

Diploma Thesis - Zürcher Hochschule Winterthur **Swiss Internet Analysis 2002**

Short Description of the Project

The Swiss Internet will be analyzed from a technical point of view, focusing on two domains:

1. Type, distribution and usage of the different Internet resources (domains, name server, web server, mail server, etc.)
2. Network performance (based on a statistical analysis of TCP connection parameters)

The raw data and logs will be collected from the Internet and parsed by tools which have to be programmed.

Time period: 2 months; September 6 – October 28, 2002

Goals of this Diploma Thesis

- - Representative overview of the Swiss Internet in 2002
- - Statistical data allowing fine tuning of routers, servers and operating systems (for example the Linux/FreeBSD kernel IP stack)
- - Results and all tools will be released publicly

Presentation of the Students

Olivier Müller und Daniel Graf are CS (computer science) students of the Swiss University of Applied Sciences of Winterthur (Zürcher Hochschule in Winterthur, ZHW). They will work on this study as a final work before graduation.



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

Responsible Professor

Prof. Hans Weibel, Professor for telecommunication and information technology at the Swiss university of applied sciences.

Contact: wei@zhwin.ch - Homepage: <http://www.zhwin.ch/~wei>

Zürcher Hochschule Winterthur - ZHW

ZHW is the largest multidepartmental college of any Swiss university of applied sciences, boasting a wide range of study programmes, services and R & D activities. It also offers special opportunities for postgraduate study and professional development in the fields of architecture, technology, computer science, chemistry, business & management and language & communication.

Partner Company

IX Europe Telehouse Facilities AG

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

IX Europe Telehouse provides the workplace and network infrastructure as well as test servers, hardware and know-how.

Supporting Organizations

- - **SWITCH** The Swiss Education & Research Network
- - **OFCOM** Swiss Federal Office of Communications
- - **ZHW** Zürcher Hochschule Winterthur

Planned Steps

- - **Phase 1** Collection of raw data
- - **Phase 2** Analysis of the raw data using a parser and database
- - **Phase 3** Results analysis
- - **Phase 4** Publication of the report

What we need

Part A: Domain Analysis

Data is already available (Source: SWITCH).

Part B1: Analysis of Web Traffic

For this part we need raw logfiles of popular websites. As far as possible the logs should be of the type „combined“ (for example with the Apache web server):

```
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“
```

These logs should contain the following information:

- - date, time
- - client hostname or IP address
- - answer code (200, 401, etc.)
- - size
- - user agent

Logs of other formats will also be helpful; we will just convert them to the „combined“ format before the parsing & analysis work.

Part B2: Analysis of Email Traffic

For the mail traffic analysis, we need mail server log files from important Swiss providers. For privacy reason, the „username“ part of the mail addresses (before the „@“) must be hidden.

Sample of qmail mail server log file:

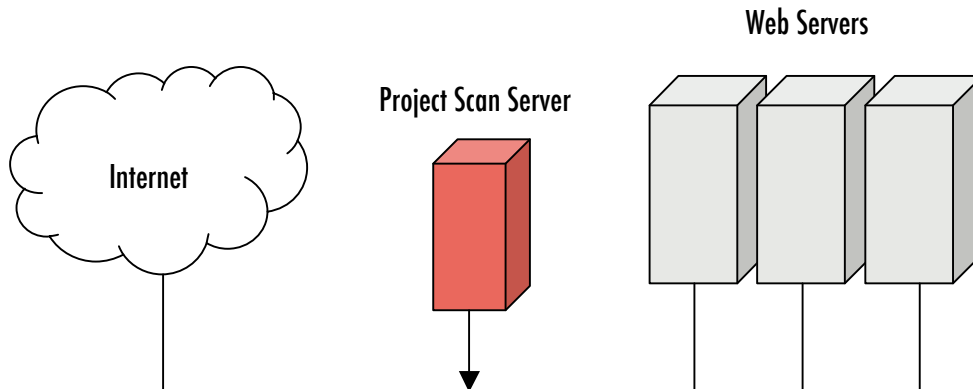
```
@400000003d7c61590defe8c4 info msg 257210: bytes 2825 from <XXXXXX@eValueMail.com> qp 24368 uid 505  
@400000003d7c61590fc047ac starting delivery 22101: msg 257210 to local XXXXX@praevis.ch  
@400000003d7c615a065c0b4c info msg 257210: bytes 5437 from <XXXXXX@Topjobs.ch> qp 24394 uid 505  
@400000003d7c615a083d851c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

Useful information:

- - date, time
- - sender and recipient domains
- - size
- - answer code (so far available)

Part C: TCP Analysis

For the second part of the diploma thesis, we will setup some „scan-servers“ near very busy webservers. `Tcpdump` will be used to catch the TCP headers of all packets during a defined time period (one week). The output will then be saved for further analysis. Only the TCP headers will be saved: the contents of the TCP packet (data, cookies, etc.) will not be observed at all.



All the TCP headers (from different providers) will then be merged together in a „pool“ of data. The analysis will not be started before data has been merged completely.

NDA – Non Disclosure Agreement

A non disclosure agreement (NDA) will be signed with the companies delivering the log files. This way you will be sure that the data will be kept secure and only be used for this study. At the end, all log files will be deleted of course.

Documents / References

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].

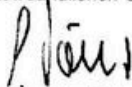
Bezüglich Statistiken mit Schwerpunkt Schweiz ist jedoch ausser den Weissbüchern von Robert Weiss [5] kaum etwas zu finden. SWITCH erhebt selbst jeden Monat die so genannte Hostcount-Statistik, die von RIPE europaweit konsolidiert wird [6]. Beide Statistiken decken jedoch nur Teilbereiche ab: das Weissbuch 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

Diploma thesis list:	http://www-t.zhwin.ch/oapa
Project 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

