University of Zurich[UZH]

Burkhard Stiller
Karoly Farkas
Fabio Hecht
Guilherme Sperb Machado
Patrick Poullie
Christos Tsiaras
Andrei Aurel Vancea
Martin Waldburger

Internet economics VI

TECHNICAL REPORT – No. IFI-2012.02

2012

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

Burkhard Stiller, Karoly Farkas, Fabio Hecht, Guilherme Sperb Machado, Patrick Poullie, Christos Tsiaras, Andrei Aurel Vancea, Martin Waldburger

Internet economics VI

Technical Report No. IFI-2012.02

Communication Systems

Department of Informatics (IFI)

University of Zurich

Binzmuehlestrasse 14, CH-8050 Zurich, Switzerland

http://www.csg.uzh.ch/

University of Zurich[UZH]

Burkhard Stiller,
Karoly Farkas, Fabio Hecht,
Guilherme Sperb Machado, Patrick Poullie,
Flavio Santos, Christos Tsiaras,
Andrei Vancea, Martin Waldburger (Eds.)

Internet Economics VI

TECHNICAL REPORT – No. IFI-2012.00XX

April 2012

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

B. Stiller et al. (Eds.): Internet Economics VI
Technical Report No. IFI-2012.00XX, April 2012
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Introduction

The Department of Informatics (IFI) of the University of Zurich, Switzerland works on research and teaching in the area of communication systems. One of the driving topics in applying communications technology is addressing investigations of their use and application under economic constraints and technical optimization measures. Therefore, during the fall term HS 2011 a new instance of the Internet Economics seminar has been prepared and students as well as supervisors worked on this topic.
Even today, Internet Economics are run rarely as a teaching unit. This observation seems to be a little in contrast to the fact that research on Internet Economics has been established as an important area in the center of technology and economics on networked environments. After some careful investigations it can be found that during the last ten years, the underlying communication technology applied for the Internet and the way electronic business transactions are performed on top of the network have changed. Although a variety of support functionality has been developed for the Internet case, the core functionality of delivering data, bits, and bytes remained unchanged. Nevertheless, changes and updates occur with respect to the use, the application area, and the technology itself. Therefore, another review of a selected number of topics has been undertaken.

Content

This new edition of the seminar entitled "Internet Economics VI" discusses a number of selected topics in the area of Internet Economics. The first talk "Internet Usage Measurement" develops the state-of-the-art in Internet usage measurement research by presenting results gathered from more than 200 papers published within 4 specific workshops in the years 2010 and 2011. The talk "Mobile Payment Systems" provides an introduction into mobile payments. It looks at different mobile payment methods, compares them, and it presents relevant business models, market drivers and constraints as well as risks of mobile payment solutions. The talk "An Economic Overview of Internet Mobile Platforms" performs an analysis along various economic dimensions of the relevant set of mobile ecosystems. It adopts the perspective of a platform provider as well as of an application developer in order to assess advantages and drawbacks of different business model approaches. The talk "ISP-friendly Content Distribution Systems" is driven by the observation that the total amount of Internet traffic has been rising extremely over the last decade. In this light, it provides an overview of ways to address this trend, namely of peer matching approaches and implementations. The talk "The Hidden Extras: The Pricing Scheme of Cloud Computing" sketches a technical use case, which is used to evaluate the five large payers in cloud computing (Amazon, Rackspace, Terremark, IBM and Windows Azure). The results of this investigation show that there are many aspects of hidden costs

in the pricing schemes of cloud providers. The talk "Cloud Computing Services (CCS): A Threat for Internet Neutrality?" adopts a cloud-centric focus as well, while it draws its attention to the two areas of enterprise cloud computing services and internet neutrality. The talk "Convenient Trust" concludes this seminar report by exploring several open issues in public-key infrastructures and in establishing a user-friendly security infrastructure.

Seminar Operation

Based on well-developed experiences of former seminars, held in different academic environments, all interested students worked on an initially offered set of papers and book chapters. Those relate to the topic titles as presented in the Table of Content below. They prepared a written essay as a clearly focused presentation, an evaluation, and a summary of those topics. These essays are included in this technical report in a separate section each. This allows for an overview on important areas of concern, sometimes business models in operation, and problems encountered.
In addition, every student prepared a slide presentation of approximately 45 minutes to present his/her findings and summaries to the audience of students attending the seminar and other interested students, research assistants, and professors. Following a general question and answer phase, a student-led discussion debated open issues and critical statements with the audience.
Local IFI support for preparing talks, reports, and their preparation by students had been granted by Karoly Farkas, Fabio Hecht, Guilherme Sperb Machado, Patrick Poullie, Flavio Santos, Christos Tsiaras, Andrei Vancea, Martin Waldburger, and Burkhard Stiller. In particular, many thanks are addressed to Martin Waldburger for his strong commitment on getting this technical report ready and published. A larger number of pre-presentation discussions have provided valuable insights in the emerging and moving field of Internet Economics, both for all students and supervisors. Many thanks to all people contributing to the success of this event, which has happened in a lively group of highly motivated and technically qualified students and people.

Zürich, April 2012

Contents

Chapter 1

Internet Usage Measurement

Robert Erdin

As the Internet evolves over time, an updated understanding of how, by whom and why it is being used becomes necessary. A lot of research is being done in this field usually under the name of Internet usage measurement, but there is no existing documentation that sums up the work of the community itself. Therefore this report takes a first step in addressing this knowledge gap and documents the state-of-the-art in Internet usage measurement by presenting results gathered from 208 Papers published within 4 workshops in the years 2010 and 2011. The results are presented in two tiers. First by providing a general overview over the work of the whole Internet measurement community where an approach is made to assign each paper to a general subarea of Internet measurement such as Internet usage measurement. This first level of analysis shows that roughly every tenth paper is of interest for Internet usage measurement. On the second level of analysis a more detailed insight in the work on Internet usage measurement is provided. It shows that two workshops are of special interest for Internet usage measurement and that papers published can be divided into large-scale papers with data of more than 10,000 users and papers with less then 100 users. It also shows that the most used technologies for gaining data are packet capturing, crawling, quality of experience and application level capturing.

Contents

1.1 Introduction and Motivation

The Internet has become one of the main communication platforms of our society, offering a vast amount of services. These services, private or public as well as commercial or free, affect almost every aspect of everyday life. Therefore it is crucial to obtain a deep-going understanding of how, by whom and why the Internet is used. This understanding is needed for decision making in technology investments, Internet application development as well as for policy making with regard to current and future networks and services. The main instrument to develop a quantified understanding of Internet usage is found in Internet usage measurements.
Internet usage measurements are done by many different parties pursuing different targets under application of a variety of methods. Furthermore the data is collected from different sets of metering points. There is no existing documentation of the work of the Internet measurement research community. Cosequently, this seminar adopts a single dedicated objective: to document the state-of-the-art in Internet usage measurement from a technological, scientific quality and ethics (privacy, confidentiality) perspective. The state-of-the-art is being evaluated by examining the publications of the research community[1] for similarities and disparities so that the current work can be illustrated and conclusions be drawn.

1.1.1 Report Structure

In order to document the state-of-the-art in Internet usage measurement several steps have to be taken to achieve this final goal. Therefore the paper is structured as follows. The first part (Section 1.2) is a detailed description of the approach chosen for this seminar. It contains information about the scope and a rough description of the conferences and workshops described in the scope as well as a description of how relevant publications will be identified and the results will be clustered later on.
The second part(Section 1.3) is a bird's-eye view on the results of the categorisation of all published papers including analysis on where relevant papers are being published. The idea of this part is to illustrate the broad distribution of different fields of interests and work within the Internet measurement community and on what topics research is being done.
The third part (Section 1.4) is focused on the work that was identified as relevant to Internet usage measurement in the second part. The methodology of the clustering is being described as well as the results of a deeper insight into the relevant papers.

1.2 Approach

To achieve the aim of this seminar – to document the state-of-the-art in Internet usage measurement – a multi-tier approach is necessary. As a first step the work of the Internet measurement community is being analysed and presented. In the second step only the work related to Internet usage measurement is being used for further analysis and more accurate conclusions. Conclusions are drawn regarding each of the following perspectives: technological, scientific quality and ethics (privacy, confidentiality) which are chosen to have a fully extensive consideration of what is included in the state-of-the-art.

This seminar is focused, but not limited, to papers published in the conferences and

[1]For feasibility reasons the research community is limited to the four conferences and workshops described in Section 1.2.1

workshops mentioned in the task (Section 1.2.1). Bibliography references in these papers may also be used if needed. Only the papers of the two latest occurrences(years 2010 and 2011) of the mentioned conferences and workshops are used for the results in this paper. A brief description of these workshops including their purposes is provided in the proceeding section.

1.2.1 Conference Proceedings Covered

A number of conferences and workshops have been selected for this seminar as they are seen as the major events where research in Internet measurement is published. The list of conferences and workshops considered here, thus, is perceived to cover the suited overview over recent work in the community. Nonetheless, it shall be noted that papers from community members may – and most probably will – be published in other events so that this set of events looked at in this seminar does not claim completeness.
In each case (except for the Workshop on Measurements up the STack) the two latest occurrences are being used (years 2010 and 2011). The Workshop on Measurements up the STack was created in 2011 and therefore only one occurrence (2011) existed when this report was written. The following paragraphs describe all four conferences and workshops.
ISMA AIMS Workshop on Active Internet Measurement The Workshop on Active Internet Measurement [175], [176] is held annually and is part of the Internet Statistics and Metrics Analysis (ISMA) workshop series hosted by CAIDA (The Cooperative Association for Internet Data Analysis). The goals are to further the understanding of the potential and limitations of active measurement research and infrastructure in the wide-area Internet, and to promote cooperative solutions and coordinated strategies to address future data needs of the network and security research communities.
Internet Measurement Conference The ACM (Association for Computing Machinery) SIGCOMM (Special Interest Group on Data Communications) Internet Measurement Conference (IMC) [177], [178] is an annual conference focusing on Internet measurement and analysis. The aim is that papers presented at the conference contribute to the current understanding of how to collect or analyze Internet measurements, or give insight into how the Internet behaves. The IMC also tries to encourage broader data sharing within the research community.
Passive and Active Measurement Conference The Passive and Active Measurement Conference (PAM) [179], [180] is held by various universities (University of Auckland, ETH Zurich, Worcester Polytechnic Institute, Endace, University of Washington, KAIST, LIP6, Case Western Reserve University) and has its focus on research and practical applications of network measurement and analysis techniques. The conference's goal is to provide a forum for current work on these measurement and analysis techniques in its early stages.
Workshop on measurement Up the STack (W-MUST) The Workshop on Measurements Up the STack (W-MUST) [181] is part of the ACM SIGCOMM Conference and has its main focus on end-host measurements to explore the behavior of networked applications and user perceptions of Internet services. These goals are often unachievable with classic Internet measurement techniques that use packet traces. SIGCOMM introduced this workshop in 2011 to fill this gap and bring the research community together to share new ideas and experiences facing the challenges of "measurement up the stack".
To sum it up, the four conferences and workshops have different key areas but also areas that overlap. All four conferences and workshops together cover measurements from a global perspective of the Internet down to measurements on single end-user devices. Active and passive measurement methodologies are equally used and measurements, independent of the scale, are being done to achieve different goals such as topology mapping, performance or security.

1.2.2 Identification of Relevant Papers

As described in the introduction of the approach (Section 1.2), this report uses a multi-tier approach whereat the section at hand provides the first level of analysis and has two purposes: first to present an overview over the work of the Internet measurement research community and second to identify all papers that are relevant to Internet usage measurement which is essential for the second part of the analysis, described in Section 1.2.3.
The publications of the conferences described in the preceding section are the source of information on which this report is based. Because of the large amount of publications the first step is to get an overview by skim-reading the abstracts of all the papers published. This is necessary to get an idea on what the different publications are about and to identify the different fields in which papers are being published. This knowledge is essential to define criteria to categorise papers later on.
In order to identify the papers related to Internet usage measurement, identified publications have to be categorised. Therefore categories must be defined with criteria when a specific paper is assigned to a category. The categories are obtained from meta information of the publications such as key words provided within the paper or the session topic where it was published in the conference as well as from the goal of the papers such as improving security or performance. It is in the nature of things that Internet usage measurement is one of these categories. Each publication is then assigned to one of these categories. If a paper matches none of the defined categories it will be assigned to the category 'undefined'.

1.2.3 Clustering of Results

The clustering of results is the second part of the multi-tier analysis described at the beginning of the approach section. It pursues the goal to document the state-of-the-art in Internet usage measurement as described in the introduction. Therefore, for this last step only the papers previously identified as work on Internet usage measurement (see Section 1.2.2) are being used. These papers are analysed for similarities and lack of information regarding technological, scientific quality and ethics (privacy, confidentiality) perspectives. A more detailed description of the methodology used for the clustering part is provided at the beginning of the Section 1.4. Depending on the results of this clustering suggestions are made where possible.

1.3 Results

The range of identified publications is very broad. There are a total of 208 papers published in 55 different categories (The conferences and workshops described in the Section 1.2.1 structure the publications into their own categories) during the years 2010 and 2011. Some categories used in the conferences and workshops are very general, such as 'security' or 'performance', and can potentially be used to categorise the papers as described in Section 1.2.2. Most of the categories papers are published in are very specific though, e.g. 'measurement of content distribution networks' or 'routing and path quality measurements'. These specific categories are not suitable for the nature of this seminar because it is not clear whether or not a paper is related to Internet usage measurement on the basis of the categories in the workshops. Furthermore 55 categories are too many to illustrate what the Internet measurement community as a whole is working on. Therefore an own, compressed set of categories has to be created.
The following sections are an approach to defining categories in order to identify the papers relevant to Internet usage measurement. Further the catogeries can be used to

illustrate the ratio of relevant papers to the other topics discussed during the considered conferences and workshops.

1.3.1 Grouping of Categories

The examination of meta information (keywords and general terms within papers as well as the category a paper was published in) and the goal of each paper lead to the conclusion that the majority of papers are related to the following six general categories:

- Internet Usage Measurement
- Security and Privacy
- Performance
- Measurement Techniques
- Topology
- Cooperation

A detailed description when a paper matches one of the determined categories is provided in the following sections. The Internet usage measurement category is the most important category since it is also used for the next level of analysis (see Section 1.4) to fulfill the aim of this seminar, to document the state-of-the-art in Internet usage measurement. Therefore the description is more detailed than the one of the other categories to make sure all relevant papers are identified.

1.3.1.1 Internet Usage Measurement

This category refers to four groups of Internet usage measurement as described subsequently:
The behaviour of a specified group of users For example users in a university campus or a company, households of an Internet service provider, consumers of Internet services such as proxy networks or content distribution networks et cetera. The behaviour analysed can vary from very general perspectives such as how often the Internet is used or what services it is used for to specific analyses e.g. the composition of websites browsed. To summarize, this category includes papers that analyse usage behaviour by measuring close to the source of the traffic.
The behaviour of users measured on a target system Typical target systems embrace a website or web service. This also includes the behaviour of users in social networks.
Conventional measurement Measurements performed on both the source and the destination are limited in expressiveness (with respect to usage behaviour) and therefore papers with a focus on new or alternative methods on how to get input from users regarding their behaviour and or satisfaction in the Internet or when using specific service are also part of Internet usage measurement.
Focus on how and where to measure traffic generated by users This research is essential for Internet usage measurement although there are no concrete results about the behaviour of users provided. The border between these papers and the papers described in 1.3.1.4 is blurry. To be listed in this category the described technique must create output that can directly be used for Internet usage measurement in a further step.

1.3.1.2 Security and Privacy

Security and privacy endorses all publications with an aim to detect security or privacy issues in a given service, system or protocol. This category also includes papers which prove a given service, system or protocol as insecure or rate its level of security. Further proposals of possible solutions to increase security or privacy in a given service, system or protocol are also part of the security and privacy category.

1.3.1.3 Performance

This category includes papers on how to measure performance of a certain service, system or protocol or how to increase it. Also research on what affects performance (in a positive or negative way) and how to improve performance by changes to protocols or topology is considered. Equally, comparison and benchmarking of comparable services, systems or protocols are topics of interest.

1.3.1.4 Measurement Techniques

Measurement techniques is a category for all work on how to measure something specific, with no other objective than the measurement itself. Papers that focus on the measurement of user-generated traffic that can possibly be used for Internet usage measurement is excluded from this category since it is handeled in 1.3.1.1.

1.3.1.5 Topology

Internet topology deals with finding the topological structure of the Internet. All work related to Internet topology such as attempts to draw maps of the Internet topology at different layers and regions of the Internet is part of this category.

1.3.1.6 Cooperation

A major issue within the research community is the cooperation between all different parties. Many goals cannot be achieved by an individual research group itself. One possible example for such a goal is global Internet topology mapping which not even a regional registrar such as RIPE NCC can achieve on its own. But not only large scale projects profit from a close collaboration between research institutes. A key success factor for many Internet measurement attempts is the data set used for the analysis which is often hard to collect. Therefore there is a high demand for a broad variety of measurement data. It is in everyone's interest to simplify access to measurement data e.g. through building data repositories and making collected data for past research available to the community. This category contains all work that has a focus on increasing cooperation within the research community as well as attempts to share and exchange data used for future work.

1.3.2 Results of Categorisation

The categorisation of all the papers and presentations published in the examined workshops gives insight in what the Internet measurement community is currently working on. It is possible to show which conferences focus on specific topics and, since it is the main focus of this paper, to show which conferences are particularly interesting for Internet usage measurement.
The distribution of papers per category as displayed in Figure 1.1 gives an overview on the work of the research community as a whole.
To determine which conferences are interesting for Internet usage measurement, the number of relevant papers per conference might be an indicator, shown in Figure 1.2.

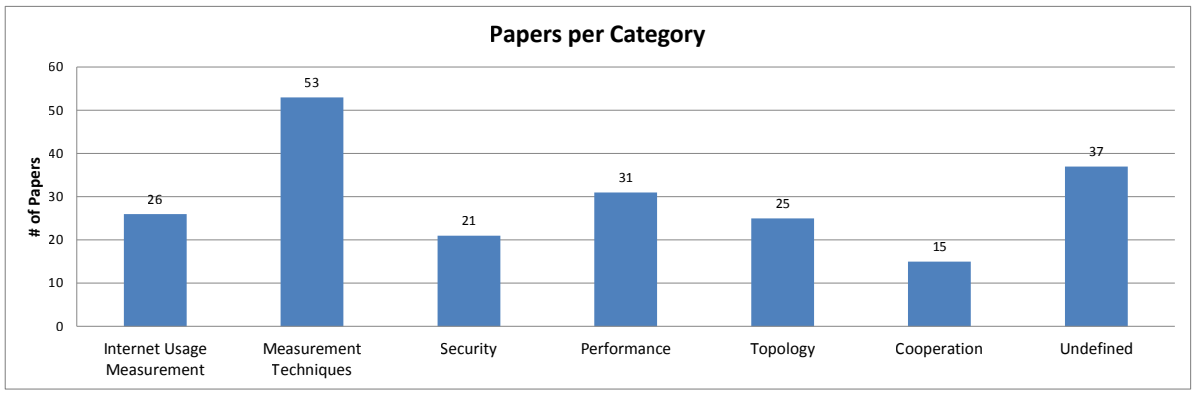


Figure 1.1: Distribution of papers per category

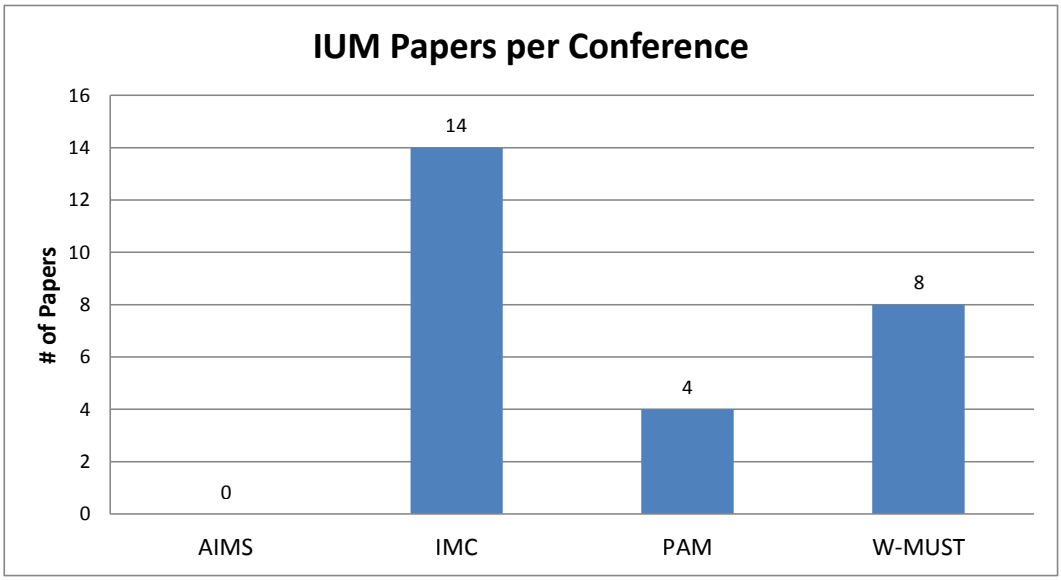


Figure 1.2: Internet usage measurement (IUM) papers per conference

However, absolute numbers draw a biased picture without the numbers of publications per workshop. Table 1.1 displays the amount of papers published by each conference. This amount of papers published varies not only because the conferences and workshops differ in terms of length, but also because the Workshop on Measurements Up the STack was first held in 2011.
Where the Internet Measurement Conference has the highest number of papers on Internet Usage Measurement, its ratio of relevant papers (0.16) is significantly lower then the ratio of the Workshop on Measurements Up the STack (0.73). The Workshop on Active Internet Measurements is the only one without any publications on Internet usage measurement at all. The reason is probably that most work in this field is done with passive methodologies.

Table 1.1: Number of papers per conference

AIMS	61
IMC	89
PAM	47
W-MUST	11
TOTAL	208

Mapping the number of papers per category to each of the examined conferences and workshops (Figure 1.3) gives a rough idea what the main focus of each event is. The Workshop on Active Internet Measurements has a strong focus on topology and cooperation because most research on Internet topology requires strong cooperation between different institutions. The Internet Measurement Conference and the Passive and Ac-

tive Measurement Conference are very diverse which includes papers on Internet usage measurement. The Workshop on Measurements Up the STack has a very clear focus on Internet usage measurement as described before.

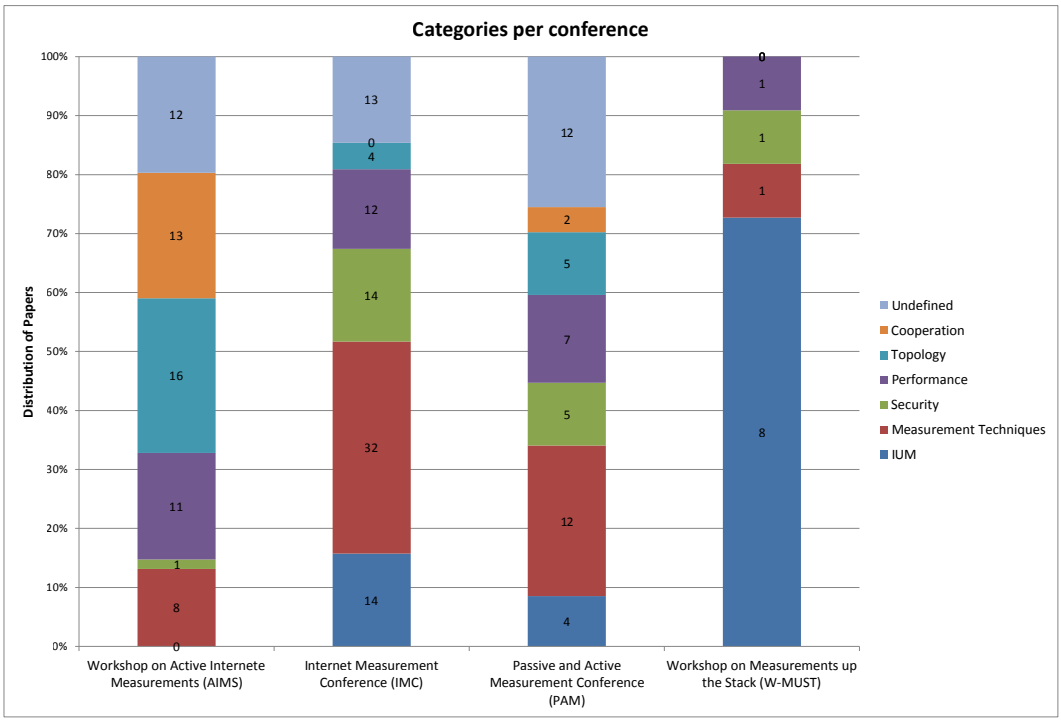


Figure 1.3: Categories per conference (IUM used as acronym for Internet usage measurement)

1.3.2.1 Non-categorised Papers

A significant amount of publications (n=37 out of 208) did not match any of the categories described in Section 1.3.1. These papers are listed as 'undefined' in the preceding charts. Several papers (n=7) had strong similaries: all were related to geolocation, but geolocation was not considered as a category of its own to simplify matters. Other papers had no obvious primary goal or did simply not match any of the identified categories.

1.4 Clustering

1.4.1 Methodology

In this part all the papers identified as relevant to Internet usage measurement are being analysed for similarities, disparities and lack of information provided. About half the papers contain more than one data set on which the research is based. In order to make the different papers comparable an attempt is made to reduce the information to one set. Depending on the paper different methodologies are applied. If a paper makes use of multiple data sets but one can be identified as the main one and the other(s) are for example used to figure out the methodology, only the information of the main data set is used for the clustering. Detailed information on how the information is being merged is provided in the following sections.

1.4.1.1 Scientific View

In this perspective papers were examined for their scientific methodology and therefore the proceeding factors were extracted from all papers in the Internet usage measurement category:

Time period and or interval of data collection. If data was not collected in one piece, the durations of all tranches are summed-up and the duration between the beginning and the end of all capturing activities is also provided. If data of different sets was collected in intervals but could basically be merged into one big set (e.g. 3 sets of 90 days each on the same capturing location, it is being listed as 1 set of 270 days), this is being done if there are no other conflicts like different amounts of users.
Users involved. The amount of users from whom data was being captured. Information regarding user numbers are provided in three different ways in the examined papers. Either the amount of users mentioned is the number of concurrent users or the number of unique users in a system. The third option is to just mention a number with no further description. Real comparability of these numbers is not given but for the purposes of this work this is accepted as sufficient. To simplify matters unique end-user devices are treated as users.
Data publishing. Whether or not data was made available to the research community after the publication. If it is not explicitly mentioned that the data is being published, the assumption is made that this is not the case.
Interaction with users involved, e.g. through surveys or QoE tools.
Description of capturing methodology. Whether or not a detailed description is provided of the methodology on how a data set was collected. For example „We crawled the social network XYZ" is not sufficient whereas a description like „It contains packetlevel traces, including link layer headers, for data sent and received by the smartphone. We collected these traces using Netlog on Windows Mobile and tcpdump on Android" is considered as a description of the capturing methodology.
Own data collection. Whether data was being collected by the researchers themselves, data was provided by a third party such as an ISP, or data was made available from other researchers or a data repository.

1.4.1.2 Technological View

To figure out the state-of-the-art from a technological perspective the interests lie in the technology used for the measurements and where in a system measurements are being made. To get this information in a comparable form, the following factors were extracted from the papers:
Technologies used for measurement. What kind of technologies were used to collect data. To keep results comparable the possible options are limited to the following: APIs (Application Programming Interface) such as Twitter or YouTube API. Packet capturing, which contains all technologies to capture either all network packets or only the packets of a specific protocol or session level capturing of network packets. Application level capturing, which contains all technologies that get information out of an application either on a client device or a server. QoE tools such as surveys or specialised software to get user feedback on a specific system.
Point of capturing. Where the individual system to capture data is installed. For the same reasons as mentioned previously the following possibilities exist: Capturing at the edge of a network, which is the border between the own or examined network and the Internet. Centralised, which covers all approaches that get data from a centralised system. Capturing on the Device. Distributed capturing systems. Mixed or other approaches.
How the data is being stored Information on how data is being stored. This is primary focus for large-scale measurement projects and therefore mainly expected to be listed in such papers.
Amount of data. The amount of data in gigabyte.

1.4.1.3 Ethics View

Internet usage measurement often uses sensitive user data for research purposes and therefore it is crucial to know how to treat this data and how to interact with users. This is of importance for the users, both aware and unaware of their participation in a study as well as for the respective researchers to protect themselves from possible legal issues. To gain knowledge about ethisc-driven standards the following factors were extracted from the papers:

User Awareness. Are users aware that their data is being used for research purposes? Even though it is likely that many users whose data was used signed (or checked a checkbox) at some point an agreement containing some small print on usage of their data for research purposes the following assumption is made. If not explicitly mentioned in the paper users are considered as unaware of the fact that their data is being used unless they actively signed up for the study in the paper.

Anonymisation. Is the data used for the research in a paper anonymised? If it is not explicitly mentioned the data is considered as not anonymised.

How long is the data stored? Is the data only stored for as long as the research is ongoing or is it stored beyond that?

Is it used for other purposes? Is the captured data only used for the research it was captured for or is it also used for further research or other purposes?

1.4.2 Scientific View

Figure 1.4 shows that papers from the Internet usage measurement category can be divided into two groups: large-scale papers with more than 10,000 users (n=13) and papers with less than 100 users (n=6). 4 papers did not provide a number on how many users were involved.

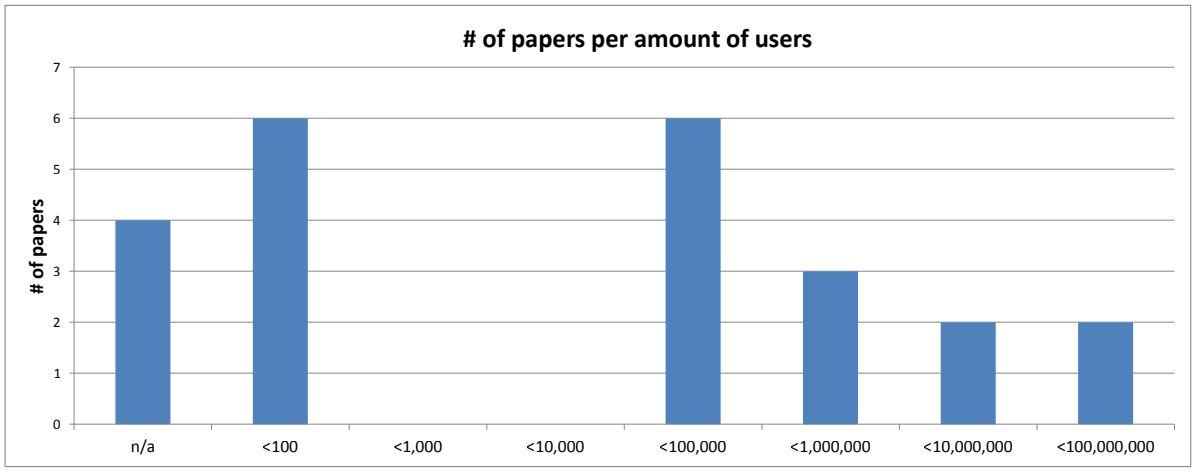


Figure 1.4: Number of Internet usage measurement papers per amount of users in study

All Internet usage measurement papers with less than 100 users were studies where participants had to be recruited and 5 out of 6 were QoE-related.

The amount of days in which the data was collected varies from 1 day to 1 year where the time period of collecting goes from 1 day to 5 years. Not surprisingly there was no user interaction in none of the large-scale measurement projects whereas researchers interacted with the participants in all 6 papers with less than 100 users. There is no paper on Internet usage measurement where the data set was made available to the research community, which is a bit surprising since it is not uncommon to do so in other fields of Internet measurement. It is even more surprising regarding the fact that researchers of 15 papers did collect the data entirely on their own, 5 papers used data partially from other sources and 3 papers relied entirely on third party data.

Large-scale Internet usage measurement papers embrace: [74] [111] [95] [167] [49] [40] [131] [107] [60] [102] [55] [59] [72].
Internet usage measurement papers with less than 100 users embrace: [45] [108] [76] [110] [133] [159].

1.4.3 Technological View

The technological methodologies in all Internet usage measurement papers are described sufficiently so that the measurement technology and the point of measurement (PoM) is clear in all papers. Figure 1.5 shows the distribution of the points of measurement. It can be said that for packet capturing technologies PoM is in almost all cases either the edge of the network, the device or a part of a distributed system. Centralised approaches are usually used for application level data capturing. Only 7 papers provide details on the amount of data of the data sets. 6 out of these 7 papers are large-scale measurement projects with amounts of data from 8 to 25 terabyte. The four most used measurement technologies (packet capturing n=10, application level capturing n=7, QoE n=5, crawling n=4) are additionally described in the following sections.

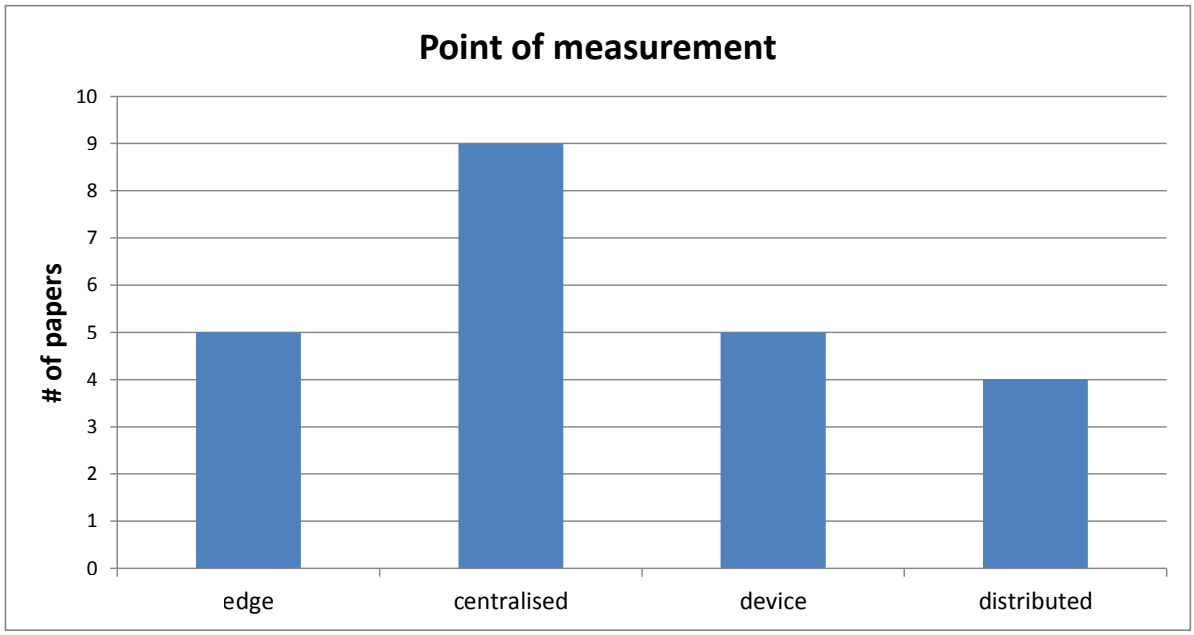


Figure 1.5: Number of Internet usage measurement papers per point of measurement

1.4.3.1 Packet Capturing

Packed capturing provides a powerful instrument for Internet usage measurement. Depending on the point of capturing the data can contain information about the behaviour of a single user or group of users such as a household, students of a university, employees of a company or users of a specific geographical region or even selected users distributed all over the world.
The downside of packet capturing is that most common capture techniques are not able to distinguish between data relevant for a given study and such that is not. Therefore in many cases only a fractional part of the raw, captured data is really being used. This drawback leads to considerable problems. On one hand the vast amount of data is hard to store. Many research projects have to use data of small time windows due to a lack of disk space. On the other hand it is very time-consuming to get all the relevant data out of the raw dump files.

Internet usage measurement papers with a technological view on packet capturing embrace: [60] [167] [72] [49] [107] [60] [102] [55] [108].

1.4.3.2 Quality of Experience

Quality of Experience (QoE) is a measure of the overall level of customer satisfaction with a service. QoE differs from Quality of Service (QoS), which is the notion that hardware and software characteristics can be measured with conventional Internet measurement techniques. In contrast, QoE expresses user satisfaction both objectively and subjectively. QoE can be used as a stand-alone technique to analyse user behaviour or satisfaction in certain situations. But many research projects use it as an enhancement to conventional measurement techniques in order to map user perception to measurable information. Therefore they try to match the provided feedback with measurable data e.g. from packet capturing to get knowledge on how a specific network issue affects end-users.
Internet usage measurement papers with a technological view on Quality of Experience embrace: [159] [133] [110] [76] [108]

1.4.3.3 Application Level Capturing

Application level capturing is often a convenient way of collecting data. Depending on the application in question, needed information is measured with less overhead compared to other approaches like packet capturing or crawling. A typical application is writing different kinds of logs e.g. a syslog.
Application level capturing can also be done by programs which are custom-built only for this purpose, for example a video streaming player that records the pointer movement of a mouse pointer and sends the respective data to the researchers.
Internet usage measurement papers with a technological view on application level capturing embrace: [60] [144] [95] [59] [72] [75] [110].

1.4.3.4 Crawling

A web crawler is a relatively simple automated program, that methodically scans or ”crawls” through a graph of linked files in the Internet. Depending on what information is needed this crawling can be limited to files of the same domain or to any other unit up to Internet-wide scanning such as a search engine does it.
With the establishment of Web 2.0 there is an enormous amount of user-generated content available in the Internet. User-generated content often implicates how specific services are being used. Information gained by crawling e.g. an online social network can give detailed insight on how this specific service is being used by its users. One of the papers in the Internet usage measurement category demonstrated how crawling can lead to remarkable numbers of users whose behaviour can be analysed by gathering 42 million user profiles and 1.66 billion appendant social links.
Internet usage measurement papers with a technological view on crawling embrace: [74] [173] [40].

1.4.4 Ethics View

The analysis of the ethics view is somewhat disillusioning. None of the users were actively informed that their data is being used for research purposes, except for the ones where users had to be recruited (which would be hard otherwise). Only 6 studies anonymised their data sets when capturing (or were forced to by the third party that was involved). In 3 of these 6 cases information was provided how the data was anonymised. Also none

of the Internet usage measurement papers mentioned how long the data will be stored and if researchers intend to use it otherwise.
Internet usage measurement papers that make use of anonymised data embrace: [74] [59] [167] [102] [75].
Internet usage measurement papers that provide information on anonymisation embrace: [74] [60] [102].

1.5 Summary and Conclusion

In this seminar, a two-tiered analysis on Internet (usage) measurement was made. As a first step the work of the Internet measurement community as a whole was being analysed and presented. In the second step only the work related to Internet usage measurement was being used for further analysis. The first level of the analysis led to the conclusion that the majority of papers (208 papers in total) are related to the following six general categories (listed by number of papers per category): Measurement Techniques, Performance, Internet Usage Measurement, Topology (same number of papers as for Internet Usage Measurement), Security and Privacy, and Cooperation. In the most important Internet measurement conferences and workshops, roughly every tenth paper in the proceedings for the years 2011 and 2010 was found to be of interest to Internet usage measurement. The first step of the analysis also pointed out that some conferences have a stronger focus on Internet usage measurement than others. In particular, the Internet Measurement Conference and the Workshop on Measurements up the STack showed an emphasis on Internet usage measurement.
The second, deeper level of analysis, in which only the 25 papers on Internet usage measurement were considered, illustrated the state-of-the-art in Internet usage measurement from a scientific, technological and ethics perspective. For each perspective all publications were analysed for similarities, disparities and lack of information provided. The scientific perspective allowed to divide the publications into two groups: large-scale papers with more than 10,000 users and papers with less than 100 users. From the technological perspective four different capturing technologies that are being used the most emerged, which are: Packet Capturing, Quality of Experience, Application level Capturing and Crawling. It can be said that for packet capturing technologies the point of measurement is in almost all cases either the edge of the network, the device or a part of a distributed system. Centralised approaches are usually used for application level data capturing.
It can be said that the information provided regarding scientific and technological approaches is sufficient in the majority of papers, which means the reader is well informed of the scientific approach of the paper as well as the technologies involved. A possible improvement for the community could be to include a structured, machine readable, part into papers where information of scientific and technological methodologies is contained so that the search for existing suitable data sets is made possible. This assumes that researchers are willing to share the data collected for their studies. Such a machine readable section could further be used to publish the hardware and software used for capturing which allows researchers to create a link between their goal and suitable technologies. It can also be said that almost all considered studies show deficiencies in privacy and confidentiality and they lack relevant information on how privacy and confidentiality was addressed. Therefore it would be desirable that conferences and workshops as well as journals enforce a minimum of information on privacy and confidentiality which each publication must contain. Such restrictions would increase the awareness of the researchers for these issues.

Bibliography

[1] Vijay Kumar Adhikari, Sourabh Jain, Zhi-Li Zhang: *YouTube Traffic Dynamics and Its Interplay with a Tier-1 ISP: An ISP Perspective*, IMC, 2010. http://doi.acm.org/10.1145/1879141.1879197

[2] Bernhard Ager, Wolfgang Mühlbauer, Georgios Smaragdakis, Steve Uhlig: *Comparing DNS Resolvers in the Wild*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p15.pdf

[3] Bernhard Ager, Wolfgang Mühlbauer, Georgios Smaragdakis, Steve Uhlig: *Web Content Cartography*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p585.pdf

[4] Mehmet Burak Akgun: *Subnet Based Internet Topology Generation*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_makgun.pdf

[5] Mohammad Al-Fares, Khaled Elmeleegy, Benjamin Reed, Igor Gashinsky: *Overclocking the Yahoo! CDN for Faster Web Page Loads*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p569.pdf

[6] Zakaria Al-Qudah, Michael Rabinovich, and Mark Allman: *Web Timeouts and Their Implications*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/22.pdf

[7] Demetris Antoniades, Evangelos P. Markatos, Constantine Dovrolis: *MOR: Monitoring and Measurements through the Onion Router*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/14.pdf

[8] Patrik Arlos and Markus Fiedler: *Influence of the Packet Size on the One-Way Delay in 3G Networks*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/7.pdf.

[9] Jordan Auge, Timur Friedman, Thomas Bourgeau: *Overview of TopHat: Interconnecting the OneLab measurement infrastructures*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_jauge.pdf

[10] Jordan Auge, Timur Friedman, Thomas Bourgeau: *Update on TopHat and measurement system interconnection*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_jauge.pdf

[11] Richard Barnes: *Some Internet Measurement Thoughts*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_rbarnes.pdf

[12] Steven Bauer, Robert Beverly, Arthur Berger: *Measuring the State of ECN Readiness in Servers, Clients,and Routers*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p171.pdf

[13] Steven Bauer, Robert Beverly: *Measuring the current state of ECN support in servers, clients, and routers*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_sbauer.pdf

[14] Fabricio Benevenuto, Tiago Rodrigues, Meeyoung Cha, Virgilio Almeida: *Characterizing User Behavior in Online Social Networks*, ACM SIGCOMM, 2011. http://dl.acm.org/citation.cfm?id=1644900

[15] Robert Beverly Arthur Berger: *Directed Probing for Efficient and Accurate Active Measurements*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_rbeverly.pdf

[16] Robert Beverly, Arthur Berger, Geoffrey Xie: *Primitives for Active Internet Topology Mapping: Toward High-Frequency Characterization*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_rbeverly.pdf

[17] Robert Beverly, Arthur Berger, Geoffrey Xie: *Primitives for Active Internet Topology Mapping: Toward High-Frequency Characterization*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p165.pdf

[18] Zachary S. Bischof, John S. Otto, Mario A. Sanchez, John P. Rula, David R. Choffnes, Fabian E. Bustamante: *Crowdsourcing ISP Characterization to The Network Edge*, W-MUST, 2011, AIMS, 2010. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p61.pdf

[19] Stevens Le Blond, Chao Zhang, Arnaud Legout, Keith Ross, Walid Dabbous: *I Know Where You Are and What You Are Sharing: Exploiting P2P Communications to Invade Users Privacy*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p45.pdf.

[20] Lothar Braun, Alexander Didebulidze, Nils Kammenhuber, Georg Carle: *Comparing and Improving Current Packet Capturing Solutions based on Commodity Hardware*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p206.pdf

[21] Pat Brundell, Andy Crabtree, Richard Mortier, Tom Rodden, Paul Tennent, Peter Tolmie: *The Network from Above and Below*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p1.pdf

[22] Fabian E. Bustamante: *Internet-wide systems need Internet-wide measurement platforms*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_fbustamante.pdf

[23] Michael Butkiewicz, Harsha V. Madhyastha, Vyas Sekar: *Understanding Website Complexity:Measurements, Metrics, and Implications*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p313.pdf

[24] Xue Cai, John Heidemann, Balachander Krishnamurthy, Walter Willinger: *Towards an AS-to-Organization Map*,IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p199.pdf

[25] Alfredo Cardigliano, Joseph Gasparakis, Francesco Fusco: *vPF-RING: Towards Wire-Speed Network Monitoring Using Virtual Machines*,IMC, 2011.http://conferences.sigcomm.org/imc/2011/docs/p533.pdf

[26] Kevin M. Carter, Richard P. Lippmann, and Stephen W. Boyer: *Temporally Oblivious Anomaly Detection on LargeNetworks Using Functional Peers*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p465.pdf

[27] Meeyoung Cha, Hamed Haddadiy, Fabricio Benevenutoz, Krishna P. Gummadi: *Measuring User Influence in Twitter: The Million Follower Fallacy*http://an.kaist.ac.kr/~mycha/docs/icwsm2010_cha.pdf

[28] Edmond W. W. Chan, Xiapu Luo, Weichao Li, Waiting W. T. Fok, Rocky K. C. Chang: *Measurement of Loss Pairs in Network Paths*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p89.pdf

[29] Edmond W. W. Chan, Xiapu Luo, Waiting W. T. Fok, Weichao Li, and Rocky K. C. Chang: *Non-cooperative Diagnosis of Submarine Cable Faults*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Chan.pdf

[30] Rocky K. C. Chang: *OneProbe: Measuring network path quality with TCP data-packet pairs*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_rchang.pdf

[31] Chia-Wei Chang1, Alexandre Gerber, Bill LinSubhabrata Sen, Oliver Spatscheck: *Network DVR: A Programmable Framework for Application-Aware Trace Collection*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/20.pdf

[32] Yingying Chen, Sourabh Jain, Vijay Kumar Adhikari, Zhi-Li Zhang: *Characterizing Roles of Front-end Servers in End-to-EndPerformance of Dynamic Content Distribution*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p559.pdf

[33] David Choffnes: *EdgeScope: Exposing the View of the Edge of the Network*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_dchoffnes.pdf

[34] Lucas Di Cioccio1, Renata Teixeira,Catherine Rosenberg, Martin May: *Home Network Performance Diagnosis*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_rteixeira.pdf.

[35] Lucas Di Cioccio1, Renata Teixeira,Catherine Rosenberg, Martin May: *HostView:Annotating end-host performance measurements with user feedback*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_djoumblatt.pdf.

[36] Italo Cunha, Renata Teixeira, Christophe Diot: *Measuring and Characterizing End-to-End Route Dynamics in the Presence of Load Balancing*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Cunha.pdf

[37] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, Antonio Pescape: *Analysis of Country-wide Internet Outages Caused by Censorship*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p1.pdf

[38] Pralhad Deshpande, Xiaoxiao Hou, and Samir R. Das: *Performance Comparison of 3G and Metro-Scale WiFi forVehicular Network Access*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p301.pdf

[39] Amogh Dhamdhere, Lee Breslau, Nick Duffield, Cheng Ee,Alexandre Gerber, Carsten Lund,Subhabrata Sen: *FlowRoute: Inferring Forwarding Table Updates UsingPassive Flow-level Measurements*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p315.pdf

[40] Yuan Ding, Yuan Du, Yingkai Hu, Zhengye Liu, Luqin Wang, Keith W. Ross, Anindya Ghose: *Broadcast Yourself: Understanding YouTube Uploaders*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p361.pdf

[41] Benoit Donnet: *Internet Topology Discovery Through mrinfo Probing*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_bdonnet.ppt

[42] Nandita Dukkipati, Matt Mathis, Yuchung Cheng, Monia Ghobadi: *Proportional Rate Reduction for TCP*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p155.pdf

[43] Brian Eriksson, Paul Barford, Rhys Bowden, Nick Duffield, Joel Sommersm, Matthew Roughan: *BasisDetect : A Model-based Network Event Detection Framework*, IMC, 2011. http://conferences.sigcomm.org/imc/2010/papers/p451.pdf

[44] Jeffrey Erman, Alexandre Gerber, K.K. Ramakrishnan, Subhabrata Sen, Oliver Spatscheck: *Over The Top Video: The Gorilla in Cellular Networks*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p127.pdf

[45] Hossein Falaki, Dimitrios Lymberopoulos, Ratul Mahajan, Srikanth Kandula, Deborah Estrin: *A First Look at Traffic on Smartphones*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p281.pdf

[46] Xun Fan, John Heidemann: *Selecting Representative IP Addresses for Internet Topology Studies*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p411.pdf

[47] Nick Feamster: *Characterizing VLAN-induced sharing in a campus network*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_nfeamster.pdf

[48] Nick Feamster, Srikanth Sundaresan, Walter de Donato, Renata Teixeira: *The Case for Measurementsfrom Home Network Gateways*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_nfeamster.pdf

[49] Alessandro Finamore, Marco Mellia, Maurizio M. Munafo, Ruben Torres, Sanjay G. Rao: *YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p345.pdf

[50] Daniel A. Freedmanxz , Tudor Marianx , Jennifer H. Leey ,Ken Birmanx , Hakim Weatherspoonx , Chris Xu: *Exact Temporal Characterization of 10 Gbps Optical Wide-Area Network*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p342.pdf

[51] Francesco Fusco, Deri Luca: *High Speed Network Traffic Analysis with Commodity Multi-core Systems*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p218.pdf.

[52] Kaustubh Gadkari, Daniel Massey, and Christos Papadopoulos: *Dynamics of Prefix Usage at an Edge Router*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Gadkari.pdf

[53] Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, Ben Y. Zhao: *Detecting and Characterizing Social Spam Campaigns*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p35.pdf

[54] Richard Gass, Christophe Diot: *An Experimental Performance Comparison of3G and Wi-Fi*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/8.pdf

[55] Aaron Gember, Ashok Anand, and Aditya Akella: *A Comparative Study of Handheld and Non-HandheldTraffic in Campus Wi-Fi Networks*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Gember.pdf

[56] Alexandre Gerber, Jeffrey Pang, Oliver Spatscheck, Shobha Venkataraman: *Speed Testing without Speed Tests: Estimating AchievableDownload Speed from Passive Measurements*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p424.pdf

[57] Denisa Ghita, Katerina Argyraki, Patrick Thiran: *Network Tomography on Correlated Links*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p225.pdf

[58] Oana Goga, Renata Teixeira: *Speed Measurements for Residential Internet Access*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_ogoga.pdf

[59] Vijay Gopalakrishnan, Rittwik Jana, K. K. Ramakrishnan, Deborah F. Swayne, Vinay A. Vaishampayan: *Understanding Couch Potatoes: Measurement and Modeling of Interactive Usage of IPTV at large scale*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p225.pdf

[60] GregorMaier, Fabian Schneider, and Anja Feldmann: *A First Look at Mobile Hand-held Device Traffic*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/17.pdf

[61] Saikat Guha, Bin Cheng, Paul Francis: *Challenges in Measuring Online Advertising Systems*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p81.pdf

[62] Thom Haddow, Sing Wang Ho, Jonathan Ledlie, Cristian Lumezanu, Moez Draief, and Peter Pietzuch: *On the Feasibility of Bandwidth Detouring*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Haddow.pdf

[63] Shuang Hao, Nick Feamster, Georgia Tech, Ramakant Pandrangi: *Monitoring the Initial DNS Behavior of Malicious Domains*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p269.pdf

[64] Seppo Hatonen, Aki Nyrhinen, Lars Eggert, Stephen Strowes, Pasi Sarolahti, Markku Kojo: *An Experimental Study of Home Gateway Characteristics*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p260.pdf

[65] Ralph Holz, Lothar Braun, Nils Kammenhuber, Georg Carle: *The SSL Landscape - A Thorough Analysis of the X.509 PKI Using Active and Passive Measurements* ,IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p427.pdf

[66] Bradley Huffaker, Amogh Dhamdhere, Marina Fomenkov: *AS Assignment for Routers*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_bhuffaker_asassignment.pdf

[67] Bradley Huffaker, Amogh Dhamdhere, Marina Fomenkov: *Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers*,PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/11.pdf

[68] Young Hyun: *Ark update and measurement case study*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_yhyun_ark.pdf

[69] Young Hyun: *Internet Topology Data Kit*, AIMS, 2011. http://www.caida.org/publications/presentations/2011/itdk/itdk.pdf

[70] Julio Ibarra: *AMPATH update*, AIMS, 2010.

[71] Julio Ibarra: *AMPATH update*, AIMS, 2010. www.caida.org/workshops/isma/1002/slides/aims1002_jibarra.pdf

[72] Sunghwan Ihm, Vivek S. Pai: *Towards Understanding Modern Web Traffic*,IMC, 2011. http://www.cs.princeton.edu/~sihm/papers/webtraffic-sigmetrics11.pdf

[73] Jelena Isacenkova, Davide Balzarotti: *Measurement and Evaluation of a Real World Deploymentof a Challenge-Response Spam Filter*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p413.pdf

[74] Jing Jiang, Christo Wilson, Xiao Wang, Peng Huang, Wenpeng Sha, Yafei Dai and Ben Y. Zhao: *Understanding Latent Interactions in Online Social Networks*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p369.pdf

[75] Yu Jin, Nick Duffield, Alexandre Gerber, Patrick Haffner, Wen-Ling Hsu, Guy Jacobson, Subhabrata Sen, Shobha Venkataraman, Zhi-Li Zhang: *Large-scale App-based Reporting of Customer Problemsin Cellular Networks: Potential and Limitations*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p13.pdf

[76] Diana Joumblatt, Renata Teixeira, Jaideep Chandrashekar, Nina Taft: *Performance of Networked Applications:The Challenges in Capturing the User's Perception*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p37.pdf

[77] Andrew J. Kalafut, Minaxi Gupta, Christopher A. Cole, Lei Chen, Nathan E. Myers: *An Empirical Study of Orphan DNS Servers in the Internet*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p308.pdf

[78] Partha Kanuparthy, Constantine Dovrolis: *DiffProbe: Detecting ISP Service Discrimination*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_pkanuparthy.ppt

[79] Partha Kanuparthy, Constantine Dovrolis: *End-to-end Methods forTraffic Shaping Detection, Performance Problem Diagnosis, Home Wireless Troubleshooting*, AIMS, 2010. http://www.caida.org/workshops/isma/1102/slides/aims1102_pkanuparthy.pdf

[80] Partha Kanuparthy, Constantine Dovrolis: *ShaperProbe: End-to-end Detection of ISP Traffic Shapingusing Active Methods*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p473.pdf

[81] Hakan Karde, Talha Öz, David Shelly, Mehmet H. Gune: *Cheleby: An Internet TopologyMapping System*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_hkardes.pdf

[82] Daniel Karrenberg: *The RIPE NCC Network Measurement Data Repository*,AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_dkarrenberg.pdf

[83] Ethan Katz-Bassett, David Choffnes, Colin Scott, Harsha Madhyastha, Arvind Krishnamurthy and Tom Anderson: *Failure Isolation in the Wide Area*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_dchoffnes.pdf

[84] Ethan Katz-Bassett, Harsha V. Madhyastha,Vijay K. Adhikari, Colin Scott, Justine Sherry, Peter van Wesep,Arvind Krishnamurthy, Thomas Anderson: *Reverse Traceroute*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_ekatzbassett.pdf

[85] Ken Keys, Young Hyun, Matthew Luckie: *Internet-Scale Alias Resolution with MIDAR*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_yhyun_midar.pdf

[86] Robert Kisteleki: *INRDB theInternet Number Resource Database*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_rkisteleki.pdf.

[87] Robert Kisteleki: *RIPE Atlas*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_rkisteleki.pdf

[88] Christian Kreibich, Nicholas Weaver, Chris Kanich, Weidong Cui, Vern Paxson: *GQ: Practical Containmentfor Measuring Modern Malware Systems*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p397.pdf

[89] Christian Kreibichy, Nicholas Weavery, Gregor Maiery, Boris Nechaev, Vern Paxson: *Experiences from Netalyzr with Engaging Users in End-System Measurement*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p25.pdf

[90] Katrina LaCurts, Hari Balakrishnan: *Measurement and Analysis of Real-World 802.11 MeshNetworks*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p123.pdf

[91] Markus Laner, Philipp Svoboda, Eduard Hasenleithner, Markus Rupp: *Dissecting 3G Uplink Delay byMeasuring in an Operational HSPA Network*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Laner.pdf

[92] Changhyun Lee, DK Lee, Yung Yi, and Sue Moon: *Operating a Network Link at 100 percent*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Lee.pdf

[93] Derek Leonard, Dmitri Loguinov: *Demystifying Service Discovery: Implementing an Internet-Wide Scanner*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p109.pdf

[94] Kyriaki Levanti, Sihyung Lee, and Hyong S. Kim: *On Reducing the Impact of Interdomain Route Changes*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Levanti.pdf

[95] Yuheng Li, Yiping Zhang, Ruixi Yuan: *Measurement and Analysis of a Large Scale CommercialMobile Internet TV System*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p209.pdf

[96] Yabing Liu, Krishna P. GummadiBalachander Krishnamurthy, Alan Mislove: *Analyzing Facebook Privacy Settings: User Expectations vs. Reality*, IMC, 2011. http://conferences.sigcomm.org/imc/2010/papers/p281.pdf

[97] Matthew Luckie, David Murrell: *Inference of False Links in Traceroute Graphs*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_mluckie.pdf.

[98] Matthew Luckie, Ben Stasiewicz: *Measuring Path MTU Discovery Behaviour*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p102.pdf

[99] Matthew Luckie: *Scamper: a Scalable and Extensible Packet Prober forActive Measurement of the Internet*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p239.pdf

[100] Cristian Lumezanu, Katherine Guo, Neil Spring, Bobby Bhattacharjee: *The Effect of Packet Loss on Redundancy Elimination in Cellular Wireless Networks*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p404.pdf

[101] Harsha V. Madhyastha: *iPlane Status*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_hmadhyastha.ppt.

[102] Gregor Maier, Fabian Schneider, and Anja Feldmann: *NAT usage in Residential Broadband Networks*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Maier.pdf

[103] Ernest McCracken: *NetViews: Dual Plane Internet Monitoring in Real Time*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_emccracken.pdf

[104] Tony McGregor, Shane Alcock, Daniel Karrenberg: *The RIPE NCC Internet Measurement Data Repository*,PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/12.pdf

[105] Tony McGregor, Shane Alcock, Vern Paxson, Mark Allman: *The RIPE NCC Internet Measurement DataRepository*,PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/12.pdf

[106] Jakub Mikians, Pere Barlet-Ros, Josep Sanjuas-Cuxart, and Josep Sole-Pareta: *A practical approach to portscan detection in very high-speed links*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Mikians.pdf

[107] Robert Miller, Warren Matthews, Constantine Dovrolis: *Internet usage at elementary, middle and high schools: A first look at K-12 traffic from two USGeorgia counties*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/16.pdf

[108] J. Scott Miller, Amit Mondal, Rahul Potharaju, Peter A. Dinda, Aleksandar Kuzmanovic: *Understanding End-user Perception of Network Problems*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p43.pdf

[109] Abedelaziz Mohaisen, Aaram Yun, Yongdae Kim: *Measuring the Mixing Time of Social Graphs*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p383.pdf

[110] Ricky K. P. Mok, Edmond W. W. Chan, Xiapu Luo, and Rocky K. C. Chang: *Inferring the QoE of HTTP Video Streaming fromUser-Viewing Activities*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p31.pdf

[111] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage and Geoffrey M. Voelker: *An Analysis of Underground Forums*, IMC, 2011. http://conferences.sigcomm.org/imc/2010/papers/p281.pdf

[112] Pascal Merindol, Benoit Donnet,Olivier Bonaventure, Jean-Jacques Pansiot: *On the Impact of Layer-2 on Node Degree Distribution*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p179.pdf

[113] Tu Ouyang, Soumya Ray, Michael Rabinovich, Mark Allman: *Can Network Characteristics Detect Spam Effectively ina Stand-Alone Enterprise?*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Ouyang.pdf

[114] Talha Oz, Hakan Kardes, Mehmet Gunes: *Subnet-level Internet Mapper*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_toz.pdf

[115] Jean-Jacques Pansiot, Pascal Merindol,Benoit Donnet, Olivier Bonaventure: *Extracting Intra-Domain Topologyfrom mrinfo Probing*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/9.pdf

[116] Jong Han Park, Dan Jen, Mohit Lad, Shane Amante, Danny McPherson, Lixia Zhang: *Investigating occurrence of duplicate updatesin BGP announcements*, IMC, 2011. http://www.pam2010.ethz.ch/papers/full-length/2.pdf.

[117] Abhinav Pathak, Ming Zhang, Y. Charlie Hu, Ratul Mahajan, Dave Maltz: *Latency Inflation with MPLS-based Traffic Engineering*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p463.pdf

[118] Abhinav Pathak, Y. AngelaWang, Cheng Huang, Albert Greenberg, Y. Charlie Hu, Randy Kern, Jin Li, Keith W. Ross: *Measuring and Evaluating TCP Splitting for CloudServices*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/5.pdf

[119] Caleb Phillips, Scott Raynel, Jamie Curtis, Sam Bartels, Douglas Sicker, Dirk Grunwald, Tony McGregor: *The Efficacy of Path Loss Models for Fixed Rural Wireless Links*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Phillips.pdf

[120] Ingmar Poese, Benjamin Frank, Bernhard Ager, Georgios Smaragdakis, Anja Feldmann: *Improving Content Delivery UsingProvider-aided Distance Information*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p22.pdf.

[121] Rahul Potharaju, Jeff Seibert, Sonia Fahmy, and Cristina Nita-Rotaru, Purdue University: *Omnify: Investigating the Visibility and Effectiveness of Copyright Monitors*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Potharaju.pdf

[122] Feng Qian, Zhaoguang Wang, Alexandre Gerber, Z. Morley Mao, Subhabrata Sen, Oliver Spatscheck: *Characterizing Radio Resource Allocation for 3G Networks*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p137.pdf

[123] Tongqing Qium, Junlan Feng, Zihui GeJia Wang, Jun (Jim) Xu, Jennifer Yate: *Listen to Me if You can: Tracking User Experience of Mobile Network on Social Media*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p288.pdf

[124] Lin Quan, John Heidemann: *On the Characteristics and Reasonsof Long-lived Internet Flows*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p444.pdf

[125] Elias Raftopoulos, Xenofontas Dimitropoulos: *Detecting, Validating and CharacterizingComputer Infections in the Wild*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p29.pdf

[126] Amir Hassan Rasti, Reza Rejaie, Walter Willinger: *Characterizing the Global Impact of P2P Overlays on the AS-Level Underlay*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/1.pdf

[127] Amir H. Rasti, Nazanin Magharei, Reza Rejaie, Walter Willinger: *Eyeball ASes: From Geography to Connectivity*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p192.pdf

[128] Shravan Rayanchu, Ashish Patro, Suman Banerjee: *Detecting Non-WiFi RF Devicesusing Commodity WiFi Hardware*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p137.pdf

[129] Edward Rhyne: *DHS S and T Cyber Security Division Overview*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_erhyne.pdf

[130] Bruno Ribeiro, Don Towsley: *Estimating and Sampling Graphs withMultidimensional Random Walks*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p390.pdf

[131] Tiago Rodrigues, Fabricio Benevenuto, Meeyoung ChaKrishna P. Gummadi, Virgilio Almeida: *On Word-of-Mouth Based Discovery of the Web*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p381.pdf

[132] Josep Sanjuas-Cuxart, Pere Barlet-Ros, Nick Duffield, Rao Kompella: *Sketching the Delay: Tracking TemporallyUncorrelated Flow-Level Latencies*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p483.pdf

[133] Raimund Schatz, Sebastian Egger: *Vienna Surfing - Assessing Mobile Broadband Quality in the Field*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p19.pdf

[134] Dominik Schatzmann, Wolfgang Mühlbauer, Thrasyvoulos Spyropoulos,Xenofontas Dimitropoulos: *Digging into HTTPS: Flow-Based Classificationof Webmail Traffic*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p322.pdf

[135] Dominik Schatzmann, Simon Leinen,Jochen Kogel, and Wolfgang Muhlbauer: *FACT: Flow-based Approachfor Connectivity Tracking*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Schatzmann.pdf

[136] Aaron Schulman, Neil Spring: *Pingin' in the Rain*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p19.pdf

[137] Yaron Schwartz, Yuval Shavitt, and Udi Weinsberg: *A Measurement Study of the Origins ofEnd-to-End Delay Variations*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/3.pdf

[138] Yaron Schwartz, Yuval Shavitt, and Udi Weinsberg: *A Measurement Study of the Origins ofEnd-to-End Delay Variations*, AIMS, 2010. http://www.pam2010.ethz.ch/papers/full-length/3.pdf

[139] Vyas Sekar, Michael K Reiter, Hui Zhang: *Revisiting the Case for a Minimalist Approachfor Network Flow Monitoring*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p328.pdf

[140] Yuji Sekiya, Kenjiro Cho: *Gulliver Project - status update in 2009*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_ysekiya.pdf

[141] Sayandeep Sen, Jongwon Yoon, Joshua Hare, Justin Ormont, Suman Banerjee: *Can They Hear Me Now?: A Case for a Client-assistedApproach to Monitoring Wide-area Wireless Networks*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p99.pdf

[142] Justine Sherry, Ethan Katz-Bassett, Mary Pimenova, Harsha V. Madhyastha, Thomas Anderson, Arvind Krishnamurthy: *Resolving IP Aliases with Prespecified Timestamps*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p172.pdf

[143] Matt Smith, Dmitri Loguinov: *Enabling High-PerformanceInternet-Wide Measurements on Windows*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/13.pdf

[144] Han Hee Song, Zihui Ge, Ajay Mahimkar, Jia Wang, Jennifer Yates, Yin Zhang, Andrea Basso, Min Chen: *Q-score: Proactive Service Quality Assessment in a Large IPTV System*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p195.pdf

[145] Brett Stone-Gross, Ryan Stevens,Richard Kemmerer, Christopher Kruegel, Giovanni Vigna, Apostolis Zarras: *Understanding Fraudulent Activities in Online Ad Exchanges*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p279.pdf

[146] Mikhail Strizhov: *Real-Time BGP Data Access*, AIMS, 2010. http://www.caida.org/workshops/isma/1102/slides/aims1102_mstrizhov.pdf

[147] Peng Sun, Minlan Yu, Michael J. Freedman, and Jennifer Rexford: *Identifying Performance Bottlenecks in CDNs through TCP-Level Monitoring*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p49.pdf.

[148] Srikanth Sundaresan, Walter de Donato,Nick Feamster, Renata Teixeira, AntonioPescape: *Benchmarking Broadband InternetPerformance*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_ssundaresan.pdf

[149] Harika Tandra: *Distributed Virtual Network Operations Center DVNOC - Towards Federated and Customer-focused Cyberinfrastructure*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_htandra.pdf

[150] Renata Teixeira, Nick Feamster: *Which factors affect access network performance?*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_rteixeira.pdf

[151] Kurt Thomasy, Chris Griery, Vern Paxsony, Dawn Song: *Suspended Accounts in Retrospect: An Analysis of Twitter Spam*, IMC, 2011. http://conferences.sigcomm.org/imc/2011/docs/p243.pdf

[152] Brian Tierney: *perfSONAR Deployment onESnet*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_btierney.pdf

[153] M. Engin Tozal, Kamil Sarac: *Network Layer Internet Topology Construction*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_ksarac_mtozal.pdf

[154] M. Engin Tozal, Kamil Sarac: *PalmTree: IP Alias Resolution Algorithm with Linear Probing Complexity*, IMC, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_mtozal_palmtree.pptx

[155] M. Engin Tozal, Kamil Sarac: *TraceNET: An Internet Topology Data Collector*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p356.pdf

[156] M. Engin Tozal, Kamil Sarac: *TraceNET: An Internet Topology Data Collector*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_mtozal_tracenet.pptx

[157] Brian Trammell, Bernhard Tellenbach, Dominik Schatzmann, Martin Burkhart: *Peeling Away Timing Error in NetFlow Data*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Trammell.pdf

[158] Gabor Vattay: *Wide side of the Internet: Benford type distributions in Internet data*, AIMS, 2011. http://www.caida.org/workshops/isma/1102/slides/aims1102_gvattay.pdf

[159] Katarzyna Wac, Selim Ickin, Jin-Hyuk Hong, Lucjan Janowski, Markus Fiedler, Anind K. Dey: *Studying the Experience of Mobile Applications Used in Different Contexts of Daily Life*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p7.pdf

[160] Xiao Sophia Wang, David Choffnes, Patrick Gage Kelley, Ben Greenstein, David Wetherall: *Measuring and Predicting Web Login Safety*, W-MUST, 2011. http://conferences.sigcomm.org/sigcomm/2011/papers/w-must/p55.pdf

[161] Xiaofei Wang, Seungbae Kim, Ted Taekyoung Kwon, Hyun-chul Kim, and Yanghee Choi: *Unveiling the BitTorrent Performance in Mobile WiMAX Networks*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Wang.pdf

[162] Masafumi Watari, Atsuo Tachibana, Shigehiro Ano: *Inferring the Origin of Routing Changes basedon Preferred Path Changes*, PAM, 2011. http://pam2011.gatech.edu/papers/pam2011--Watari.pdf.

[163] Rhiannon Weaver: *A Probabilistic Population Study of theConficker-C Botnet*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/19.pdf

[164] Chris Wilcox, Christos Papadopoulos: *Correlating Spam Acitivity with IP Address Characterisics*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_cpapadopoulos.pdf

[165] Craig Wills, Mark Claypool, Artur Janc, Alan Ritacco: *Development of a User-Centered Network Measurement Platform*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_cwills.pdf

[166] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, Geoff Huston: *Internet Background Radiation Revisited*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p62.pdf.

[167] Qiang Xu, Jeffrey Erman, Alexandre Gerber, Z. Morley Mao, Jeffrey Pang, Shobha Venkataraman: *Identifying Diverse Usage Behaviors of Smartphone Apps*, IMC, 2011. http://conferences.sigcomm.org/imc/2

[168] Sandeep Yadav, Ashwath Kumar Krishna Reddy, A.L. Narasimha Reddy, Supranamaya Ranjan: *Detecting Algorithmically Generated Malicious Domain Names*, IMC, 2010. http://dx.doi.org/10.1145/1879141.1879148

[169] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, Yafei Dai: *Uncovering Social Network Sybils in the Wild*, IMC, 2011. http://doi.acm.org/10.1145/2068816.2068841

[170] Matt Zekauskas: *perfSONAR Overview*, AIMS, 2010. http://www.caida.org/workshops/isma/1002/slides/aims1002_mzekauskas.pdf

[171] Yu Zhang, Ricardo Oliveira, Hongli Zhang, Lixia Zhang: *Quantifying the Pitfalls of Traceroute in AS Connectivity Inference*, PAM, 2010. http://www.pam2010.ethz.ch/papers/full-length/10.pdf

[172] Chao Michael Zhang, Vern Paxson: *Detecting and Analyzing Automated Activity on Twitter*, PAM, 2011. http://dx.doi.org/10.1007/978-3-642-19260-9_11

[173] Renje Zhou, Samamon Khemmarat, Lixin Gao: *The Impact of YouTube Recommendation System on Video Views*, IMC, 2010. http://conferences.sigcomm.org/imc/2010/papers/p404.pdf

[174] Jia Zhou, Yanhua Li, Vijay Kumar Adhikari, Zhi-Li Zhang: *Counting YouTube Videos via Random Prefix Sampling*, IMC, 2011. http://dx.doi.org/10.1145/2068816.2068851

[175] Cooperative Association for Internet Data Nalysis: ISMA 2010 AIMS-3 Workshop on Active Internet Measurements. 2010, http://www.caida.org/workshops/isma/1002/

[176] Cooperative Association for Internet Data Analysis: ISMA 2011 AIMS-3 Workshop on Active Internet Measurements. 2011, http://www.caida.org/workshops/isma/1102/

[177] Internet Measurement Conference: IMC Conference. 2010, http://conferences.sigcomm.org/imc/2010/imc-papers.html

[178] Internet Measurement Conference: IMC Conference. 2011, http://conferences.sigcomm.org/imc/2011/program.htm

[179] Passive and Active Measurement Conference: PAM2010. 2010, http://www.pam2010.ethz.ch/

[180] Passive and Active Measurement Conference: PAM2011. 2011, http://pam2011.gatech.edu/

[181] Workshop on Measurements Up the STack: W-MUST. 2011, http://conferences.sigcomm.org/sigcomm/2011/workshops/W-MUST/

1.6 Annex

1.6.1 Papers in Categories

1.6.1.1 Internet Usage Measurement

Title	Conference	Year	Source
A First Look at Traffic on Smartphones	IMC	2010	[45]
Listen to Me if You can: Tracking User Experience of Mobile Network on Social Media	IMC	2010	[123]
Understanding Latent Interactions in Online Social Networks	IMC	2010	[74]
The Impact of YouTube Recommendation System on Video Views	IMC	2011	[173]
Analyzing Facebook Privacy Settings: User Expectations vs. Reality	IMC	2011	[96]
An Analysis of Underground Forums	IMC	2011	[111]
Q-score: Proactive Service Quality Assessment in a Large IPTV System	IMC	2011	[144]
Measurement and Analysis of a Large Scale Commercial Mobile Internet TV System	IMC	2011	[95]
Understanding Couch Potatoes: Measurement and Modeling of Interactive Usage of IPTV at large scale	IMC	2011	[59]
Towards Understanding Modern Web Traffic	IMC	2011	[72]
Identifying Diverse Usage Behaviors of Smartphone Apps	IMC	2011	[167]
YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience	IMC	2011	[49]
Broadcast Yourself: Understanding YouTube Uploaders	IMC	2011	[40]
On Word-of-Mouth Based Discovery of the Web	IMC	2011	[131]
Internet usage at elementary, middle and high schools: A first look at K-12 traffic from two US Georgia counties	IMC	2011	[107]
A First Look at Mobile Hand-held Device Traffic	PAM	2011	[60]
NAT usage in Residential Broadband Networks	PAM	2011	[102]
A Comparative Study of Handheld and Non-Handheld Traffic in Campus Wi-Fi Networks	PAM	2011	[55]
The Network from Above and Below	W-MUST	2011	[21]
Studying the Experience of Mobile Applications Used in Different Contexts of Daily Life	W-MUST	2011	[159]
Large-scale App-based Reporting of Customer Problems in Cellular Networks: Potential and Limitations	W-MUST	2011	[75]
Vienna Surfing - Assessing Mobile Broadband Quality in the Field	W-MUST	2011	[133]
Experiences from Netalyzr with Engaging Users in End-System Measurement	W-MUST	2011	[89]
Inferring the QoE of HTTP Video Streaming from User-Viewing Activities	W-MUST	2011	[110]
Performance of Networked Applications: The Challenges in Capturing the User's Perception	W-MUST	2011	[76]
Understanding End-user Perception of Network Problems	W-MUST	2011	[108]

1.6.1.2 Measurement Techniques

Title	Conference	Year	Source
A First Look at Traffic on Smartphones	IMC	2010	[45]

Internet-wide systems need Internet-wide measurement platforms	AIMS	2010	[22]
Reverse Traceroute	AIMS	2010	[84]
End-to-end Methods for Traffic Shaping Detection, Performance Problem Diagnosis, Home Wireless Troubleshooting	AIMS	2011	[79]
Real-Time BGP Data Access	AIMS	2011	[146]
NetViews: Dual Plane Internet Monitoring in Real Time	AIMS	2011	[103]
Failure Isolation in the Wide Area	AIMS	2011	[83]
Measuring the current state of ECN support in servers, clients, and routers	AIMS	2011	[13]
OneProbe: Measuring network path quality with TCP data-packet pairs	AIMS	2011	[30]
OneProbe: Challenges in Measuring Online Advertising Systems	IMC	2010	[61]
Measurement of Loss Pairs in Network Paths	IMC	2010	[28]
Measuring Path MTU Discovery Behaviour	IMC	2010	[98]
Demystifying Service Discovery: Implementing an Internet-Wide Scanner	IMC	2010	[93]
Measurement and Analysis of Real-World 802.11 Mesh Networks	IMC	2010	[90]
Resolving IP Aliases with Prespecified Timestamps	IMC	2010	[142]
On the Impact of Layer-2 on Node Degree Distribution	IMC	2010	[112]
Comparing and Improving Current Packet Capturing Solutions based on Commodity Hardware	IMC	2010	[20]
High Speed Network Traffic Analysis with Commodity Multi-core Systems	IMC	2010	[51]
Network Tomography on Correlated Links	IMC	2010	[57]
Scamper: a Scalable and Extensible Packet Prober for Active Measurement of the Internet	IMC	2010	[99]
FlowRoute: Inferring Forwarding Table Updates Using Passive Flow-level Measurements	IMC	2010	[39]
Digging into HTTPS: Flow-Based Classification of Webmail Traffic	IMC	2010	[134]
Revisiting the Case for a Minimalist Approach for Network Flow Monitoring	IMC	2010	[139]
Exact Temporal Characterization of 10 Gbps Optical Wide-Area Network	IMC	2010	[50]
TraceNET: An Internet Topology Data Collector	IMC	2010	[155]
Measuring the Mixing Time of Social Graphs	IMC	2010	[109]
Estimating and Sampling Graphs with Multidimensional Random Walks	IMC	2010	[130]
Speed Testing without Speed Tests: Estimating Achievable Download Speed from Passive Measurements	IMC	2010	[56]
On the Characteristics and Reasons of Long-lived Internet Flows	IMC	2010	[124]
BasisDetect : A Model-based Network Event Detection Framework	IMC	2010	[43]
Temporally Oblivious Anomaly Detection on Large Networks Using Functional Peers	IMC	2010	[26]

Pingin' in the Rain	IMC	2011	[136]
Can They Hear Me Now?: A Case for a Client-assisted Approach to Monitoring Wide-area Wireless Networks	IMC	2011	[141]
Detecting Non-WiFi RF Devices using Commodity WiFi Hardware	IMC	2011	[128]
Measuring the State of ECN Readiness in Servers, Clients, and Routers	IMC	2011	[13]
Understanding Website Complexity: Measurements, Metrics, and Implications	IMC	2011	[23]
Counting YouTube Videos via Random Prefix Sampling	IMC	2011	[174]
ShaperProbe: End-to-end Detection of ISP Traffic Shaping using Active Methods	IMC	2011	[80]
Sketching the Delay: Tracking Temporally Uncorrelated Flow-Level Latencies	IMC	2011	[132]
vPF-Ring: Towards Wire-Speed Network Monitoring Using Virtual Machines	IMC	2011	[25]
Web Content Cartography	IMC	2011	[3]
Investigating occurrence of duplicate updates in BGP announcements	IMC	2011	[116]
A Measurement Study of the Origins of End-to-End Delay Variation	PAM	2010	[137]
Influence of the Packet Size on the One-Way Delay in 3G Networks	PAM	2010	[8]
Enabling High-Performance Internet-Wide Measurements on Windows	PAM	2010	[143]
Network DVR: A Programmable Framework for Application-Aware Trace Collection	PAM	2010	[31]
Dynamics of Prefix Usage at an Edge Router	PAM	2011	[31]
Detecting and Analyzing Automated Activity on Twitter	PAM	2011	[172]
Dissecting 3G Uplink Delay by Measuring in an Operational HSPA Network	PAM	2011	[91]
Inferring the Origin of Routing Changes based on Preferred Path Changes	PAM	2011	[162]
Peeling Away Timing Error in NetFlow Data	PAM	2011	[157]
FACT: Flow-based Approach for Connectivity Tracking	PAM	2011	[135]
Non-cooperative Diagnosis of Submarine Cable Faults	PAM	2011	[29]
Crowdsourcing ISP Characterization to The Network Edge	W-MUST	2011	[18]

1.6.1.3 Security

Title	Conference	Year	Source
Correlating Spam Acitivity with IP Address Characterisics	AIMS	2010	[164]
Detecting and Characterizing Social Spam Campaigns	IMC	2010	[53]
Detecting Algorithmically Generated Malicious Domain Names	IMC	2010	[168]
Internet Background Radiation Revisited	IMC	2010	[166]

An Empirical Study of Orphan DNS Servers in the Internet	IMC	2010	[77]
Analysis of Country-wide Internet Outages Caused by Censorship	IMC	2011	[37]
Detecting, Validating and Characterizing Computer Infections in the Wild	IMC	2011	[125]
I Know Where You Are and What You Are Sharing: Exploiting P2P Communications to Invade Users Privacy	IMC	2011	[19]
Suspended Accounts in Retrospect: An Analysis of Twitter Spam	IMC	2011	[151]
Uncovering Social Network Sybils in the Wild	IMC	2011	[169]
Monitoring the Initial DNS Behavior of Malicious Domains	IMC	2011	[63]
Understanding Fraudulent Activities in Online Ad Exchanges	IMC	2011	[145]
GQ: Practical Containment for Measuring Modern Malware Systems	IMC	2011	[88]
Measurement and Evaluation of a Real World Deployment of a Challenge-Response Spam Filter	IMC	2011	[73]
The SSLLandscape - A Thorough Analysis of the X.509 PKI Using Active and Passive Measurements	IMC	2011	[65]
A Probabilistic Population Study of the Conficker-C Botnet	PAM	2010	[163]
Web Timeouts and Their Implications	PAM	2010	[6]
Can Network Characteristics Detect Spam Effectively in a Stand-Alone Enterprise?	PAM	2011	[113]
A practical approach to portscan detection in very high-speed links	PAM	2011	[106]
Omnify: Investigating the Visibility and Effectiveness of Copyright Monitors	PAM	2011	[121]
Measuring and Predicting Web Login Safety	PAM	2011	[160]

1.6.1.4 Performance

Title	Conference	Year	Source
Characterizing VLAN-induced sharing in a campus network	AIMS	2010	[47]
Which factors affect access network performance?	AIMS	2010	[150]
A Measurement Study of the Origins of End-to-End Delay Variations	AIMS	2010	[138]
Gulliver Project - status update in 2009	AIMS	2010	[140]
DiffProbe: Detecting ISP Service Discrimination	AIMS	2010	[78]
Home Network Performance Diagnosis	AIMS	2011	[34]
The Case for Measurements from Home Network Gateways	AIMS	2011	[48]
Benchmarking Broadband Internet Performance	AIMS	2011	[148]
Speed Measurements for Residential Internet Access	AIMS	2011	[58]
HostView: Annotating end-host performance measurements with user feedback	AIMS	2011	[35]
Comparing DNS Resolvers in the Wild	IMC	2010	[2]

Improving Content Delivery Using Provider-aided Distance Information	IMC	2010	[120]
Characterizing Radio Resource Allocation for 3G Networks	IMC	2010	[122]
An Experimental Study of Home Gateway Characteristics	IMC	2010	[64]
The Effect of Packet Loss on Redundancy Elimination in Cellular Wireless Networks	IMC	2010	[100]
Performance Comparison of 3G and Metro-Scale WiFi for Vehicular Network Access	IMC	2010	[38]
YouTube Traffic Dynamics and Its Interplay with a Tier-1 ISP: An ISP Perspective	IMC	2010	[1]
Over The Top Video: The Gorilla in Cellular Networks	IMC	2011	[44]
Proportional Rate Reduction for TCP	IMC	2011	[42]
Latency Inflation with MPLS-based Traffic Engineering	IMC	2011	[117]
Characterizing Roles of Front-end Servers in End-to-End Performance of Dynamic Content Distribution	IMC	2011	[32]
Overclocking the Yahoo! CDN for Faster Web Page Loads	IMC	2011	[5]
Measuring and Evaluating TCP Splitting for Cloud Services	PAM	2010	[118]
An Experimental Performance Comparison of 3G and Wi-Fi	PAM	2010	[54]
Operating a Network Link at 100 percent	PAM	2011	[92]
The Efficacy of Path Loss Models for Fixed Rural Wireless Links	PAM	2011	[119]
On the Feasibility of Bandwidth Detouring	PAM	2011	[62]
On Reducing the Impact of Interdomain Route Changes	PAM	2011	[94]
Unveiling the BitTorrent Performance in Mobile WiMAX Networks	PAM	2011	[161]
Identifying Performance Bottlenecks in CDNs through TCP-Level Monitoring	W-MUST	2011	[147]

1.6.1.5 Topology

Title	Conference	Year	Source
Directed Probing for Efficient and Accurate Active Measurements	AIMS	2010	[15]
Internet Topology Discovery Through mrinfo Probing	AIMS	2010	[41]
Overview of TopHat: Interconnecting the OneLab measurement infrastructures	AIMS	2010	[9]
iPlane Status	AIMS	2010	[101]
Ark update and measurement case study	AIMS	2010	[68]
PalmTree: IP Alias Resolution Algorithm with Linear Probing Complexity	AIMS	2010	[154]
Internet-Scale Alias Resolution with MIDAR	AIMS	2010	[85]
Inference of False Links in Traceroute Graphs	AIMS	2010	[85]
AS Assignment for Routers	AIMS	2010	[66]
Subnet-level Internet Mapper	AIMS	2010	[114]
Network Layer Internet Topology Construction	AIMS	2011	[153]

Subnet Based Internet Topology Generation	AIMS	2011	[4]
Primitives for Active Internet Topology Mapping: Toward High-Frequency Characterization	AIMS	2011	[16]
Wide side of the Internet: Benford type distributions in Internet data	AIMS	2011	[158]
Cheleby: An Internet Topology Mapping System	AIMS	2011	[81]
Primitives for Active Internet Topology Mapping: Toward High-Frequency Characterization	IMC	2010	[17]
Eyeball ASes: From Geography to Connectivity	IMC	2010	[127]
Towards an AS-to-Organization Map	IMC	2010	[24]
Selecting Representative IP Addresses for Internet Topology Studies	IMC	2010	[46]
Characterizing the Global Impact of P2P Overlays on the AS-Level Underlay	PAM	2010	[126]
Extracting Intra-Domain Topology from mrinfo Probing	PAM	2010	[115]
Quantifying the Pitfalls of Traceroute in AS Connectivity Inference	PAM	2010	[171]
Toward Topology Dualism: Improving the Accuracy of AS Annotations for Routers	PAM	2010	[67]
Measuring and Characterizing End-to-End Route Dynamics in the Presence of Load Balancing	PAM	2010	[36]

1.6.1.6 Cooperation

Title	Conference	Year	Source
AMPATH update	AIMS	2010	[70]
Development of a User-Centered Network Measurement Platform	AIMS	2010	[165]
	AIMS	2010	[170]
INRDB the Internet Number Resource Database	AIMS	2010	[86]
The RIPE NCC Network Measurement Data Repository	AIMS	2010	[82]
EdgeScope: Exposing the View of the Edge of the Networky	AIMS	2010	[33]
Internet Topology Data Kit	AIMS	2011	[69]
Some Internet Measurement Thoughts	AIMS	2011	[11]
DHS S and T Cyber Security Division Overview	AIMS	2011	[129]
Distributed Virtual Network Operations Center DVNOC - Towards Federated and Customer-focused Cyberinfrastructure	AIMS	2011	[149]
perfSONAR Deployment on ESnet	AIMS	2011	[152]
Update on TopHat and measurement system interconnection	AIMS	2011	[10]
RIPE Atlas	AIMS	2011	[87]
The RIPE NCC Internet Measurement Data Repository	PAM	2010	[104]
MOR: Monitoring and Measurements through the Onion Router	PAM	2010	[7]

Chapter 2

Mobile Payment Systems

Michael Blöchlinger

This report is an introduction into mobile payment. It shows different mobile payment methods with their advantages and disadvantages. Different business models, market drivers and constraints as well as risks of mobile payment solutions are presented. Furthermore this report gives an overview on Near Field Communication (NFC) and the fields of application of this technology. Finally market penetration and upcoming future of mobile payment are discussed.

Contents

2.1 Introduction

The modern world is a fast moving environment. Information technology enabled us to do things more efficient and faster. An inapprehensible network of connected information systems evolved during the last decades. Part of this modernisation process are payment methods. Today, online purchases and therefore online payment transactions are very popular. However Internet payment is not the only technologically enhanced payment method. Mobile payment (m-payment) is an actively discussed topic and there is quite some development in this area. This report shows different methods for m-payment and their corresponding advantages and disadvantages. Furthermore business models are discussed and the risks of these new methods are into perspective. Also an overview on Near Field Technology is presented. Finally the report evaluates market drivers and constraints, penetration of the market and presents forecasts on m-payment in the future.

2.2 Terminology

2.2.1 Micro and Macro Payment

Micro payment systems are capable of handling arbitrarily small amounts of money [15]. Micro payment is an electronic payment transaction in the range of about $10 to about one tenth of a cent. The transaction travels over the Internet or public network infrastructure. Macro payment is a payment transaction above $10.

2.2.2 E-Commerce

Electronic commerce is the paperless exchange of business information using electronic data interchange (EDI), e-mail, electronic bulletin boards, fax transmissions, and electronic funds transfer. It refers to Internet shopping, online stock and bond transactions, the downloading and selling of soft merchandise (software, documents, graphics, music, etc.), and business-to-business transactions [4].

2.2.3 M-Commerce

Mobile commerce is the buying and selling of goods and services through wireless handheld devices such as cellular telephone and personal digital assistants (PDAs) [25]. It is therefore a form of e-commerce in the sense of a electronic business transaction. M-commerce is also called next-generation e-commerce, defined by using mobile devices with internet capabilty to conduct business.

2.2.4 Point of Sale

Point of sale (POS) is the location where a transaction occurs. To conduct a POS payment, the buyer has to be physically present at the store where the payment transaction is made.

2.2.5 Mobile Payment

Mobile payment is defined as a payment that is carried out with a handheld device such as a mobile phone or a PDA (personal digital assistant) [14]. Payment involves a direct or indirect exchange of monetary values between parties. Handheld devices can be used at real POS, in e-commerce and in m-commerce [11]. According to the mobile payment forum [17], payment has evolved from the physical exchange of notes and coins, to writing checks, and to transferring payment card details either in person or at distance, over the

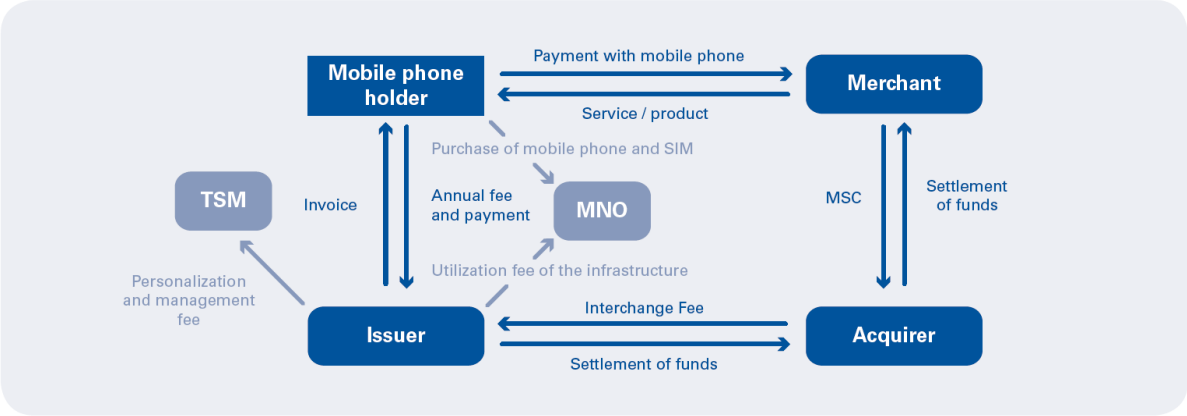


Figure 2.1: Six-party-scheme, possible scenario for mobile contactless payment in Switzerland [13]

phone or the Internet. This evolution has involved a shift from the physical transference of tangible tokens of value to an exchange of information between parties. The emergence of e-commerce has further digitized the payment process, where payment details are sent over open networks with no physical contact between the buyer and the seller [17].

2.3 Business Model

For an m-payment ecosystem to be successful in Switzerland, many partners have to work together. Therefore the alignment of activities and processes is essential. Figure 2.1 shows a possible mobile contactless payment scenario in Switzerland. A total of six parties are involved in this ecosystem. For this scenario the secure element lies is on the SIM card. Further information on secure elements can be found in Section 2.7.
The mobile phone holder purchases a SIM card from the mobile network operator (MNO). The trusted service manager (TSM) personalises the NFC-enabled SIM card with the payment application. The issuer is responsible for the execution of the payment transaction. The income for the issuers consists of the annual fees from the user and the interchange fee from the acquirer. Depending on the number of payment transactions, the issuer has to pay a license fee to the card scheme (Visa, MasterCard). The issuer also have to pay a chip management fee to the MNO and a personalisation and management fee to the TSM. All transactions are carried out by the acquirer and forwarded to the issuer. The acquirer receives a merchant service charge (MSC) but also has to pay a license fee to the card scheme.
A business model has to answer questions like: What are the roles and responsibilities of the key players? What is the financing strategy of the mobile payment transactions? What are the customer perceptions and needs?
Although this scenario is already rather complex, it is only a simplified illustration of possible key players for the mobile payment market in the future. Since the technology for contactless m-payment is new in Switzerland, many questions remain still unanswered. M-payment solutions are associated with immense technical and organisational complexity [13].

2.3.1 Business Model Comparison

Several actors play different roles in a m-payment system such as banks, operators or service providers. Each actor has to consider functionality, cost, security and benefit issues. The success of the m-payment service is based on the close interaction of the these roles. It is therefore important to select a suitable m-payment business model that

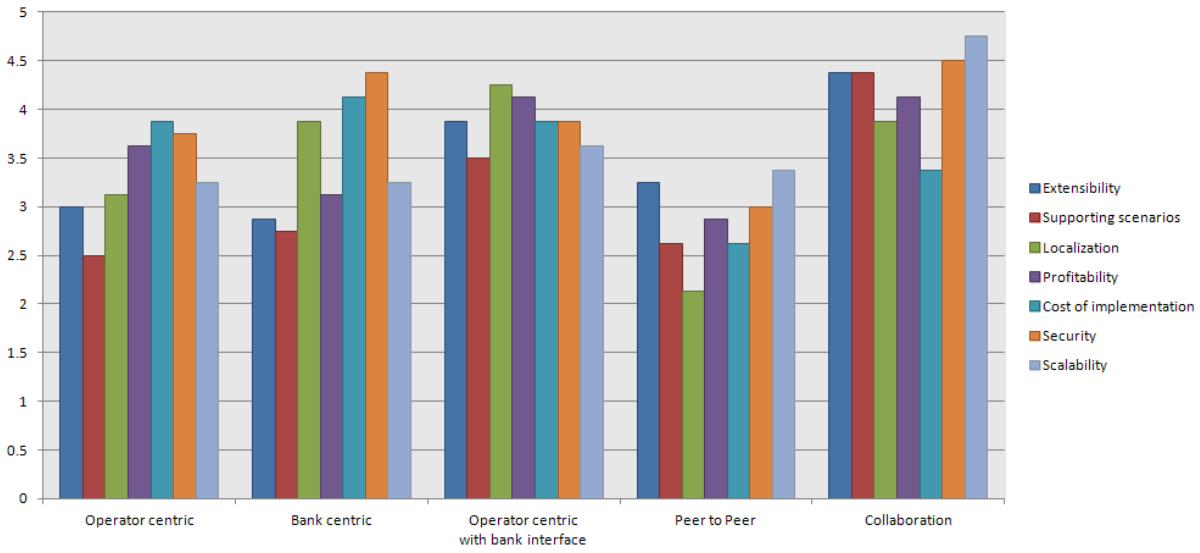


Figure 2.2: Comparison of different business models based on seven criteria

any actor is able to achieve its own purposes. Since there are many stakeholders, it is likely to encounter different or even conflictive stakeholder' requirements. Hence creating a business model implies balancing these requirements and mostly is a complex problem. A research team lead by Fatemeh Asghari at the Qom University in Tehran [6] proposed an approach on the selection of a business model based on the multi criteria decision making (MCDM) method. In order to apply the MCDM method, his team has surveyed five different mobile payment business models shown in Table 2.1.
Seven criteria for applying MCDM method were defined. Figure 2.2 shows the results of the comparison done by the Iranian research team. They interviewed serveral experts in the m-payment domain and let these experts rate the business models with points. The business models were rated in terms of extensibility, supporting scenarios, localization, profitability, cost of implementation security and scalability. According to their criteria, the overall best choice is the collaboration model. The worst model is the peer-to-peer model. However their goal was to find a suitable business model for m-payment solutions in Iran. The results could be different in other locations according to their characteristics.

2.4 Market Penetration

Figure 2.3 shows a selection of different mobile payment projects worldwide [19].
Netherlands, Amsterdam 2006: Gemalto developed a mobile payment application together with JBC credit card company. The application is called Mobile J/Speedy which was installed on Nokia6136 mobile phones. ViVotech manufactured the reader/writer devices to communicate with the mobile phone using NFC technology. KPN was involved as a mobile carrier and CCV Holland were leading the processing business organization.
Germany 2005: RMV (Rhein-Main Verkehrsverbund) developed bus ticketing system based on Nokia3220 mobile phones. Nokia was manufacturing the terminals and Vodafone participated as a system integrator.
France 2005: A project was started in Cean City where they installed NFC contactless payment terminals at supermarkets to realize cashless payment. Further m-payment for parking tickets was made possible through virtual tickets on the mobile phone. Samsung D500 were used as NFC capable phones. France Telecom participated as carrier, Orange as wireless network operator and Samsung as terminal manufacturer. Finally the

Table 2.1: Different Business models for m-payment solutions

Business Model	Description
Operator centric model	The operator is responsible for the production and management of the m-payment service. Since financial institutes do not participate in the payment process, the two possible payment methods are prepaid and direct carrier billing. Therefore this model is unsuitable for macro-payments.
Bank centric model	The bank is responsible for the production and management of the m-payment service. The operator does not participate in the payment process. However if the bank uses a SIM-based application technology for their mobile application, the bank has to pay rental fees, thus the operator benefits as well. Since payments are made through bank accounts, both micro and macro payments are possible.
Operator centric with bank interface model	The operator is still in control of the business, but now bank are participating. The business model combines the main features of the operator centric and the bank centric model. Both micro and macro payments are supported. Micro payments transactions are handled through direct carrier billing and prepaid phones no the other hand macro payments are possible through bank accounts. So the main difference between this model and the previously discussed models is the additional feature of a single interface for communicating with several accounts on different banks.
Peer-To-Peer model	There is a substantial difference between the peer-to-peer model an the other models. In this model, an independent third party is managing the mobile payment service using the existing infrastructure of banks and operators. Therefore the customer needs a mobile phone and a bank account to make a transaction. Both micro and macro payments are possible. One famous service provides is PayPal.
Collaboration model	This model states, that there is a service manager which is responsible for service management and collaborates with operator and bank. Each party focuses on its primary functions and competences. Meaning the operator provides the infrastructure and the financial institutes enable payment transactions. Income is generated through transaction fees.

Figure 2.3: Overview of the NFC projects worldwide [18]

project included a test with cinema posters, where the customer could download cinema information on the mobile phone by placing the phone over the poster.
Taiwan 2005: In Taipei mobile payment based on BenQ mobile phones was introduced in transportation ticketing systems, retail outlets and banks. The project was a cooperation between Philips as semiconductor manufacturer, the Communications Industries Development Council of the Finance Department and the Taiwanese government.
United states 2007: Citibank New York started a project in where customers could use a Nokia6136 mobile phone as a credit card. NFC Technology was used for contactless payment. MasterCard was involved as credit card issuer and Cingular Wireless as mobile carrier.
Since 2005 mobile payment projects can be found worldwide. Different countries are conducting tests with NFC technology and try to find useful fields of application. Contactless m-payment is certainly suitable for ticketing systems, where fast and cashless payment is required.

2.5 Market Drivers and Constraints

According to KPMG status report [13], 40% of all transactions in Switzerland today are cashless. Compared to the Nordic European countries, Switzerland still has a high cash share. However it is still less than Germany with 75% cash payments share. Experts say that the reduction of the cash payment share is an important driver for introducing mobile contactless payment in Switzerland. The handling costs of cash transactions peaked at CHF 2.2 billion in 2007 and generated more costs in percent of the turnover than, for example payment with the Maestro debit card [7].
Nowadays high costs of banking transactions, makes the e-commerce and m-commerce essential tools for routine financial operations. Considering the significant growth in mobile phone use by customers and its high accessibility, this device can be considered as the most suitable tool in payment category.
According to Bill Gajda customer benefits clearly are obvious. The combination of modern mobile technology and electronic payments holds the promise of an enhanced shopping experience [8]. Merchants are therefore able to offer their customers convenience and control of payment options. Smartphones with Internet access let the consumers research purchases, compare prices or even share their favorite products on social media networks. On the other hand there are advantages for merchants as well. For example merchants can distribute digital coupons for a product promotion. The customer receives the coupon

Table 2.2: Advantages and risks of mobile contactless payment [13]

Advantages	Risks
Increased speed	Security concerns
Increased convenience	Missing international standards
Innovative payment channel	Lack of commercial business model
Image advantages	Lack of customer acceptance
Link to other NFC services such as mobile contactless ticketing	Customer confidence in the new technology and the technology provider
Increase of spontaneity	Customer advantage not clear
Increase internal efficiency (e.g. personnel)	Mobile contactless payment is not perceived as a long-term project that has to be planned years ahead
Decrease of physical cash	Merchant advantages not clear
Generation of new revenues	Slow diffusion –> late ROI
Cost savings	Lack of merchant acceptance
Increase in consumption	Technological complexity
Novel solutions can be offered (innovative factor)	Loss of mobile phone
	The availability of NFC mobile phones

Figure 2.4: Summary of market drivers and constraint for m-payment

and shares it with friends or the coupon is directly integrated into his digital wallet. This means merchants can reach their customers through multiple touch points simultaneously. The fact that modern mobile phones often have integrated positioning system opens a new dimension to marketing. Questions like where customers most likely are to buy certain products are highly interesting.
In the context of the survey for the Swiss Status Report [13], advantages and risks for contactless m-payment in Switzerland were elicited from different domain experts. Table 2.2 shows possible advantages and risks using NFC technology for the Swiss market.
Seen worldwide, Juniper research evaluated market drivers and constraints [9]. Figure 2.4 shows general, global market drivers and constraints for m-payment.
The driver and constraints of the Swiss market clearly overlaps with global situation. Speed, convenience and physical cash reduction are drivers for m-payment. On the other hand security concerns, lack of devices, technology standards and a clear business models are strong constraints to the Swiss market and m-payment worldwide.

2.6 Overview of Current M-Payment Methods

The development of m-payment methods is based upon mobile telecommunication technology. In the early stage there was a success in selling mobile contents and services such as logos and ring tones. The adaption of m-payment services however has not been that rapid. There are several ways to categorize m-payment methods. Figure 2.5 shows a

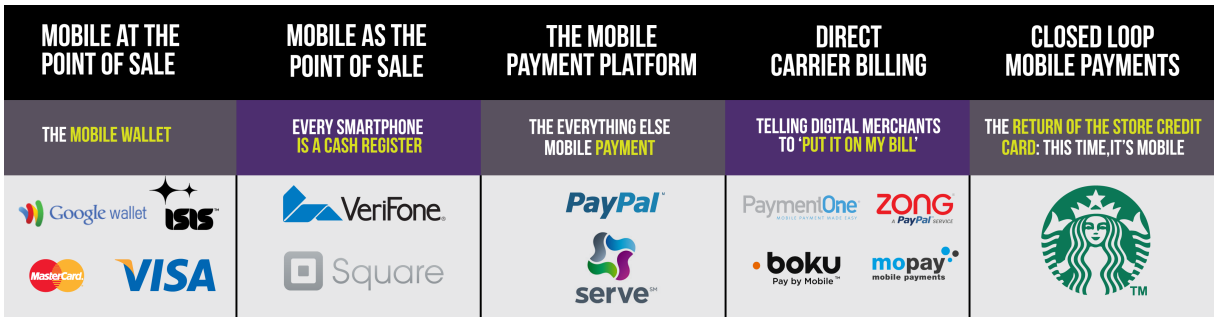


Figure 2.5: Different m-payment methods by mobilepaymentstoday [16]

reasonable overview on the different categories. The following sections will explain these methods in more detail.

2.6.1 Mobile at the Point of Sale

This method enables the customer to pay with a mobile phone at the point of sale (POS). To complete a transaction, customers must be able to synchronize with the merchant system.
Mobile at the point of sale is useful for micro payments when the consumer does not have any coins left. On the other hand the disadvantage of this method is that the mobile phone has to be able to communicate with the merchant system. This means that the phone has to have a NFC (Near Field Communication) or a RFID (Radio Frequency Identification) chip installed. Alternatively infrared technology could be used to transfer the protected information of a pre-selected card (debit, credit, retailer loyalty and pre-paid card) to the m-payment device. Another disadvantage is that the shop owner has to install a POS payment system. In the e-commerce environment, mobile payments at the physical point of sale and are also known as proximity payments.
However not only mobile phones are able to hold a NFC chip. Mastercard offers a service called PayPass. To use their service, the customer holds his/her NFC enabled credit card close to the payment device so that the payment information could be transfered. Another company called Mint offers a mobile payment application that enables the customer to make POS payment. To process the payment the consumer has to enter the phone number of the merchant Mint device. If the transaction was successful, the customer receives a confirmation via SMS. To make us of this service, the customer has to create an account in advance. Mobipay on the other hand uses either a mobile phone, an identification number or a Mobipay barcode which the consumer has to present to the merchant. After the merchant has entered the amount to be paid in the terminal, the consumer has to authorize the transaction with a personal PIN.
On special kind of POS m-payment are e-wallets. An electronic wallet is an encrypted storage medium holding credit card and other financial information that can be used to complete electronic transactions without re-entering the stored data at the time of the transaction [12]. G-plus [10] compared different e-wallet solutions and evaluated pro and cons for the services. Google Wallet is the first m-payment platform for Android phone users. Google provides the Nexus S 4G phone which is already NFC enabled. The good thing about Google Wallet is that the customer can use preexisting coupons and savings from Google Offers. The solution supports MasterCard, Citi, First Data, Sprint and Google as payment network partners. At the time G-Plus did the comparison, VISA had not been supported. Google announced though that in the future, VISA and American Express will be supported [27].
VISA on the other hand is working on a e-wallet solution itself. Visa Wallet is expected to handle multiple cards and payment options though many financial networks. The

Figure 2.6: SquareUp from Square showing the adapter for the iPhone [26]

application should run on most NFC enabled mobile phones. The advantage is that VISA has 50+ years of experience in payment processing. A downside is the lack of MasterCard support. SERVE is a payment platform by American Express. Therefore it clearly focuses on the American Express customers. Users of SERVE can send money securely between two devices. Also with SERVE is it unknown if VISA oder MasterCard will be supported. The last solution G-plus compared is a service from ISIS. ISIS is a coalition between AT&T, Verizon Wireless and T-Mobile. The solution should run on any NFC capable devices supported by the three carriers. An advantage of this service is, that users can pay with multiple credit and debit cards, which can be stored in the e-wallet application.

2.6.2 Mobile as the Point of Sale

This method is more on the merchant side. Each mobile phone can act as a cash register. Therefore the merchant is using a mobile device to process credit cards payments. The obvious advantage of this method is the possibility to handle credit card payments nearly everywhere and anytime. Also the costs are rather low, because the merchant only needs an adapter for the mobile phone. Figure 2.6 shows a solution called SquareUp from Square.

2.6.3 Mobile Payment Platforms

In Figure 2.5 they call this method the everything else mobile payment. Services like PayPal offer a great variety of payment methods. Peer-to-peer payment using mobile phones is possible or purchases at online shops. PayPal also offers a service called text to buy. Basically this is mobile payment via SMS but backed up by the payment platform. The advantage of these platforms is the variety of payment possibilities they offer.
Figure 2.7 show the steps for PayPal service text-to-buy. (1) Customers sees 'Text to Buy' in an ad for an item he or she wants to purchase. (2) The customer sends a text message with the item code to the number shown. (3) PayPal will call or text the customer back to confirm your payment.(4) The item is shipped to the customer.

Figure 2.7: PayPal mobile service text to buy [21]

2.6.4 Direct Carrier Billing

Basically this method is know for buying ringtones, logos, games and other digital stuff by putting the charges on the phone bill. This method is an alternative way to make payment transactions using the mobile phone. For example the customers gives the phone number to the shop and will be charged on carrier phone bill. Therefore no credit card is required. The advantage of this method is a secure way for buying goods on the Internet or a shop. Additionally the shop owner does not have to invest in special equipment for m-payment. There are two possible ways to use direct carrier billing. Either by calling a premium line service or by charging via SMS. Using SMS based payment, the customer is only charged after reception of a confirmation. The issue with premium call lines is that they charge the customer when he/she calls and receives the transaction code. If the customer does not use the code correctly, he/she has to call again and for that reason will be charged twice. This issues is solved with SMS mobile payment but due to fixed premium SMS rates, only a limited amount of money can be transferred. Therefore this methods is only suitable for micro payments [28].
Using SMS mobile payment to pay with a mobile phone requires the following steps:

1. enter the mobile number on the website of the online shop,
2. receive a text message with a transaction code,
3. enter the code on the website and get a final confirmation.

The transaction has now securely ended and customer gets charged on the mobile phone bill. This is a convenient way to pay without a credit card and the whole process takes less than 20 seconds. Examples of this method are mopay, boku or PaymentOne. Direct carrier billing is widespread and used worldwide. In Switzerland the company E-24 offers together with Postfinance m-payment solutions for parking tickets [5].
As Figure 2.8 shows, there are two possible ways to pay the parking tickt. A owener of a Postfinance account can call the Postfinance number on the bottom and enter the location id 26 and the parking lot number. To charge the ticket on the Postfinance account, the customers mobile phone number has to be registered for m-payment in advance. All other parkers can us direct carrier billing and call the corresponding 0900 pay numbers to pay. In this case the parking hours will be charged on the phone bill.

2.6.5 Closed Loop Mobile Payment

If a company develops its own payment solution this is called closed loop. Starbucks for example offers their customers to pay with the mobile phone. To use this service, the customer has to download an Starbucks application on the smart phone. Starbucks offers prepaid cards which can be used to load money onto the mobile phone. The customer can then use the mobile phone to pay at the shop. The advantage of closed loop payment solutions is flexibility. The provider can design the application to meet the customers needs. Furthermore Starbucks can gather information on the purchasing behavior. For

Figure 2.8: M-payment solution for parking tickets by E-24 [5]

the customer it is an alternative way to pay. Queuing time in the shops can be reduces as well. The downside clearly is the costs. Developing and maintaining an own payment system is very expensive compared to other payment methods.

2.7 Near Field Communication NFC

Near Field Communication is a short rang communication technology. It is based on the radio-frequency identification technology, short RFID. NFC was developed by NXP Semiconductors (formerly Philips Semiconductors) and Sony. The technology is operating at 13.56 MHz and is able to manage communication distances up to 10 cm. Since NFC is based on RFID it is also called second generation standard for RFID technology. NFC is compatible with FeliCa, a standard developed by Sony.
While NFC is more human centric, RFID supports distances up to 3m and is therefore more suitable for item tracking in logistics.
As Figure 2.9 shows, there are two categories of NFC standards. The first category is communication between a passive NFC chip called tag and an active device to read or write information from or onto the chip. The tag can be embedded in a movie poster, an identity card or a device. The second category is data communication. This means that two devices actively transmit and receive data using the NFC technology.
Figure 2.10 shows different types of NFC chips. The USB device is like a USB storage stick. The second type is included in a Securedisk card. Therefore if a SD-card with a NFC chip is installed in a mobile phone, it enables the devices to use NFC technology. Using this chip-type, the distance is even more limited. Further there is the built-in module type. This is ideal for integrating the NFC chip on a circuit board. Because this chip type needs an external antenna, the communication distance is up to 10 cm. Communication speed is rather slow. Therefore NFC is useful for transmitting small amounts of data. Videostreaming or transfering large files is not advisable.
There are three solutions for integrating NFC technology in a mobile phone [13]. The important fact hereby is where the secure element lies. Secure element is hereby a term for the mobile contactless payment data. This data can be stored in three places. (a) The secure element can be implemented in the SIM card. The SIM card / UICC (universally integrated circuit card) solution has the advantage, that it is worldwide usable and mobile phone independent. Further the solution is beneficial for the mobile network operator

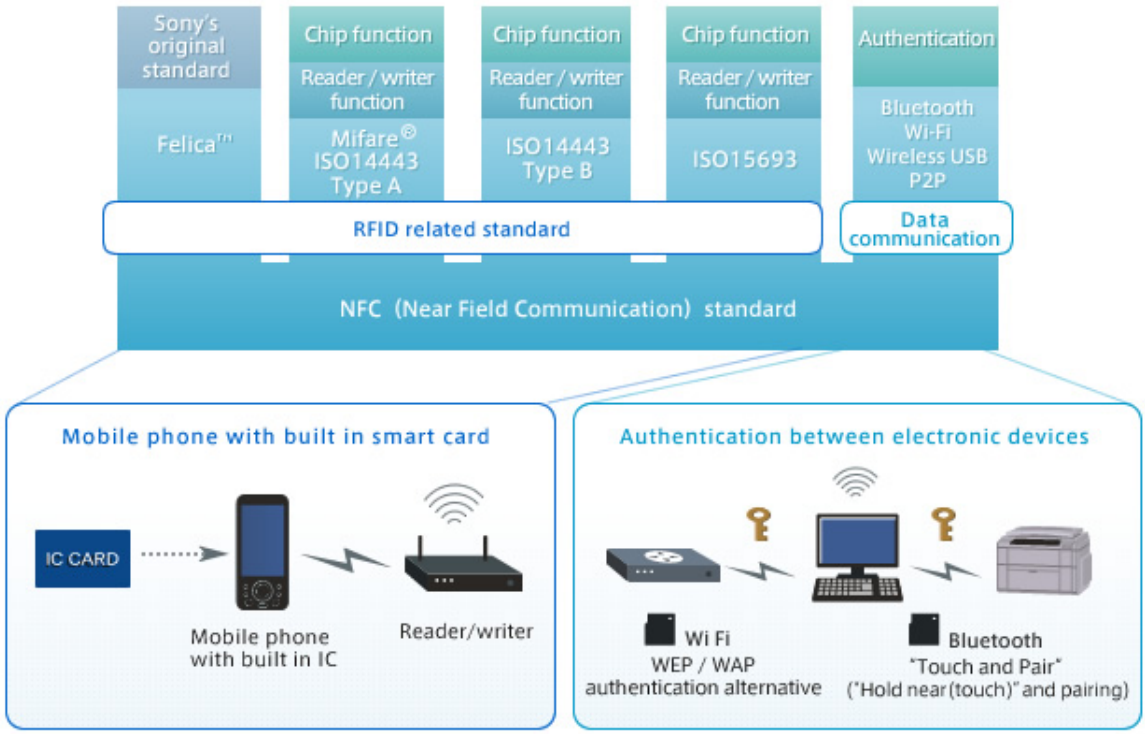


Figure 2.9: Overview of the NFC standards [20]

	USB Type	miniSD Type	Built-in module
Type	USB Type NFC reader	miniSD Type NFC module	Built-in module Type
Communication Standard	ISO18092, FeliCa™ Mifare® Famiy Compliant to ISO14443 Type-A, ISO14443-4(T=CL)		
Operation Frequency	13.56MHz HF Band		
Communication Speed	106/212/424/kbps		
I/F	USB2.0 full speed interface	SD I/O	Serial
Power Consumption	115mW (idle)	140mW (idle)	115mW (idle)
	550mW (in use)	580mW (in use)	550mW (in use)
Input Power Supply	5V (USB supply)	3.3V	5V
Communication Distance	Upto 2cm	Upto 1cm	Upto 10cm
External Dimension	60mm×20mm×10mm	20mm×21.5mm×1.4mm	25mm×16mm×1.4mm (exl. Ant)

Figure 2.10: Overview of the NFC chip types

because of the ownership of the SIM card, service fees and customer contact. (b) Second possible solution is using memory cards. Therefore the device has to offer a card slot. The memory card is exchangeable and individually purchasable. The service provider mainly benefits from this solution because each service provider can offer its own solution. (c) Third solution is the embodied chip. The chip is obviously not exchangeable and the customer needs to buy a mobile phone with the chip already installed. The ownership of the chip is beneficial for the mobile phone manufacturer.
NFC technology can be used in wide field of application:

- health monitoring,
- access control (door-keys),
- pairing devices,
- credentials for WiFi networks login,
- check-ins: Foursquare, Latitude, etc.
- mobile tickets for trains, planes, mass transit,
- initiate a video chat or join a conference call,
- cata sharing between phones: contacts, meetings, credentials,
- attendance control.

Bostinnovation asked three NFC experts about possible successful NFC applications in the future [3]. Brent Bowen of INSIDE Secure (provider of chips) does not believe that m-payment will be the most potent field of application for NFC technology. According to Bowen NFC will have its breakthrough in the area of social media networks. Vik Pavate of MIT spin-off Kovio on the other hand sees great potential in the $500B advertising industry. Ivan Lazarev who is the owner of ITN International says that NFC technology is most likely to become successful in B2B space.
The experts' opinions differ greatly. NFC is likely to become widespread and well accepted in different fields of application in the near future. The technology is dependent on NFC capable mobile devices though. Currently the number of NFC enabled phones on the market is very limited, this clearly is one of the main constraints for the technology.

2.8 Risks of M-Payment Methods

Most certainly new challenges concerning security arise whenever a new payment technology is introduced to the market. Fortunately though, security concerns of m-payment methods are similar to the ones already known and addressed by the payment industry. Addressing those threats should be a shared responsibility of all stakeholders.
In the field of m-payment the protection of personal data that is either stored in or flows through a mobile device is critical. Personal data includes amongst others PINs, security codes, passwords etc. Customers could think that transferring personal data over a wireless network makes them vulnerable to theft. For that reason it is more difficult for m-payments service providers to assuage consumer concerns about security and privacy.

2.8.1 Point of Sale M-Payments

Point of sale m-payments are based on the EMW standard. EMV stands for Europay, MasterCard and VISA, a global standard for inter-operation of integrated circuit cards (IC cards) and IC card capable point of sale terminals and automated teller machines (ATMs), for authenticating credit and debit card transactions [29].
Therefore if a mobile phone has an EMV-approved chip installed the same end-to-end security offered by a smartcard-enabled payment is ensured. Payments with EMV chips face the fewest security challanges. Types of POS payments that rely on barcodes or other means of payment face a significant challenge in delivering a secure, efficient and cost-effective solution.
Standardization and integration is important for contactless payment. There are many new mobile devices every year, each offering different ways to access payment information stored on the chip. There is a lack of technology standards in the industry. Problems occur if the manufacturer of the mobile device, the payment chip manufacturer and the mobile networks that distribute and enable the devices do not work together. Defensive measures to secure the entire value chain must be established by actors across the industry [8].

2.8.2 Remote M-Payments

Current smartphones are able to execute all types of applications. For example instant messaging social media applications, games or even online banking and trading solutions. Unfortunately the ability to execute applications makes the devices vulnerable to viruses and malware as well. Today not many viruses and malware are targeting mobile platforms. But accoring to Bill Gajda that is about to change once there is increased adoption and penetration of mobile payments by consumers. Unsurprisingly there is antivirus software for smartphones available on the market today.
The key differences and challenges for mobile phone-based eCommerce transaction are:

1. Software: While PC-based eCommerce is based on standardized Web software through Microsoft Windows, MacOS or Linux operating systems, the world of phone-based eCommerce looks different. Mobile operating systems are still evolving rapidly with frequent changes. Additionally, Andriod for example, comes with a wide variety of underlying hardware architectures.

2. Internet connection: In the PC world the risk of an attack is limited to the amount of time the computer is online. However with smartphones the time of exposure is greatly increased because normally, the devices are not switched off, even while we sleep.

3. Scams: Phishing attacks trick victims into divulging personal information. Scammers can easily apply these strategies to the mobile channel. Short messages can be used to commit a fraud. In the mobile area these attacks are called smishing(SMS text phishing) and vishing (voice phishing).

To ensure the safety of mobile payments, industry leaders must address these major security issues. The potential value of m-payment for merchants and customers is tremendous, but security is the precondition to benefit from this technology in the future.
The Payment Card Industry Security Standards Council (PCI SSC) provides with the Data Security Standard (DSS) an actionable framework for developing a robust payment card data security process. The framework addresses prevention, detection and appropriate reaction to security incidents [22]. This standard requires compliance to all entities that process, transmit or store payment information. Figure 2.11 shows the tools of PCI DSS compliance and self-assessment.

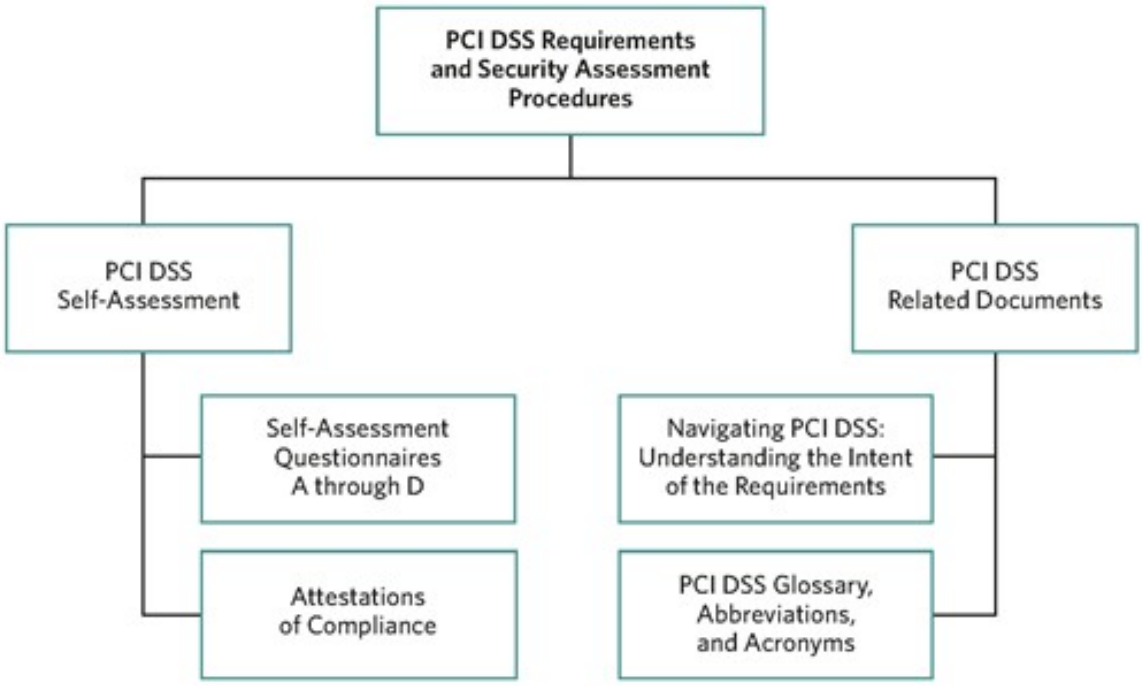


Figure 2.11: PCI DSS assessment procedures [23]

Furthermore the council maintains a Payment Application Data Security Standard (PA-DSS) which helps software vendors and others develop secure payment applications. Although these standard are not initially developed for the mobile space, the fundamental principles would be equally applicable.

2.9 Example Application of M-Payment

Figure 2.12 shows the steps of a payment process from Postfinance [24]. Let us assume the customer wants to buy a product at a shop and likes to pay via sms using the Postfinance mobile payment method.

1. The customer sends a sms with a keyword to the corresponding number provided by the service provider (shop owner).
2. The application service provider (ASP) sends an authorization request to Postfinance.
3. If the requested amount is on the customer's bank account, the amount is reserved on the account.
4. The ASP sends a confirmation to the customer and the show owner.
5. The customer receives the product.
6. The shop owner receives the money in the following days.
7. The ASP charges the transaction to the shop owner

The difference between the service provider (SP) and the application service provides (ASP) is quite obvious. The SP offers a product or a service to the customer and the ASP acts as an interface between SP and Postfinance. For example the ASP provides the phone number and line to send messages to. The ASP then charges to SP for each payment transaction.

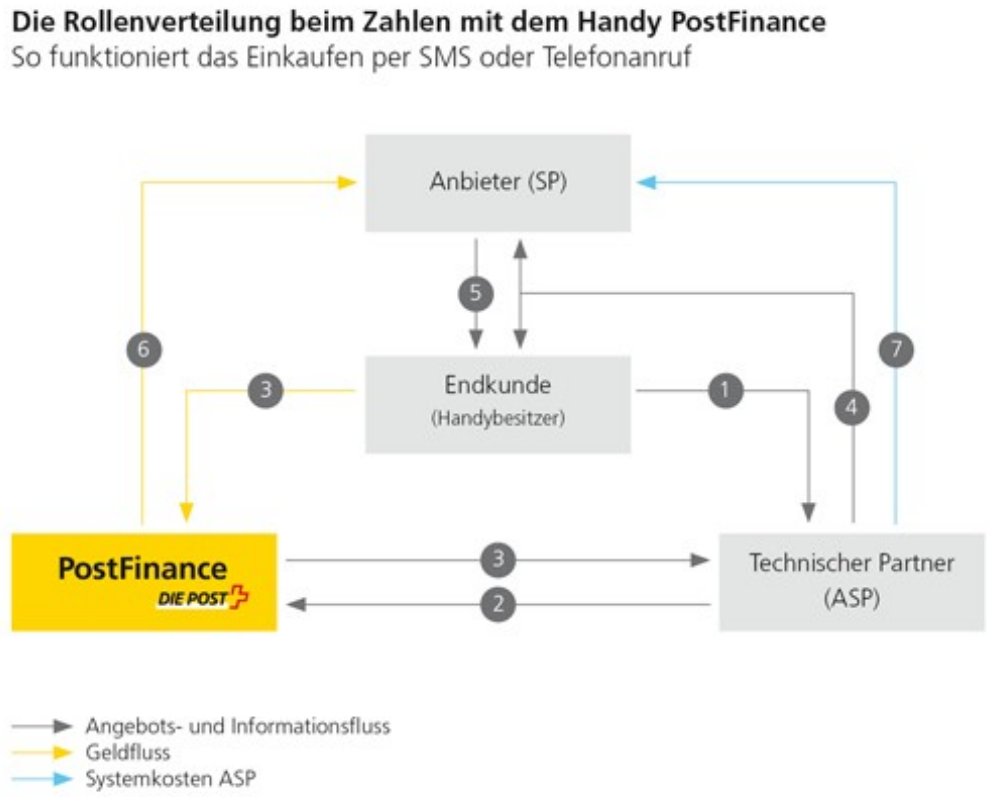


Figure 2.12: Postfinance steps of the payment process [24]

According to Postfinance, the use of mobile payment methods in Switzerland is depending on technology. Currently in Switzerland NFC is not that widespread and there are only a few NFC phones available on the market. Postfinance therefore uses m-payment with sms and phone calls, which they call remote payment. Remote payment is suitable for flyers and ads. It is a new channel of sales for service providers. Donations in the public social aid sector, tickets for concerts, movies or skiing areas are applicable to m-payment. Finally automats are suitable for m-payment, if the customers does not have any coins left.

2.10 Future of M-Payment

According to Nokia, m-payment reached the plateau of productivity, meaning that it is not a overhyped phenomenon. M-payment has a lot of potential and there are various fields of application. As Figure 2.13 shows, spending via smartphone clearly increases. From 2011 until 2015 the volume of global transaction increases nearly five times [2].
The overall value is tremendous. In 2005, the British marketing research firm Juniper Research predicted that total transactions via mobile devices would be $155 million that year and top $10 billion by the end of the decade. Not only did mobile payments exceed that forecast tenfold, reaching $100 billion in 2010, but the total for digital and physical goods are expected to reach $630 billion by 2014. The future of mobile payments is robust. Consumers are poised to realize enormous benefits, and merchants to gain unrivaled opportunity. But in order for any of this promise to be fulfilled, the fundamental issue of security will need to be vigilantly addressed [8].
The forecast from G-Plus in Figure 2.14 shows that by 2013 one in five mobile phones will be capable of using NFC technology. Only one year later, Google predicts that half

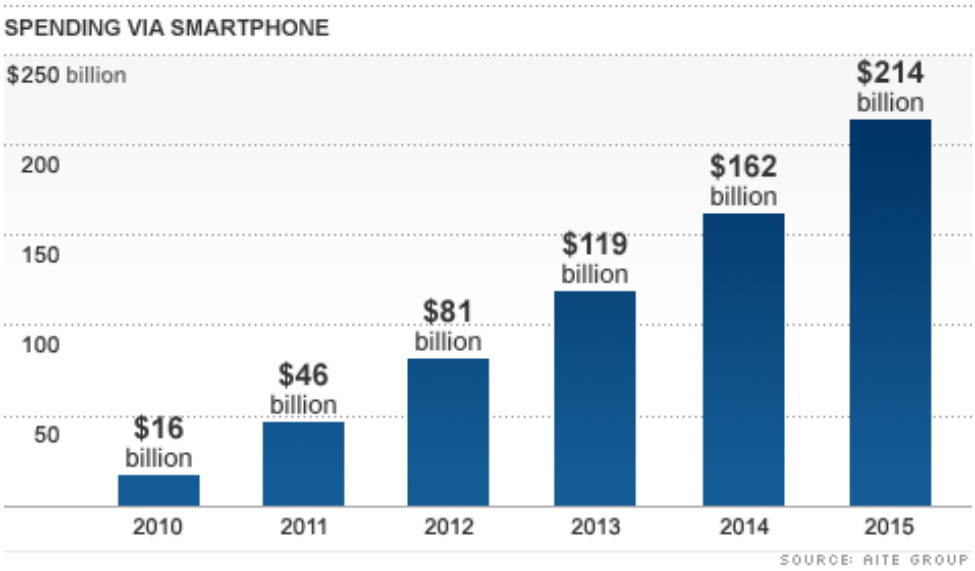


Figure 2.13: Forecast on spending via smart phone by Aite Group [1]

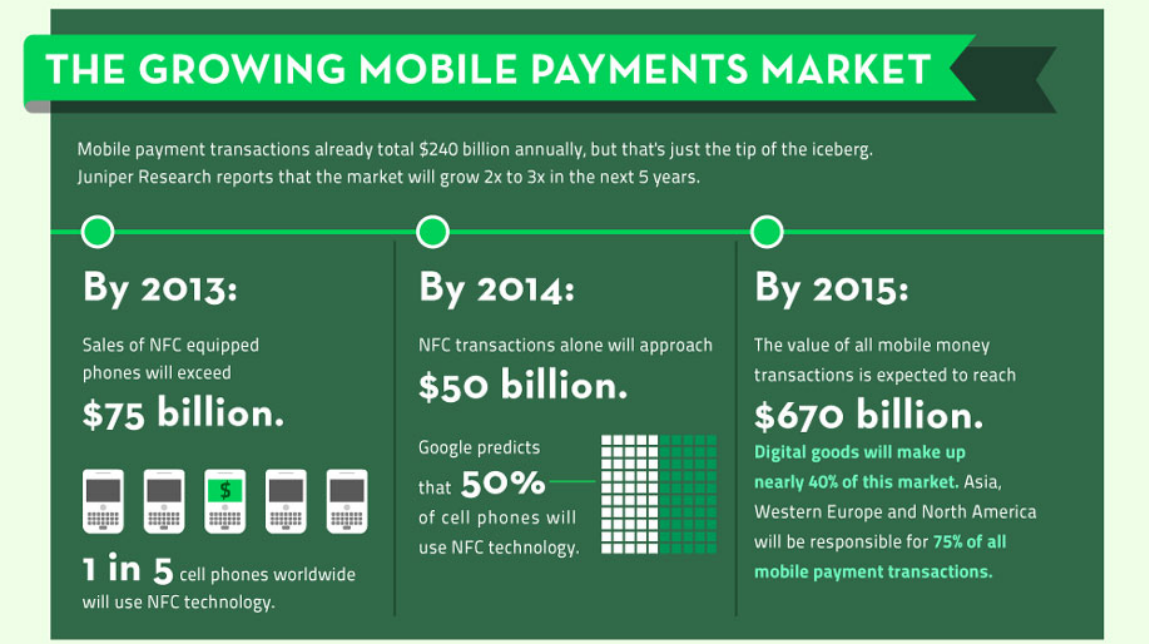


Figure 2.14: Forecast from G-Plus concerning NFC technology [10]

the mobile phone are NFC enabled and the volume of payment transaction using NFC is approaching $50 billion.
Forecast numbers should be treated with respect. Alone the difference between the forecast from the Yankee Group ($1 trillion) and Juniper Reseach ($670 billion) worth of global transactions by 2015 is $330 billion. Despite the margin in the forecasts the numbers clearly show a tendency which is upwards. M-payment is gaining more ground and will certainly continue penetrating the markets if the development proceeds as in the previous months.

2.11 Summary and Conclusion

This report showed different m-payment methods with their corresponding advantages and disadvantages. We discussed market drivers and constraints and showed different business models. Furtermore, facts about Near Field Communication were presented and

possible fields of application of NFC technology. Finally we discussed risks of m-payment, penetration of the market and the future of m-payment.
Considering that mobile payment is not a hype anymore but reached the plateau of productivity, the forecasts predict a bright future for m-payment application. Statistics of NFC enabled mobile phone on the market predict a massive increase of devices over the next few years. Since the availability of devices is a great constraint to m-payment, we can assume that the penetration of market concerning m-payment solutions will greatly increase. Research companies like Juniper Research or Yankee Group ended up with numbers for the volume of projected global m-payment transactions by 2015 between $670 billion and $1 trillion. Although there is big margin it is save to say that with an increasing number of devices there will be an increase in m-payment transactions in the future.

Bibliography

[1] Aite group: Forecast http://www.olmsteadwilliams.com/thebigmouthblog/2011/02/02/the-end-of-credit-cards-is-coming, last visited 14.11.2011.

[2] Aite group: Spending via phone statistics http://yesiamcheap.com/2011/01/the-end-of-credit-cards-is-coming, last visited 16.11.2011.

[3] Bostinnovation: Statements on NFC applications http://bostinnovation.com/2010/12/14/nfc-enabling-mobile-payments-the-internet-of-things-and-the-next-wave last visited 16.11.2011.

[4] businesstown.com: The Definition of E-Commerce http://www.businesstown.com/internet/ecomm-definition.asp, last visited 19.10.2011.

[5] E-24: Mobile Parkplatz-Zahlungslösung http://www.e-24.ch/page-parkplatz-benuetzen.htm, last visited 20.11.2011.

[6] F.Asghari, A.A.Amidian, J.Muhammadi, H.Rabiee: A Fuzzy ELECTRE Approach For Evaluating Mobile Payment Business Models; IEEE technical report (978-1-4244-8507-9 2010 IEEE).

[7] FEW-HSG: Die Kosten des Bargeldes, Study, 2007.

[8] Bill Gajda: Managing the Risks and Security Threats of Mobile Payments, Lydian Journal February 2011, http://pymnts.com/assets/Lydian_Journal/LydianJournalMarchRiskSec.pdf.

[9] Alan Goode: Mobile Payment Strategies and Markets 2007-2011, Whitepaper, Juniper Research 2007, http://www.wirelessmobile-jobsboard.com/pdf/MobilePaymentswhitepaper.pdf Last visited 2011-10-10.

[10] G-plus: Goodbye wallets, Infograpic https://www.gplus.com/Infographic/INFOGRAPHIC-Goodbye-Wallets-How-Mobile-Payments, last visited 20.11.2011.

[11] C. Hort, S. Gross, E. Fleisch: Critical success factors of mobile payment, Switzerland, 2002, p 1 - 74.

[12] InverstorWords: Definition of e-wallet http://www.investorwords.com/1681/electronic_wallet.html, last visited 20.11.2011.

[13] KPMG ETH Zürich: Mobile contactless payment and mobile ticketing, Status report 2010, http://www.kpmg.com/ch/de/library/articles-publications/seiten/mobile-contactless-payment-und-mobile-ticketing.aspx, last visited 20.12.2011.

[14] M. Krueger: The future of M-Payments, business options and policy issues, Seville, Spain, 2001, p 1.

[15] T. Michel: What is Micro-payment? http://www.w3.org/ECommerce/Micropayments, last visited 16.11.2011.

[16] mobilepaymentstoday.com: Mobile Payment Infographic http://www.mobilepaymentstoday.com/blog/6295/The-most-important-mobile-payment-infographic-Ever, last visited 22.11.2011.

[17] Mobile payment forum: white paper enabling secure, interoperable and user-friendly mobile payments, Mobile payment forum, Wakefield, Massachusetts, 2002, p 3.

[18] NFC-World: NFC casestudies http://www.nfc-world.com/en/cases/index.html, last visited 02.12.2011.

[19] NFC-World: NFC projects worldwide http://www.nfc-world.com/en/cases/index.html, last visited 14.11.2011.

[20] NFC-World: NFC Technology http://www.nfc-world.com/en/about/index.html, last visited 13.11.2011.

[21] PayPal: TextToBuy https://www.paypal.com/cgi-bin/webscr?cmd=xpt/Marketing/mobile/MobileT2B-outside, last visited 22.12.2011.

[22] PCI SSC Data Security Standards Overview https://www.pcisecuritystandards.org/security_standards/, last visited 16.10.2011.

[23] PCI: DSS Self-Assessment Questionnaire, https://www.pcisecuritystandards.org/merchants/self_assessment_form.php, last visited 23.12.2011.

[24] Postfinance: NFC Technology http://www.postfinance.ch/de/biz/prod/pay/debsolution/mobile/detail.html, last visited 13.11.2011.

[25] Searchmobilecomputing: The Definition of M-Commerce http://searchmobilecomputing.techtarget.com/definition/m-commerce, last visited 19.11.2011.

[26] Square: SquareUp http://www.medialightbox.com/blog/2011/02/futuristic-iphone-accessories-available-today/squareup/, last visited 23.12.2011.

[27] Techcrunch: Visa support for Google Wallet http://techcrunch.com/2011/09/19/google-wallet-sprint/, last visited 20.11.2011.

[28] Emilie Valcourt, Jean-Marc Robert and Francis Beaulieu: Investigating mobile payment: supporting technologies, methods, and use; IEEE technical report (0-7803-9182-9/05/2005 IEEE).

[29] wikipedia.org: Definition of EMV, http://en.wikipedia.org/wiki/EMV, last visited 16.10.2011.

Chapter 3

An Economic Overview of Internet Mobile Platforms

Nicolas Bär

With the emergence of smartphones, the Internet mobile business becomes highly important to all participants in the mobile market. The platform providers repositioned themselves in the market and build a platform with a strong ecosystem including device manufacturers, network operators and application developers. The Internet mobile business offers platform providers and application developers new opportunities to drive their business. Apple (iOS), Google (Android) and Microsoft (Windows Phone OS) are the main platform providers. Each one has its own strategy to emerge with the surrounding ecosystem in oerder to become the market leader. Google floods the market with an open operating system adopted by a wide variety of device manufactures and generates profit through online services. While Apples iPhone is an exclusive product with a diverse marketplace. Microsoft is lacking in a substantial market share, but is extending its ecosystem through a strategic partnership with Nokia. On top of the platform application developers are able to build applications and distribute these to the customer through the platform providers marketplace. The minimal functionality on mobile operating systems create a need for third-party developers and can only be compensated by a strong interrelation between developers and platform providers. To build a strong collaboration of the two parties, dedication- and constraint-based mechanisms have to be considered. As the importance of developers commitment rise, platform providers have to support economic, social and resource factors and offer a certain degree of flexibility. This paper analyzes the economic dimensions from the perspective of platform providers and application developers and highlights the advantages and disadvantages of different business models.

Contents

3.1 Introduction

In the past few years the growth of the mobile market was primary driven by smartphone sales. With the introduction of the iPhone in January 2007 Apple created a whole new Internet experience on the smartphone [2]. It redefined the smartphone product category and opened a new market for Internet services and personal computing in the mobile industry. This led to new market entrants such as Google and the present players had to reposition themselves. With the emergence of smartphones the Internet mobile market becomes highly important for network operators, application developers, device manufacturers and platform providers.
This paper first of all describes the main platform providers and their interrelations within the Internet mobile ecosystem. Then the economic models used by the platform providers and application developers will be described and the advantages and disadvantages as well as challenges and opportunities will be outlined.

3.1.1 Platform Providers

The platform providers covered in this paper are Apple, Google and Microsoft. Their business origins, platform specification and mobile services will be described briefly.
The business origins of **Apple** lie in the computer industry. Apple is one of the main players in this industry and has a strong focus on graphical user interfaces and interaction design. In 2001 Apple entered the mobile music player market with its successful product iPod [3]. In the first press release about the iPhone Apple stated that the iPhone is "[...] a revolutionary mobile phone, a widescreen iPod with touch controls, and a breakthrough Internet communications device with desktop-class email, web browsing, searching and maps-into one small and lightweight handheld device" [2]. Apple used their knowledge from the computer and music player industry to build a new generation of smartphone. It changed the mobile market in two ways. Firstly, with a full functional browser on the iPhone, the users were no longer bound to network operator specific content and applications [10]. As a result the network operators lost power within the ecosystem and the boundaries between the Internet and the mobile device collapsed. Secondly, Apple created a decently open SDK for mobile application developers to build a very strong application provider base [10]. Many experts in the industry see this strong application provider base as a key driver of success for the iPhone [39]. At the introduction of the iPhone, Apple did not point out any plans of publishing a framework for mobile application developers. Only in 2008 it released the first version of the SDK [4]. The operating system iOS is running on the iPhone and powers the iPod, iPad and Apple TV. The iOS SDK is a framework to build applications on top of the iOS and supports objective-c based code. Apple is offering a few online services. One of them is the AppStore, which is the marketplace for developers to provide the application to the end-user. On the other hand the end-user can browse the applications in the AppStore and install the application with one click. In July 2011 Apple announced that over 15 billion applications have been downloaded from the AppStore by more than 200 million iPhone, iPod touch and iPad users [6]. Another online service of Apple is the iCloud. The iCloud helps the user to keep all the data on different devices in sync, by uploading it from one device to a space in the cloud and distribute it to the other devices [7]. Besides that Apple offers the iTunes store to download music, movies and podcasts [8].
Google announced against all rumors about a Google phone the Open Handset Alliance and the first open platform for mobile devices called Android in November 2007 [20]. The business origins of Google lie in Internet services. Googles main products are its search engine and advertising services. In addition they offer various other Internet services e.g. Google Finance, Google+, Google Code, etc. [21]. Google entered the mobile market from

its leading position in computer centric Internet services. Since they had no experience in phone and mobile platform development, they had to grow knowledge and did so by silently acquiring the startup Android Inc. [13]. Then Google created the Internet consortium based on the Android platform, which is free to license and, besides a few Google applications, open. The platform was attractive for a great number of mobile device manufacturers, which were not specialized in software development and the open platform also appealed to mobile operators, semiconductor builders, software companies and commercialization companies. Driven by this attraction, the Open Handset Alliance grew to a strong network with 84 members [36]. The Android OS powers various tablets and smartphones from different manufacturers. It provides basic phone and Internet functionalities along with the integrated Google services. Android offers a free SDK for application developers and supports the languages Java, C and C++. The SDK is available to developers for free and there is support to integrate the Android SDK in widely spread IDE's e.g. Eclipse. Google launched a $10 million Android Developer Challenge to attract developers and build an application provider base [22]. The Android platform increased fast in market share and the openness is appealing to all kind of participants in the market. In November 2011 Google announced that it has activated 200 million Android phones [14]. On top of the Android platform Google is integrating all their different online services to fit the new medium and is offering an application marketplace.
Microsoft has a solid background in software and operating system development. It's the leading company in the operating system market. Microsoft entered the mobile market in 2000 with the PocketPC platform. In 2003 it announced the next generation of the PocketPC called Windows Mobile and had a reasonable spreading in the market. When iOS and Android was released in 2007 the market share started to shrink, therefore it concentrated on building a new platform called Windows Phone OS, which adopts new technologies to build a better user experience on the Internet. The Windows Phone OS was released to manufacturing in September 2010 [31]. Device manufacturers can buy licenses to adopt the platform on their devices. The platform comes with basic phone functionality, state-of-the-art Internet integration and supports the Microsoft Office. On top of the platform it offers application developers a SDK for free and the framework supports code in XNA, Silverlight and VisualBasic. The SDK is integrated in the Visual Studio IDE. Microsoft offers an application marketplace called Zune, which has to be used by application developers to provide their applications to the end-user. In addition Zune offers a wide variety of music and movies to the end-user. Microsoft offers Internet services such as the Bing search engine and Office365. These services are well integrated in the Windows Phone OS platform.

3.2 Platform Provider Economic Perspective

The platform providers Apple, Google and Microsoft offer similar functionality on their platforms, but the strategies and business models of these companies are in many ways different. This section will first analyse the strategies and models by certain factors, then discuss the opportunities and challenges in the market and at the end a conclusion and assessment on the individual platform provider is given.

3.2.1 Platform Strategies and Business Models

The strategies and business models of the platform providers can be described by the factor customer lock-in, value capture and the strength of the network, which is determinded by the licensing model and the ecosystem of the platform provider. The following subchapters analyze these factors and compare the different approaches of the platform providers.

Rahul and Basole [10] introduced a set of technological layers to overcome the complexity of platform strategies and their different technology. They broke down the technological layers - the stack - in the categories online services, storefront, native apps, OS and handset. These categories will provide the ground for the discussion of the different factors.

3.2.1.1 Customer Lock-In

The factor customer lock-in describes the degree of how sticky the product mix is to the customer and how big the switching costs to competitors are. Moreover there can be other market barriers. The value capture can happen on a completely assorted layer in the stack than the lock-in.
All platform providers are locking-in the customer on the OS layer for obvious reasons. When a customer gets used to the platform in terms of interaction and graphical user interface, the switching costs are high, because it will take time and effort to understand a new platform in the same way. Moreover the configuration and the installed additional applications can be impossible to migrate to any other platform. In addition Apple locks in the customer on the handset and storefront. The iPhone is the only device running iOS, therefore whenever a customer gets locked-in to the OS it will as a consequence be locked-in on the handset. The storefront of Apple combines mobile applications, music and videos and is as a single platform with a combination of different medias and an outstanding ease of use. The product mix of iTunes and AppStore is sticky, because it connects the different medias with all Apple devices and offers a user-friendly frontend [10]. Microsoft on the other hand does not provide any phone, but is licensing it to different manufacturers. Therefore there is no lock-in on the handset. In addition to the OS Microsoft locks-in the customer on the native apps such as office, facebook integration and kinect [10]. The online services of Microsoft are growing, but the switching costs are in general low in this layer and therefore it doesn't classify as a lock-in. This is also the case for the Google online services. Google has various online services, but none of them have high switching costs. For example the search engine is the most used one on the Internet, but another search engine is only one click away. However the overall product mix of Googles online services are outstanding and could promise a lock-in, but Google has not yet brought a solution to combine these services to one great hook. The question here is, if they will fulfill the quest for synergy with the new launched Google+ service and build a solution to lock-in the customer based on a mix of services and generate high switching costs [23]. The customer lock-in for Google is apart from the OS layer on the native apps layer. The strong integration of Googles services in the Android platform is outstanding and can't be done the same way on other platforms. As a result a user, who is used to deal with Googles online services on an Android, will have high switching costs [10].

3.2.1.2 Value Capture

Unlike Apple and Microsoft, Google is primary a service provider with various different products. Googles biggest asset is a enormous database with information and the ability to provide this information along with advertisements to the costumer in a quick way [10]. Google is capturing value in the highest level of the stack - the online services. This makes Google almost independent from the lower layers in the stack, due it generates revenue by providing content dependent advertisements along with the content distributed to customers from the cloud. With the open licensing model of Android, Google is not capturing value on the device or operating system, but whenever an Android smartphone

is plugged to the Internet and the user is searching on Google, it will show advertisements and therefore increase its revenue stream.
Microsoft is capturing value in a similar way to Google by offering online services like Bing search or Windows Live and distribute the content along with advertisement. In addition they sell licenses of the Windows Phone OS to device manufacturers for adopting the platform. While Apple does not license the platform iOS to other device manufacturers, but is selling the handset as an end-product to customers. Besides that it has a storefront which is a significant source of revenues [10].

3.2.1.3 Open vs. Closed

Googles Android platform is the most open one of the three competitors. The kernel of the Android linux is published under the GNU Public License version 2 and the rest of the OS code is published under the Apache License version 2.0 [24]. All members of the Open Handset Alliance are extending the code under the terms of the GPL and publish their work. Device manufacturers can either adopt the published source code of Android without or with the closed source top-level Google applications like the Android marketplace, sync, etc. If the second alternative is chosen, the device manufacturer has to make sure, that the device complies with the Google compatibility requirements [26]. The Android licensing model is open for third party-party developers to contribute to the kernel as well as to develop mobile applications on the top level for no cost. The SDK is provided for free and there is a decent documentation published. In contrast to iOS and Windows Phone OS the Android does not have any security mechanism to stop mobile application providers to deliver their apps to the platform [25], but there is a registration fee of $25 to gain access to the marketplace.
Apples iOS platform is the most closed platform of the three competitors. The platforms operating system is maintained by Apple and the code is closed source. There is no licensing model for device manufacturers. As a result Apples iPhone is the only phone powered by iOS. However Apple is offering an SDK for third-party developers to build applications on top of the iOS. The SDK and the corresponding IDE can be obtained for a yearly fee of $100 in form of a development certificate, which also grants access to the AppStore [9]. Nevertheless a third-party application has to pass certain security mechanisms to be available on the AppStore and Apple is preventing the iPhone user from installing applications from other sources then the AppStore. Apple is strictly controlling all software running on the iOS.
Microsofts licensing model is somewhere between the other competitors. The Windows Phone OS is closed source, but can be licensed by device manufacturers. The license is not free as the Android license, but in contrast Microsoft is offering support for device manufacturers and helps to adopt the operating system. In terms of third-party applications, Microsoft added support for mobile development to the Visual Studio Express edition, which is free of charge [32]. Application developers can only provide their product through the Microsoft Zune marketplace and applications have to pass a certain security mechanism and tests to be published on the marketplace [33].

3.2.1.4 Ecosystem

The ecosystem of the Internet mobile market consists of five participants namely platform providers, device manufacturers, network operators, mobile application developers and customers. With the emergence of the Internet mobile market the platform provider repositioned themselves and activities were redistributed. For example the network operator has lost power to the application developer, because platform providers extended their activities and build more flexible solutions to provide application to the customer.

As a result of shifts and extensions on activities of the participants, the Internet mobile industry is described by Basole [11] as a complex ecosystem with multiple interrelations of companies in different segments. The market share and the reachability of a platform is determined by the network of a platform provider and its interrelation with differente participants. All three platforms allow developers to provide applications for the platform, therefore this dimension will not be further explained in this section. As well the intermediaries between a participant and the customer will not be explained, as they have low impacts on the ecosystem.
Since Apple is not allowing device manufacturers to adopt the iOS on their devices, the only way to distribute the iOS platform with the iPhone is by building relations with network operators or to sell it directly. When the iPhone was introduced in 2007, Apple only had a few contracts with network operators, allowing them to sell the iPhone. This can be seen as a marketing strategy to make the iPhone an exclusive good. After the initial release Apple expanded their relations with network operators and were able to increase the number of sold iPhones significantly [11]. In contrast Android was adopted by many device manufacturers like Samsung, LG, HTC, Sony Erricsson etc. and their strong supply management boosted the market share of Android. According to Gartner [19] Samsung has the second biggest market share behind Nokia in terms of devices sold with 16.3% in the first quarter of 2011. The third biggest market share has LG with 5.7%. The impact of Samsung and LG to use Android as the main platform on their devices is enormously. Android and iOS are according to Gartner [19] the obvious winners in terms of market share in 2011. In the second quarter of 2011 the market share of the two platforms doubled to nearly 62%. This sum is divided in 43.4% market share for Android and 18.2% for iOS. Considering that Android was introduced a half year later then the iOS platform, Googles open licensing strategy encouraged to flood the market with Android powered devices.
Microsoft with eleven years background in the mobile industry has in contrast only a market share of 1.6% according to Gartner [19]. Their strategy to license the platform to device manufacturers did not yet lead to a success in the market. Nevertheless Microsoft has strong network in the market and interrelations with different device manufacturers [11]. In February 2011 Microsoft [34] announced plans for a broad strategy partnership with Nokia to build a new global ecosystem. Nokia has an efficient supply chain and vast connections to point of sales around the globe. Gartner [18] predicts a market share of 19.5% for microsoft in 2015, due to the alliance with Nokia.

3.2.2 Opportunities and Challenges

This section will provide a few ideas on what challenges the platform providers and the market as a whole could face in the future and what opportunities exist.
Until now the biggest impacts on the market were driven by technological inventions. Apple for example entered the market with a high-end product and changed the consumer behavior. As a result new platforms arose and the market started to change. As there is no market leader defined yet, there could still be space for a new platform with an improved technology [10]. For instance Facebook could enter the market with a new platform and use their strong customer base to gain market share. This would complement their Internet service and integrate the social network experience. As well the business model shows similarities to the one of Google in terms of advertisement.
Since the introduction of Android, Microsoft has tried to sue device manufacturers for patent infringements. This lead to a serie of patent infringement lawsuits. In 2010 Microsoft sued HTC for using their technology patents on different Android powered devices and in 2011 they started a lawsuit against Motorola. Microsoft is offering a licensing model for device manufacturers to not infringe their patents. For example HTC took a

license covering its Android devices. On the other hand are companies like Foxconn not willing to gain a license [35]. Since there is a growing number of patent infringements, the platform providers and device manufacturers are building a stock of patents to stay competitive and pass over patent licenses. The question is whether Google can come up with a model to prevent Microsoft from suing device manufacturers. Nonetheless Microsoft is gaining market share with patent infringements by agreeing with manufacturers, that they will adopt the Windows Phone OS on devices.

3.2.3 Conclusion

Apple entered the high-end sector of the market with a more functional product. It integrated the knowledge from the computer and music player industry and created synergies to differentiate through technology. The most significant value of the iPhone is the ability to cross boundaries between the mobile device and the Internet. Apple has a pricing power on the device, because it is not providing licenses to other device manufacturers. In contrast it is losing market share to cheaper alternatives powered by Android. Gartner [18] predicts a loss in market share for the iOS by 2015. However the close strategy of Apple is open enough to attract application developers. The iOS has with more then 425 thousand applications in the AppStore a very diverse stock of applications and hit the 15 billion application downloads in November 2011 [6]. The restrictions on the AppStore is seen by many developers as a problem, because the application checks before the application is published lead to shorter time to market. These restrictions make application developers think about switching to another platform [41]. The application developers are complementing the iOS in numerous ways and the stability of this community should therefore be intended. Apple is offering an user-friendly integration of other Apple products such as computers and music players to the customer. This can be a considerable advantage to other platforms, due to the need of synchronized data and information. On top of this Apple is expanding its online services and is building solutions in the cloud to create value in other layers of the stack.
Google is mainly interested in generating more Internet traffic on mobile devices to create more revenue from advertisers. Its open strategy encourages this by providing device manufacturers a platform at no licensing cost. Regarding the market share, Google is flooding the market with Android devices and is predicted by Gartner [18] to become the market leader in the near future. Since the value capture happens in the highest level of the stack, Google is shifting the value away from the lower layers by providing the platform for free. The Open Handset Alliance represents the strongest ecosystem in the market and is ensuring a quality product on low costs. Google is lacking a customer lock-in on the online services, but currently this does not seem like a problem, since services like search and gmail are used by a colossal amount of customers. In addition Google generates profit from other platforms offering integration with its online services. For example iOS has the Google search engine integrated. Google has announced to acquire Motorola Mobility by the end of 2011 or beginning of 2012 [27]. This will lead to an even stronger ecosystem and enables Google to capture value on the handset as well. Since the open platform attracts application developers, Google has introduced the AdMob service, which offers application developers to integrate advertisements in their applications. The AdMob service is firstly helping the developers to create more revenue and secondly expanding the advertisement platform of Google and therefore creates advertisement revenues.
Microsoft is using a classic platform strategy, which shows similarities to its business model used in the computer industry. It offers licenses to device manufactures and has therefore a pricing power. The market share of the Windows Phone OS in 2010 and 2011 are not significant, but based on the partnership with Nokia the ecosystem can be expanded with more interrelations. Microsoft needs this strong partner to overcome the

gap to iOS and Android in terms of market share. In addition the patent infringement lawsuits on device manufacturers lead to licensing contracts to adopt the Windows Phone OS and generates revenue by the cost of the licensing model for patent infringements. Microsofts Office product integration on the Windows Phone OS is creating synergies with the Office suite and the introduction of Office365 tops the product mix off. With the Office365 and the Bing search engine Microsoft is capturing value on the highest layer in the stack like Google. Yet both products are not beating the Google products in terms of market share. Gartner predicts an increase of market share for the Windows Phone OS of almost 500% by 2015 [18]. At the first glance this number seems quite opportunistic, but the strong influence of Microsoft in the network and the partnership with Nokia should not be underestimated.

3.3 Application Developer Economic Perspective

An elemental concept of the mobile Internet business is to provide an effective way for application developers to reach customers and deliver mobile application through a concentrated market. Firstly, the impacts of the strong binding between platform providers and developers through the application marketplace on the ecosystem is discussed. Finally, the strategies behind different pricing models and the corresponding support from platform providers are analysed.

3.3.1 Application Delivery

The marketplaces of Apple, Google and Microsoft offer the application developers a platform to distribute their applications in an effective way. The marketplace handles the payment process, transfer to the end-user device and install procedure. Due to this technology, the application developer can focus on the application building process. In case of Apple and Microsoft application providers are forced by security constraints to use the marketplace of the platform provider. The integration of third-party developers in the business value chain of platform providers in the Internet mobile market is a crucial factor of success. Therefore it is of high importance to understand the underlying factors on the relationship between the two parties. Guo et al [28] analyzed the mobile marketing platforms as a customer and merchant interaction within the platform as a framework. Since this framework is representing a platform provider, a two sided model will not include the interests of the platform provider, which are influencing the customer and developer. Besides the commission on transactions in the store, there are different factors affecting this relationship between platform providers and developers.
Kim, Kim and Lee [30] identified factors to influence the developer to provide applications frequently in a dual model approach of dedication- and constraint-based mechanisms. The first point consists of benefit-sharing attractiveness, market demand, perceived usefulness of development tools, perceived usefulness of online forums and review process fairness. The second part is defined as learning cost and set-up cost.
The factors of dedication-based mechanism are oriented on economic, resource and social content. The economic content refers to the benefits of each party in a monetary way. In the case of an application market, the application developer will grow trust in the relationship as substantial value is gained. This encourages stable and long-term participation in the marketplace. As an example the factor benefit-sharing attractiveness is in the marketplace the ratio of value for each party. Apple, Google and Microsoft take a commission of 30% on each payment. The static ratio of three (platform provider) to seven (application provider) is questionable as both are seeking as much profit as possible from a transaction [41]. The marked demand is the factor to measure the customers need

for applications from third-party developers. Due to minimal functionality provided by the platform provider there's a great demand for applications in the marketplace. But the demand is determined by the intention of the consumer, which is proven by marketing concepts to be variable [28]. The perceived usefulness of development tools such as SDKs is a resource factor and contains the ability to code in an efficient way. Platform providers have to find a good balance to provide a useful SDK to the developer and comply to their licensing model and their openness. For instance Apples source code is not published and the developers are very limited by using the SDK. This limitation could lead to more security, but it also restricts the developer. In a survey of VisionMobile [41] low cost development tools, quick to code and prototyping were under the top five criteria for choosing a platform. Both criteria have an immediate impact on initialization or switching costs. If a developer is not familiar with the programming language, low cost development tools will likely be useful [30]. In addition the perceived usefulness of online forums is especially important for unexperienced developers, since it is a good way to learn from others and get different views on a problem. The review process fairness is considered a social factor and is somehow controversial. For example Apple does publish a guideline for applications, but the rejection rules are not clear, because of a lack of detail. The developer is not able to be sure, if the application will be published in the marketplace by time of writing the code and the impact of rejected application on developers is increasing with any hour of work on the application. For many application developers interviewed by VisionMobile [41] is the rejection a great frustration and as well a reason to switch to another platform.
Constraint-based mechanisms describe factors such as learning costs and set-up costs, which are encouraging a relation even though one party is not content with the conditions [30]. Learning costs contain the knowledge acquired by the developer to be able to produce applications for a specific platform. This knowledge is useless when switching to another platform, since platform providers use different programming languages, technology and design patterns. The time and money invested to learn platform specific requirements can exceed the discontent with the platform and hinder the developer to switch. There are conflicting goals in the context of a platform provider. Although, platform providers want to attract developers from other platforms and therefore should support low switching costs, they want developers to not leave the own portal. Blackberry published in May 2011 the support of Android applications in a future release of their tablet product called playbook to extend its ecosystem [38]. Setup-costs are generated, if for example a developer has to buy an Apple computer to use the iOS SDK or contracts and certificates have to be obtained [30]. The physical and mental investments in an environment to develop on a certain platform are considered setup-costs. These costs can not be transfered when switching to another platform and therefore act as a contraint.

3.3.2 Application Pricing

Application developers have a few options to charge customer for the application within platform marketplaces. The used pricing and revenue options should meet the expectations and intentions of the customer. Platform providers offer a variety of different pricing models and handle the corresponding money transaction. Developers get paid within a certain amount of days or weeks depending on the agreement. Whereas developers had to implement different pricing models on their own before marketplaces were used, these models are available for no development cost. The following list provides categories and use cases of the basic models as used in the Apple AppStore, Android marketplace and Zune marketplace.
Pay-per-download

The developer publishes the application for a fixed price in the marketplace. If the user accepts the price and downloads the application, the developer will get the amount minus the commission taken by the platform provider. In case the application is published for 1$ in the Apple AppStore and the commission is therefore 30%, the developer will get 0.70$. Pay-per-download is especially useful, if the application is a solution to a specific problem or if there is no long-term involvement of the developer. There is only one transaction between the developer and a unique customer. After purchasing the application, the customer can use it without any restrictions and keep it for a life-time.

Subscription

An application is paid for a certain amount of time. Apple for example supports in-app-payment, which is a process to charge the user under certain conditions after downloading the application. It is also possible to charge different prices on download than after a period of time. As an example the developer publishes an application with a download price of 3$ and charges after one year for every following year 5$. It would also be possible to offer larger time-periods for a less yearly fee. A subscription pricing model would be suited if the developer offers a long-term service, which has to be maintained. As an example a book recommendation service with a central database. The database will grow by time and the infrastructure needs to be maintained as well as complexity will rise.

Freemium

Freemium means an application is initially for free, but after certain conditions the customer has to pay for using the service. As known from computer software, a restricted mobile application could be provided for free, but the user has to pay for the activation of all functionalities. In addition it would be possible to introduce a module based pricing model and the customer could purchase different modules. For example, a developer could publish a racing game with one car for free and the user could buy new cars for the game.

Free with advertisement

Advertisements on websites are common in the world wide web to generate profit from websites users can browse without any payment. The same is applicable on mobile applications. Developers can use advertisement services to show banners within the application and generate profit. If the users intention to spend money on the context of a certain application is low, advertisements can be integrated to create value. Since revenue is generated when a user is using the application, it should be an application to spend time on. For example, a news feed reader, where a user spends time reading article would fit into this model.

Free

Free applications are the most attractive for customers, but application developer will not get any profit. Therefore this model is applicable for businesses offering mobile applications as an integration to their existing business or as vertical integration. Furthermore developers could provide customers free apps to build a relation to the customer and gain a strong image.

Gartners [17] prediction on money transfer through the application marketplaces states $2.5 billion in 2009, $8.15 billion in 2010 and $17.7 billion in 2011. The increase from 2010 to 2011 is justified by the increasing trust of customers to mobile payment processes. This is another factor, developers have to consider, when determining a pricing strategy. These payment processes are platform dependent. In addition the different pricing models are determined by the technology offered from the platform providers.

3.3.3 Conclusion

The interrelation between platform providers and application developers is particularly important for both sides. Platform providers rely on applications provided by third-party developers to increase the functionality on the device and as well generate profit from

their content. The developer relies on the technology of the platform provider and the platforms market. In this relation the developer is bound to the conditions of a platform provider. The commission taken by platform providers for example is fix. The developer has a low chance to argue upon it. It is questionable if this commission based model with a fixed price is in line with the dedication-based factors. Since the applications provided by third-party developers become more important, a more dynamic commission model could be considered. For example a rewarding system for successful application developers, which allows a lower commission, since these developers are of high importance to the platform providers. The bargaining power of application developers on the commission is very low, because the power of the mass is no bundled. If application developers would consider building an interest group to argue on commission and as well on other aspects of the relation, the relationship would become more dynamic and not only platform provider determined. There is a need for lower constraint-based factors as well. Companies are building frameworks, that allow developers to port the application to different platforms in an efficient way. Although there are different pricing models, the application provider is bound to the conditions on these different processes and services. The application developers content on the relation to the platform provider is of high interest to the platform provider and therefore a more flexible way of integration should be analyzed.

Bibliography

[1] Mohsen Anvaari, Slinger Jansen: *Evaluating architectural openness in mobile software platforms*, ECSA 2010 Proceedings of the Fourth European Conference on Software Architecture: Companion Volume, 23. August 2010. http://dl.acm.org/citation.cfm?id=1842775

[2] Apple Inc.: *Apple Reinvents the Phone with iPhone*, http://www.apple.com/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html, January, 2007

[3] Apple Inc.: *Apple Presents iPod*, http://www.apple.com/pr/library/2001/10/23Apple-Presents-iPod.html, October 2001

[4] Apple Inc.: *iPhone SDK Downloads Top 250,000*, http://www.apple.com/pr/library/2008/06/09iPhone-SDK-Downloads-Top-250-000.html, June, 2008

[5] Apple Inc.: *About Creating Your First iOS App*, http://developer.apple.com/library/ios/#documentation/iphone/conceptual/iPhone101/Articles/00_Introduction.html, October, 2011

[6] Apple Inc.: *Apple App Store Downloads Top 15 Billion*, http://www.apple.com/pr/library/2011/07/07Apples-App-Store-Downloads-Top-15-Billion.html, July, 2007

[7] Apple Inc.: *Apple Introduces iCloud*, http://www.apple.com/pr/library/2011/06/06Apple-Introduces-iCloud.html, June, 2006

[8] Apple Inc.: *Apple Introduces iTunes - World's Best and Easiest To Use Jukebox Software*, http://www.apple.com/pr/library/2001/01/09Apple-Introduces-iTunes-Worlds-Best-and-Easiest-To-Use-Jukebox-Software.html, January, 2001

[9] Apple Inc.: *About iOS Development Team Administration*, http://developer.apple.com/library/ios/#documentation/ToolsLanguages/Conceptual/DevPortalGuide/Introduction/Introduction.html#//apple_ref/doc/uid/TP40011159, October, 2011

[10] Rahul C. Basole, Jürgen Karla: *On the Evolution of Mobile Platform Ecosystem Structure and Strategy*, Business and Information Systems Engineering: Vol. 3: Iss. 5, 313-322, 30 August 2011. http://aisel.aisnet.org/bise/vol3/iss5/6/

[11] Rahul C. Basole: *Structural Analysis and Visualization of Ecosystems: A Study of Mobile Device Platforms*, AMCIS 2009 Proceedings, Paper 292, 2009. http://aisel.aisnet.org/amcis2009/292

[12] Giovanni Camponovo, Yves Pigneur: *Business model analysis applied to mobile business*, ICEIS 2003, Angers, 2003. http://www.mics.org/micsPublicationsDetail.php?pubno=318

[13] Bloomberg: *Google Buys Android for Its Mobile Arsenal*, http://www.businessweek.com/technology/content/aug2005/tc20050817_0949_tc024.htm, August, 2007

[14] Forbes: *Google Activates 200 Million Android Phones, Closes Gap With Apple*, http://www.forbes.com/sites/mobiledia/2011/11/18/google-activates-200-million-android-phones-closes-gap-with-apple/, Novemeber, 2007

[15] Renė Früh, Daniel Kesch, Stephan Plüss: *Mobile Computing Business Opportunities and Business Models from the Perspective of an IT Service Provider*, Business Engineering, I, 117-155, DOI: 10.1007/3 540 27664 5 6, 2005. http://www.springerlink.com/content/v8860u230uu14w41/

[16] Joshua S. Gans: *Mobile Application Pricing*, Melbourne Business School and Microsoft Research, 23 Mai 2011. http://ssrn.com/abstract=1850667

[17] Gartner Inc.: *Gartner Says Worldwide Online Application Store Revenue Forecast to Surpass $15 billion in 2011*, http://www.gartner.com/it/page.jsp?id=1529214, November, 2011

[18] Gartner Inc.: *Gartner Says Android to Command Nearly Half of Worldwide Smartphone Operating System Market by Year-End 2012*, http://www.gartner.com/it/page.jsp?id=1622614, April, 2011

[19] Gartner Inc.: *Gartner Says Sales of Mobile Devices in Second Quarter of 2011 Grew 16.5 Percent Year-on-Year; Smartphone Sales Grew 74 Percent*, http://www.gartner.com/it/page.jsp?id=1764714, August, 2011

[20] Google Inc.: *Where's my Gphone?*, http://googleblog.blogspot.com/2007/11/wheres-my-gphone.html, November, 2007

[21] Google Inc.: *Everything Google*, http://www.google.com/about/products/, November, 2011

[22] Google Inc.: *Android Developer Challenge*, http://code.google.com/android/adc/, November, 2007

[23] Google Inc.: *A quick look at Google+*, http://www.google.com/+/learnmore/, November, 2011

[24] Google Inc.: *Licenses*, http://source.android.com/source/licenses.html, November, 2011

[25] Google Inc.: *Philosophy*, http://source.android.com/about/philosophy.html, November, 2011

[26] Google Inc.: *Android Compatibility*, http://source.android.com/compatibility/index.html, November, 2011

[27] Google Inc.: *Google to Acquire Motorola Mobility*, http://investor.google.com/releases/2011/0815.html, August, 2011

[28] Xunhua Guo, Yannan Zhao, Yan Jin, Nan Zhang: *Theorizing a two-sided adoption model for mobile marketing platforms*, ICIS 2010 Proceedings, Paper 128, 2010. http://aisel.aisnet.org/icis2010_submissions/128

[29] Martin Kenney, Bryan Pon: *Structuring the smartphone industry: Is the mobile internet OS platform the key?*, ETLA discussion paper 1238, 3 January 2011. http://hdl.handle.net/10419/44498

[30] Hyung Jin Kim, Inchan Kim, Ho Geun Lee: *The Success Factors for App Store-Like Platform Businesses from the Perspective of Third-Party Developers: An Empirical Study Based on A Dual Model Framework*, PACIS 2010 Proceedings, Paper 60, 2010. http://aisel.aisnet.org/pacis2010/60

[31] Microsoft: *Windows Phone 7 - Released To Manufacturing*, http://windowsteamblog.com/windows_phone/b/windowsphone/archive/2010/09/01/windows-phone-7-released-to-manufacturing.aspx, September, 2010

[32] Microsoft.: *Visual Studio 2010 Express for Windows Phone*, http://msdn.microsoft.com/en-us/library/ff630878(v=VS.92).aspx, September, 2011

[33] Microsoft.: *Developing and Publishing Applications Overview for Windows Phone Marketplace*, http://msdn.microsoft.com/en-us/library/ff941089(v=vs.92).aspx, October, 2011

[34] Microsoft: *Nokia and Microsoft Announce Plans for a Broad Strategic Partnership to Build a New Global Mobile Ecosystem*, http://www.microsoft.com/presspass/press/2011/feb11/02-11partnership.mspx, February, 2011

[35] Microsoft: *Android Patent Infringement: Licensing is the Solution*, http://blogs.technet.com/b/microsoft_on_the_issues/archive/2011/03/21/android-patent-infringement-licensing-is-the-solution.aspx, March, 2011

[36] Open Handset Alliance: *Members*, http://www.openhandsetalliance.com/oha_members.html, November, 2011

[37] Esko Penttinen, Matti Rossi, Virpi Kristiina Tuunainen: *Mobile Games: Analyzing the Needs and Values of the Consumers*, Journal of Information Technology Theory and Application, Volume 11, Issue 1, pp. 5-22, March 2010. http://aisel.aisnet.org/jitta/vol11/iss1/2

[38] Research in Motion: *http://press.rim.com/release.jsp?id=4935*, RIMExpandsApplicationEcosystemforBlackBerryPlayBook, Mai, 2011

[39] Martinez Salazar, Marco Martinez: *The Mobile Phone User: Identifying Top Mobile Applications*, SMC University, 1 April 2010. http://ssrn.com/abstract=1691210

[40] Thomas F. Stafford, Michelle Belton, Terry Nelson: *Exploring Dimensions of Mobile Information Technology Dependence*, ICIS 2010 Proceedings, Paper 179, 2010. http://aisel.aisnet.org/icis2010_submissions/179

[41] Vision Mobile Ltd.: *Developer Economics 2011 How developers and brands are making money in the mobile app economy*, http://www.visionmobile.com/devecon.php, June, 2011

Chapter 4

ISP-friendly Content Distribution Systems

Daniel Meier

The total amount of Internet traffic has been rising extremely over the last decade. While a big part of this surge in traffic can be allocated to social networks and media-sharing portals, the main part of Internet traffic today is generated by multiple Peer-to-Peer (P2P) platforms. Those technologies provide a wide range of reliable and scalable services like data sharing, voice-over-IP and video streaming. Therefore, Internet Service Providers (ISPs) had to extend their infrastructure and buy more IP-transit capacity. This generally resulted in higher operating costs for the ISPs, so they tried to shape or even block P2P related traffic in their networks. The result was a "cat-and-mouse game" between P2P developers using new obfuscation technologies and the ISPs trying to detect and minimize P2P traffic. At the end, this resulted in a broad discussion about network neutrality. Current researchers are looking for alternative ways to tackle the traffic problem, while also improving the content delivery quality. The main principle behind the newer approaches is the reduction of inter- and intra-ISP traffic. This report provides an overview on actual peer matching approaches and implementations to determine their availability.

Contents

4.1 Introduction and Problem Statement

Over the last decade, the way how people use the Internet changed enormously. These days' people exchange all kind of information through various platforms, such as social networks and media-sharing portals. Going along with the increased usage of those platforms during the last couple of years, the overall Internet traffic has risen in a very steep manner. Interesting examples to demonstrate the sharp rise in overall traffic are the mean and peak value traffic statistics of large Internet Exchanges. Therefore, taking the German Commercial Internet Exchange (DE-CIX) as illustration for the rise in traffic is legitimate since it is one of the largest exchanges worldwide. Examining the yearly traffic graph for the last 800 days constitutes peak traffic values around 1Tbit/s in the summer of 2009. Two years later, the observed peak traffic values were around 4Tbit/s [5]. It is indisputable that social networks and the various media-sharing portals are responsible for a large quota of Internet traffic today. However, the number one Internet traffic producers are Peer-to-Peer platforms. Those platforms can be used for many different purposes, such as data sharing, voice-over-IP services and also for video streaming. The most popular P2P usage is file sharing, especially through the BitTorrent protocol. According to different studies, P2P traffic represents more than 50% of the Internet traffic. The Ipoque Internet study states that P2P file sharing generates 55% of the whole Internet traffic in Southern Europe. For Eastern Europe, Ipoque observed a traffic quota of up to 70% [12]. The enormous rise of Internet traffic has been leading to inevitable new antagonisms between the interests of ISPs and their subscribers. Usually, subscribers prefer the best possible service quality in terms of bandwidth, response time and routing. Those whishes are clearly comprehendible, however the private end users also intend on paying the lowest service price possible. Conversely, there are the ISPs which very often provide Internet access services as flat rate subscription schemes. To accomplish their goals, the providers have to found their pricing schemes on mixed calculations based on the required average bandwidth or the estimated traffic per subscriber. This approach worked fine until a couple of years ago when P2P systems became popular and the traffic volume increased. With the enormous rise of P2P traffic, the providers began or at least tried to influence the traffic flows on their networks. The reason for this behavior was the matter of fact that additional P2P traffic costs reduced their profit considerably. Most ISPs did not experience too many difficulties increasing their backbone capacity and clearing higher bandwidth profiles, especially if they provided best effort or traffic limited services to their subscribers. The primary concern for ISPs was the inter-ISP traffic caused by P2P applications which led to costly IP-transit traffic. In one or another way, the ISPs had to cope with their "traffic problem", otherwise their business would generate constant losses. In a nutshell, the ISPs began actively influencing P2P traffic flows. Traffic shaping and simple port blocking on well-known P2P ports were their first attempts. As reaction, the P2P developers started using random ports as countermeasure. This started a "cat-and-mouse game" between ISPs and P2P developers. Unintentionally, the providers also initiated a widespread discussion about network neutrality. The principle behind the network neutrality concept is that providers cannot impose any restrictions on Internet access based on the content, sites and platforms the subscriber invokes. Therefore, the ISPs had to find new ways to optimize inter- and intra-ISP traffic flows as well as reducing the costs on their networks.

As consequence, traffic engineering became more and more important for all network operators. Besides the private sector, the academic community also became interested in optimizing network traffic flows. The following sections provide an introduction into different peer matching approaches for P2P systems with the primary focus on the BitTorrent protocol. Optimized peer matching provides the possibility to reduce the amount of inter and intra-ISP traffic and therefore reduces the traffic costs for providers. After

introducing simple and more advanced peer matching approaches, the focus of this report lies on actual implementations and concepts.

4.2 Improving Peer Matching

One of the crucial aspects in regard to P2P system performance is the peer selection mechanism. The design choices for peer matching algorithms have a broad impact for the end users as well as for ISPs. End users prefer low latency and high bandwidth downloads for P2P file sharing and video streaming. Therefore, the end user is typically satisfied if there are enough (fast) peers available. In contrast, there are the providers which try to minimize the resulting traffic and the thereby incurred costs. Intra-ISP traffic itself is usually not the primary concern, because this traffic occurs in the providers' own network. For that reason, it is not considered to be a primary cost driver. Exceptions to this generally valid statement are leased last mile connections, where the providers are charged by backhaul connectivity providers. The primary cost driver for providers, especially in the P2P use case, is the inter-ISP traffic that causes IP-transit costs. Hence the ISPs are trying to reduce such costly traffic.
The basic BitTorrent protocol uses *random peer matching* as default algorithm, which leads to clearly discrepant interests for the providers and their subscribers. Introducing biased peer matching algorithms provides benefits for both involved parties. In the beginnings of biased peer matching for BitTorrent, the developers' primary focus was on optimizing the download performance for BitTorrent clients. The first approaches were based on comparatively simple techniques. Those methods were simple to implement and did not require a lot of computational resources. In practice all those "simple methods" had their shortcomings in one or another way. Therefore, the resulting performance improvements were very often small or in the worst case even caused degradation of the whole P2P system performance. The following paragraph introduces some of the most popular simple approaches for biased peer matching.

Geographic location biased peer matching: This peer matching approach leverages information about the geographic location of peers. The information about the locations can be extracted from IP geolocation databases. Depending on the granularity of the database, the peer matching can be biased towards using hosts from the same continent, country or city. Using this approach can reduce the latency between the peers and can result in improved download rates. A clear downside of this static approach is the ignorance of the underlying network topology itself. This means that perhaps slower peers in terms of upload capacity could be selected or congested network paths are preferred. Another issue is the fact that geographically near peers from another provider are preferred over geographically slightly more distant sources that are within the same provider network.

Autonomous System biased peer matching: This approach is based on the Autonomous System Number (ASN) of network operators. Usually ISPs have one unique ASN and therefore all of their subscribers share the same provider ASN. Based on very simple ASN lookups, the hosts within the same ASN are selected to improve the locality of peers. While this approach seems very simple and effective, there are a couple of issues. Some large providers maintain more than one ASN (e.g. Comcast seems to maintain over 40 different ASNs [2]) and therefore a simple ASN lookup is insufficient. Another problem is the fact, that this approach only improves download rates and traffic locality if there are enough seeding peers on the providers' network. Therefore, this approach relies on having enough peers available on the

same provider network, otherwise there is no direct reduction of inter-ISP traffic possible.

IP prefix biased peer matching: The Internet Assigned Numbers Authority (IANA) distributed IPv4 addresses in large blocks (the block size was /8 in CIDR notation) to the Regional Internet Registries (RIRs). Those RIRs are responsible for the IP address allocation in the different "parts of the world" like the Réseaux IP Européens (RIPE) for Europe and the American Registry for Internet Numbers (ARIN) for North America. Therefore, IP prefix biased matching allows preferring peers from the same Internet registries. More granular bias is possible for providers that own large and continuous network blocks. Since the shortage of IPv4 addresses, this seems getting less likely due to the redistribution of small network blocks. Due to the fact that providers usually pay for IP-transit bandwidth and not for transit distance (only in rare cases), this approach only leads to improvements for the end user and no feasible inter-ISP traffic reduction.

Direct measurement based approaches: Approaches that rely only on direct measurements typically use ICMP packets like "Echo Reply" and "Traceroute". The principle behind this approach is to prefer low ping and low hop count peers. Besides the simplicity of this approach, the drawbacks are arguable. Some researchers consider that the probing packets approach generates more than reasonable overhead, especially if there are a lot of probes involved. Another problem arises with nodes on the network paths that block ICMP messages (e.g. firewalls), which can render the resulting measurements useless. Generally this approach is considered time-consuming, due to waiting times for the results of complete traceroutes. Therefore, this approach can improve the end user experience but usually does not deliberately reduce the ISP traffic load.

The described approaches are usually very simple in terms of computational complexity and resource usage. Due to the fact that they all have their specific drawbacks, more evolved network positioning systems and peer matching techniques have been developed. Earlier approaches proposed *landmark based systems* like the Global Network Positioning system and *landmark-free systems* like Vivaldi. Those systems demonstrated clear improvements over the simple biased peer matching approaches, at least for the end users. As disadvantage, the new systems relied on the generation of an explicit model and therefore required more computational resources. Despite the progresses on the end user side, the ISPs were still troubled by the occurring P2P traffic flows. The latest developments of biased peer matching systems are explicitly considering the ISPs requirements. Currently, there are two primary approaches to create ISP-friendly peer matching systems. The first ones are *provider-aided approaches* where the ISP is explicitly involved in peer matching. The second ones are *client-side approaches* where different methods are used to create views of the network topology. Especially notable are the approaches where already existing network views from large Content Delivery Networks (CDNs) are reused. The following paragraph gives an overview of the latest approaches.

Landmark based systems: Landmark based systems are trying to estimate the network distance between two peers based on a small set of distributed hosts called landmarks. Founding on the measurements of the inter-landmark distances, it is possible to model the Internet as an n-dimensional geometric space (Euclidean space). On that space, the position of each host is characterized by geometric coordinates. Therefore, the geometric distance between hosts is used to predict network distances. The Global Network Positioning (GNP) system is a well-known implementation of such a landmark based approach. Based on extensive Internet experiments, the

researchers concluded that a 7-dimensional Euclidean space can predict Internet distances among globally distributed hosts in 90% of the cases with less than 50% error [6].

Landmark-free systems: Landmark-free systems try to generate a fully decentralized computation of network locations. The information is encoded in a low-dimensional Euclidean space [3]. Systems based on this approach are trying to predict the network distances based on various ping measurements that are used to generate a synthetic coordinate system. A general drawback of landmark-free systems is the fact, that they are mostly based on Internet latencies and therefore violate the triangle inequality [4]. One of the most popular implementations is the Vivaldi network positioning system plug-in distributed by the Vuze (formerly Azureus) BitTorrent client.

Provider-aided systems: Provider-aided approaches require the involvement respectively the support of the providers. Such approaches can base on different concepts like altering the P2P control data stream to bias peer selection towards hosts within the same network. Another approach relies on the providers, which explicitly facilitate network information. Both approaches have the mutuality that they require explicit provider support in the area of hardware and information allocation. Currently, there are multiple efforts to establish a standard for provider supported network information. Nevertheless, those approaches can also result in unknown legal implications, since P2P systems can be used for the distribution of copyrighted material.

CDN-based Relative Positioning (CRP) systems: CRP systems are (re)using already existing information about the actual network topology. This information is gathered from large CDNs. The principle behind is to leverage the network views that CDNs generated as well as using indirectly their globally distributed (mirror) clusters. Hosts with the same low latency against such a CDN cluster are considered being located nearby each other.

4.3 Provider-aided Approaches

Provider-aided approaches depend on ISP provided infrastructure or subscribers using ISP provided network information. The first approach requires ISP operated BitTorrent trackers and dedicated hardware to redirect client tracker queries to the ISP tracker. The second approach relies on the concept that the ISP provides some kind of portal with network related information to the subscribers. Both approaches work at least in theory, but they have their drawbacks in real-world applications. All approaches require a mutual trust basis between ISPs and their subscribers which is not given per se. Using ISP operated trackers can lead to several legal implications, because the providers could be directly involved in copyright infringements. The following sections introduce ISP Managed Peer-to-Peer (IMP) and the Transparent Network Tracker (TnT) as example projects for ISP deployed tracker approaches. The Provider Portal for Applications (P4P) project is a generic example of using provider supplied network information, while the IETF Application-Layer Traffic Optimization (ALTO) is a work in progress project on a future standard for provider supplied network information.

4.3.1 TnT and IMP

IMP [9] is the predecessor of TnT [10]. Both concepts were proposed by the same authors and TnT includes the latest advancements in the area of ISP tracker approaches.

The underlying principle is that edge routers of ISP networks are extended with high-performance network processors (NP). Those NPs are dedicated devices to detect tracker queries from the subscribers and then redirect them to the ISP operated tracker. The tracker replaces the default random peer selection of BitTorrent with a peer matching that prefers peers within the local network. This approach leads to the advantage that no protocol changes are required. Therefore, this solution is independent from the BitTorrent client a subscriber uses. On the drawback side, there is clearly the fact that this approach requires the ISP to operate dedicated hardware. Figure 4.1 visualizes how TnT handles tracker query redirections.

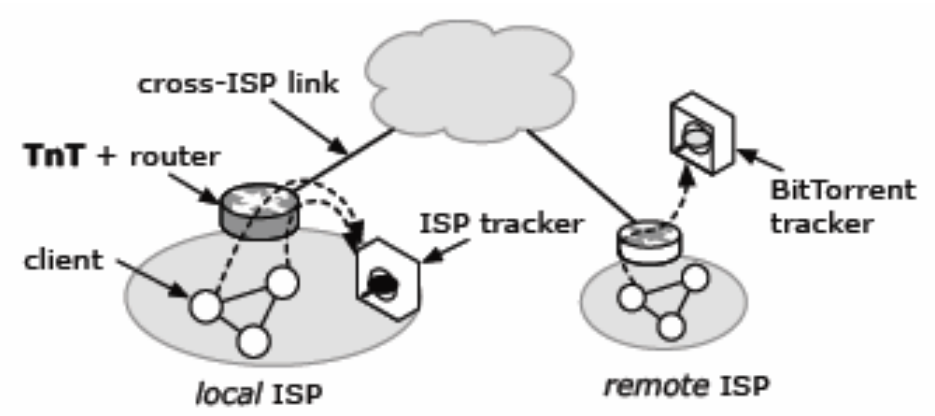


Figure 4.1: TnT tracker query redirection [10].

At the moment, there is only a proof of concept implementation of TnT available. Therefore, those systems have only been evaluated in small academic research networks or test-beds. Those usually do not reflect real-world BitTorrent swarm behavior. Nevertheless during the evaluations, the inter- and intra-ISP traffic was notably reduced. The end user download rates were also enhanced. Besides those benefits, there are not ignorable shortcomings like the fact that the ISPs have to deploy dedicated hardware. Especially NPs can become very expensive if they are required in great quantities. Another serious problem is the fact that ISPs could be involved in the distribution of copyrighted material, which results in legal implications. At the moment, it is not clear up to which degree a tracker operator can be hold responsible for the distribution of copyrighted material. Last but not least, the TnT approach is pretty much useless when the subscribers use encryption and/or nonstandard ports on their BitTorrent clients or any other obfuscation method. It is utterly impossible or economically inefficient to analyze the complete data streams at edge routers in real time.

4.3.2 P4P

The P4P project [13] represents a voluntary open standard that is based on network provider cooperation. The P4P working group core members include companies like AT&T, Cisco, Pando Networks, Verizon as well as the Washington University and the Yale University. The main principle behind P4P is that network providers facilitate explicit information, guidelines and capabilities of their network to emerging P2P applications. Internally, the ISPs calculate so called p-distance values which the subscribers can use for optimized peer matching. The values are provided through a provider operated portal called *iTracker*. Besides the existing P2P networks, the subscribers run an *appTracker* that registers the clients with the iTracker. This leads to divided traffic control responsibilities between applications and network providers. Therefore the end users are able to choose if they want to use the network provider facilitated information. Figure 4.2 shows the iTracker interfaces and the according information flow.

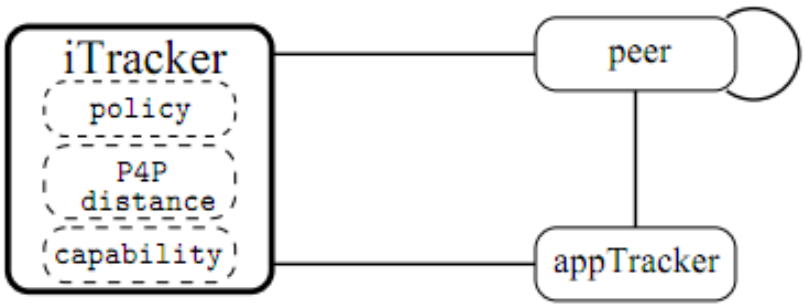


Figure 4.2: P4P iTracker interfaces and information flow [13].

Global P4P field tests from Pando Networks showed increased delivery speeds up to 235% across US cable networks and up to 898% for international broadband networks. Further tests showed that the inter-ISP traffic was reduced by an average of 34% based on values up to 44% for US networks and 75% for international networks [11]. While the evaluated values seem very promising, there is again the risk of distributing copyrighted material. As solution for this problem, P4P also provides mechanisms to prevent or at least narrow down the distribution of copyrighted material.

4.3.3 ALTO

The IETF ALTO [8] is a work in progress project. Basically the ALTO working group uses concepts that are similar to P4P, but more advanced and applicable for generic P2P use cases. The working group solely focuses on the communication protocol between applications and ALTO servers. Therefore, the primary considerations of the working group are based on preferable and avoidable IP ranges, ranked lists of requested peers, information about topological proximity and approximate geolocation [8]. Since ALTO is still in the drafting phase, there are mostly specifications and use cases available.

4.4 Client-side Approaches

The client-side approaches are generally easier to realize compared to ISP-aided approaches. Client-side solutions can be distributed as plug-ins for existing P2P systems. Another benefit is that there is no direct ISP involvement required, therefore the providers are not facing any direct legal implications. Actual client-side approaches have different concepts for optimal peer matching algorithms. Those algorithms can range from models that require very low computational resource usage up to very complex and computational intense models. The following paragraph presents two interesting client-side approaches. ISPF-(Lite) is a very complex approach that leverages publicly available information to generate an overall network topology. ONO in contrast reuses CDN based network views.

4.4.1 ISPF(-Lite)

ISPF-(Lite) [7] is an ISP-friendly peer matching algorithm. The general approach of ISPF is to infer the underlying network topology and create a distance oracle that minimizes the intra- and inter-ISP costs. The algorithm itself distinguishes between *ISP distance* and *PoP distance*. ISP distance reflects the inter-ISP routes, while PoP distance focuses on the intra-ISP routes. By aggregating the metrics, ISPF creates a complete distance oracle for ISP-friendly peer matching. Figure 4.3 illustrates an example network topology. When receiver R requests partial files from the senders S1, S2 and S3, inter-ISP traffic

cannot be avoided. A closer look on the PoP locations shows, that there is optimization potential on intra-ISP traffic.

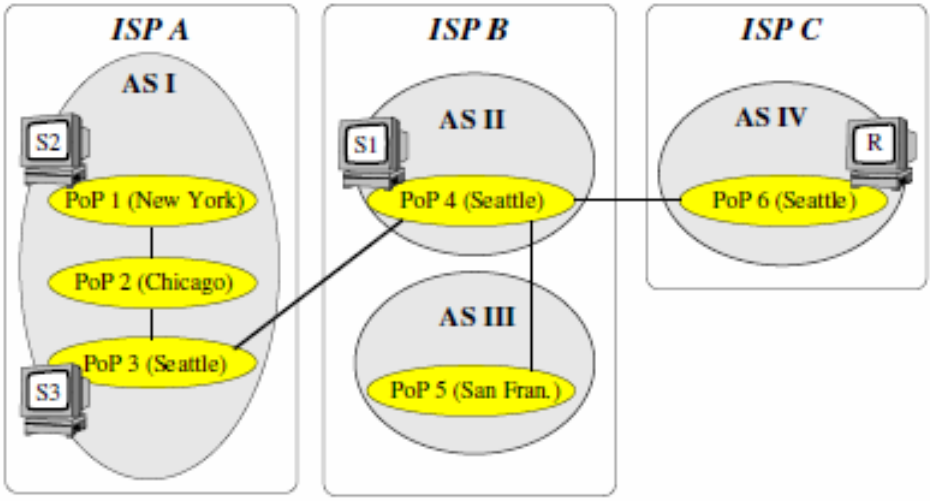


Figure 4.3: Illustrative network topology [7].

To compute ISP and PoP distances, the algorithm leverages publicly available information about BGP tables, BGP updates and IP geolocation databases. The ISP distance is based on the count of inter-ISP links along the inferred AS path between two peers. This procedure creates internally an AS graph. Estimating the PoP distance requires an algorithm that is a bit more complex. At the beginning the algorithm uses IP geolocation databases to cluster all IP prefixes (gathered from BGP tables) into one or more PoPs. To generate an updated view of the actual PoP topology for each AS, the algorithm also accounts for BGP updates. At the end, the PoP topology is combined with the AS graph of the ISP distance. Creating such a model for a distance oracle requires a lot of computational resources. Besides raw CPU time, a not optimized distance oracle requires also a lot of memory. Storing more than 25,000 ASes alone requires around 625MB [7]. To create a more efficient and concise distance oracle, ISPF only stores the ISP distance between any two *transit ASes*. This allows reducing the memory usage to 10MB while also achieving an *O(1)* lookup time. The authors propose the usage of ISPF-Lite when the computational resources are scarce. Especially the PoP computation is very demanding, since there are too many PoPs on the Internet. To reduce this burden, ISPF-Lite sorts tied potential senders on the length of the shared IP prefix with the receiver. The logic behind this approach assumes that the longer the shared IP prefix between two peers is, the more likely they are located within the same network. This works nearly cost free since IP prefix matching is a very simple operation. Figure 4.4 shows the components of the ISP-friendly distance oracle.
ISPF(-Lite) was evaluated based on trace-driven tests. The data sets for the evaluations have been gathered from collected torrents (isoHunt) and from CBC/Radio-Canada. During the tests, ISPF outperformed random peer matching up to six times [7]. The researchers also observed significant traffic reductions. In opposite to the promises made, ISPF did not yet appear in public. While the ISPF approach seems very promising, there are also drawbacks. A problem is that the algorithm introduces a high computational complexity that results in inevitable offline preprocessing. In practice, this means that it is not possible to calculate the model in real time.

4.4.2 ONO

The basic approach of ONO [2] is to recycle network views that have been generated by large CDNs. Therefore, the peer matching oracle relies on public information sources from

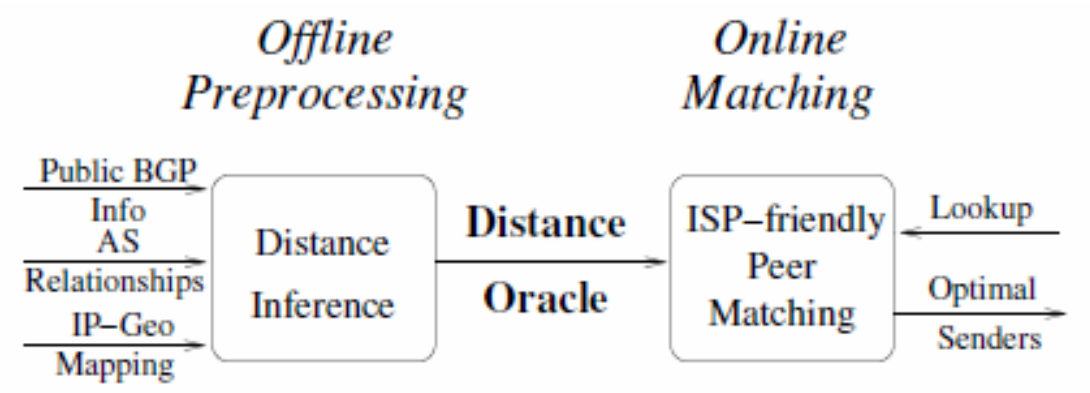


Figure 4.4: Components of the ISP-friendly distance oracle [7].

CDNs like Akamai and Limelight Networks. Such information is usually gathered through the resulting DNS redirections by accessing CDN-hosted Internet sites. A big advantage of this approach is that DNS redirections reflect the actual Internet topology as well as ISP policies. To keep track of the DNS redirections to CDN clusters, ONO maintains internally ratio maps. If two hosts have the same ratio map values, the general assumption is that the path between them should only cross a small number of networks. Based on the formulation of ratio maps, the cosine similarity of the ratio maps between two different hosts can be calculated. If the resulting value is greater than a specific threshold, the peer is recommended in terms of peer matching. This leads to inter-ISP traffic reduction. Because CDNs and their DNS redirections are primary latency driven, ONO follows this paradigm. The generic principle behind such CRP systems is the assumption that low latency hosts are likely to be close to each other. Therefore, they are likely to have the same ISP and based on that fact, they should produce less inter-ISP traffic.
ONO has been extensively tested in test-beds and under real-world conditions (over 120,000 subscribers for the Vuze plug-in). During the test-bed evaluations, the inter-ISP traffic was reduced in over 33% of the time and ONO also led to a two order lower latency. Compared with random peer selection, the loss rate was lowered around 30%. The download rate improved around 32% and in case of large bandwidth environments, the download rates were increased up to 207% on average [2]. While those numbers alone seem very impressing, there is the downside that those values are just test-bed measurements which do not reflect real world conditions. Currently there is a broader discussion in the academic community about the value and validity of test-bed based results [1].

4.5 Summary and Conclusions

The approach of creating ISP-friendly content delivery systems is without any doubt very reasonable. Both end users and providers gain benefits from optimized peer matching algorithms, especially when comparing with random peer matching. Better peer matching results in a better end user experience in terms of faster downloads and better stream quality. Conversely, providers are struggling with high P2P traffic volumes these days. Especially the huge amount of inter-ISP traffic results in higher operating costs. Due to low margins and a huge competition in the ISP market, providers cannot adapt their pricing schemes from one day to another. Optimized peer matching provides a great ability to reduce P2P related traffic loads and therefore clearly benefits the ISPs.
Different approaches for ISP-friendly content distribution systems can be considered as possible tools for traffic engineering purposes. Therefore, the approaches presented in this report can help network operators to manage their traffic flows. Assuming that the overall Internet traffic rises in a similar manner like over the last years, improved traffic

engineering methods will be crucial for ISPs and their success. On one side there are the provided-aided approaches and on the other side there are client-side solutions. Provider-aided solutions like P4P and most notably ALTO can really produce benefits for all involved parties. The fact that an IETF working group addresses those problems, shows that such a solution is not only wished, but necessary. The ALTO problem statement is noted down in RFC 5693, therefore this standard proposal is really taken serious. Other provider-aided approaches like usage of ISP trackers seem to be less likely in real-world environments. Reasons for this are the initial hardware investments as well as the acceptance of having the ISPs interfering with data streams. The weightiest counter-arguments to this approach are clearly the pending legal issues of aiding in the distribution of copyrighted material. Therefore, client-side approaches seem to be the solution for the near future, since they can be easily applied to a lot of P2P clients. The largest challenge on this approach seems to be the multitude of P2P systems and clients. This makes it difficult to establish a large installed base. A general aspect that needs to be accounted for is the fact, that all the measurements and evaluated performance improvements were mostly accomplished under test-bed conditions. Those test-bed conditions display only a specific use case. This makes it hard to nearly impossible to state the real benefits gained in inter- and intra-ISP traffic reduction as well as performance gains for end users. Obviously, this problem lies in the nature of P2P file sharing systems. The amount of seeding peers, their link speed and the geographic distribution is never the same. The only way to solve this problem, at least for comparison purposes between different peer matching algorithms, would be an averaged and preferably real-world reflecting testing standard.

As general conclusion, it can be stated that ISP-friendly content distribution systems are beneficial for ISPs and their subscribers. It is quite feasible that ALTO will have a huge impact in traffic engineering if it establishes as standard. Even if the methods of information exchange will alter in other directions in the future, there is still the need of better tools for traffic engineering that ALTO could provide.

Bibliography

[1] David R. Choffnes, Fabián E. Bustamante: *Pitfalls for Testbed Evaluations of Internet Systems*, ACM SIGCOMM Computer Communication Review, ACM, New York, NY, USA, 2010, Volume 40 Issue 2.

[2] David R. Choffnes, Fabián E. Bustamante: *Taming the Torrent: A Practical Approach to Reducing Cross-ISP Traffic in Peer-to-Peer Systems*, Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication (SIGCOMM 2008), ACM , New York, NY, USA, 2008, 363-374.

[3] David R. Choffnes, Mario A. Sánchez, Fabián E. Bustamante: *Network Positioning from the Edge: An empirical study of the effectiveness of network positioning in P2P systems*, Proceedings of the 29th Conference on Information Communications (INFOCOM 2010), IEEE Press, Piscatway, NJ, USA, 2010, 291-295.

[4] Frank Dabek, Russ Cox, Frank Kaashoek, Robert Morris: *Vivaldi: A Decentralized Network Coordinate System*, Proceedings of the ACM SIGCOMM 2004 Conference (SIGCOMM 2004), ACM , Portland, OR, USA, 2004.

[5] DE-CIX Website: *DE-CIX - Traffic Statistics*. http://www.de-cix.net/content/network/Traffic-Statistics.html, last visited: 19th November 2011.

[6] GNP Website: *Global Network Positioning (GNP)*. http://www.cs.rice.edu/ẽugeneng/research/gnp/, last visited: 19th November 2011.

[7] Cheng-Hsin Hsu, Mohamed Hefeeda: *ISP-Friendly Peer Matching without ISP Collaboration*, Proceedings of the 2008 ACM CoNEXT Conference (CoNEXT 2008), ACM , New York, NY, USA, 2008, Article 75.

[8] IETF Website: *Application-Layer Traffic Optimization (alto) - Documents*. http://datatracker.ietf.org/wg/alto/, last visited: 19th November 2011.

[9] Shakir James, Patrick Crowley: *ISP Managed Peer-to-peer*, Proceedings of the 5th ACM/IEEE Symposium on Architectures for Networking and Communication Systems (ANCS 2009), ACM , New York, NY, USA, 2009, 167-168.

[10] Shakir James, Patrick Crowley: *TnT: Transparent Network Tracker for P2P applications*, Proceedings of the 6th ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS 2010), ACM , New York, NY, USA, 2010, Article 28.

[11] Pando Networks Website: *Pando Networks Releases Results of Global P4P Field Test to Improve Peer-to-Peer Performance in Broadband Networks*. http://www.pandonetworks.com/node/81, last visited: 19th November 2011.

[12] Hendrik Schulze, Klaus Mochalski: *Internet Study 2008/2009*, Ipoque, 2009. http://www.ipoque.com/sites/default/files/mediafiles/documents/internet-study-2008-2009.pdf, last visited: 19th November 2011.

[13] Haiyong Xie, Y. Richard Yang, Arvind Krishnamurthy, Yabin Liu, Avi Silberschatz: *P4P: Provider Portal for Applications*, Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication (SIGCOMM 2008), ACM, New York, NY, USA, 2008.

Chapter 5

The Hidden Extras: The Pricing Scheme of Cloud Computing

Stephane Rufer

Cloud computing is an ambiguous term, with varying definitions of what exactly it entails. In the following paper the definition provided by th NIST is used as a framework for conceptualizing cloud computing. With this definition in mind, the scope of cloud computing is expanded to include the aspects of charging and pricing, and how they apply to cloud computing. From this a technical use case is provided, which is used to evaluate the five large payers in cloud computing (Amazon, Rackspace, Terremark, IBM and Windows Azure). The results of this investigation shows that there are many aspects to hidden costs in the pricing schemes of cloud providers. These costs are not only of technical nature (load balancing), but also in terms of pricing models, which suggests that arbitrage between different cloud providers can be executed to exploit certain hidden costs.

Contents

5.1 Introduction

Cloud computing, like its name suggests, is a very cloudy subject. This opaqueness already begins with the definition. There is no clear and universal definition for cloud computing, even though there are many approaches. Some say that cloud computing is an entirely different approach to computing and infrastructure development. Others argue that cloud computing has been around for many years in the form of grid computing, and is just a new twist on old technology.
Indeed cloud computing is concept that has been hyped to extremes over the last couple years. The Gartner Hype Cycle of 2011 put cloud computing at the peak of inflated exceptions [4]. As a result there are a pleora of publications that deal with this topic. Like cloud computing itself, the research on the topic is vast as well as opaque, with little data and hard facts concerning the costs and pricing structures of cloud providers.
To be able to elucidate the hidden costs of cloud computing, one must first define what cloud computing entails. A comprehensive definition of cloud computing helps to lead to this end and forms the framework within which one can investigate charging and pricing. This terminology is central for understanding the functionality of pricing schemes in the cloud computing paradigm, as well as for pinpointing specific hidden costs in these models. Next, a selection of the key market players are analyzed, with the intention of finding a common denominator in their pricing schemes from which a meaningful use case can be created. Out of this common conception and with the help of the established theoretical framework a use case is derived, which is then applied to the specific providers. In addition to this general use case an example scenario for a hidden cost is applied, which helps to visualized where hidden costs can lie and how they can be used to the benefit of the cloud consumer. Finally, the findings resulting from the application of the use case are summarized and integrated into a congruent statement showing the difficulties encountered with pricing schemes in cloud computing as well as potential hidden costs.

5.2 Related Work

In A View of Cloud Computing, Michael Armbrust *et al.* define a globally applicable framework for cloud computing based on the ideas formulated in the NIST Cloud Computing Reference Architecture. This is the basis on which the definition of cloud computing used in this paper is founded in. With this basis of cloud computing, the ideas on accounting in dynamic, scalable systems in An Integrated Accounting and Charging Architecture for Mobile Grids by Cristian Morariu, Martin Waldburger, Burkhard Stiller can be synthesized and expanded to fit into the paradigm elucidated in the paper In Pricing and Charging for QoS by Safiullah Faizullah and Ivan Marsic. These papers contribute to the understanding of pricing and charging in the field of Information Technology. Additionally they give an outlook into the cloud computing space. It is these element on which the definitions in this paper are predicated on. The implications of these theoretical views in an applied practical situation is analyzed by Hongyi Wang *et al.* in Distributed Systems Meet Economics: Pricing in the Cloud. The concepts expressed in this paper form the basis for analyzing hidden costs within the pricing schemes of cloud computing providers.

5.3 Cloud Computing: A Definition

Cloud computing is a very opaque area with many different definitions and concepts of what exactly cloud computing is. With these varying concepts and ideas, it is necessary to first develop a global understanding of cloud computing. One needs to define a baseline

definition, to then build on and expand into the realm of pricing in the cloud. Armbrust et al. [7] assert that the three key aspects of a cloud computing service are the appearance of infinite, on demand commuting resources, the elimination of upfront commitment and the ability to pay for resources as they are needed. These are the generic requirements that must be fulfilled for a service to be a cloud service.
In terms of the service structure at the topmost level, cloud computing can be separated in to three distinct categories. Each of these categories addresses a separate consumer need, as well as providing a variety of services [9].

- **SaaS (Software as a Service):** Is the cloud based delivery of complete software applications. These applications run on infrastructure fully managed by the SaaS provider. Charging is usually on a subscription basis targeting an end user, where configuration is limited to application settings. In short SaaS provides an **application stack** *e.g.* GMail.

- **PaaS (Platform as a Service):** Is the delivery of a visualized runtime platform that has a software stack for developing applications or application services (programing language, libraries, etc.). The infrastructure as well as the platform is run and managed by the service vendor. With PaaS the customer is dependent on the technology of the service. PaaS provides a full **software stack** for developers *e.g.* Google App Engine.

- **IaaS (Infrastructure as a Service):** Is the delivery of the raw computing infrastructure such as servers and storage. The underlying hardware is visualized, thus providing the transparency of a service upon which applications can be built. The infrastructure is managed by the vendor, but the customer has full customization control (all the way down to the OS running on the hardware). This means that the service is technology independent. IaaS is concerned with delivering **computing infrastructure and storage** *e.g.* Amazon EC2

Generally speaking one can assert that SaaS represents the high level layer of cloud computing, where the end user is in direct contact with the cloud service. Everything is managed by the vendor. PaaS represents the middle tier, were a software library is provided on which the customer builds an application for the end user. The vendor manages the entire infrastructure as well as the software platform, with the customer managing the application and the connected data. Consequently IaaS is the most low level cloud service, where the customer has full control over the application all the way down to the hardware level. The vendor manages the actual raw hardware as well as it visualization, but the customer controls the whole software environment on which applications for end users are built. An important fact one must keep in mind is that in all cases, the customer does not physically know where the application is running. Expansion and contraction of computing power is fully transparent.
This tier-like categorization helps to give a general overview of the consumer facing services, but does not cover the intricate details of the cloud computing environment, as well as the relationships of various actors within an around the cloud computing market. The NIST (National Institute of Standards and Technology) offers the most comprehensive and generic definition of cloud computing. The NIST identifies five different actors in the cloud computing space. Further, the NIST defines a role a set of activities and functions for each of these actors. This is summarized in the table 5.1 below.

Table 5.1: Actors in the cloud computing space

Actor	Function	Activities

Cloud Consumer	Person or organization that uses the services of a cloud provider.	• Browses service catalogs • Requests services • Sets up contracts (SLAs) • Uses services.
Cloud Provider	Person or organization that makes a service available.	• Installs infrastructure and software • Maintains services and infrastructure • Supports services • Manages infrastructure and software
Cloud Auditor	Independent person or organization that examines and evaluates cloud services.	• Makes assessments of cloud providers • Verifies security controls, privacy impact and performance of service provider • Publishes assessment results
Cloud Broker	Person or organization that mediates the relationship between cloud providers and cloud consumers in respect to use, performance and delivery	• Enhances services • Aggregates services of multiple cloud providers • Conducts service arbitrage, *i.e.* providing flexible and opportunistic pricing
Cloud Carrier	Person or organization that functions as an intermediary between cloud providers and cloud consumers in respect to delivery of services (connectivity and transport)	Provide access to services by means of network and telecommunication technology

Figure 5.1 shows a general abstraction of how all these actors interact as well as how their activities and functions relate and interface with one another.

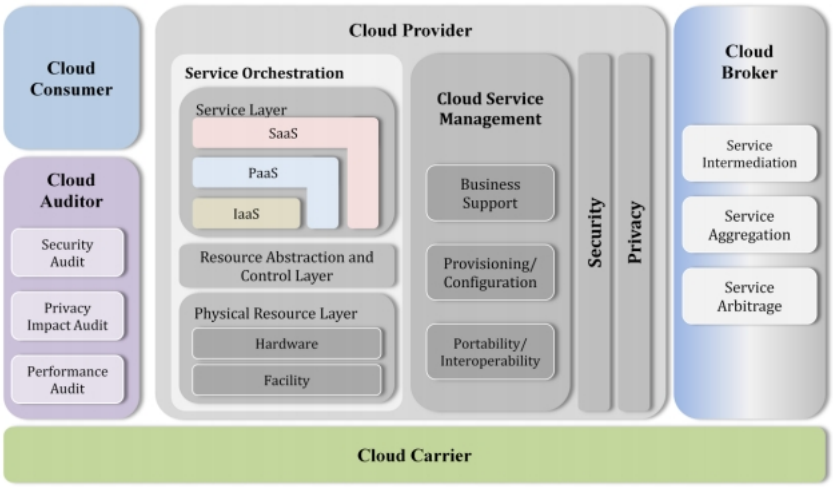


Figure 5.1: An overview of the cloud computing architecture as described by the NIST [3]

The third dimension of cloud computing consists of the method of deployment [10]. As with the different tiers of service models in cloud computing, there is a variety of ways a cloud application can be deployed.

- **Private Cloud:** The private cloud is much like the concept of a private subnet of a company. The resources of the cloud are available only to authorized users of one entity. The actual physical location of the hardware can be on or off premises, which indicates a high need for security on the side of the vendor, if the cloud is not in house.
- **Community Cloud:** The community cloud is also a private cloud in the sense that only specific consumers are eligible to use the service. However it differs from the private cloud in that multiple users from multiple entities access the cloud infrastructure.
- **Public Cloud:** The cloud infrastructure is open for general public use and may be managed by any entity.
- **Hybrid Cloud:** A combination of two or more of the aforementioned cloud deployment methods, where some sort of technology allows data and application portability between clouds.

The actual real world cloud computing service providers combine various elements of these three dimensions, which then make up the services they provide to customers. In practice this blurs the line between the different aspects of each dimension, as well as the boundaries of the dimensions amongst themselves. It is precisely this dynamism that makes categorizing a cloud service so difficult. Additionally it fosters the confusion in creating an abstract definition of what cloud computing is.

5.4 Charging/Pricing Terminology in IT

Charging: Charging is about the actual object being charged, thus a technical measure. It is concerned with the units of trade, which in this case are the metrics (packet rate, delay etc.).

Pricing: Pricing is about how is the object is being charged. Pricing plays a key role in the market environment, as it bridges the gap between charging and what is billed to the customer. It controls how the actors (customer and provider) in the market conduct themselves to reach maximum utility.
In principle charging is simple, a countable element is recorded and aggregated, the result of which is then the input of some function that defines the pricing schema. This simple process becomes exceedingly complex as the environment within which the charging and pricing occurs becomes more advanced. In the case of cloud computing flexibility, unpredictability and versatility, all key elements of the cloud service result in substantial difficulties for charging and pricing. As stated in [2] an inherently dynamic system needs to be supported by a context-based charging model. This implies that not only the service, but also the charging mechanism needs to be transparent. This means that a user is charged the same price regardless of when, where and how the process is executed. The internal technical workings of the cloud do not interest the user on a technical level and thus should also not have any influence on the accounting level of an interaction with the cloud. Additionally, this is a key aspect of personal and social pricing fairness in respect to the cloud environment [5]. Lastly, the billing service should not only offer the same degree of transparency as the service itself, but also the same functions of service and domain aggregation. [2] shows how a generic architecture can be built, which is on par with the cloud service it supports in respect to transparency, mobility and aggregation.
Charging is a defined, fine-grained framework based on technical indicators that make up the parameters of Quality of Service (QoS) (source). These concrete technical parameters are provided by metrics, such as packet rate, packet size, delay etc. Metrics are specific internal process, that depend upon the particular application, whereas charging is a global generic model that encompasses metrics. This raw data is the basis on which the actual price billed to the consumer is calculated via the pricing scheme (pricing function).
As basic economics teaches us, pricing plays a key role in the market. It functions as an efficient allocator of scare resources. Pricing has the function of bridging the gap between the user optimizing an application on the basis of incurred cost and the provider designing an infrastructure that maximizes profit on the basis of a pricing schema [5]. For this to be possible the user of a service needs to be able to infer how the underlying mechanisms work (*i.e.* the pricing scheme must be transparent). Concurrently, the resulting cost of the usage of a resource should be equal to the utility it provides to the customer. This all factors into how a resource is used and allocated.
Safiullah Faizullah and Ivan Marsic show the following generic formula that illustrates the key elements of how charging and pricing interact, resulting in the cost of a specific service [12]:

$$C_{traffictype}(\sigma_{QoS}) = \alpha_{traffictype} * P(\sigma_{QoS}) + \beta * R(\sigma_{QoS}) + \gamma \tag{5.1}$$

In the formula 5.1 $C_{traffictype}(\sigma_{QoS})$ is the total cost of a specific traffic type. $P(\sigma_{QoS})$ is the pricing function and $R(\sigma_{QoS})$ is the function for the resource reservation charge. σ_{QoS} is the Quality of Service defined by the metrics that are gathered by the metering system. $\alpha_{traffictype}$, β and γ are coefficients. $\alpha_{traffictype}$ is for usage charges, β is the charge for reservation and γ represents the fixed access charge.
This formula shows that the activity of charging involves the metering and aggregation of various QoS parameters that are then applied to a pricing scheme, which in turn should enable the service provider to price the customer in a fair and context insensitive way. This shows that charging relies heavily on metrics, which implies the the method of capturing such raw data is of great importance. The only variables that influence the final cost are access charges (γ), charges for resource reservation (β) and charges for a higher service guarantee ($\alpha_{traffictype}$). The actual price of the underlying metric ($P(\sigma_{QoS})$) remains

constant. A key element for such a mechanism is that the method used is as scalable as the underlying service being charged. The price per atomic unit of a metric, as well as the cost of recording it should remain the same regardless of the actual scale of the service that is being demanded by the customer.
Considering all the points raised above, one can see an issue the transcends all charging and pricing schemes related to cloud computing. Namely the fact that there is a gap between the billing and usage of a resource. This raises the question if real time charging and subsequent "real time" billing of a resource is even possible. One can also conclude that the transparency of a distributed system results in the intransparancy of the price for using this system. The user is no longer aware of the underlying hardware being used and thus cannot, or only insufficiently benchmark the price of this used resource. This results in the paradox that a technically transparent system becomes inherently opaque in economic terms.

5.5 Comparative Analysis of Cloud Computing Providers

In the cloud computing market there are a multitude of different vendors. With the general boom in cloud computing a vast number of startups have emerged with new methods both in terms of technology and pricing, challenging the established players. With such a wide spectrum it is difficult to properly gain an overview of the entire market. For this reason a subset of the market has been chosen including all the key vendors, which will be analyzed in detail.

5.5.1 Amazon

The usage of Amazon Web Services includes a "Free Tier", which is allocated on a per month basis for the duration of the first year of a new customer. Amazon offers three different types of instances according to user needs. These types are on demand, reserved and spot instances.
On demand instances are priced solely on a per hour bases (*i.e.* only the used resources are billed). This follows the schema of a traditional pay as you go service in that partial instance hours are billed as a full hour, much like the pay as you go services of mobile service providers.
For a reserved instance the customer pays a one time fee, which can be limited to a one or three year term and in turn the hourly rate for instance usage is reduced significantly. The customer may then choose to whether to run the instance or not, in which case the one time fee becomes a sunk cost and there are no usage charges. The difference between an on demand and reserved instance, is that the customer makes a long-term commitment by paying a one time fee.
Spot instances allow customers to place bids on unused resources in the Amazon cloud. If a customer's maximum bid is above the spot price, the resource is allocated to this user and the current spot price is billed, if not the request is not served. The spot price is determined by supply and demand in the Amazon cloud, whereby Amazon specifies a lower limit under which the spot price never falls.
In addition to this general pricing breakdown, they segment these instance types further by the size and power of the instance as well as the region of the datacenter where the instance is located. There are six different subcategories of instances within each of the three basic types. Within each of these subcategories there are some that are refined even more by size *e.g.* standard instances are subdivided into small, large and extra large instances. This subdivision is homogeneous across all the basic instance types. Much like the size and power segmentation of instances, the regional segmentation follows the same

blueprint. A customer can choose to run an instance in the Virginia, Oregon, Northern California, Ireland, Singapore or Tokyo datacenter. The pricing is differentiated on all levels (instance type and subcategory) across all regions.
Additionally Amazon charges outgoing bandwidth on a per tier basis. The first GB is part of the "Free Tier", after which each GB is charged within the respective tier. As with instance types, bandwidth charges are also segmented according to the datacenter region. All incoming bandwidth is free of charge, as well as data transfers between instances and other Amazon Web Services (AWS) located in the same region is free if they are located in the same availability zone. An availability zone is a further technical segmentation of a region to provide greater fault tolerance. Between instances and AWS services in different availability zones in the same region, a regional data transfer charge or $0.01 per GB is levied. Data transfered between all AWS services in different regions is charged at the normal data transfer rates.
A certain amount of block storage is included per instance. The size of this block store varies depending on the instance type, with some instance types not including any block storage, which would require the Amazon Elastic Block Store.

5.5.2 Terremark

Terremark has 3 different types of cloud services. One of them is Enterprise Cloud, which provides dedicated cloud resources for enterprises. The vCloud Datacenter service provides hybrid cloud services also geared towards enterprises. vCloud Express it the full scalable cloud computing service provided by Terremark. It is this service for which the pricing scheme will be detailed below.
Terremark splits its server offerings into unlicensed and licensed servers. Licensed servers are servers with a pre-installed images of Windows, for which a separate license fee is charged. Unlicensed servers are blank servers, on which any operating system can be installed. The hourly rate charged is dependent on the amount of virtual processors, RAM and the server type (licensed or unlicensed).
Internet bandwidth (transfers in and out of the Terrmark cloud) are each charged at $0.17 per GB respectively. Bandwidth between servers within the Terremark cloud is not charged.
There is some block storage included in the server price, but the amount is not specified. One can conclude from the information provided that storage is de facto not included in the server price. Storage is charged per GB and is prorated on a hourly basis. This means that the amount of storage used is measured every hour and charged at that measured rate. There is no charge or I/O operations on files in the block storage.
A specialty of Terremark is that they offer "Internet Services" at an hourly rate or $0.01. An example of an Internet Service are load balancers, to which an unlimited amount of servers can be hooked up to for the same hourly rate.

5.5.3 IBM

IBM has a cloud offering geared towards enterprise customers. IBM virtual machine instances are separated into 32 bit and 64 bit instances, with separate reserved instance pricing for each. They offer four different service types (Copper, Bronze, Silver, and Gold) for 32 bit instances and an addition Platinum service for 64 bit instance. Each of this different services includes an allotted amount of virtual CPUs, RAM and block storage. The pricing per instance is further segmented by the OS image running on the virtual machine. A customer can choose from either a Red Hat, SUSE or Windows image, which strongly limits the possibility but augments integration with other IBM products.

For data transfer IBM charges incoming and outgoing bandwidth in tiers, that are measured in TB. Within each of these tiers the pricing is specified on a per GB basis. The status of data transfer between two instances is unclear.
Support can be bought as a package for an upfront fee. When a support service is used there is an additional charge in addition to the upfront fee. The support service is charged on a per hour rate, that varies depending on the virtual machine's OS.

5.5.4 Microsoft

Microsoft's cloud service is called Windows Azure. There are many various services ranging from computing to content distribution networks. The service that is of interest here is the computing service. There are five different virtual machine sizes ranging from one to eight cores. Windows Azure virtual machines include not only local block storage but also local storage in web and worker roles (specific roles for web development and web applications). Additionally, each instance type is connected to the Internet at different bandwidth speed. The actual bandwidth consumed is priced at the normal bandwidth rate on a per GB basis. The separation of actions into different roles makes evaluation difficult, as it is unclear exactly what operations run in which role.
Windows Azure charges for outgoing, but not incoming bandwidth. The outgoing bandwidth is further segmented into two separate regions. Traffic to North America and Europe is charged at a different rate than Asia Pacific, but data transfer within a subregion and the Windows Azure platform is free of charge.

5.5.5 Rackspace

Like many other cloud providers Rackspace offers two core service types. On a global level they offer licensed (Windows) servers and unlicensed Linux based servers. These two server types are then split up into various categories according to RAM and disk size. With the Rackspace service, one server always equals four virtual CPUs on Linux, there is no CPU expansion offer on a server instance level. In the case of a Windows image there are differences between the different offers in terms of CPU power provided.
As is the case with most cloud providers that are not geared towards enterprise customers, Rackspace charges only outgoing and not incoming bandwidth. The status and pricing of intracloud bandwidth is not clear from viewing the Rackspace offering. But since they note that bandwidth is calculated on a per server instance level, one can imply that this means that intracloud communication between seperate instances is charged at the normal bandwidth rate.
Rackspace's cloud service is unique in that they offer a managed service level at an hourly rate without an initial support service charge. Additionally Rackspace integrates its cloud services with its existing hosting offerings, allowing customers to easily implement a hybrid cloud solution on the Rackspace infrastructure. On the other hand Rackspace includes a $100 account fee per month, regardless of service usage. This account fee is levied as soon as at least one server is active. In return Rackspace implements a basic bonus-malus system, where they credit up to 100 percent of usage charges, if they do not meet the QoS parameters specified in their SLA.

5.5.6 Conclusions

Generally, the analysis of pricing by cloud computing providers shows that the information presented to the customer is both vague and unsuited for comparison to other cloud service providers. Additionally, the flexibility provided by cloud computing results in the complexity of the offering, leading to the necessity of a greater investment of time for

evaluation from the side of the user. Cloud providers do little to alleviate this issue. If fact, by obfuscating and burying pricing details within their sites, cloud providers cause additional confusion and increase ambiguity in an already cloudy market. Additionally the lack of a common metric results in providers quoting their service in divergent ways, which makes offerings difficult to compare. The language used to describe the services tends to be complicated and ambiguous, making it unclear what charges are precisely being billed in a given scenario.

Amazon: Amazon offers a comprehensive service that allows various levels of scaling and customizing that it directed at a wide array of consumers. Enterprises, small businesses and even private developers can configure the service to meet there needs. This amount of flexibility and reconfigurability nevertheless comes at a price. Understanding the pricing model and the different aspects and details of the service are obfuscated or difficult to understand without a considerable investment of time. The positive side of this is that the elements being charged are fairly clear.

Terremark: Terremark is clearly geared towards enterprise customers, with a limited degree of configurability and concentration on standardized services. The general pricing is clear-cut and straightforward, but some charging details are omitted, such as included system storage size as well as VPU (virtual processing unit) power.

IBM: With their limited OS support IBM specifically caters to enterprise customers. The pricing scheme of their cloud service fits in with the pricing models of other products, by using a common naming scheme. The limited customization of the service aids understanding and fits into the standardization schemes of larger corporations. The common language helps to create a common understanding, but limits the usability of the service to the intended target group (enterprise consumers).

Microsoft: Windows Azure sports close integration with the Windows environment, while proving a wide array of services. The structure and variety of services, makes the comprehension of the pricing scheme a daunting task. Additionally, the platform introduces bandwidth speed issues, by differentiating instance by the speed to which they are connected to the Internet. Each service is describe in detail as well as providing the technical specifications included in the price, but the use of language renders many of these details and statements to be futile due to ambiguity.

Rackspace: The Rackspace cloud offering is mostly clear, compact and comprehensible, allowing for a reasonable amount of customization. The inclusion of support services with low initial fees, shows that Rackspace is targeting small businesses and private developers, in addition to the fact that instances are limited in their processing power. Although there is an account fee for all usage, the including of a bonus-malus system establishes a certain notion of pricing fairness.

5.6 Cloud Computing Use Case

The four key elements in all cloud computing price schemes are CPU utilization, network utilization (bandwidth), RAM and storage. In more sophisticated scenarios, load balancers and caching are also issues, but are not part of the basic use case used to evaluate different cloud computing providers. With the inclusions of such technical and case specific parameters the evaluation becomes extremely complex, if not impossible to evaluate without actually simulating the case on the actual infrastructure. To be able to evaluate the cost incurred using each platform, a generic application needs to be specified for which metrics can be derived. These raw metrics are then the foundation on when the pricing models can each be evaluated. The application type that has been chosen is the use case of a web application, as this is the prominent and generalized use case in which cloud computing is utilized. It also offers itself as a good model for defining specific metrics

of utilization that can be easily scaled to show differences in resource usage, as well as specifying different technical procedures and show how they influence performance and cost.
Finding actual numbers of metrics that show the utilization of different resources in varying scenarios has proved to be a challenging task. Most use cases that have been specified are on a high level, only showing the interactions between actors in the cloud and no actual concrete processes and the resulting resource usage. The numbers in the following use case have been derived from the information Google engineers have gathered from usage statistics of Google App Engine [15]. These are metrics that, according to Google are able to support 5 million page views per month, which is a good benchmark of a standard webpage with dynamic content. Additionally, for bandwidth consumption recent web data from Google was used to calculate the bandwidth needed to serve 5 million page views with an average size of 377 KB (uncompressed, full document size including all resources) [14]. The ratio of in versus outgoing bandwidth has been set a 1:3, as this seems to be a safe assumption, seeing that most webmaster forums state a long term ratio of 1:4 or 1:5 to be sufficient. Where applicable, the option of one CPU running on a 32 bit architecture was chosen.

Table 5.2: Metrics for the Cloud Computing use case

Metric	Quantity
CPU Utilization	720 hours
Network Utilization (bandwidth)	600 GB in 1798 GB out
RAM	2 GB
Storage	30 GB

Such basic and clear cut numbers tempt one to assume that an application of these metrics to the pricing models of the providers is simple. But even in such a simple example the actual calculation of performance received and costs incurred becomes ever more challenging as the use case is expanded and viewed in detail. Evidently this use case contains many of the same inherent problems that are subject of complex cases and actual real world implementations. One fundamental problem is the dependency on the underlying power of the hardware. In many cases the actual power of the CPU is either insufficiently specified or not at all. On a low level the choice of CPU architecture can have a significant effect on the cost/performance ratio. Moreover, RAM latency and general system latency is not transparently specified by providers, which makes (pre-purchase) evaluation exceedingly difficult. Not only are there implications for such direct problems, but the effects of indirect hidden costs also need to be considered.
One scenario that expressively demonstrates how the different metrics relate in terms of subsequent cost is when gzipping of HTML pages is introduced. In this case an HTML page that is assumed to be 377 KB (HTML file with resources attached) can either be transported to the end user in the raw or compressed format. For compression a typical single core CPU takes 32 ms of computation time to gzip the data, resulting in a compressed output that is 30% smaller than the original raw content. The benchmark used to create this raw data was a Pentium 4 2.8 GHz running on Ubuntu 10.04. Now a comparison can be made between the total cost of serving the content in the raw condition versus the total cost of serving the compressed content. In such a scenario only the actually incurred costs are considered. The opportunity costs on the end users side, arising from the increased latency of having to gzip the content are not considered. Of course such opportunity costs can become critical in applications where low latency is paramount. With such a slight alteration to the original use case, the relationship between the different metrics and providers can be shown. It is not unthinkable that changes in

the parameters of the application results in different outcomes in terms of which provider is more cost efficient for the customer of a cloud service.

5.7 Application of Use Case to Providers

The metrics of the use case defined above can be applied to the pricing models of the different cloud service providers to illustrate the pricing model and determine the efficiency of each provider in terms of costs in the specific scenario of the use case. The use case provides a general indication of which region of the pricing schema we are actuating within. It should be noted, that for the application of the use case, static metrics such as storage and RAM must generally be considered first. Most providers have various tiers of performance where CPU utilization is fully elastic, while storage and RAM are only minimally elastic. This circumstance adds an additional level of complexity that needs to be considered in the overall evaluation of the costs versus performance. Additionally it is important to note that for the application, only the prices of unlicensed servers (Unix based) have been considered for the providers that make a distinction between unlicensed and licensed servers.

5.7.1 Amazon

Amazon has a wide array of cloud web services ranging from computing to storage and load balancing. They even have specific instances for specialized computing tasks such as MapReduce. This allows for a great level of customization, but also brings with it complexity, as well as the time consuming task of evaluating which solution is best for the specific case that is to be implemented. Additionally, one must consider the ramifications of scale as well as changing requirements. Amazon has many different tiers of hardware to choose from, but the underlying power is obfuscated. Also the amount of customization on the level of CPU power of one virtual instance is limited. As a whole, Amazon views cloud computing through a different paradigm. They view the resource from a macro perspective, not the micro perspective of one virtual instance.
The following table shows the values when applying the use case to Amazon:

Table 5.3: Cloud Computing use case Amazon

Metric	Unit Price	Total Price
CPU Utilization	$0.085 per hour	$61.20
Network Utilization (bandwidth)	Inbound traffic free $0.0 for first GB $0.12 per GB up to 10 TB	$215.64
RAM	1.7 GB included in VPU price	
Storage	160 GB included in VPU price	
	Total	$276.84

Much like in the beginnings of the automobile, where the definition of a “horse power” was created to aid people in understanding and judging the underlying power of an automobile, Amazon has defined an “EC2 Compute Unit” to aid developers in assessing the power of their VPUs. An “EC2 Compute Unit” signifies the power of a processor regardless of the actual underlying hardware. Amazon defines one EC2 Compute Unit as the capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor. It is important to note that the small standard on-demand instance was used as a benchmark for the Amazon services, as well as the US East datacenter where the pricing is the lowest. The RAM and virtual CPUs available per instance on Amazon depends on the instance type chosen, which would be

1.5 GB and 1 VPU in this case. Even though this value is 500 MB short of the benchmark, the next largest instance offers 7.5 GB of RAM and 4 VPUs, which would significantly alter the results and hamper the comparison with other providers even more.
For the sake of comparability and fairness in terms of benchmarking, the next available option that fits all criterion has also been considered. This instance runs 4 VPUs on a 64 bit architecture, with the same traffic costs:

Table 5.4: Cloud Computing use case Amazon

Metric	Unit Price	Total Price
CPU Utilization	$0.34 per hour	$244.8
RAM	7.5 GB included in VPU price	
Storage	850 GB included in VPU price	
	Total	$460.44

5.7.2 Terremark

Terremark is more the traditional hoster. They still have hosting solutions besides there cloud computing business. Recently they were acquired by Verizon, one of the leading telecommunication companies in the United States [13]. Terremark allows many macro parameters to be specified separately, but does not provide the details of the hardware power behind these abstractions.
The following table shows the values when applying the use case to Terremark:

Table 5.5: Cloud Computing use case Terremark

Metric	Unit Price	Total Price
CPU Utilization	$0.120 per hour	$86.40
Network Utilization (bandwidth)	$0.17 per GB in and out	$407.66
RAM	2 GB included in VPU price	
Storage	$0.25 per month per GB	$7.50
	Total	$501.56

In the case of Terremark there is no internal system storage that is included in the price. All storage is charged on a per GB per month basis, whereby the metering occurs on an hourly basis. Also, it should be noted that the power of 1 VPU is not specified, so a comparison can only occur under the assumption that the underlying technology is on par with a typical consumer grade CPU available today.

5.7.3 IBM

The following table shows the values when applying the use case to IBM:

Table 5.6: Cloud Computing use case IBM

Metric	Unit Price	Total Price
CPU Utilization	$0.095 per hour	$68.4
Network Utilization (bandwidth)	$0.15 per GB in and out for first 10 TB	$359.70
RAM	2 GB included in VPU price	
Storage	60 GB included in VPU price	
	Total	$428.10

Note, for the CPU utilization, the SUSE Linux OS option was chosen, as this is the cheapest Unix based option. IBM rates their VPUs at 1.25 GHz, but does not specify the architecture that this benchmark is based on.

5.7.4 Microsoft

With the Windows Azure product, Microsoft offers a product that is not limited to the cloud computing market. The service offers combinations with content delivery networks and various other services. The compute service has been used exclusively for the application of the use case.
The following table shows the values applying the use case to Microsoft:

Table 5.7: Cloud Computing use case Microsoft

Metric	Unit Price	Total Price
CPU Utilization	$0.12 per hour	$86.40
Network Utilization (bandwidth)	$0.155 per GB outbound	$278.69
RAM	1.75 GB included in VPU price	
Storage	165 GB included in VPU price	
	Total	$365.09

As with Amazon, the Windows instance only provides 1.75 GB of RAM with the VPU (Small instance type), which needs to be considered when comparing the providers. Windows Azure only offers its own Windows Azure Guest OS, which is based on Windows Server 2008 R2. Additionally the outbound data transfer pricing from the Windows Azure platform depends on the destination, thus the average of the two prices was taken as a benchmark.
The next higher offering includes 3.5 GB of RAM and 2 VPUs:

Table 5.8: Cloud Computing use case Microsoft

Metric	Unit Price	Total Price
CPU Utilization	$0.24 per hour	$172.8
Network Utilization (bandwidth)	$0.155 per GB outbound	$278.69
RAM	3.5 GB included in VPU price	
Storage	340 GB included in VPU price	
	Total	$451.49

5.7.5 Rackspace

Rackspace, much like Terremark also still has one mainstay in traditional hosting services. This also gives them the unique opportunity to provide hybrid cloud services to customers. The following table shows the values when applying the use case to Rackspace:

Table 5.9: Cloud Computing use case Rackspace

Metric	Unit Price	Total Price
CPU Utilization	$0.12 per hour	$86.40
Network Utilization (bandwidth)	$0.18 per GB outbound	$323.64
RAM	2 GB included in VPU price	
Storage	80 GB included in VPU price	
	Total	$510.04

Rackspace also includes a $100 flat fee per account, which is not included in the hourly rates. On the pricing page they do not specify how many VPUs are included in each offering. This information is only found buried in one of their FAQs. Additionally, Rackspace does not specify the underlying power of the processing unit of the instances.

5.7.6 Scaling of Price and Resources

When surpassing the confines of the use case and applying the element of scalability, the picture shown by the following graph emerges.

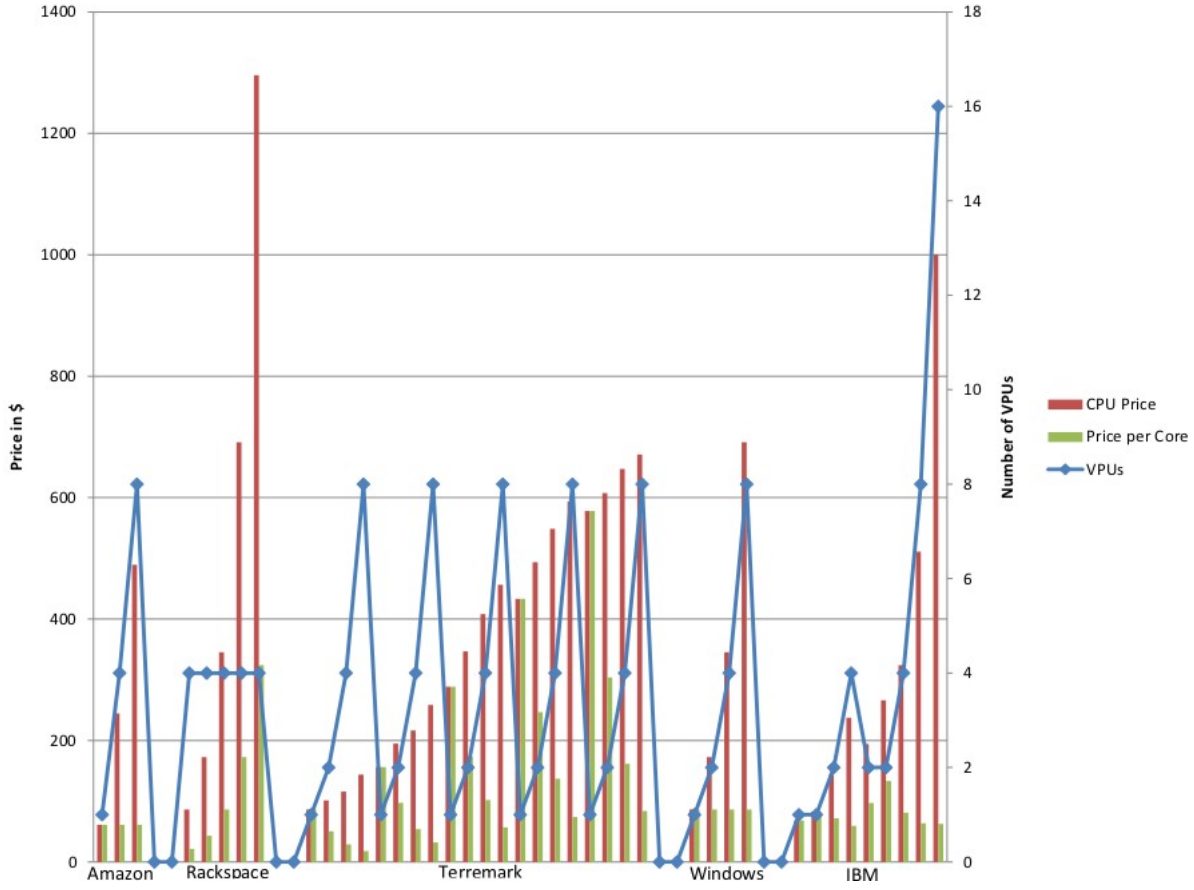


Figure 5.2: Comparison of the scaling costs of different providers for there unlicensed Unix cloud services

The numbers behind this graph were created by applying the raw metrics of CPU hours and bandwidth usage from the use case to the various product offerings of the providers. These two metrics have been chosen to represent the price scaling, due to the fact that they are the metrics that show the most commonality across the pricing schemes of all the providers. Such metrics as RAM and storage are usually included in the price for CPU hours, as this represents the full hourly price of one instance type using the specified hardware.

The ensuing images shows the different pricing structures of the cloud providers. Providers like Amazon and Windows Azure exhibit linear price scaling, as the price per VPU core stays constant as the underlying hardware increases in power. Terremark has a more fine grained offering, where the user can choose the number of VPUs and the amount of RAM with a relatively high degree of freedom, which results in the price per core dropping as the customer adds more VPUs and keeps RAM constant. Consequently, when the user switches to the next higher option in terms of RAM power, the relative price jumps up before it drops again. It must also be noted here that Terremark charges system storage separately, so this parameter must be considered on its own, whereas for the other

providers it is included in the offering. The IBM offering shows a twofold pricing trend. One the one hand the price per core rises for both the 32 and 64 bit offing and then drops as the hardware power is increased. This indicates that IBM is especially attractive for smaller scale and very large scale, but in between the relative costs of power are high. The most interesting price development is exhibited by Rackspace, which offers four VPUs for every level of their product (on Unix systems). This means that the relative price per core rises as the next higher offering is chosen. This would indicate that smaller deployments are relatively more cost efficient then larger ones.
This shows the emergence of another aspect of a hidden cost. Not all providers scale up their hardware in the same fashion, some scale both VPU, RAM and storage on all levels, others scale RAM but not VPU power on certain levels and yet others scale only RAM and VPU power and eliminate storage from the equation by making it separate cost. This makes the objective and numerical comparison difficult, unless one creates a measure that accounts for the different configurations of power vis-a-vis price. One option is to specify a ratio of power to price.

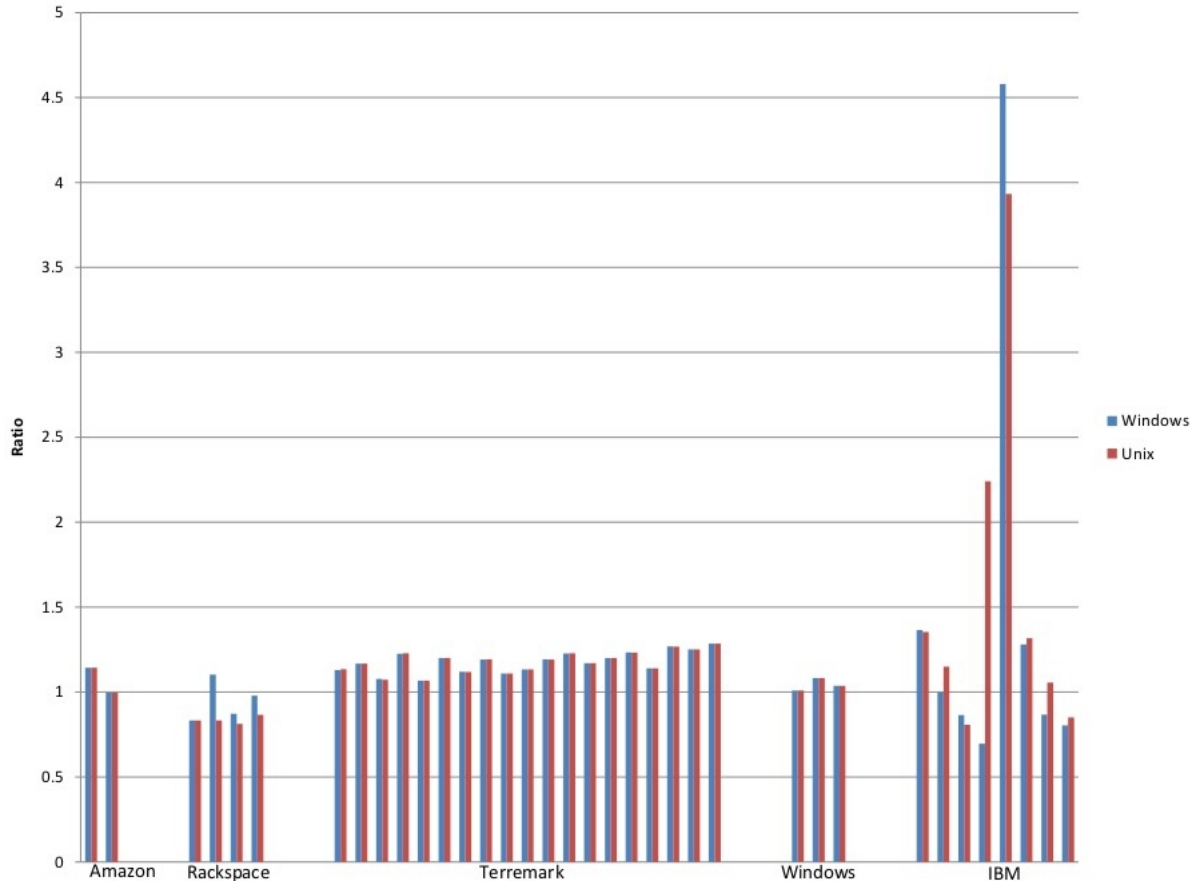


Figure 5.3: Power to price ratio of the different cloud providers and their offered services

This shows a similar picture to the one above, but additionally clarifies the intuitions that were implicitly expressed before. One can see that Amazon and Windows Azure display a ratio of about 1:1 in terms of power to price. In the case of IBM the assumption that small deployments are relatively cheaper that mid sized deployments holds, but for very large deployments (64 bit platinum offering) the relative cost rises. This effect could be the result of the costs of scaling. Regarding Terremark, the ratio is constantly over 1:1 which indicates that more power does not result in relatively more costs, meaning there are no costs for scale. Contrarily, Rackspace has a ratio that is constantly under 1:1, showing that scale has a price that is rolled over to the cloud user. From the figure above, one can also see that, while Windows deployments are more expensive in absolute terms compared to Unix it scales in a similar fashion. In the case of Rackspace as well as

IBM there is no difference in scaling of Unix and Windows offerings within themselves, but when comparing the two operating systems, one can make out significant differences which points towards hidden costs.

5.8 Hidden Costs of Cloud Computing

To illustrate where hidden costs can lie in terms of the technical limitations of metrics and actual method of implementation, the scenario of gzipping data before transport offers itself as a simple, yet expressive example. The costs of gzipping are twofold. On the one hand there are the real costs of the usage of CPU time and on the other the opportunity costs of the added time needed to process a request. Additionally, it shows the relationship between two different metrics (CPU time and bandwidth usage). Thus it indirectly expresses hidden costs that arise in terms of technical implementation, as well as from economic considerations of responsiveness of the system as a whole. Additionally different software running on the same systems, but scaling in contrasting fashion can constitute a hidden cost as well.
The table below shows the amount that can be saved when applying gzip compression to the use case specified above in the case of a Unix powered instance. Additionally, it shows what the savings would be, even if an additional instance would have to be used to execute the compression. Such a scenario could be the case, if the existing instances are already running at 100% capacity.

Table 5.10: Savings when gzip compression is applied to content

	Raw Bandwidth Price	Compressed Bandwidth Price	Savings
Amazon	$215.64	$150.912	$64.728
Rackspace	$323.64	$226.548	$97.092
Terremark	$407.66	$315.962	$91.698
Windows Azure	$278.69	$195.083	$83.607
IBM	$359.7	$278.79	$80.91

Even when it is factored in that an instance could be running at 100%, which would result in the need of an additional instance to take care of the compression task, the savings still far outweigh the marginal costs of an instance, even if this instance is the highest priced product the provider offers. This case shows that especially when bandwidth costs are high, compression and efficiency can bring tangible savings. It also shows that the unit bandwidth costs across the board are comparatively higher than the unit cost of a CPU hour. This is the case even for instances with high power hardware and high hourly CPU usage costs.
It has also become evident that pure cloud computing solutions are less cost effective than their managed hosting counterpart. This is because an instance is billed even if the resource is not being used. In order to not be billed, the server needs to be in the shutdown state. A hybrid solution, where a cloud service is used to handle burst in requests is the more cost effective solution. This can be shown by comparing a hosted solution to its cloud computing equivalent. In this example Hetzner as a hoster was used as a benchmark. The Hetzner solution that fits the original use case is their Root Server X3 offering priced at 39 Euros per month (including unlimited traffic), which is about $51.

Table 5.11: Cloud providers compared to Hetzner

Provider	Total Price	Price without Bandwidth Charges
Hetzner	$51	$51

Amazon	$276.84	$61.20
Rackspace	$510.04	$86.40
Terremark	$501.56	$86.40
Windows Azure	$365.09	$86.40
IBM	$428.10	$68.4

As can be seen from this table, scalability has its price. The average premium that needs to be paid for such flexibility is a factor of eight. Considered in more detail, the largest factor in this price difference is the cost of bandwidth. Moreover, because currently cloud computing providers do not meter CPU hours in terms of cycles, but the actual wall clock time for which the CPU is used (*i.e.* when the operation system is running), developers have the incentive to use as much of the CPU cycles as possible. This leads to developers optimizing their applications only for certain parameters that reduce costs and not for overall performance. It could very well be that a more performant application in terms of costs results in an application that is "wasting" certain other resources, which is not in the interest of the cloud provider. Furthermore, this could also effect the the other customers using the service by mitigating their utility, resulting in an unfair pricing situation.
In the case of hidden costs resulting from different software platforms within the realm of the same offering, the verdict is multifaceted. 5.3 shows the scaling of power to price for both Unix and Windows offerings. Amazon, Terremark and Windows exhibit identical scaling of both platforms, whereby with Windows Azure this is the case because it exclusively offers a Windows derivative OS. On the other hand Rackspace and IBM do not scale both platforms in the same manner. In the case of Rackspace this attributable to the fact the Rackspace provides 4 CPUs for all Unix deployments, while in the case of Windows there is an element of scaling on the CPU level. As such this can be considered a marginal hidden cost, as it can be traced back to diverging hardware configurations. Contrarily, IBM offers the same underlying hardware regardless of the software platform running on top. In this case the difference between the two configurations of OSs shows signs of consequential hidden costs. These hidden costs are incurred on both systems, depending on the level of scale the application is deployed at. For example, the 64 bit Copper offering on the Unix platform is significantly cheaper in relative terms, as the price of this offering compared to the next cheaper one is much lower in terms of the additional power received. For the next higher offering (64 bit Bronze) the tables are turned and the Windows platform offers superior power for the price paid.

5.9 Summary and Conclusion

The raison d'etre of cloud computing can be boiled down to one of its key features. Cloud computing allows the transfer of the infrastructure building and administration risk from the cloud customer to the provider of the cloud service. The cloud service provider is then able to disperse and spread this risk through infrastructure scale and the scale of the customer base, as well as the scale of customer demand for computing power. This of course brings the rise of a different risk: The risk of service utilization. Hongyi Wang *et al.* show that the cost of different executions of the same application that demands the same performance and results in the same output, can diverge up to 40%. A deviation of 40% indicates high volatility in costs, which is an economic risk for the customer [5]. This is a new type of risk that needs to be closely considered by the customer, as it alters such statistics as price per page view, execution etc. It is a hidden cost that cannot be controlled by the customer, as Hongyi Wang *et al.* have shown that the performance anomalies happen arbitrarily.

Besides these new risks and emerging systemic issues resulting from distributed computing and the cloud computing architecture, the economic models employed by the cloud computing providers introduce additional complexity and difficulties. One can see that the various pricing schemes of the providers results in deployments having a variable level of cost efficiency depending on scale and the provider used. This indicates that one could conduct arbitrage, which legitimizes the existence of cloud brokers such as Rightscale. Moreover, technical parameters can have a significant impact on the resulting cost of a running application. So, not only do cloud providers afford little pricing transparency and hamper comparability, but their pricing schemes allow for cost by exploiting certain aspect of how metering is done. Of course this necessitates detailed information on how billing and metering is done, which cannot be obtained without subscribing to the service and experimenting.

A key aspect that can be harvested from the analysis of the use case, especially when considering certain special scenarios is that bandwidth is a significant cost factor. It stands to reason that this is even more prominent if complex solutions such as load balancing are used. This asserts the fact that when dealing with cloud implementations a developer must not only keep in mind the global performance of an application, but also the performance of its atomic units, especially in view of metrics that are subject to billing.

Bibliography

[1] Amazon, *Amazon EC2 Pricing*, Pricing Page, 2011. http://aws.amazon.com/ec2/pricing/

[2] Cristian Morariu, Martin Waldburger, Burkhard Stiller, *An Integrated Accounting and Charging Architecture for Mobile Grids*, Third International Workshop on Networks for Grid Applications, October 2006.

[3] Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger and Dawn Lea: *NIST Cloud Computing Reference Architecture*, NIST Special Publication, Gaithersburg, September, 2011. http://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/ReferenceArchitectureTaxonomy/NIST_SP_500-292_-_090611.pdf

[4] Gartner, *Hype Cycle for Cloud Computing*, David Mitchell Smith Publication, 2011

[5] Hongyi Wang, Qingfeng Jing, Rishan Chen, Bingsheng He, Zhengping Qian and Lidong Zhou: *Distributed Systems Meet Economics: Pricing in the Cloud*, Conference Proceedings, Berkeley, 2010.

[6] IBM, *IBM SmartCloud Enterprise*, Pricing Page, 2011. http://www-935.ibm.com/services/us/en/cloud-enterprise/tab-pricing-licensing-options.html

[7] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia: *A View of Cloud Computing*, Article, Communications of the ACM, Vol. 53, No. 4, April, 2010

[8] Microsoft, *Windows Azure Pricing Overview*, Pricing Page, 2011. http://www.windowsazure.com/en-us/pricing/details/

[9] OpenCrowd: *Cloud Taxonomy*, New York, 2010. http://cloudtaxonomy.opencrowd.com/

[10] Peter Mell, Timothy Granc: *The NIST Definition of Cloud Computing*, NIST Special Publication, Gaithersburg, September, 2011. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[11] Rackspace, *Rackspace Cloud Servers*, Pricing Page, 2011. http://www.rackspace.com/cloud/cloud_hosting_products/servers/pricing/

[12] Safiullah Faizullah and Ivan Marsic: *Pricing and Charging for QoS*, Conference Proceedings, IEEE, 2005

[13] Spencer E. Ante and Cari Tuna, *Verizon Buys Terremark*, News article, January, 27, 2011. http://online.wsj.com/article/SB10001424052748703399204576108641018258046.html

[14] Sreeram Ramachandran: *Web metrics: Size and number of resources*, Google, Palo Alto, May 2010. http://code.google.com/speed/articles/web-metrics.html

[15] Stringbuffer.com, *The per page-view cost of hosting a reasonably efficient GAE/J application*, Blog entry, Sunday, May 3, 2009. http://blog.stringbuffer.com/2009/05/per-page-view-cost-of-hosting-resonably.html

[16] Terremark, *VCloud Express*, Pricing Page, 2011. http://vcloudexpress.terremark.com/pricing.aspx

Chapter 6

Business Models of Community Networks

Alice Hong Mei Mei

Since the written report for this seminar talk did not fully meet all formal requirements, the report was not included.

Chapter 7

Cloud Computing Services (CCS): A Threat for Internet Neutrality?

Beat Kuster

This paper introduces the reader to the two concepts of Enterprise Cloud Computing Services and Internet Neutrality. These two topics are related to each other as critics of Internet Neutrality could rely on different characteristics of Cloud Computing Services to argue against Internet Neutrality. Therefore an overview of main criticism on Internet Neutrality is given. The resulting points are reviewed under the aspect of characteristic behavior of Cloud Computing Services. Here the paper focus on the bandwidth consumption of aforementioned services, because it is shown that bandwidth consumption is one of the most criticized point regarding Internet Neutrality. It is found that the bandwidth consumption of Cloud Computing Services can on it's own not be used to argue against Internet Neutrality. It is too uncertain whether an overall surge in internet bandwidth consumption can be tracked back to only Cloud Computing Services.

Contents

7.1 Definitions

This paper will use the following terms defined as in this section:
Cloud Computing: A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service level agreements established through negotiation between the service provider and consumer [9].
Internet Neutrality: Internet neutrality represents the idea that Internet users are entitled to service that does not discriminate on the basis of source, destination, or ownership of Internet traffic [3].
Best Effort Networking: Best Effort Networking means that there is no guarantee for the quality of delivery for a data package.
ISP: An Internet Service Provider makes a selected number of networks available to his customers based on a contract. The contract will at least cover remuneration and service level description.
Tier 1 ISP: An ISP is described as a Tier 1 network provider if they have access to the entire Internet routing table through their peering relationships [8].
Service Provider: In this paper, every player in the market of internet economics which is not a Cloud Service Provider is generally addressed as a Service Provider.
CSP: A Cloud Service Provider makes selected services (CCS) available to his customers which they can access over the internet.
CCS: Cloud Computing Services are provided over the internet whereas the range of this services can vary between complete Information Technology (IT) infrastructure and selected software.
IaaS: Infrastructure as a Service provides the customer with IT services which are on the infrastructure level. Common examples of infrastructure services include computation or storage.
PaaS: Platform as a Service provides the customer with a hosted platform whereas the underlying infrastructure is hidden and the platform is maintained by the CSP. Common examples of platforms include collaboration and web development.
SaaS: Software as a Service or hosted software provides the customer with a ready-to-use software package. The software package is licensed, hosted and maintained by the CSP. Common examples of hosted software include productivity or Costumer Relationship Management (CRM) applications.

7.2 Introduction

If it is heard that CCS could pose a threat to internet neutrality one may ask how a service can threat an idea? Of course CCS itself do not have the power to threat the idea of internet neutrality, only ISPs can restrict or totally refuse internet neutrality to customers. This is because ISPs do own the network infrastructure on which it becomes possible to provide an internet service. With this power given by their infrastructure they are free to provide an internet service which may or may not comply to the internet neutrality definition.
Historically, the internet has been a place where every data packet was treated the same from the TCP protocol. Or at least this is the common perception, there are also other views which doubt that there ever has been a leveled playing field at all [1]. But nonetheless, treating every packet the same is seen as one of the most important factors for the innovation thriving internet as we know it today. During its rise through the last two decades countless innovative ways of making business have emerged and some of them are still flourishing today. Well known example include trading platforms like eBay Inc.

or content provider like Amazon Inc.. The division of work seemed clear, on both side of the internet resides a participant of one of the aforementioned business cases which are both consumers of an internet service provided by an ISP. Through the rise of huge companys who rely exclusively on doing virtual business it became clear that there is a great opportunity to create profit of the internet. There weren't only two customers of internet services anymore, nowadays there is a split into service providing internet customers and service consuming internet customers. These services can consist of anything from news, books, virtual marketplaces to video. This emerge of commercial service providers which capitalize directly on their services differs greatly from the old scheme of the internet where services were provided from both sides equally and the internet was only the medium to exchange it. That change has also been noticed by the ISPs and they are jealously looking at the ever increasing revenue generated from service providers.
Given by the ISPs power through their infrastructure and their obvious discomfort being only the "dumb pipe" to flourishing business models, why don't they charge service providers for accessing their customer base? They could do this in various ways, which is described in a separate chapter. The answer lies in existing or looming regulation from the state. States have clearly recognized the positive impact of the internet on their respective economies. Not only are service providers like eBay Inc. good taxpayers but the availability of their services can also enhance the competitiveness of the companys within the national economy [2]. A good example is the finance industry where access to certain online trading platforms is crucial to stay competitive. To ensure access to the internet for all interested parties, there are laws which regulate the way ISPs can do business. The degree of regulation is different in every country, but the awareness for this topic is definitely raised. Therefore the ISPs cannot act too aggressively in discrimination of internet traffic eg. charge specific traffic without being at risk provoking stricter regulation.
This paper aims to assess the threat Cloud Computing Services pose to Internet Neutrality.The arguments which are brought forward by the ISPs against Internet Neutrality will be summarized in the first section. Second, an overview of business models ISPs could engage which would conflict or end Internet Neutrality is presented and how they are related to CCS. In the third part it is to be determined if certain aspects of CCS do support the arguments of the ISPs and therefore would threatening Internet Neutrality itself.

7.3 Arguments against Internet Neutrality from ISP-side

Internet Neutrality in it's working form, imposed by law or through competition, constrains ISPs in their decision making and makes them opposing Internet Neutrality. This is due to the fact that they cannot engage in the business models described in the last chapter. But also some economists and policy maker are opposing Internet Neutrality, researchers summarize them as deregulationists [3]. Deregulationists positions should be carefully examined as it is not always clear what incentive is behind their point of views. This means that ISPs spend enormous sums to influence the public, up to $100 million according to some sources [3].
1. Internet Neutrality causes rising bandwidth consumption
From a Content Providers point of view it makes sense to engage as much as possible customers as much as possible with his services [7]. Because if his business model is solid, he can generate revenue with every interaction. This can be as example by advertisement or direct fees. These interactions over the internet by the Content Providers customers demand bandwidth. As of common sense this should be beneficiary for ISPs but they do not agree. Because they object the revenue made of the internet is not divided in

a fair manner (see argument 2), ISPs are not eager to invest in new infrastructure to support higher demand. For even further cost optimizations ISPs would like to use traffic management and justify this with rising bandwidth demand [5].
2. Revenue split is not fair
As seen above, ISPs are against more network traffic because they feel that the extra revenue is not splitted in a fair way. It is argued that despite the fact the internet provides a wealth of possible business models, only certain players in the market can exploit them. The others are forced to act as providers of commodity goods; in the case of the ISPs this would mean acting as a "dumb pipe" to deliver data packages without a possibility to engage in higher margin business. With this perception on the side of the ISPs, they do not believe in being able to capitalize on a additional dollar revenue from internet economics as other market players can [4].
3. Internet Neutrality hinders technical and economical innovation
According to deregulationists not enforcing Internet Neutrality would allow much more differentiated business models. Through the use of Quality of Service (QoS) measurements there are many new offerings possible for ISPs not only in case of raw internet access but also in content which customers could subscribe to. In addition ISPs could also integrate their product lines more vertically because they do not have to fear ex post regulation [3]. All this arguments combined could lead to more investments from ISP-side.

7.4 ISP business models conflicting with Internet Neutrality

This section explains the three most important business models which ISPs can apply to generate more revenue but are not compatible with Internet Neutrality. This is achieved not only by charging the service customer to access the internet but also the service provider to get access to the ISPs customer base. This access can be further differentiated through quality levels. Furthermore ISPs can take measurements to lower the perceived quality of competing products in order to boost it's own product line.
Differentiation through service quality
Service provider pay an ISP to get access to the internet and therefore the ability to reach it's customers. In most cases however, the service provider and his customer do not rely on the same ISP to get internet access. Thus the service provider may be charged for the traffic by his own ISP but the customer-side ISP can only collect the monthly access-fee from his customers. In combination with competition-enforced measurements like bandwidth oversubscribing, the customer-side ISP can get into trouble if there are multiple popular service providers using his bandwidth. Additionally, if the service provider is located in an internet segment which is for the customer-side ISP only reachable through a tier-1 ISP-network then traffic charges may apply. To address this issue customer-side ISPs could restrict access to it's customer base to different quality levels. This means that if an service provider is willing to pay a fee his traffic will be prioritized against other service providers [7]. Or the other way around is not slowed down whereas all other service providers are throttled.
Restricting access to own customers
A more rigid approach for an ISP is it to enter direct negotiations with Content Providers about paying for access to his customers and use his customers as kind of a pledge. This is possible as the ISP is the only market player through which his customers can be reached at the time. If a Content Provider would not accept to pay for the access, he must convince his potential customers to change their ISP. But it is more likely that customers switch to another Content Provider instead and the ISP do know that. In economic theory this is know as termination fee in a two-sided market which is not always desirable. From an

Internet Neutrality point of view this practice is violating the zero-price rule because the originator of a package would be charged [7].
Favour own services
Due to the fact that most infrastructure-owning ISPs originate from incumbents, they have other business units like voice telephony or cable television. In order to grab a bigger market share to increase revenue it is obvious one could degrade the available bandwidth to competing companies until their service is unusable. A good example is Voice-over-IP (VoIP) where an ISP also has a voice telephony business. Throttling the data of the competitor could lead to serious quality problems with his VoIP service and resulting in customers switching to the voice service of the ISP [7]. In the United States, Comcast Corporation has been forced to settle a lawsuit because of alleged bandwidth throttling. Comcast has been suspected and later also acknowledged to throttle some traffic types, especially P2P traffic. At the same time, they stated in the according Terms of Service that their own VoIP service is not affected by this traffic management.

7.5 Bandwidth consumption of CCS

After it has been showed that with the current regulations in place much of the ISPs argumentation against Internet Neutrality is based on the increasing demand for bandwidth, this section covers bandwidth consumption of CCS. CCS can be bandwidth hungry and therefore might pose a capacity problem to ISPs. ISPs thus would be forced to invest more in their network infrastructure without participating equally on the extra network traffic. The following section shows what characteristics the different CCS have regarding computation and therefore also bandwidth usage. In a use case where most of the computation is done at one single network node, the bandwidth consumption is smaller compared to a distributed computing model. There, the computational steps are divided between multiple nodes and the synchronization of intermediate data between these nodes consumes bandwidth. Therefore it is critical to understand the underlying computation model of a CCS workload to assess it's impact.

7.5.1 IaaS

Which are typical workloads on IaaS and how bandwidth intense are they?
IT infrastructure services are the most broad approach to cloud computing. The server hardware is provided and maintained by the CSP. The infrastructure at the client site is reduced to hardware devices which require direct input from employees eg. PC's, Printers. Servers providing distributed services as email or collaboration are virtualized and run on the infrastructure of the CSP. Therefore every interaction of a client machine with data hold available uses bandwidth between the enterprise to the CSP. If the virtualized server is acting as a web server for the customer of the enterprise, the bandwidth consumption is more comparable to the non-CCS case but depends on the location of the CCS data center. Nevertheless IaaS has to be viewed as much more bandwidth intense than their non-CCS counterparts.

7.5.1.1 IaaS Provider

Amazon AWS
Amazon Web Services (AWS) consist of a wide range of remote accessible web services. The two best known and also most heavily used of these services are called Elastic Cloud Compute (EC2) and Simple Storage Service (S3). With these two services it is possible to entirely replace a common IT infrastructure within a small- or medium-sized enterprise.

The basic functionality an IT infrastructure has to provide is the ability to compute and store the computed data, this is achieved by EC2 and S3.
IBM
IBM's cloud offerings are called SmartCloud Enterprise and often viewed as one the most mature but also most complex CCS in the current market. Their IaaS offerings focus more on providing integrated solutions including operating systems and IBM software solutions. The difference to PaaS is determined by the "one-time"-help character of their offerings: The infrastructure is delivered up and running but the customer has to take care of the maintenance. Therefore the CCS are targeted at a more sophisticated clientele which not only need to substitute common server hardware but also need to solve problems addressed by IBMs own software offerings. These workloads include collaboration services, web application delivery and process engineering platforms. All of the aforementioned workloads are quite data intense but the most bandwidth intense use case is data analytics.

7.5.2 PaaS

Which are typical workloads on PaaS and how bandwidth intense are they?
PaaS offerings consist not only of a operating system but also of higher level applications like databases or other middleware. In contrast to IaaS these applications are maintained by the PaaS provider. It is mostly used to develop and deliver web applications for customers or internal need of an enterprise. Regarding bandwidth consumption of PaaS there is to differentiate between two use cases. In one use case the platform replaces a system which is used within the customers enterprise itself like a testing environment or a database. In the other case it replaces a customer-facing system such as a web shop. This is important because in the first case there is more network traffic generated than before and in the second case this depends on both the location of the PaaS and the customer.

7.5.2.1 PaaS Provider

Microsoft Windows Azure
Windows Azure is the PaaS offering from Microsoft Corp. Within this platform it is possible to develop, host and manage applications which are then accessible through the internet. Additionally virtual machines can be uploaded and operated in the cloud. Data can be stored in binary form or at an SQL database. From a bandwidth perspective the Content Delivery Network (CDN) offering is relevant. Windows Azure replicates the data which is used by the hosted application across different data centers, therefore possibly reducing the length of the data route. Another offering with potentially high impact on bandwidth demand is the High Performance Computing (HPC) through connecting a HPC version of Windows Server to the cloud. In this use case the bandwidth demand depends heavily on the data / computation ratio. The data / computation ratio is calculated from the hours of computing a given amount of data can invoke.
Google AppEngine
AppEngine is a product of Google Inc. and aims to make the special characteristics of Google's infrastructure available to the wider public. This contains a development environment, which when used correctly, should guarantee a customers web application the same high availability as the internal used infrastructure at Google has. Developing and hosting an application on AppEngine requires the usage of defined programming languages and Application Programming Interfaces (API). Below a certain threshold the computation and bandwidth is free, but this threshold is very low. Every hosted application which consumes bandwidth above this threshold is billed for it and thus developers are provided with an incentive to save bandwidth.

7.5.3 SaaS

Which are typical workloads on SaaS and how bandwidth intense are they?
In the SaaS model the computation of an end user program is done directly in the cloud, only interactions with the program happen at the customer device. SaaS is accessed in most cases through a web browser. There are other possibilities to provide a Graphical User Interface (GUI) for the customer to access a SaaS offering but they are not widely used [6]. The bandwidth intensity of SaaS depends heavily on the data type used. This means that providing any audio or video capabilities can multiply the bandwidth demand of a SaaS offering.

7.5.3.1 SaaS Provider

Salesforce Sales Cloud
Salesforce.com Inc. offers its CRM suite as SaaS. They claim to be the number one provider for cloud based CRM software. Their Sales Cloud service is able to move the sales process completely to the cloud. This means that every step from generating interest from the customer until the closing of a deal can be supported by their SaaS offering. But that also implies that if this CRM solution is used consequent every work step generates traffic and therefore consumes bandwidth. The limitation here is the amount of data an employee of the sales department can create by himself. Everything the employee saves in the cloud is created by him during the sales process through input on a web interface, there are no huge amounts of artificially created or historical data involved. Also a voice or video conference functionality is missing, further reducing the potential bandwidth consumption.
Google Apps
Google Apps are different SaaS offerings provided by Google Inc and bundled for enterprise customers. They offer a wide range of services from collaboration (email, instant messaging, calendar, voice and video chat) to productivity tools (document creation and handling). The self-declared goal of Google Inc. is to move most workloads to CSPs and Google Apps is the pillar of that strategy which is most visible by the employees of an enterprise. Client-side software like Microsoft Office and Outlook can be replaced, also basic file sharing functionality provided by Sharepoint is covered by Google Apps. The movement of this type of workloads does not mean an automatic surge in bandwidth demand for a modern enterprise. There are use cases where bandwidth intense actions, especially mailing with big attachments, can be reduced by using SaaS. This is explained by filtering and hosting the whole email data at the CSP and many cases only viewing of the data through a web interface is conducted. In the case of Google Apps the bandwidth impact of such actions is further reduced through optimizations.

7.6 Conclusion

After evaluating the technical and economical environment of the internet today, summarizing arguments against Internet Neutrality, listing possible business models for ISPs and discussing the bandwidth related aspects of CCS it is now to conclude if CCS pose a threat to Internet Neutrality.

7.6.1 The ISP and deregulationists side

There are made some general objections to Internet Neutrality by a diverse group of stakeholders. This group consists mostly of companies (ISPs) but also of politicians and economic scientists. The motivation of ISPs and politicians is obviously maximizing their

own or their election campaign backers profit. As for the economic research on this topic the impact on general welfare is highly uncertain [7]. Altogether it does not present a very convincing picture.
In the current regulation environment ISPs do not see how they can equally participate at revenue growth derived from rising bandwidth demand. This is an indicator why ISPs often highlight a looming bandwidth shortage and their inability to finance any large scale infrastructure upgrade. There are different business models known which do address these concerns. Nevertheless they are not compatible with Internet Neutrality as defined today. To influence policy makers and the wider public ISPs invest a large amount of energy to show that it is not possible to deal with the fast rising bandwidth demand with their current compensation.

7.6.2 The bandwidth problem

In the absence of better arguments, ISPs do stretch the enormous growth in network traffic and that there has to be more compensation. But it is not all that clear if a shortage of bandwidth really exists. There is evidence that it exists enough spare network capacity, at least for the foreseeable future [5].
The focus of this paper was to determine if CCS in a typical enterprise use case do pose a threat to Internet Neutrality. Although is very difficult to estimate the extra bandwidth demand CCS will generate, it can be said that it is very unlikely to cause a bandwidth shortage only by CCS. Therefore it is unlikely that ISPs will try to weaken Internet Neutrality arguing with CCS.

Bibliography

[1] J. Crowcroft: *Net neutrality: the technical side of the debate: a white paper*, Newsletter, ACM SIGCOMM Computer Communication Review Volume 37 Issue 1, Pages 49 - 56, January, 2007. http://www.cs.bham.ac.uk/~sxz845/Cloud-Adoption.pdf.

[2] D. Harhoff et al.: *Research, Innovation and Technological Performance in Germany*, Report, Commission of Experts for Research and Innovation, 2011. http://www.e-fi.de/fileadmin/Gutachten/EFI_2011_en_final.pdf.

[3] S. Jordan: *Implications of internet architecture on net neutrality*, Article, ACM Trans. Internet Technol., Pages 28, May, 2009. http://doi.acm.org/10.1145/1516539.1516540.

[4] A. Maier: *Interview with CEO of British Telecom*, Article, Financial Times Germany, 16.07.2008. Visited at 17.11.2011 http://www.ftd.de/it-medien/it-telekommunikation/:gebuehrenpflicht-auf-der-datenautobahn-briten-bieten-google-die-stirn/386651.html.

[5] M. Pope and J.P. Shim: *The Looming Bandwidth Crunch - Legitimate Crisis, or Cyberspace Chicken Little?*, Article, Communications of the Association for Information Systems Volume 27 Issue 1, Article 43, 2010. http://aisel.aisnet.org/cais/vol27/iss1/43.

[6] Progress Software Corp.: *SaaS User Interface*, Whitepaper, Visited at 21.11.2011. http://www.progress.com/docs/whitepapers/public/SaaS/SaaS-User-Interface.pdf.

[7] F. Schuett: *Network Neutrality: A Survey of the Economic Literature*, Article, Review of Network Economics Volume 9 Issue 2, Article 1, 2010. http://www.bepress.com/rne/vol9/iss2/1.

[8] M. Winther: *Tier 1 ISPs: What They Are and Why They Are Important*, IDC Whitepaper. Visited at 24.11.2011. http://www.us.ntt.com/fileadmin/NTT-America/media/pdf/about-us/resources/IDC_Tier1_ISPs.pdf.

[9] S. Zardari and R. Bahsoon: *Cloud Adoption: A Goal-Oriented Requirements Engineering Approach*, In proceedings of the IEEE/ACM International Workshop on Software Engineering for Cloud Computing, the ACM/IEEE 33rd International Conference on Software Engineering (ICSE), Pages 7, 2011. http://www.cs.bham.ac.uk/~sxz845/Cloud-Adoption.pdf.

Chapter 8

Convenient Trust

Karthick Sundararajan

Nowadays, most of the organizations have information systems to fulfill the demands of their business requirements. Once this system is connected to the Internet, it is vulnerable to network attacks. In order to avoid such attacks, the organization needs to enhance the security infrastructure of the information system. The standards of public-key infrastructures like X.509/PKIX are implemented by industries to enhance security for server authentication. However, for user to user authentication, there are still no well implemented standards even though some implementations like web of trust exist. This seminar report explores the several open problems in public-key infrastructures and in establishing the user-friendly security infrastructure. First, we begin with a brief introduction to security attacks and argue about the importance of security infrastructures. Second, we also discuss about public-key cryptography, public-key infrastructures (PKI), trust models, and certificate revocations. Third, we will see briefly about various implementations of PKI, including X.509/PKIX, PGP encryption, and KeyChains. Finally, we conclude with problems of centralized and decentralized PKIs in technical as well as administrative aspects and proposed solutions for these problems.

Contents

8.1 Introduction

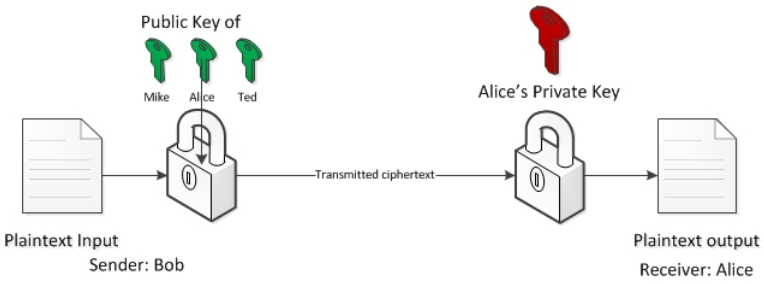
In the last two decades, as the Internet grew the security threats have also grown. Particularly in the last two years, there were massive hack attempts on most popular companies including the Sony PlayStations network [17] and MasterCard credit card solution provider [12]. In addition to these attacks, very often the banking and financial industries were being attacked. When a system is exposed to the Internet, we need to consider critical network security related issues. Since business information which maybe trade secrets, business plans, financial details, etc., are valuable, securing the communication is very important.
In this report we will discuss about "Convenient Trust", where convenient means comfortably suited for an entity and trust means confident or faith in it. So "Convenient Trust" is used in the sense of providing comfortable trust between two ends i.e. convenient trust between two users on a computer network.
We begin with a brief introduction about public-key cryptography and then move on to PKI, function of PKI, centralized and decentralized approaches, trust models, and certificate revocation of PKI. Afterwards, we will discuss about various implementations including X.509/PKIX, PGP, and KeyChains and along with the various problems encountered in those implementations by analysing into both administrative and technical aspects. Finally, the conclusion part illustrates pros and cons of various approaches with the proposed solutions.

8.2 Public-Key Cryptography

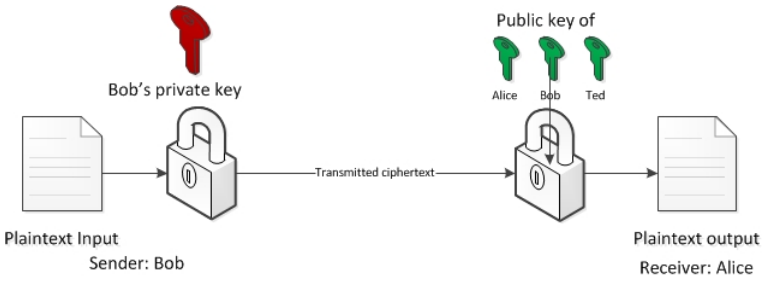
Cryptography is a Greek word which means hidden or secret. It is a set of methods to establish secure communication between two entities by converting the plain message into an unreadable form called cypher text. Cryptographic techniques are broadly classified into two categories namely symmetric key cryptography and asymmetric key cryptography. In symmetric key cryptography, sender and receiver use the same key for encryption and decryption. However, in asymmetric key cryptography, two different keys (often called key pair) are used namely public and private key. If one key is used for encryption, only the other key can be used for decryption and vice versa. The public key from the key pair is made publicly available to all users and the private key is owned only by the key-owner. In this approach, all the participants can access public keys through key distribution centre. Asymmetric key cryptography scheme is also called as public-key cryptography, and it can be classified into two different schemes such as encryption with public key and encryption with private key as depicted in Figure 8.1. In the following subsections, we will look at the public-key encryption, digital signature, applications and algorithms for implementing public-key cryptography.

8.2.1 Public-Key Encryption

The public-key encryption scheme is secure as long as the user protects his/her private key. In addition, the public-key algorithms are based on a mathematical function; it works in such a way that the user can easily generate the key pair, but it is very difficult for someone to find out the private key with the knowledge of the public key.
If Bob wants to send a secret message to Alice, he encrypts the message with Alice's public key and transmits it to Alice. At the receiver's end, Alice decrypts the message with her private key, and no other recipient can decrypt this message since only Alice knows the private key.



(a) Encryption with public key



(b) Encryption with private key

Figure 8.1: Public-Key Cryptography

8.2.2 Digital Signature

The digital signature is used to achieve message integrity. In practice, the entire message will not be encrypted with private key. Instead, the message of any size is passed into cryptographic hash function, which produces a fixed-length output sequence called hash value. Afterwards, this hash value is encrypted with sender's private key to produce a digital signature. The main benefit of cryptographic hash function is that hash value will change if anyone modifies the plain text. Finally, the digital signature and the sender's certificate are attached to the data and transmitted to the receiver. See Figure 8.2 illustrates the signing and verification process of the digital signature.
At the receiver's end, the plain text is again passed into the hash algorithm which generates the fixed-length hash. Meanwhile, the receiver decrypts the signature with the sender's public key to compare with generated hash. If both hashes are equal, the signature is valid and message has not been altered by anyone. Thus, message integrity is achieved by using digital signature mechanism.
The main problem in public-key cryptography is the user cannot verify the public key that belongs to a particular person i.e. mapping of public key with the key-owner. This problem may be result in man in middle attack; it is a kind of eavesdropping where the attacker creates independent connection with the targets and makes them to believe that they are directly talking with each other. However, in reality, the attacker interprets the communication between them, and then he/she can control the entire conversation.

8.2.3 Application and Algorithms

The public-key cryptography scheme is mainly used to achieve three main goals:

- Sending the private message,
- Creating the digital signature and
- Exchanging the public key.

Table 8.1 depicts the comparison of algorithms with application.

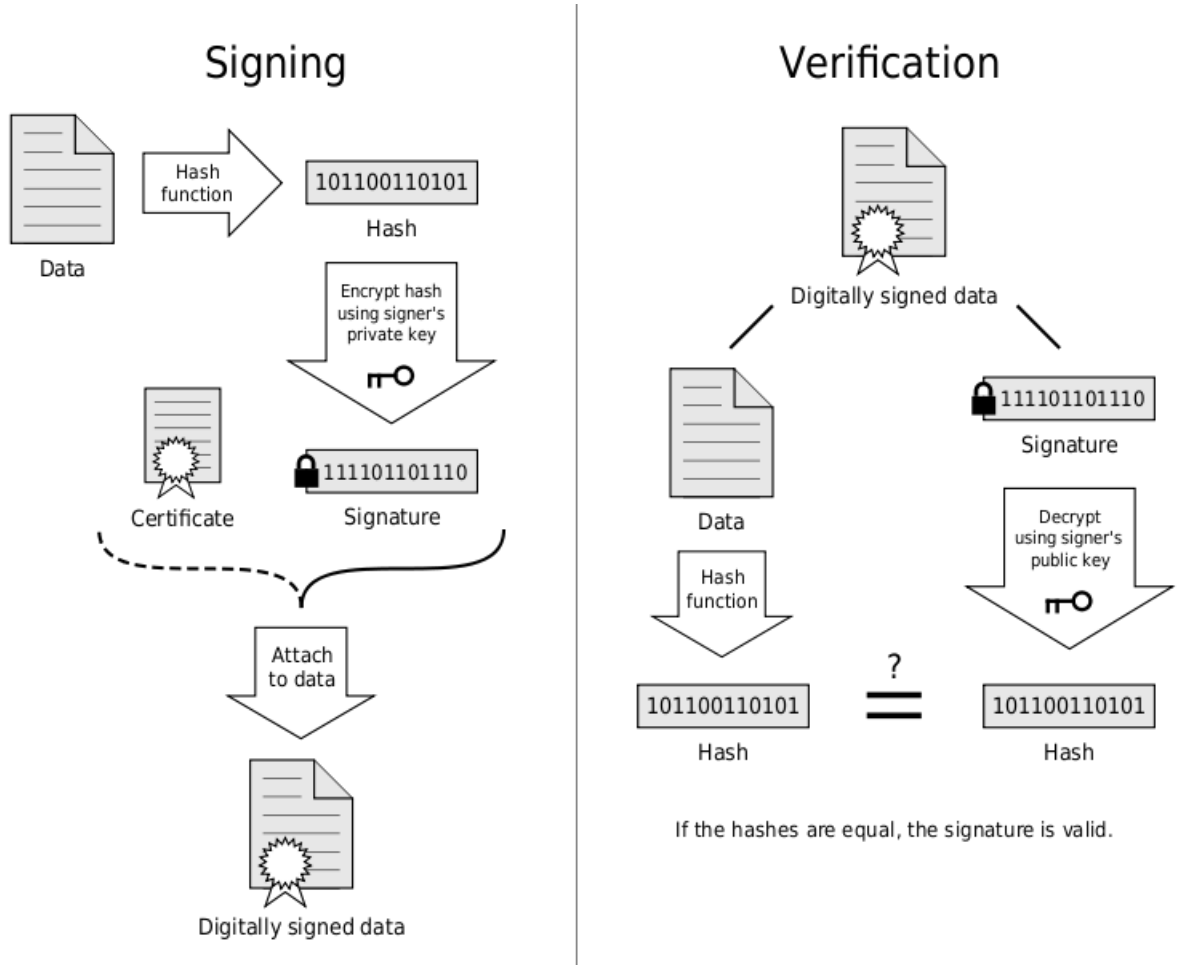


Figure 8.2: Digital Signature [18]

8.3 Public-Key Infrastructure

According to RFC 2822, public-key infrastructure is defined as a set of people, polices, procedures, hardware and software, that is needed to create, manage, store, distribute and revoke digital certificates based on asymmetric cryptography. The objective of this PKI is to protect and distribute the information required in widely spread networks (commonly the Internet), in which end users and resources are residing at different locations. In particular, it solves the key-owner mapping problem with the help of the public key certificate in which it contains the public key, user ID of the key-owner, and the signature of a trusted key signing party.

8.3.1 What is a PKI?

Nowadays, in a formal business transaction, the customers and sellers are depending on Credit Cards (e.g. VISA/Master Card) in order to complete their financial transaction. During this transaction, the seller needs to authenticate the customer with a signature or with additional identification such as national ID cards. In addition, the seller believes that information on credit card is valid and the payment will be received. Meanwhile, customer can refuse to pay if the seller failed to provide product or service. In this case, the credit card provider is a third party trusted by both customer and seller. The PKI is based on trusted third party often represented as TTP.
A PKI is the combination of encryption technologies, software and services, which enable the organization to protect communication and business transactions on their network. It incorporates the concepts of public-key cryptography, digital certificates, and certificate

Table 8.1: Applications and Algorithms for Public-Key Cryptography [15]

Algorithm	Encryption/Decryption	Digital Signature	Key Exchange
RSA	Yes	Yes	Yes
Diffe-Hellman	No	No	Yes
DSS	No	Yes	No
Elliptic curve	Yes	Yes	Yes
ElGamal	Yes	Yes	Yes

authority (CA) into a complete network. In the next section, we will discuss functions of PKI, centralized and decentralized approaches, trust models, and certificate revocation.

8.3.2 Function of PKI

In public-key encryption, the sender encrypts the secret message with receiver's public key, and receiver can only decrypt this message with his own private key (See Figure 8.1 (a)). In this scenario, an attacker can change the identity of public key, so both end users are not able to trust each other. The PKI overcomes these problems with the help of a TTP. Both sender and receiver trust the third party who issues the digital certificates which say that the public key belongs to the respective person. The main function of PKI is to create trust between end users by issuing certificates to them. The certificate contains the public key of the user, user id, and a digital signature of the certificate issuer.

8.3.3 Centralized and Decentralized Approaches

In PKI the certificates can be either generated by central or by distributed entities. The choices are based on the security policy of an organization [16].
A typical centralized PKI has the TTP which creates the certificates for users and distributes them. In this approach, the CA is the TTP. Here, the certificate is generated by a central server and transmitted to the certificate requester. For example, VeriSign[1] is a TTP who issues the digital certificates for websites.
In decentralized approaches PKI functionalities are distributed and managed by peer systems. In this case, the peer system controls and manages the operation of PKI. Here, the certificate is generated by other trusted participants in peer network. There are number of approaches to discover a certificate in the trusted graph which we will discuss in the upcoming sections.

8.3.4 Trust Models

The PKI can be built based on different trust models. In a closed environment (i.e. small organization) the tracking and verification of a certificate is very simple, since it needs single root CA. However in an organization there are needs to communicate with external resources such as client, vendor, customers, and associates. Since CA from another organization is not accepted by them due to lack of trust, establishing trust is more difficult.
In general, most of the organizations use one or more trust models to interact with external persons. The trust models can be classified into three different groups which are:

- Direct Trust,
- Hierarchy trust and

[1]Symantec has acquired VeriSign, http://www.verisign.com/

- Web of trust.

8.3.4.1 Direct Trust

The direct trust is a simple trust model in which users trust each other by verifying their public key, as they know where this key is from. A typical cryptographic mechanism functions in this way.

8.3.4.2 Hierarchical Trust

In the hierarchical trust model, there is more than one CA where trust is extended in a hierarchical way. X509 PKI uses this trust model. Here, the most trustworthy entity is root CA, which issues the certificates to subordinate CA. Later, these subordinate CAs can issue certificates to their customers. The hierarchy of CA in this model establishes the chain of certificates. When an entity wants to validate the public key of an opponent entity, it must provide its own certificates along with certificates of all other CA's in its certificate chain. In this process opponent's certificates is considered as valid if and only if an entity finds a CA that is trustworthy at an appropriate level in the hierarchy.
For example, consider the Figure 8.3; here both entities A and B trust the $ROOT\ CA$, entities B and C trust $CA2$. Here, entity A can validate C's certificate along with certificate chain from $CA2$ to $ROOT\ CA$.

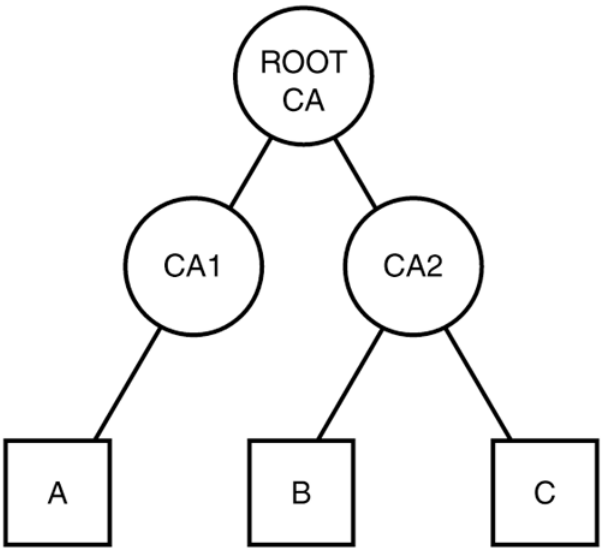


Figure 8.3: Hierarchical Trust [6]

8.3.4.3 Web of Trust

The web of trust model is a cumulative trust model which includes both direct trust and hierarchy model. The model is developed by Philip R. "Phil" Zimmermann Jr in 1991, and his idea is, "more information is better". He mentioned that, "A certificate might be trusted directly, or trusted in some chain going back to a directly trusted root certificate (the meta-introducer), or by some group of introducers". This model is designed for implementing the PKI in a decentralized infrastructure, i.e. for Peer to Peer systems.
Consider the Figure 8.4. Here, Alice and Carol trust each other, similarly Bob and Dave trust each other by direct trust, and Carol trusts Bob. Then, Alice can trust Dave, since Alice trusts Bob through Carol; Bob and Dave both trust each other (i.e. trust path would be Alice -> Carol -> Bob -> Dave). If Alice has more than one certificate paths to

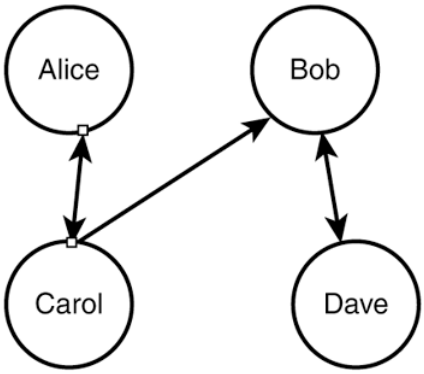


Figure 8.4: Web of Trust - Alice Bob Scenario [2]

validate Bob's key, web of trust uses the sum of the trustworthiness level on those paths towards Bob.
When Alice needs to validate Bob's public key certificate, at that time she finds Carol's signature on Bob's certificate. Since Alice already validated Carol's certificate, so Alice can trust Bob's certificate.
This scenario indicated as the dotted line shown in Figure 8.5.

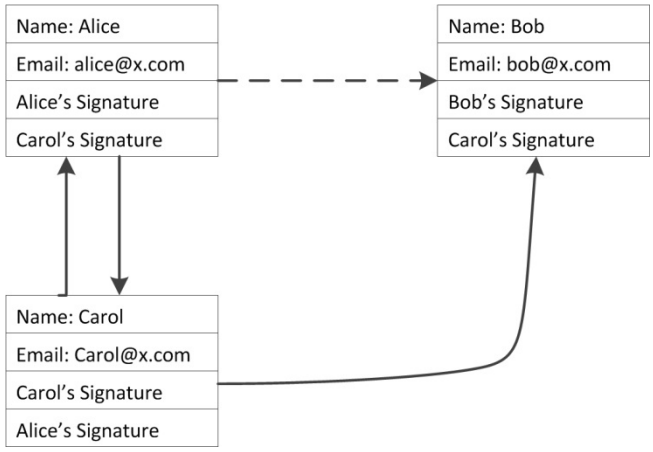


Figure 8.5: Web of Trust - Trustworthiness of Alice and Bob

8.3.5 Certificate Revocation

In general, when the certificates are created, they are scheduled with a validity period, i.e. start date/time, and end date/time. After the expiration date, public key in certificates is not authenticated and it is dangerous to use these certificates. There are several reasons for revoking a certificate, such as the user may loss the private key or the user may forget passphrase for his private key or an employee may shifted to other organization [11].
Any individual who had already signed the expired certificate can remove his or her own signature from that certificate. It shows that users no longer trust on the public key and identification of that certificate. In X.509 certificate, the revocation of a signature is the same as the revocation of certificates, since it has only one signature of CA. In PGP, the certificates are revoked only by the owner of the certificate or by someone designated as a revoker (only certificate owner can designate someone as revoker). In practice designated revoker can revoke the certificates even if the passphrase of private key is lost. In contrast, for X.509 certificate, a revoke is possible only by the certificate issuer.

8.4 Implementations

8.4.1 Public-Key Infrastructure X.509 (PKIX)

The public-key infrastructure X.509 is a centralized approach of PKI and it is a formal model, based on X.509 certificates which can be deployed as certificate based architecture on Internet. The IETF working group X.509/PKIX has been the driving force for setting up this formal model. The internet applications such as WWW, electronic mail, user authentication, and IPsec can make use of PKIX [11]. The Figure 8.6 illustrates the simplified view of architectural model.

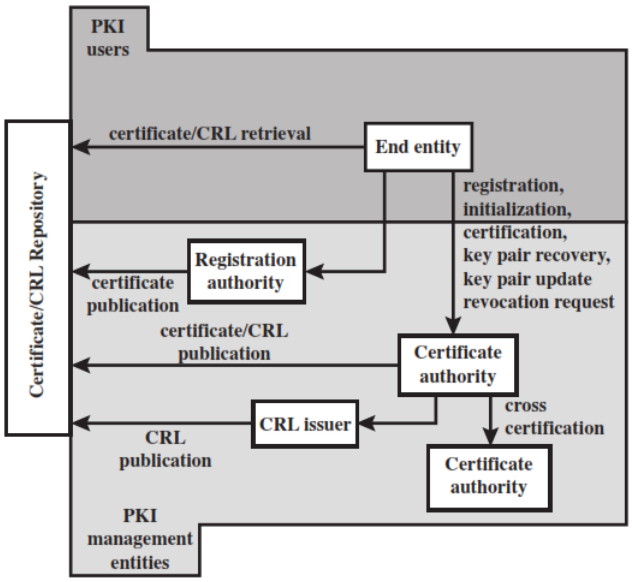


Figure 8.6: PKIX ArcitecturalModel [15]

8.4.1.1 Components of PKIX

End Entity: It represents the end users or the devices (servers, routers or other network components) identified in subject field of PKI certificates. This entity consumes the service offered by management entities of PKI and it also relies on other components such as CA and RA for obtaining the digital certificates.

Certificate Authority: The CA is a management entity and vital component of PKI which issues digital certificates and indicates revocation status of the certificates. It supports number of administrative tasks and it may delegate some of the tasks to registration authorities.

Registration Authority: The RA is an optional component of PKI which shares administrative function from the CA and mostly it verifies the certificate contents such as name of applicants, public key, and algorithms from end entities. The CA identifies RA by name with public key and it trusts the information provided by RA.

CRL Issuer: The CRL stands for Certificate Revocation List. This is an optional but expedient component, which generates CRL. The CRL is a list of revoked certificates which CA can delegate to CRL issuer to publish.

Repository: It is a central or collection of distributed systems on which other PKI entities are dependant on for storing certificates, certificates verification and revocation status. It needs to inter-operate with other services of PKI for retrieving the certificates and the CRLs [14].

8.4.1.2 PKIX Management Operations

As we see in the Figure 8.6 the PKIX has a number of management functions supported by management protocols which are as follows.

End Entity Initialization: The initialization is the first step for an end entity to deal with PKI management entities and it provides information about PKI support functions. Commonly, the client system is required to initialize a secure communication with management entities such as CA/RA. End entities are introduced to CA/RA in order to get new certificate. When this process succeeds, CA will issue a new certificate for the end entity and stores it in the public repository.

Proof of Possession (POP) of Private Key: The proof of possession of private key is a very important step in PKI management operation. Here, a PKI management entity validates the binding between an end entity and a key pair. It is mandatory for CAs/RAs to enforce POP, because many non-PKIX operational protocols do not explicitly check the binding between the end entity and the private key [10]. For example, email protocols do not check this binding. If this binding is not verified by the CA/RA, the certificates in PKI are less meaningful.

Key Pair Recovery: If an entity loses the decryption key, it is impossible to recover encrypted data. The loss of key is equivalent to forgotten password/PIN or corrupted disk. So it is important to provide a mechanism to recover the lost decryption key. This can be accomplished by key pair recovery service offered by CA, which enables the user to restore the encryption/decryption keys from authorized backup facility.

Key Pair Update and Revocation Request: In general, when a user lost his/her private key or when the certificate expires, they should have a new key pair. In order that the CA can issue a new certificate, this new public key needs to be updated. At the same time, the old certificate has to be added in the certificate revocation list.

Cross Certification: In general, PKI may have more than one CA and information between those two CA's is exchanged by establishing cross certification. The cross certificate is a certificate issued by one CA which signs the public key of another CA. In this process, the requester CA initiates cross-certification request "ccr" with a fresh random number and sends to responder CA. Here messages are protected by Machine Authentication Code (MAC), which is a short piece of code used to authenticate the message. The responder CA validates message, saves the random number generated by requestor CA and it generates a new random number (responder random number). After that, it creates a new certificate which contains requester CA's public key signed with the responder CA's private key, this process is called cross certification. The responder CA sends a cross certification response "ccp" message back to requestor CA. Finally, the requestor CA validates this "cpp" message and responds with "certConf" message.

8.4.2 Pretty Good Privacy

The PGP is a hybrid cryptosystem which combines the features of symmetric and public-key cryptographic technique. Philip R. "Phil" Zimmermann Jr. created this PGP encryption system in 1991 and it is used to encrypt text, email, files, directories, and even storage disk. PGP compresses the plain text before the encryption, which reduces the message size and the round trip time in network. Moreover, compression of plain text resists against cryptanalytic attacks by reducing similar patterns that can be found in uncompressed plain text.
PGP uses web of trust model rather than hierarchical model. Here public key of a particular entity is verified by the signature signed by one or more peer entities. PGP also supports certificates like PKI; it recognizes both certificates X.509 and PGP certificates. The main difference is, the X.509 certificate can have only one signature, but PGP can have one or more signatures. Another benefit of PGP is that it offers the digital signature feature. In addition, PGP uses complex cryptographic hash algorithm, so modification or changes to plain text along with the signature is very difficult. Any modification to plain text will result in failure of digital signature verification process. PGP combines public key and conventional encryption techniques thereby benefitting from both security and efficiency. In the following subsection we will discuss about the encryption and decryption mechanism of PGP.

8.4.2.1 Encryption and Decryption

PGP encryption is tricky part; it creates the session key with the help of true random number generator (TRNG) which generates the random number based on an entropy source. The source is referred from the physical environment of a computer such as mouse movements, disk electrical activity, keystroke timing patterns, and values of system clocks. This session key act as one time secret key and it is used for encrypting the compressed plain text using very secure and fast conventional encryption algorithm (i.e. symmetric key encryption). After that, the session key is encrypted using the recipient's public key and then this encrypted session key along with the cipher text is transferred to the receiver. This ensures that the receiver who has corresponding private key only can decrypt to get the one time session key. The Figure 8.7 illustrates the PGP encryption mechanism.

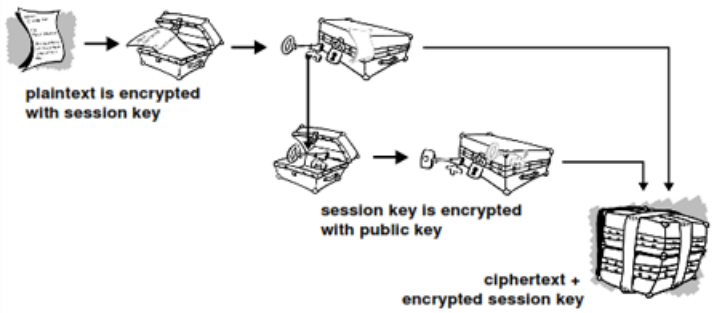


Figure 8.7: PGP Encryption [7]

In PGP decryption the receiver uses the private key for decrypting the one-time session key. Then receiver uses the session key and decrypts the message by using the conventional decryption algorithm. The Figure 8.8 represents the PGP decryption scheme.

8.4.3 KeyChains

The KeyChains is a completely decentralized PKI approach based on the PGP web of trust model [9], which is used for exchanging secure email without CA. In this model user

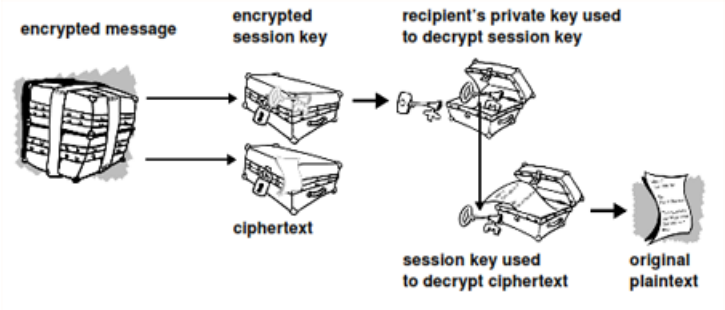


Figure 8.8: PGP Decryption [7]

verifies the certificate either by using direct certification or by using the certificate chain. The PGP uses centralized key servers to store and retrieve the certificates. So, scalability is possible by replicating the key server which requires significant amount of hardware resources and infrastructure. This increases the cost and complexity for managing the replicated key servers. The KeyChains is completely decentralized PKI, which scales when the number of users increase.

8.4.3.1 Approach

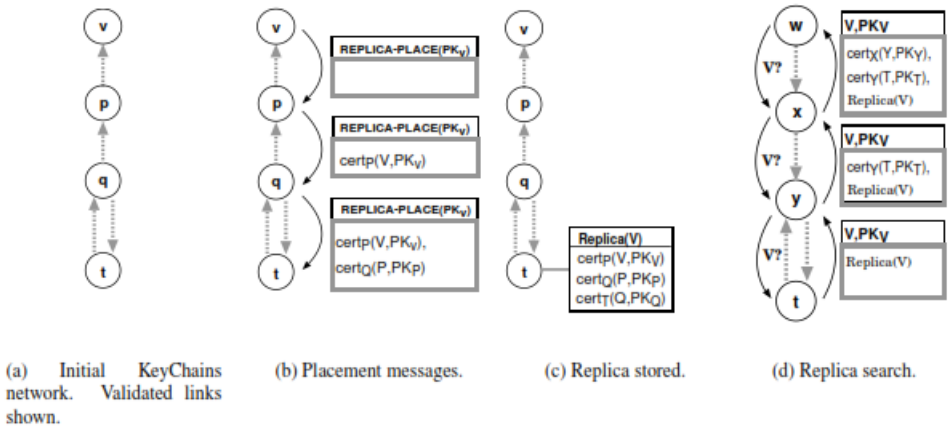
The KeyChains is built on top of the PGP web of trust model; it uses distributed mechanism for users to store and retrieve their public keys. This KeyChains uses key location protocol, which is modified version of Local Minima Search (LMS). LMS is an existing object lookup protocol, which provides provable performance and guaranteed when it is run over the arbitrary unstructured networks [9]. It is specially designed to discover and retrieve the certificate chain, and not only the user's public key. KeyChains uses PGP certificate and provides the PKI system functionality such as publish, search, and validate public keys without central servers.
The key idea in KeyChains is by using PGP certificate graph (consider there is a trust edge from A to B if $cert_A(B,\ PK_B)$ exist), the PKI store and retrieve operations are performed with the help of LMS. The LMS is a lookup protocol returns the certificate chain from the initial peer to target peer when it finds the required public key. In order to find the public keys and certificate chains faster, it requires several modifications in existing LMS protocols. Next section describes about implementation of KeyChains.

8.4.3.2 Implementation using LMS

The semantics of LMS are similar to the Dynamic Hash Table (DHT). In DHT peers and objects are mapped into an identifier space using consistent hashing, and objects are stored at peers, determined by the distance between objects' and peers' identifiers in this space [9]. Peers within h hops of the network are identified by using their identifier. A peer sends a number of probes into network to perform store and search operation. These probes are forwarded within the local minimum (i.e. within fixed length in h hops) along undirected links between peers. The local minimum is selected randomly. The performance of this protocol depends on the number of searches and storing replicas.

The KeyChains: The LMS runs over absolute topologies, so the features of LMS are more suitable for implementing a PKI. It can run over peer to peer system where the topology replicates web of trust certificate graphs. Here the principle (i.e. user) and the peer are distinct; a peer may have one or more principles. The KeyChains does not enforce any trust related information between peers and principles. It uses the principle as out-of-band mechanism to store/retrieve certificate form peers. Let us consider a certificate

from A to B is represented by $cert_A(B,\ PK_B)$ and it means A certifies the statement, "B's public key is PK_B" with its signature. In this graph, A to B directed edge exist only when certificate $cert_A(B,\ PK_B)$ exist.



(a) Initial KeyChains network. Validated links shown. (b) Placement messages. (c) Replica stored. (d) Replica search.

Figure 8.9: KeyChains - Replica And Search [9]

LMS Adaption for Trust Graph: In general, the LMS supports undirected links to the peers, but the trust with certificates are represented in directed link. So, this requires modification of undirected links of LMS to directed link for certification requirements and this process specified in [9]. When a message is forwarded from one node to another node, the links point to a particular direction and corresponding certificates are added to the chain.

Placing Replicas: The certificate of a principle is replicated in trusted graph nodes within local minimum. See the Figure 8.9, the principle V in peer v invokes PKI store operation by sending REPLICA-PLACE probe within local minimum. If local minimum of nodes receives duplicate storage request, it rejects that request and notifies v. Each failed operation doubles the length of local minimum to send a new probe and it continues until the specified threshold is reached. The chain is constructed along the path where probe v is forwarded and corresponding certificates of edges are appended to the message see Figure 8.9 (b). Here probes are forwarded along with incoming links. Since, the certificate need to point towards V, the peer p forwards the probe to q if and only if certificate $cert_q(P,\ PK_p)$ exist in trusted graph. Finally, the replica with public key of V and the constructed path is stored at the destination t.

Finding Replicas: The retrieve operation of PKI is invoked by sending SEARCH-PROBE message within local minimum. For example, while forwarding SEARCH-PROBE from w, the corresponding certificates of traversed edge are appended to message and probes are forwarded in outgoing links (i.e. p forward probe to peer q if and only if $cert_p(Q,\ PK_q)$ exist in certificate graph) see Figure 8.9 (d). The destination peer t holds the requested key of peer v and when peer t receives search probe, it responds with values of the key along with certificate chain from t to v. In addition, SEARCH-PROBE has the certificate chain from initiator w to t. In result, both chains are combined together to get chain from w to v. Hence the search operation is performed by deterministic forwarding the probes within local minimum.

Revoking public key: The revoking public key process is cumbersome in PGP web of trust approach. The KeyChains solve this problem very easily by using a two pronged

approach. The peer who replicates the public key keeps tracking on the other peers that store the key. When a peer revokes its public key, it requests all other peers who have replicas of its key to delete it. Hence the search will not find any revoked public key. However, the malicious peer can cache the revoked public key and do fraudulent activities with it. This situation can be avoided by prompting the requested key that has been revoked, when a peer periodically searches for the revocation statements of previously retrieved key.

8.5 Problems

In the previous sections, we have discussed various approaches to implement user convenient trust. Each approach has its own trade-offs over the other. In the upcoming section we will discuss several problems involved in PKIs, in both technical and administrative aspects.

8.5.1 Technical perspective

In a technical perspective, the most common problem is to distribute the certificate revocation list (i.e. blacklist of certificates) at regular time intervals and to establish initial trust between PKI entities. In addition to this, there are various problems in both centralized and decentralized based architectures. Let us discuss about these problems in detail.

Trusting Unknown CA: In general, computer systems are shipped with a list of certificates by OS manufacturers and/or by web browsers (To view them in Firefox go to Option -> Advanced -> View Certificates -> Authorities). This list of certificates is belongs to CAs, they are trusted and used for the certificate verification process. Most of the end users are not aware of this certificate list imported to their system and they just trust those CAs. Moreover, if the browser does not have a particular trusted certificate which the user might need while browsing, the browser displays a popup message and asks "Do you trust this CA?" with options "Trust" and "Don't Trust" [5]. In this case, how does the end user decide on whether to choose trust or don't trust option? And how does the end user trust a CA? In most of the cases, user blindly clicks "Trust" option, and then the browser trusts that CA for a particular session. If the CA is not trustable enough, in that particular single session, the hacker could listen to the secured communication. Hence trusting an unknown CA becomes a main problem.

Protecting Secret/Private Key: The protection of secret / private key is another main problem in both symmetric key cryptography and asymmetric key cryptography. However, if a user loses a key in symmetric key mechanism, he/she only needs to inform other end user who shares the secret key. But in PKI this is even more complex, because here many users share a public key and they all have to be informed about the stolen/loss of private key. As we previously discussed, this process is accomplished by sending a revoke message to CA and then CA delegates to CRL issuer for issuing revocation list. If the attacker gains the private key, he can interpret every message from/to key-owner and he can also create genuine digital signature of key-owner. Normally, the private key is protected by a passphrase, where the user is recommended to give a strong passphrase of maximum length with a combination of alphanumerical and special characters. If the passphrase is weaker it can also be easily hacked. Thus protection of secret/private key is a problem in both symmetric and asymmetric mechanism.

Security of CA: The security of CA system is very important in PKI, since it contains sensitive information of customers. Moreover, the CA is a central system which has many possibilities for a single point of failure (i.e. if a CA goes offline or its machine goes down the entire security infrastructure will get affected). In addition, a typical CA has one or more root public keys and if CA is not secured, hacker can add his/her public key into the root certificate system of CA [4]. After that, he/she can create genuine certificates which will be treated as the certificate exactly issued by CA. Even more, the verification process of this fake certificate succeeds, since the hacker's public key is already in root certificate system. So, whenever a person communicates with his/her friend using that fake certificate, the hacker can interpret and/or modify the message. Hence, the insecurity of CA systems is another vital problem in centralized approaches.

Updating Certificates: Updating certificate fields are cumbersome in certificate based cryptography and it requires re-issue of certificate. The most commonly used certificates are X.509 and PGP certificates which has several fields, where user often needs to update those fields. For example, in S/MIME signed/encrypted email, the users have to specify their email id in certificate. When they change their email address frequently than the certificate, the certificate would have to be re-issued. Moreover, in a typical organization, ownership of certificate often changes for some administrative reasons. At this point, certain fields of certificates need to be updated and published again. Thus in certificate based cryptography the burdensome issue is, each and every change made to the certificate requires a re-issue.

CRL Problem: In a CA based system, issuing the revocation certificate list (i.e. Blacklisted certificates) on a regular time intervals is very challenging. As we previously discussed in Section 8.3.5, there are several reasons for revoking certificates, and in certificate revocation process the corresponding certificate will be added into the revocation list. In practice, there are several problems involved in issuing CRL which includes: distributing CRL is expensive, checking and verification process takes longer time and causes inconvenience to end user. Moreover, critical application requires more accurate and real-time certificate status information. In order to guarantee the timely status update, CRL system has to update the revocation list as frequently as possible. This increases the load on server and network traffic, when it issues the list once every minute. Alternatively, reducing the frequency of CRL update to an hour or day does not provide timely revocation for critical applications. In addition, CRL doesn't have proper pricing strategy for the customers. When a CA issues a certificate, it charges the user a fee, and the amount the CA charges is typically tied to how much checking it does before issuing the certificate [13]. But user expects CA to issue revocation status for free. In this case both CA and user cannot specify how often certificate will be validated and what degree of latency will be acceptable. This result in CA to pay more attention on issuing CRL since, creating CRL and publishing requires processing time and substantial amount of bandwidth. Hence distributing revocation list is another main problem in the technical aspect of PKI.

8.5.2 Administrative perspective

In administrative perspective, the deployment of PKI is complex for creating certification policy. The certificate policies are business rules used in PKI for implementing and it can be enumerated in various documents such as Certificate Policy (CP) and Certificate Practice Policy (CPS) [1]. Standards such as X.509 v3 certificate allows user to express policy mapping, but it is complex to use them securely [3]. The policy mapping is a field in X.509 certificate which allow user to specify certificate policies. In addition to this, we

have various problems in administrative aspects while implementing PKI and they are as follows.

Not User-Friendly: The secret key in symmetric mechanisms and private key in asymmetric mechanisms is not user-friendly. In both mechanisms, secret key and private key has to be stored in a secure place, mostly they are stored in local system. At this point, when a user upgrades their computer they need to move their secret/private key to new machine and he/she have to make sure that key is properly deleted in their old system. But, in most of the cases they may forget to move their key(s) into the new machine. In this case, user needs to generate new key(s)/key pair(s) and inform to other end user(s). As we previously discussed about updating key is even more difficult in the case of PKI. Therefore, secret/private key is not user-friendly.

Certificate Chain Problem: The construction of certificate path from one entity to another in PKI has some problems. A typical certificate path may have an iteration of loops when an entity finds multiple certificate paths (with different semantics) to its target. In extreme cases, the semantics of a certificate can change across different iterations of loops [13]. Moreover, dealing with certificate chain also results in certificate identification problem (i.e. identifying certificate in which directory?). A typical PKI requires locating the certificate and its status information not just in a single repository but also in multiple repositories. In addition, when length of certificate chain is increased, trust between two entities may be weakened. Because, the intermediate entities may restrict to utilize their hardware resources for constructing certificate chain or some entities may blindly sign the certificate (i.e. without proper verification of identity). Hence these problems weaken the certificate chain.

Expensive Certificates: The certificates are expensive for users, when they request convenient form of certificate such as smart cards and if the user chooses service monopoly CA. The smart card is pocket size integrated circuit card made up of plastic material. It contains the microprocessor elements and volatile memory components, which are used to store highly secured information. For using smart cards, the PKI enrolment process requires end entities to provide high-assurance on their client device implementation [8]. The client device (smart card reader) needs to provide high level of accountability and functionalities to be interoperable with PKI methods. Moreover, in order to issue a certificate, CA needs to ensure that corresponding smart card's client device is secured enough. Next, the service monopoly CA may charge extra fee to issue certificates. For example, certain geographic regions might have monopoly CA for some political reasons and it may lead to increase the price of certificates. Thus the convenient form of requirement on client side and the service monopoly CA increase the price of certificates.

Cross Domain Trust: The implementation of PKI in heterogeneous domains might be difficult. Employees in an organization need to communicate outside the company such as clients, vendors, and suppliers. At this point, the CA between two organizations needs to share certificate information with each other. In some PKIs, CA provides sufficient access to their entities for establishing communication with remote entities, here trust relationship is configured with remote entities. If user's trust is primarily local, then remote configuration is considered to be not practical or not useful [8]. Moreover, the trust relationship will be complex in this situation. Additionally, the policy translation across domain boundaries has significant challenges, since policy elements are different from one organization to another, and it is difficult to negotiate policies between two

organizations. Hence the implementation and policy translation of PKI in cross domain is a key issue.

Lack of Awareness: The lack of awareness for the people to understand basic functionalities of PKI is another problem. In a typical organization, non-technical employees such as marketing manager, financial manager and other management executives are not aware about the functional details of PKI. For example, some non-technical employees may think digital signatures means a real handwritten signature, and when they need to verify digital signature they might look for a real handwritten signatures instead of the cryptographic digital signature. Training programs are conducted for educating the functionalities of security mechanism, even though it is difficult for the non-technical employees to understand. Most of the security mechanisms and its functionalities are complex in nature. Thus the people need to know at least the overview of security mechanism and the basic principles behind it.

Insecure Client Device: Most of the organizations desire to have convenient form of security infrastructure, such as smart card which needs the client device (smart card reader) to operate. As we previously discussed, the security of such a device is very important. The process of encapsulating private key into smart card is very common technique and it can be understood by everyone. If a hacker finds the client device of smart card is weaker, then he can easily access the information stored in smart card via that device instead of hacking the complex security algorithm. When the user needs convenient form of smart card based solution, they also need to consider security of client devices. Hence, using the insecure client device turns out to be one of the critical issues.

8.5.3 Proposed solution

In previous sections we discussed about the problems involved in the implementation of PKI. The main benefit of centralized PKIX/X.509 is absolute trust and this partially satisfies the enterprise requirements. However, in this approach scaling infrastructure is expensive and complex, because it requires additional hardware resources and the complexity increases linearly while new CA is added in the hierarchical level of trust. In contrast, the decentralized approach is easily scalable and any number of peers (entities) can be added or removed. On the other hand, as we discussed, validating trust on a lengthy certificate chain is difficult when certain intermediate peers limits the permission to store or process certificates.
In my perspective a combination of both centralized and decentralized approach could solve those problems which we discussed earlier. The proposed solution is hybrid PKI i.e. implementation of PKI based on hybrid P2P system architecture. In hybrid PKI, each root CA's have its own entities similar to hierarchical model and it can be connected with two or more autonomous root CA in a decentralized manner. Here trust between those autonomous root CAs can be established, based on PGP web of trust mechanism. In addition, the certificate of a root CA is stored and retrieved using modified version of LMS in decentralized peers (CA) rather than centralized key servers, as we discussed on KeyChains implementation. However, the certificates of end entities are not completely decentralized and they are maintained by corresponding autonomous root CA which is similar to PKIX/X.509 implementation.
Figure 8.10 illustrates the idea of hybrid PKI. Here root CAs are RCA1, RCA2, RCA3 and RCA4; they are connected in decentralized manner and each root CAs have one or more end entities or subordinate CAs.

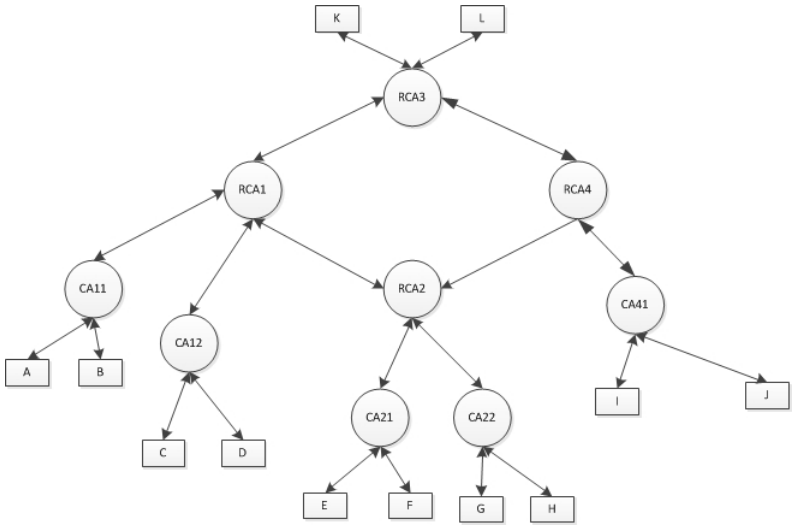


Figure 8.10: Hybrid PKI

This approach has benefits of both KeyChains and PKIX/X.509 implementation. Any autonomous CA can join or disjoin at any time and it is scalable without the complexities. Each autonomous root CA can enforce security policy within their root level.
In addition to this hybrid PKI, another proposed solution for better user convenience is a smart card with biometric scanner. Here the smart card reader will have an integrated fingerprint reader. This reader will take the value produced by a user's finger print instead of the PIN number. This solution has three main benefits which are as follows:

- **Portability:** User can carry the smart card with him and use it on machines facilitated with this new smart card reader.
- **Convenient:** User doesn't need to remember the PIN number always.
- **Security:** Smart card is locked with biometric value which is comparably more secure than typical PIN number.

Therefore this solution would be more convenient and more secure for end users. However, those above two proposed solutions, smart card with biometric finger print reader and hybrid PKI are independent. Those solutions can be implemented either jointly or individually based on requirements.

8.6 Conclusion

In this seminar report we have discussed about the public-key cryptography, public-key infrastructures, and their various implementations of centralized and decentralized approaches including PKIX/X.509, PGP encryption, and KeyChains. Also, we have seen the problems involved in technical and administrative aspects in implementing the centralized and decentralized PKI approaches. Both approaches have benefits as well as drawbacks. Some organizations prefer CA to be centralized and some of them prefer decentralized approaches as a best fit for them. The PGP encryption technique is used for protecting emails, files and hard disks. The PGP web of trust is most suitable for establishing PKI without the central controls like CA. However, it has centralized key servers for sharing public keys, and the main problem of key server is single point of failure. The KeyChains approach is completely decentralized approach that illustrates the mechanism to publish and search not only the certificate but also the certificate chain efficiently. In centralized PKI, CA as TTP offers absolute trust. Here KeyChains trade-offs on absolute assurance for greater scalability. Finally, we discussed about hybrid PKI which

combines PKIX/X.509 certificate based implementation and KeyChains; the smart card with biometric finger print reader is more secure and it provides better user convenient trust.

Like the Moore's law which states that, "The number of transistors per square inch on integrated circuits had doubled every year since the integrated circuit was invented", the processing speed of a computer is increasing every year which results in advanced researches on computing, such as grid, cloud, and ubiquitous could computing which could easily break complex security algorithms even faster. When inventing the highly secured and complex algorithms, the way to break those algorithms is also being founded. Besides those pros and cons of various encryption techniques and security infrastructures, the choice of PKI depends on the security polices of organization and their convenience.

Bibliography

[1] Amir Jafri and June Leung An OASIS PKI White Paper: *PKI Deployment - Business Issues*, 2005.

[2] Christopher Steel, Ramesh Nagappan and Ray Lai: *Core security patterns: best practices and strategies for J2EE, Web services, and identity management*, pages 217-219, 2005, `http://flylib.com/books/en/3.211.1.41/1/`.

[3] Cvrcek D.: *Real-World Problems of PKI Hierarchy*, (Accessed:) January 3, 2012, `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1.3068&rep=rep1&type=pdf`.

[4] Ellison FC. and Schneier B.: *Ten Risks of PKI: What You're not Being Told about Public Key Infrastructure*, Computer Security Journal volume 16, November, 2000.

[5] Ed Skoudis: *Search Security: Can a certificate authority be trusted?*, 2007, `http://searchsecurity.techtarget.com/answer/Can-a-certificate-authority-be-trusted`.

[6] Flenner R., Abbott M. and Boubez T: *Java P2P unleashed*, 2003, pages 217-219, `http://flylib.com/books/en/2.430.1.83/1/`.

[7] International PGP Home Page: *An Introduction to Cryptography*, 2000, `ftp://ftp.pgpi.org/pub/pgp/7.0/docs/english/IntroToCrypto.pdf`.

[8] John Linn, RSA Laboratories and Bedford: *An Examination of Asserted PKI Issues and Proposed Alternatives*, 2004.

[9] Morselli R., Bhattacharjee B., Katz J., Marsh M.A.: *KeyChains: A Decentralized Public-Key Infrastructure*, Technical Report, UM Computer Science Department, CS-TR-4788, UMIACS-TR-2006-12, Mar 2006.

[10] Network Working Group: *Request for Comments 4210 - Internet X.509 Public Key Infrastructure Certificate Management Protocol (CMP)*, Sep 2005, `http://tools.ietf.org/html/rfc4210`.

[11] Network Working Group: *Request for Comments 5280*, May 2008, `http://tools.ietf.org/html/rfc5280`.

[12] PCWorld: *MasterCard Affected in Attacks Over WikiLeaks*, December 2010, `http://www.pcworld.com/article/213039/mastercard_affected_in_attacks_over_wikileaks.html`.

[13] Peter Gutmann: *PKI - It's Not Dead, Just Resting Cover feature Security*, 2002.

[14] Richard Kuhn D., Vincent C. Hu, Timothy Polk W., and Shu-Jen Chang: *Introduction to Public Key Technology and the Federal PKI Infrastructure*, Feb 2001.

[15] Stallings W.: *Network Security Essentials Application and Standards*, Book, fourth edition, Pearson education, 2011.

[16] Techotopia: *An Overview of Public Key Infrastructures (PKI)*, (Accessed:) January 3, 2012, http://www.techotopia.com/index.php/An_Overview_of_Public_Key_Infrastructures_(PKI).

[17] Time Techland: *PlayStation Network Attacks May Cost Sony $170+ Million*, May 2011, http://techland.time.com/2011/05/23/playstation-network-attacks-may-cost-sony-170-million/.

[18] Wikipedia: *Digital signature*, (Accessed:) January 3, 2012, http://en.wikipedia.org/wiki/Digital_signature.