zeitbereich gelöst werden. Von links nach rechts nimmt der Grad der Parallelität zu und auch die Anzahl der Arbeitspakete steigt.

Single-Core-Prozessoren arbeiten sequenziell und können immer nur einen Arbeitsauftrag zu einem bestimmten Zeitpunkt abarbeiten; diese werden nacheinander gestartet. Die *embarrassingly parallel* abzuarbeitenden Arbeitsaufträge sind absolut autonom und können ohne Kommunikation zur Außenwelt gestartet, gelöst und abgeschlossen werden. In diesem Fall können zehn Arbeitsaufträge tatsächlich in einem Zeitfenster gelöst werden, wenn zehn Rechner vorhanden sind, auch Multi-Core-Prozessoren können hier sehr rasch zu einer Lösung kommen. Der letzte Fall ist ein wenig komplizierter und muss in der Realität auch etwas genauer betrachtet werden. Die Abbildung zeigt mehrere parallel ausgeführte Arbeitsaufträge; die gestrichelten Linien sollen die Kommunikation zwischen diesen darstellen. Was ist allerdings zu tun, wenn ein Arbeitspaket aus unbekannten Gründen niemals ein Ergebnis liefert, aber von anderen Aufträgen für die weiteren Berechnungsschritte benötigt wird? In diesem Fall liegt ein klassischer *Deadlock* vor (zu Deutsch Verklemmung).

Für unsere Betrachtungen in diesem Buch ist diese Problematik nicht von Belang. Mit BOINC können zwar alle drei möglichen Lösungswege umgesetzt werden, allerdings ist BOINC nur für die ersten beiden Prinzipien gedacht und konzipiert.

2.3 Einsatzbereiche

Wir haben schon erwähnt, dass BOINC generell für Probleme verwendet werden kann, bei denen keine Kommunikation zwischen einzelnen Rechnerknoten nötig ist und die Aufgabenstellungen eine hohe Granularität besitzen. Dies können Anwendungen sein, bei denen große Datenmengen bearbeitet werden müssen oder lange dauernde Berechnungen, die viele Iterationen durchführen:

Film-Sequenzen/Rendering Eine einfache und relativ schnell umzusetzende Problemstellung ist die Bearbeitung von Film-Sequenzen oder das Rendern von Einzelbildern. Eine Beispielapplikation für die Bearbeitung von Film-Sequenzen finden Sie in Kap. 12 [143]. Wenn ein kompletter Film zum Beispiel mit bestimmten Filtern bearbeitet werden soll, so kann der Film in einzelne Film-Sequenzen umgewandelt werden. Diese Film-Sequenzen können dann in einzelne Archive gepackt werden, um daraus – im BOINC-Bereich – Workunits zu erstellen.

Parameterstudien In Kap. 14 zeigen wir, wie man Parameterstudien mit Hilfe von Legacy-Applikationen – u. a. COMSOL Multiphysics – erstellen kann. Bei Parameterstudien kann man ein Simulationsmodell nehmen und mit unterschiedlichen Parametersätzen füttern, z. B. kann die Geschwindigkeit eines Flugzeugs unter bestimmten Luftdruckverhältnissen variiert werden.

Bionische Methoden Ameisenalgorithmen, oder allg. ausgedrückt Algorithmen der Schwarmintelligenz bzw. Partikelschwarmoptimierung, fallen in den Bereich der Bionischen Methoden der Optimierung und suchen das Optimum einer Problemstellung [5]. Dabei wird der Versuch unternommen, ein Minimum oder Maximum zu finden oder zumindest so lange eine Optimierung durchzuführen, bis ein vorher definiertes Abbruchkriterium erfüllt ist. Letzteres dient dazu, dass die Berechnung nicht endlos lange läuft, wenn sie niemals gegen ein Ergebnis konvergiert. Die Durchführung solcher Algorithmen beinhaltet oft mehrere tausend Iterationen mit den Ergebnissen aus der vorherigen Iteration und können daher perfekt als eigenständige Aufgaben auf Einzelrechnern ausgeführt werden.

Fahrpläne und Stundenplan Die automatische Erstellung von Fahrplänen (Straßenbahn, Busse, etc.) oder von Stundenplänen (Müllabfuhr, Schule, etc.) kann durch bionische Methoden oder durch andere Ansätze, allerdings wieder in mehreren tausend Iterationen, gelöst werden. Die Ergebnismenge wird am Ende auf das Vorhandensein einer optimalen Lösung durchsucht.

Stochastik Verhält sich wie die beiden vorherigen Aufgabenstellungen.

2.4 Forschergruppen und ihre Erfolge

Durch Public-Resource Computing (PRC) wurden schon Erfolge gefeiert, die sicherlich auch auf High-Performance-Systemen (HPC) möglich gewesen wären, allerdings nicht mit solch einem geringen Teil an Kosten. Die University of Westminster [180] in London hat einen virtuellen Supercomputer initialisiert; basierend auf BOINC. Die geschätzten Einsparungen belaufen sich auf rund £500.000 in vier Jahren.

2.4.1 SETI@home: Sind wir allein im All?

SETI@home ist das zweite Projekt, das als PRC-Projekt (Public-Resource Computing) aufgebaut wurde, und hat sich der Suche nach außerirdischen Radiowellen-signalen für den Beweis der Existenz von intelligentem Leben außerhalb der Erde

verschrieben. Der Beweis für die Existenz solchen Lebens ist bis heute leider noch nicht geliefert worden. Gegründet wurde das Projekt Mitte 1999, und es ist heute neben den zwei weiteren Projekten MilkyWay@home [150] und Einstein@home (vgl. Abschn. 2.4.2) eines der Projekte, welche sich mit aufgefangenen Signalen aus dem Weltall beschäftigen. Nichtsdestoweniger hat SETI@home den ersten Schritt gemacht und die erforderliche Software geliefert, so dass weitere Wissenschaftler ihre aufwendigen Berechnungen und ihr Datenaufkommen an Freiwillige verteilen können und ein wissenschaftlicher Fortschritt ermöglicht wird. Durch SETI@home wurde das PRC-Rechnen quasi salonfähig.

2.4.2 *Einstein@home: Weltraumforschung*

Das Projekt Einstein@home [120] beschäftigt sich mit der Suche nach einem Beweis der von Albert Einstein im Jahre 1905 aufgestellten speziellen Relativitätstheorie. Diese Relativitätstheorie besagt, dass wir in einem Universum gefüllt mit Gravitationswellen leben. Explodierende Sterne, zusammenstoßende schwarze Löcher und andere gewaltige Zerstörungen erschaffen solche Gravitationswellen, welche den Raum und die Zeit beeinflussen [120]. Bisher wurde noch kein Beweis für das Vorhandensein von Gravitationswellen durch Einstein@home geliefert.

Dies ist nicht die einzige Aufgabe von Einstein@home; seit März 2009 wird auch nach Radiopulsaren gesucht [72]. Für diese Suche werden Daten vom Arecibo-Observatorium genutzt – eventuell kennen Sie das dazugehörige Radioteleskop aus dem James-Bond-Film *GoldenEye* und dem Kinofilm *Contact* [152]. Es handelt sich dabei um ein Teleskop mit einem beachtlichen Durchmesser von 300 m und einer Höhe von 150 m. Radiopulsare sind die extremsten Objekte im Universum, es handelt sich dabei um rasend schnell drehende Neutronensterne, die entweder einen weiteren Neutronenstern oder ein schwarzes Loch umkreisen. Die Verwendung von BOINC und der damit nutzbaren Computerleistung, von den Anfängen im Jahr 2009 mit rund 200.000 Teilnehmern, ermöglichte erst die Suche nach sich noch schneller drehenden Radiopulsaren. Das hängt damit zusammen, dass mehr Daten analysiert werden müssen, um die Suche zu verfeinern; je feiner, desto schnellere Radiopulsare können entdeckt werden. Bis heute¹ sind zehn neue Radiopulsare von Einstein@home entdeckt und bestätigt worden [121]. Als Freiwilliger am Einstein@home-Projekt wird man namentlich in der Danksagung der wissenschaftlichen Publikation erwähnt. Diese ungeschriebene Regel in besonderen Maße wird von den BOINC-Teilnehmern schon fast erwartet, aber auch von den BOINC-Projekten respektiert. Die Motivation für die Teilnahme an den jeweiligen Projekten wird dadurch beflügelt.

¹ Stand am 27. September 2011.

2.4.3 *Visu@IGrid: Model-Driven-Engineering*

Mit Hilfe der Unified Modeling Language (UML) und der Möglichkeit, sogenannte UML-Profile zu erstellen, soll in näherer Zukunft der Prozess von der ersten Idee eines BOINC-Projekts bis hin zu einer fertigen ausführbaren BOINC-Installation und Lösung von Problemstellungen soweit es geht durch Modelle vereinfacht werden. Erste Schritte zeigen, dass dieser Ansatz vielversprechend ist [17–19]. Aufbauend auf der Grundstruktur der BOINC-Programmierschnittstelle (s. Kap. 7) und der BOINC-Infrastruktur (s. Abschn. 4.1) wird mit Hilfe eines Top-down-Ansatzes die Struktur analysiert und in ein UML-Modell umgewandelt, so dass aus diesem Modell folgende Komponenten automatisch durch einen Code-Generator generiert werden [52, 53]:

Entwicklungsumgebung Eine IDE (Integrated Development Environment) wird mit Hilfe des spezifischen UML-Modells generiert, so dass nur Funktionalitäten und Schnittstellen genutzt werden können, welche benötigt und durch BOINC zur Verfügung gestellt werden. Es handelt sich dabei um eine IDE, die der Domäne eines Entwicklers angepasst ist und nur kontextbezogene Modellierungselemente beinhaltet. Dadurch kann die Entwicklung auf ein Minimum der Entwicklungszeit und Wartung reduziert werden. Kleinere Änderungen bei der Konfiguration eines BOINC-Projektes oder einer Applikation werden automatisch geprüft und aufgelöst, so dass keine Fehler aufkommen können, nur weil womöglich eine weitere zu ändernde Konfiguration vergessen wurde abzuändern.

Applikation Die wissenschaftliche Applikation wird durch Standarddiagramme der UML modelliert, z. B. Sequenzdiagramme oder Statusdiagramme. So kann eine grafische Lösung einer Applikation erfolgen und externe Werkzeuge mit Unterstützung von XMI (XML Metadata Interchange) – um Daten unter verschiedenen Entwicklungsumgebungen austauschen zu können [54] – genutzt werden, u. a. IBMs Rational Rhapsody [136].

2.4.4 *Dem Dieb auf der Spur*

Mit Hilfe von BOINC konnten zwei Computer wiedergefunden werden. In den Jahren 2007 [76, 127, 167] und 2010 [168] wurde jeweils ein Laptop entwendet. Im ersten Fall hatte der Ehemann der Besitzerin Kimberly Melin eine Installation vom BOINC-Manager durchgeführt und SETI@home als Projekt hinzugefügt. Melin hat daraufhin die Datenbankeinträge von SETI@home überprüft und konnte nach einer Weile sehen, dass neue Berechnungsergebnisse abgespeichert wurden. Bei jeder Kommunikation mit einem BOINC-Projekt wird die IP-Adresse mitgesendet und protokolliert. Gut für Melin, denn sie konnte daraufhin die IP-Adresse an die örtliche Polizei weitergeben, und die konnte in Folge dessen die reale Adresse des verwendeten Internetanschlusses ausfindig machen. Der Rechner wurde wiedergefunden und glücklicherweise waren noch alle Daten von Melin vorhanden, plus zusätz-

liche Musikdateien. Dabei handelte es sich wohl um Dateien des Diebs. Die Suche, das Auffinden und Zurückbringen dauerten rund zweieinhalb Wochen.

Im zweiten Fall betraf es einen BOINC-Benutzer aus Deutschland. Ein iMac wurde in einer Nacht aus den Räumlichkeiten einer Firma entwendet, dies wurde am nächsten Morgen bemerkt und der Kriminalpolizei gemeldet. In den nächsten Tagen hat der Rechner sich fünfmal beim WUProb@home-Projekt [191] mit der Meldung von Berechnungsergebnissen gemeldet. Das Diebstahlopfer hat das verfolgt und die IP-Adresse an die Kriminalpolizei weitergeleitet. Daraufhin hat diese einen Durchsuchungsbefehl beantragt und hat das Notebook sicherstellen können. Zwischen der Feststellung des Diebstahls, der Sicherstellung und Rückgabe des Notebooks an den Besitzer sind rund neun Tage vergangen. Neben der wissenschaftlichen Tätigkeit von BOINC-Projekten dienen diese anscheinend auch wunderbar der Detektivarbeit.

BOINC

Hochleistungsrechnen mit Berkeley Open Infrastructure
for Network Computing

Ries, C.B.

2012, XIV, 370 S. 82 Abb., Hardcover

ISBN: 978-3-642-23382-1