

Exposé zur Studienarbeit – „Drahtlos in den Wolken – Sensor Network Simulation mit Hilfe von Cloud-Computing-Architekturen“

(Feb 2009 – Juni 2009)

Betreuung: Timo Gläßer, Prof. Ulf Leser
Bearbeitung: Erik Dießler

1. Motivation

Drahtlose Sensornetzwerke haben in den letzten Jahren eine rasante Entwicklung zu verzeichnen. Die Vernetzung von Maschinen und Haushaltsgeräten, Zählerablesung mit Hilfe drahtloser Sensornetzwerke, die Idee Lärmpegel und die Geräuschentwicklung in Städten und städtischen Großräumen [SV07] zu überwachen sowie die Nutzung im Bereich Katastrophenmanagement [EDMI07] zeigen, dass die Möglichkeiten des Einsatzes sehr vielseitig sind.

Diese Entwicklungen, die sinkenden Kosten sowie die verbesserten technischen Möglichkeiten, wie beispielsweise längere Lebensdauer, effizientere Kommunikation und Selbstorganisation, machen den Einsatz drahtloser Sensornetzwerke im Bereich von vielen Tausend bis hin zu Millionen einzelner Sensorknoten denkbar. So könnten ganze städtische Großräume in erdbebengefährdeten Gebieten mit Sensoren versehen werden, um im Katastrophenfall schnell Informationen über das Ausmaß der Schäden zu erhalten.

Trotz der geringeren Kosten pro Sensor, kann der Aufbau großer Sensornetzwerke nicht zu unterschätzende Kosten verursachen und Risiken bergen. Es gilt also, mögliche Probleme, die zu Fehlschlägen führen könnten, schon im Vorfeld zu eliminieren. Eine Möglichkeit, sowohl das Verhalten als auch mögliche Probleme abschätzen zu können, bieten Simulatoren wie JiST/SWANS [JS06].

Die Simulation großer Sensornetzwerke mit beispielsweise 1.000.000 Sensorknoten verlangt, bei Verwendung der aktuellen Simulatoren, nach umfangreichen Hardwareressourcen, wenn das Problem in akzeptabler Zeit berechnet werden soll.

Swans benötigt laut [B08] etwa vier Stunden zur Simulation eines Netzwerkes, das aus 400.000 Sensorknoten besteht. Dabei werden, unter Verwendung einer 2.8GHz Maschine mit 2GB Speicher, das Zone Routing Protocol eingesetzt und zehn Nachbarknoten pro Sensor simuliert.

Es würden also zusätzliche Kosten für die Erweiterung der Infrastruktur auf Hardwareebene anfallen, um das Problem zeitnah zu berechnen.

An dieser Stelle bietet sich die Idee des Cloud Computing an. Es ermöglicht, bei nur verhältnismäßig geringen Kosten, zusätzliche Ressourcen, wie Rechenzeit und Speicher, entsprechend den Bedürfnissen des zu bearbeitenden Problems zu nutzen.

Könnte somit nicht eine Kombination aus vorhandenen Simulatoren und Cloud Computing einen Lösungsansatz darstellen, mit dessen Hilfe sowohl große und komplexe Sensornetzwerke zeitnah simuliert, als auch die Kosten, trotz hoher Hardwareanforderungen, gering gehalten werden können?

2. Zielstellung

An diesem Punkt soll die Studienarbeit ansetzen. Ziel ist es, zu untersuchen, ob und in welchem Rahmen eine existierende Simulationssoftware für drahtlose Sensornetzwerke an eine Cloud Architektur angepasst werden kann.

Im Sinne der Bestimmung von guten und sinnvollen Verteilungsstrategien soll zunächst die Prüfung, wofür die meisten Ressourcen und Rechenzeit durch den Simulator verbraucht werden, im Vordergrund stehen. Auf Basis der gewonnenen Informationen können Ansätze zur Lösung des Problems entwickelt werden.

So könnte es günstiger sein, nur einzelne Teilberechnungen auszulagern, wie zum Beispiel das Bestimmen der erreichbaren Nachbarknoten in dynamischen Sensornetzwerken. Ist dies sinnvoll und möglich, so könnte MapReduce [DG04] an dieser Stelle zum Einsatz kommen.

Alternativ könnte das Netzwerk in Teile zerlegt werden, deren Simulation dann parallel erfolgt.

Letztlich müssen die entwickelten Ideen auf ihre Machbarkeit, bezüglich der Auslagerung in eine Cloud Architektur, geprüft werden. Danach gilt es den Ansatz auszuwählen, welcher den meisten Erfolg verspricht. Idealerweise werden somit die Voraussetzungen für einen Prototypen geschaffen, mit dessen Hilfe die Effizienz des Ansatzes gezeigt werden könnte.

3. Vorgehensweise

Als Simulationssoftware soll Jist/Swans eingesetzt werden. Als Cloud Architektur könnte zum Beispiel Amazon EC2 [AZ07] zum Einsatz kommen.

In einem ersten Schritt soll eine Einarbeitung in die Thematik erfolgen. Hierbei stehen der Simulator und das Konzept MapReduce im Vordergrund. Hadoop [HAD08] ist eine OpenSource Implementierung des google-Konzeptes mit MapReduce und somit ebenfalls interessant für das Problem. Außerdem gilt es, die Möglichkeiten und die Funktionsweisen von Amazon EC2 zu erarbeiten.

Der zweite Teil der Arbeit umfasst in erster Linie theoretische Betrachtungen. Das Ziel ist es, Ansätze zur Lösung der Problemstellung zu finden, diese zu bewerten und jenen, der am meisten Erfolg verspricht, auszuarbeiten.

Im letzten Schritt erfolgen die Ausarbeitung und Verschriftlichung der Ergebnisse.

4. Verwandte Arbeiten

Die Suche nach verwandten Arbeiten brachte keine direkten Ergebnisse. Es scheint also bisher keine Veröffentlichungen zu geben, welche sich mit dem Thema der Studienarbeit auseinandersetzen.

Es gibt jedoch durchaus Beispiele, die zeigen, dass Cloud Computing und insbesondere das MapReduce-Konzept in Versuchen Anwendung finden, deren Ziel es ist, Simulatoren performanter zu gestalten.

So gibt es Arbeiten im Bereich Earthquake Simulation [SRTL+08], welche die Möglichkeiten von MapReduce untersuchen. Allerdings wird in diesem Falle zu dem Schluss gekommen, dass ein überwiegender Teil von MapReduce nicht sinnvoll für die zu bearbeitenden Probleme einsetzbar war.

Eine weitere Arbeit beschäftigt sich allgemeiner mit den Möglichkeiten, die MapReduce im Zusammenhang mit Hadoop für vorhandene HighPerformance Probleme bietet. Hierbei liegt einer der Schwerpunkte auch auf den Datenmengen und den damit verbundenen Transferkosten in Clustern [MSBJ+08]. Es werden wissenschaftliche Anwendungen aus den Bereichen Astrophysik und Bioinformatik untersucht, um zu prüfen, inwieweit Hadoop und MapReduce verwendbar und deren Einsatz von Vorteil ist.

Nicht mehr sehr eng mit dem Thema verbunden gibt es eine weitere Arbeit, die sich mit den Möglichkeiten und der Performance von Amazon EC2 im Bereich von High Performance Problemen beschäftigt [EH08] und dies an Hand der Simulation von klimatischen Modellen prüft. Als Ergebnis wird, im Vergleich zu Clustern, die geringere Zuverlässigkeit und nicht ganz vergleichbare Performanz erkannt, jedoch die geringeren Kosten sowie die hohe Skalierbarkeit hervorgehoben.

5. Literatur

- [AZ07] *Projekt Webseite: <http://aws.amazon.com/ec2/>, 2007*
- [B08] Rimon Barr – SWANS – Scalable Wireless Ad hoc Network Simulator User Guide – *Projekt Website: <http://jist.ece.cornell.edu/docs.html>*
- [DG04] Jeffrey Dean and Sanjay Ghemawat – MapReduce: Simplified Data Processing on Large Clusters, Google, Inc. 2004, In *Proceedings of the 6th Symposium on Operating Systems Design and Implementation, December 6–8, 2004, San Francisco, CA*
- [EDMI07] EDIM – Earthquake Disaster Information System for the Marmara Region, Turkey, *Projekt Webseite: <http://www.cedim.de/EDIM.php>, 2007*
- [EH08] Constantinos Evangelinos and Chris N. Hill – Cloud Computing for parallel Scientific HPC Applications: Feasibility of running Coupled Atmosphere-Ocean Climate Models on Amazon’s EC2. [Extended Abstract] , Department of Earth, Atmospheric and Planetary Sciences, MIT 77 Massachusetts Ave. Cambridge, MA 02139, USA, 2008
- [HAD08] Hadoop – *Projekt Webseite: <http://hadoop.apache.org/core/>, 2008*
- [JS06] JiST / SWANS – Java in Simulation Time / Scalable Wireless Ad hoc Network Simulator – *Projekt Webseite: <http://jist.ece.cornell.edu/>, 2006*
- [MSBJ+08] Grant Mackey, Saba Sehrish, John Bent, Julio Lopez, Salman Habib, Jun Wang, Introducing Map-Reduce to High End Computing, University of Central Florida, Los Alamos National Lab, Carnegie Mellon University, 2008, veröffentlicht auf 3rd *Petascale Data Storage Workshop Supercomputing '08 (<http://www.pdsi-scidac.org/events/PDSW08/index.html>)*
- [SIDnet] SIDnet-SWANS – A Simulator and Integrated Development Platform for Sensor Networks Applications,
Projekt Webseite: <http://www.ece.northwestern.edu/~ocg474/SIDnet.html>, 2008
- [SRTL+08] Steven W. Schlosser, Michael P. Ryan, Ricardo Taborda, Julio Lopez, David R. O’Hallaron, Jacobo Bielak, Materialized community ground models for large-scale earthquake simulation, Intel Research Pittsburgh, Carnegie Mellon University, 2008, In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing, 2008, Austin, Texas November 15 - 21, 2008*

- [SV07] Silvia Santini, Andrea Vitaletti - Wireless Sensor Networks for Environmental Noise Monitoring. *In 6. Fachgespräch Sensornetzwerke Seiten 98-101, Distributed Systems Group RWHT Aachen University, 2007*
- [TLP05] Ben L. Titzer, Daniel K. Lee, Jens Palsberg – Avrora: Scalable Sensor Network Simulation with Precise Timing, UCLA, Cornell University, 2005, *In Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on Publication Date: 15 April 2005 On page(s): 477- 482*