Computer Science

Christian Henke

Evaluation of Hash Functions for Multipoint Sampling in IP Networks

Diploma Thesis

Examicus

Evaluation of Hash Functions for Multipoint Sampling in IP Networks

Diplomarbeit
am Fraunhofer Institut Fokus
betreut vom Fachgebiet Telekommunikationsnetze
Technische Universität Berlin

vorgelegt von
Christian Henke

am 10. 04. 2008

Betreuer Fraunhofer Fokus:	Dr.-Ing. Tanja Zseby
	Dipl.-Ing. Carsten Schmoll
Betreuer TU Berlin:	Prof. Dr.-Ing. Adam Wolisz

Christian Henke
Matrikelnummer: 205724

Abstract

Network Measurements play an essential role in operating and developing today's Internet. A variety of measurement applications demand for multipoint network measurements, e.g. service providers need to validate their delay guarantees from Service Level Agreements and network engineers have incentives to track where packets are changed, reordered, lost or delayed. Multipoint measurements create an immense amount of measurement data which demands for high resource measurement infrastructure. Data selection techniques, like sampling and filtering, provide efficient solutions for reducing resource consumption while still maintaining sufficient information about the metrics of interest. But not all selection techniques are suitable for multipoint measurements; only deterministic filtering allows a synchronized selection of packets at multiple observation points. Nevertheless a filter bases its selection decision on the packet content and hence is suspect to bias, i.e the selected subset is not representative for the whole population. Hash-based selection is a filtering method that tries to emulate random selection in order to obtain a representative sample for accurate estimations of traffic characteristics.

The subject of the thesis is to assess which hash function and which packet content should be used for hash-based selection to obtain a seemingly random and unbiased selection of packets. This thesis empirically analyzes 25 hash functions and different packet content combinations on their suitability for hash-based selection. Experiments are based on a collection of 7 real traffic groups from different networks.

Zusammenfassung

Netzwerk Messungen sind essentiell für die Verwaltung und Weiterentwicklung des heutigen Internets. Es existieren immer mehr Messapplikation die Daten vom mehreren Punkten benötigen, z.B. müssen Internet Provider ihre Vereinbarungen über maximale Übertragungsverzögerung verifizieren und Netzwerk Entwickler möchten wissen, wo im Netzwerk Internetpakete verloren oder verzögert werden. Mehrpunktmessungen haben aber den Nachteil, dass sie eine große Menge von Messdaten produzieren, deren Verwaltung und Speicherung hohe Messinfrastrukturkosten mit sich ziehen. Es existieren einige Stichprobenmethoden die es ermöglichen die Datenmenge zu reduzieren und die Messdaten zu schätzen. Nicht alle Stichprobenverfahren sind geeignet für Mehrpunktmessungen, nur deterministische Filter gewährleisten, dass an jedem Messpunkt die gleiche Stichprobe genommen wird. Allerdings ist die Selektionentscheidung für ein Paket bei einem Filter deterministisch abhängig von dessen Inhalt und von daher ist nicht gewährleistet, dass es sich bei den selektierten Paketen um eine repräsentative Stichprobe handelt. Ein auf einer Hashfunktion basierendes Filterverfahren soll dazu verwandt werden eine Zufallstichprobe nachzubilden um möglichst genaue Schätzungen der Messdaten zu ermöglichen.
In dieser Arbeit wird analysiert welche Hash Funktion und welche Paketinhalte für dieses Verfahren verwendet sollen, um eine möglichst unverzerrte Schätzung zu erhalten. Es werden 25 Hash Funktionen und verschiedene Paketinhalte auf ihre Eignung für das Selektionsverfahren untersucht.

Acknowledgment

First of all I would like to thank Tanja Zseby and Carsten Schmoll for giving me the opportunity to write this diploma thesis in a research centered topic at Fraunhofer Fokus. Their subtle supervision gave me confidence in my own ideas and helped me improve my researching ability. I am also very grateful that Prof. Dr.-Ing. Adam Wolisz pointed out critical issues as the adviser at the TU Berlin. Many thanks go to my colleagues Nikolaos Chatzis and Thomas Hirsch for their help and friendship. When I was struggling to the mist of C++ pointer or droughts of ideas they showed me the right way out of misery.
Special thanks go to Nora Eisemann who I could talk to about statistical matters. With her knowledge she helped me to tweak equations and make the calculations more worthwhile.
Last but not least I would like to thank my family. My father's joy about me doing research and my mother's support encouraged me to delve into the work. Thanks to my beloved Katja, the person behind myself, who gives me strength when no one else can.

Contents