



Diplomarbeit “Internet Coordinates”

Beginn: 14. Juni 2004

Ende: 13. Oktober 2004

Dieses Dokument gibt den Rahmen der Diplomarbeit von Daniel Bruggesser vor. Abweichungen oder Änderungen sind in gegenseitiger Absprache möglich.

Es gibt in letzter Zeit immer mehr Arbeiten (z.B. [1, 2]), die behaupten, dass man das ganze Internet auf eine einfache Metrik abbilden kann (zum Beispiel jeden Host/Router auf einen Punkt in einer Ebene), und dass diese Metrik dann die wirklichen Delays im Internet gut wiedergibt (d.h. dass z.B. die Distanz zweier Hosts/Router in der Ebene ziemlich genau ihrem echten Delay im Internet entspricht.)

In einem ersten Ansatz wollen wir in dieser DA folgende Punkte untersuchen (die Zeitangaben sind als ungefähre Richtwerte anzusehen):

1. Zusammenstellung/Zusammenfassung aller Arbeiten auf diesem Gebiet.
2. Kritische Würdigung dieser Arbeiten. [1-2 Wochen bis hier]
3. Wie kriegt man möglichst viele gute Messungen der echten Delays im Internet? Kann man selbst von verschiedenen Orten aus Messungen vornehmen (über web services wie traceroute.org, looking glass, etc.), oder gibt es dazu schon Datenbanken, oder kann man z.B. Planet Lab (siehe auch [3]) benutzen, oder was auch immer? Siehe auch: NLANR AMP (Active Measurement Project), CAIDA Skitter und Sockeye. [1 Woche]
4. Wie bettet man diese Knoten in eine Metrik ein? In welche Metrik? Soll man sie gleich auf die Welt-(Kugel)-Oberfläche abbilden, oder gibt es etwas besseres? Oder gibt es gar etwas viel Einfacheres, das auch schon fast so gut ist? (Wir haben bei uns in der Gruppe jede Menge Know-How, was solche ”guten” Abbildungen betrifft.) Legt man gewisse Punkte einfach fest (kennt man z.B. ihre Koordinaten auf der Weltkugel?), oder sind alle unbekannt (Stichwort Virtuelle Koordinaten)?
5. Wie kann man die eigene Arbeit evaluieren? Wie gut sind die Koordinaten nun wirklich? (ähnlich wie Punkt 3.)
6. Anwendungen: Ist das alles mal gemacht kann man viele tolle Dinge damit anstellen...
 - Die DNS-Rootserver als Anycast-Server so hinstellen, dass die maximale oder durchschnittliche Latenz optimiert wird.
 - Man könnte P2P/Overlay-Netze basteln, die leicht und ohne vorheriges Ping immer rausfinden welche Kopie am nächsten ist.
 - Interessant wäre natürlich auch eine Art von Load Balancing, oder ganz generell einen Dichtegraphen zu finden, der einem sagt wo in der Welt die meisten P2P-User, oder die meisten Webseiten oder was auch immer sind (mit z.B. den Gnutella-Daten von Bea).
 - Vielleicht könnte man auch mal das Verhältnis zwischen AS und Hosts/Routern genauer untersuchen.
 - Für ASe wäre es sicher ein nützlicher Input zu Routing- und Peer-Auswahlverfahren.
 - Eventuell eignen sich die Koordinaten auch für folgendes Problem: viele Provider versuchen, ihre Links gleichmässig auszulasten. Dadurch ergeben sich aber potentiell längere Pfade für die Routen. Wie steht es mit dem daraus entstehendem Stretch (im Vergleich zum (geographischen) Optimum)?

Es gibt auch alternative bzw. ergänzende Ideen, die man verfolgen kann.

- Bei den meisten Arbeiten existiert die implizite Annahme, dass die Routen symmetrisch sind, weil die kartesianische (oder was auch immer) Entfernung A-B ja immer gleich der Entfernung B-A ist. Ist diese Annahme gerechtfertigt oder bedarf es hier genauerer Untersuchung?
- Eigene Ideen.

Literatur

- [1] T.S.E. Ng, H. Zhang. Predicting Internet Network Distance with Coordinates-Based Approaches. *INFOCOM 2002*.
- [2] L. Tang, M. Crovella. Geometric Exploration of the Landmark Selection Problem. *PAM 2004*.
- [3] S. Banerjee, T. Griffin, M. Pias. The Interdomain Connectivity of PlanetLab Nodes. *PAM 2004*.

Allgemeines

- Viele der aktuellen Papers sind online, entweder bei den Autoren selber (via Google), oder aber auf anderen Seiten wie CiteSeer oder DBLP verfügbar. Es gibt aber auch Bereiche, für die man zahlen muss (ACM Portal, IEEE xplore, SpringerLink), worauf man aber mit den ETH Rechnern (basierend auf IP Adresse) freien Zugriff hat. Wenn du dort was brauchst, mail mir einfach den link und ich schick dir das Dokument.
- Selbstständiges Arbeiten ist Voraussetzung.
- Es ist eine Zwischen- und eine Schlusspräsentation vorgesehen (20-30 Minuten).
- Der schriftliche Teil der Arbeit umfasst zwei Dokumente:
 - Einen Bericht (30 bis 50 Seiten, Sprache wählbar), welcher über die Arbeit und die Resultate Auskunft gibt. Dieser Bericht soll unter anderem auch eine kritische Beurteilung der eigenen Arbeit enthalten.
 - Einen Forschungsbericht (10 Seiten, in Englisch), welcher die Arbeit und die Resultate kompakter und unter einem wissenschaftlichen Gesichtspunkt darstellt.

Kontaktpersonen

- | | |
|----------------------|---|
| 1. Regina O'Dell | <code>bregina@inf.ethz.ch</code> , HRS G8, phone +41 1 63 24776 |
| 2. Roger Wattenhofer | <code>wattenhofer@inf.ethz.ch</code> , HRS G5, phone +41 1 63 26312 |
| 3. Simon Leinen | <code>simon@limmat.switch.ch</code> |