

Diplomarbeit der Zürcher Hochschule Winterthur **Swiss Internet Analysis 2002**

Kurzbeschreibung der Arbeit

Die Internetlandschaft Schweiz soll vom technischen Standpunkt aus mittels einer Primärerhebung umfassend analysiert und dargestellt werden.

Über zwei Bereiche sollen Zahlen vorgelegt werden können:

1. Typ, Verteilung und Nutzung verschiedener Internetressourcen (Domains, Name Server, Web Server, Mail Exchanges, etc.)
2. Netzwerkperformance (aufgrund einer statistischen Auswertung von TCP-Verbindungsparametern)

Die Rohdaten werden aus dem Internet gewonnen und durch zu entwickelnde Tools gemäss vorgegebener Metriken verdichtet.

Arbeitsdauer: 2 Monate; 6. September – 28. Oktober 2002

Ziel und Nutzen der Diplomarbeit

- - Repräsentativer Überblick über das Schweizer Internet
- - Daten zur Optimierung von Routern, Servern und Betriebssystemen (unter anderem Linux/FreeBSD Kernel IP Stack)
- - Auswertung und Analyse-Tools werden der Öffentlichkeit zur Verfügung gestellt

Vorstellung der Studenten

Olivier Müller und Daniel Graf sind Studenten an der Zürcher Hochschule in Winterthur (ZHAW). Sie sind im dritten und letzten Studienjahr des Studiengangs „Kommunikation und Informatik“ und werden sich mit dieser Diplomarbeit befassen.



Olivier MÜLLER

Hueberstrasse 9
8304 Wallisellen

Tel.: 01 883 19 41
Gsm: 079 625 96 25

om@omnis.ch
PGP Key: 0x0E84D2EA



Daniel GRAF

Trottenweg 1
8545 Rickenbach

Tel.: 052 337 30 18
Gsm: 079 688 77 83

grafdani@zhwin.ch

Verantwortlicher Dozent

Prof. Hans Weibel, Dozent für Kommunikations- und Nachrichtentechnik an der Zürcher Hochschule Winterthur. Kontakt: wei@zhwin.ch - Homepage: <http://www.zhwin.ch/~wei>

Zürcher Hochschule Winterthur

Die Zürcher Hochschule Winterthur ist Mitglied der Zürcher Fachhochschule. Sie ist aus der Fusion des 1874 gegründeten ersten Technikums der Schweiz, dem ehemaligen Technikum Winterthur, und der ebenfalls ältesten HWV der Schweiz, der Zürcher Höheren Wirtschafts- und Verwaltungsschule Winterthur, hervorgegangen. Die ZHW ist mit 12 verschiedenen Studiengängen und rund 2100 Studierenden die grösste Mehrsparten-Fachhochschule der Schweiz. Der praxisorientierte Unterricht basiert auf einer intensiven Zusammenarbeit mit der Industrie und der Wirtschaft.

Firmenpartner

IX Europe Telehouse Facilities AG

Herr André Oppermann, VR Präsident, oppermann@tix.ch

IX Europe Telehouse wird die Netzwerk- und Scan-Infrastruktur zur Verfügung stellen, sowie Testservers, Know-how und Räumlichkeiten.

Unterstützende Institutionen

- - **SWITCH** The Swiss Education & Research Network
- - **BAKOM** Bundesamt für Kommunikation, Abteilung Statistik
- - **ZHW** Zürcher Hochschule Winterthur

Vorgehensweise

- - **Phase 1** Sammlung der Rohdaten (Logfiles)
- - **Phase 2** Rohdaten mit einem Parser analysieren und in Datenbank speichern
- - **Phase 3** Daten auswerten
- - **Phase 4** Berichterstellung und Publikation

Was wir benötigen

Teil A: Domain Analyse

Die Daten sind bereits vorhanden (Quelle: SWITCH).

Teil B1: Web-Verkehr Analyse

Für diesen Teil der Arbeit benötigen wir rohe Webserver-Logfiles von einerseits populären und andererseits providerunabhängigen Schweizer Websites.

Idealerweise sind Logs vom Typ „combined“ (zum Beispiel mit Apache Web Server) zu verwenden:

```
127.0.0.1 - - [05/Aug/2002:18:58:58 +0200] „GET /~om/phpMyAdmin/main.php3 HTTP/1.0“ 200 15890
„http://127.0.0.1/~om/phpMyAdmin/“ „Lynx/2.8.4rel.1 libwww-FM/2.14“
```

Diese Logs enthalten folgende Informationen:

- - Datum, Zeit
- - Client Hostname oder IP Adresse
- - Antwort Code (200, 401, etc.)
- - Grösse
- - User Agent

Logs in anderen Formaten sind nach gewissen Anpassungen und Konvertierungen, die wir durchführen könnten, ebenfalls nutzbar.

Teil B2: E-Mail-Verkehr Analyse

Für die Mail Exchange Analyse benötigen wir Logfiles von den Mailservern einiger wichtiger Provider. Diese Daten werden aus Datenschutzgründen „anonymisiert“ (Teil vor dem „@“ löschen).

Beispiel von gmail Mail-Server Logfiles:

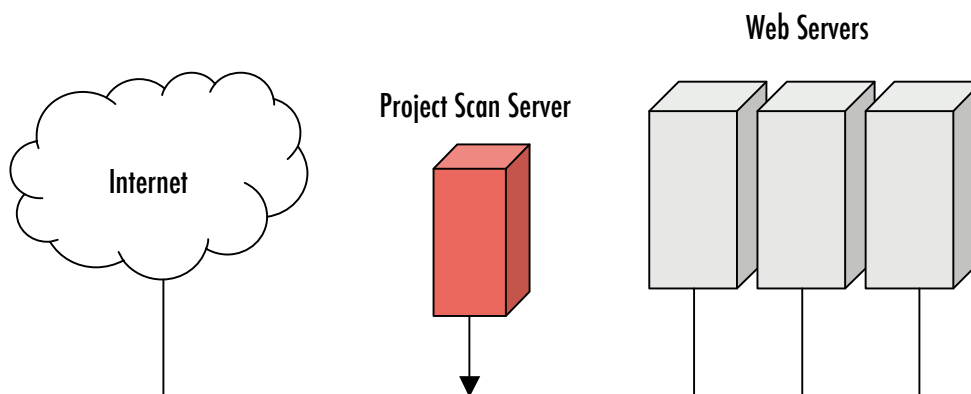
```
@400000003d7c61590defe8c4 info msg 257210: bytes 2825 from <XXXXX@eValueMail.com> qp 24368 uid 505
@400000003d7c61590fc047ac starting delivery 22101: msg 257210 to local XXXXX@praevis.ch
@400000003d7c615a065c0b4c info msg 257210: bytes 5437 from <XXXXX@Topjobs.ch> qp 24394 uid 505
@400000003d7c615a083d851c starting delivery 22102: msg 257210 to local XXXXX@telehouse.ch
@400000003d7c615c2f7d5364 info msg 257210: bytes 6924 from <XXXXX@listserv.heise.de> qp
@400000003d7c615c30e79d7c starting delivery 22103: msg 257210 to local XXXXX@omnis.ch
```

Relevante Informationen:

- - Datum, Zeit
- - Absender und Empfänger Domain
- - Grösse
- - Antwort Code (falls vorhanden)

Teil C: TCP Analyse

Der zweite Teil der Diplomarbeit verlangt den Einsatz von einem Scan-Server, der am gleichen Ethernet Anschluss wie der Webserver angeschlossen wird. Mit `tcpdump` werden während einer definierten Zeitspanne (idealerweise eine Woche) die TCP Headers zur späteren Auswertung gespeichert. Es werden also nur die TCP/IP Verbindungsmerkmale studiert: Der Inhalt von TCP Paketen (Webserver Daten, Cookies) wird weder registriert noch betrachtet.



Alle gesammelten Daten (aus verschiedenen Quellen) werden vereinheitlicht (überall gleiche Formate) und in einen gemeinsamen „Pool“ zusammengeführt. Erst dann werden die Daten zusammen analysiert. Es werden keine Einzelauswertungen über die Datenlieferanten erstellt.

NDA – Non Disclosure Agreement

Mit allen teilnehmenden Firmen wird ein NDA unterzeichnet. Somit haben sie die Sicherheit, dass die einzelnen Daten und Logfiles niemals veröffentlicht werden. Nur die Auswertung von allen Daten *zusammen* wird publiziert.

Am Schluss der Studie werden selbstverständlich auch alle Rohdaten und Logfiles gelöscht.

Dokumente / Referenzen

Support Letter Switch:

SWITCH
The Swiss Education & Research Network

Zürich, 22.8.2002

Unterstützung für die Diplomarbeit ‚Swiss Internet Analysis‘

Seit den Anfängen des Internet besteht ein Mangel an statistischen Daten. Neben der Auslastung und Performance von Übertragungsnetzen möchte man erfahren, wieviele Nutzer das Internet hat und wie es benutzt wird. Zur Frage der Nutzung: Das Internet Software Consortium publiziert periodisch den Internet Domain Survey auf [1], welcher bisher als einziger internationaler Report angesehen werden musste und der mittlerweile einen Zeitraum von mehreren Jahren abdeckt. Auch neuere Firmen wie Domain Worldwide [2] und Registrar Stats [3] publizieren Statistiken, die über die Verwendung des Internet Aufschluss geben und Zooknic veröffentlich Karten über dessen Verbreitung auf den verschiedenen Kontinenten (How many On-line), siehe [4].

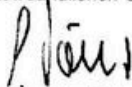
Bezüglich Statistiken mit Schwerpunkt Schweiz ist jedoch ausser den Weissbüchern von Robert Weiss [5] kaum etwas zu finden. SWITCH erhebt selbst jeden Monat die so genannte Hostcount-Statistik, die von RIPE europaweit konsolidiert wird [6]. Beide Statistiken decken jedoch nur Teilbereiche ab: das Weissbuch die verkaufte Hard- und Software und der Hostcount die zählbaren Rechner und Start of Authority Records unter CH.

Daher ist jede Arbeit, die Aufschluss über die Verwendung des Internet in der Schweiz und die Performance der Netze gibt, zu begrüssen. Die Diplomarbeit ‚Swiss Internet Analysis‘ der Zürcher Hochschule Winterthur soll die statistische Situation mit Schwerpunkt Schweiz verbessern.

Um die Erhebungen durchführen zu können, benötigen die Diplomanden Daniel Graf und Olivier Müller Zugang zu diversen Logfiles und müssen die Übertragungswege mit internet-spezifischen Tools analysieren.

Wir empfehlen den um Beteiligung angefragten Organisationen, das Projekt zu unterstützen und den damit betrauten Diplomanden die dafür notwendigen Daten zu liefern und ihnen die Analyse der Internet-Infrastruktur zu erlauben.

Mit freundlichen Grüssen



Dr. Constantin Toenz, Direktor

Referenzen:

- [1] <http://www.isc.org/ds/>
- [2] <http://www.domainworldwide.com/>
- [3] <http://www.registrarstats.com/>
- [4] <http://www.zooknic.com/Users/index.html>
- [5] <http://www.robertweiss.ch/wb2002pr.html>
- [6] <http://www.ripe.net/ripenncc/pub-services/stats/hostcount/index.html>

Support Letter ZHW:



Zürcher
Hochschule
Winterthur

Department
Informatik,
Kommunikation und
Elektrotechnik

Prof. Hans Weibel

Postfach 806
CH-8401 Winterthur

Tel. direkt: 052 267 74 44
Fax direkt: 052 268 73 01
E-mail:
hans.weibel@zhw.ch

Telefonzentrale
Sekretariat
Tel: 052 267 71 71
Fax: 052 268 71 71

Winterthur, 19.7.2002
Unser Zeichen: wei

Internet Analysis CH 2002

Sehr geehrte Damen und Herren

Das Internet wandelt sich ständig - nicht nur bezüglich Diensten und Inhalten, sondern auch in der verwendeten Technik. Leistungsfähigere Hardware, neue Software im Client- und Serverbereich sowie die vermehrte Nutzung breitbandiger und asymmetrischer Zugangstechniken (über 200'000 ADSL und Cable TV Access Channels in der Schweiz) ändern auch die Verkehrsmuster.

Unterstützt von der IXEurope Telehouse Facilities AG wollen wir im Rahmen der Diplomarbeit **Internet Analysis CH 2002** gewisse statistische Merkmale dieser Verkehrsmuster erfassen. Die Resultate sollen den Herstellern und Betreibern von Internet-Infrastruktur helfen, ihre Systeme der aktuellen Situation angepasst zu optimieren und zu „tunen“. Die Resultate der Arbeit werden veröffentlicht, so dass sie von allen, die sich für ein leistungsfähiges und stabiles Internet einsetzen, genutzt werden können.

Bei der Erhebung der Daten werden wir uns an die datenschutzrechtlichen Bestimmungen halten und weitere mit Ihnen getroffene Vereinbarungen respektieren. Die erfassten Rohdaten werden nach Anonymisierung, Aggregation und Verdichtung vernichtet. Veröffentlicht werden nur Daten, die statistische Aussage machen.

Die ZHW ist Mitglied
der Fachhochschule Zürich

Ich bitte Sie, das Projekt und die damit betrauten Diplomanden zu unterstützen. Bei Fragen stehe ich Ihnen gerne zur Verfügung.

Mit freundlichen Grüßen

Zürcher Hochschule Winterthur



Prof. Hans Weibel
Leiter Fachgruppe Kommunikation

Links

Ausschreibung der Diplomarbeit:	http://www-t.zhwin.ch/oapa
Projekt Homepage:	http://www.swiss-internet-analysis.org
IX Europe Telehouse Facilities AG:	http://www.tix.ch
SWITCH:	http://www.switch.ch
BAKOM:	http://www.bakom.ch
Zürcher Hochschule Winterthur:	http://www.zhwin.ch

