University of Zurich
Department of Informatics

Burkhard Stiller (Edt.)

Joint EC-GIN, EMANICS, and SmoothIT Workshop on "Economic Traffic Management" (Proceedings)

TECHNICAL REPORT — No. ifi-2008.10

August 2008

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

Burkhard Stiller (Edt.):
Technical Report No. ifi-2008.10, August 2008
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zürich
Binzmühlestrasse 14, CH—8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Table of Content

Economic Traffic Management — Summary of the Joint EC-GIN, EMANICS, and SmoothIT Workshop

Burkhard Stiller (Edt.)[1]

[1] University of Zürich, CSG@IFI and ETH Zürich, TIK, Switzerland, E-Mail: stiller@ifi.uzh.ch

Contributions by: *Thomas Bocek, Peter Racz, Tobias Hoßfeld, Michael Welzl, Juan Fernandez-Palacios, Sergios Soursos, Murtaza Yousaf, Cristian Morariu, Martin Waldburger, Nicolas Liebau, Amruth J. Kumar, Javier Rubio-Loyola, Miroslaw Kantor, Burkhard Stiller*

1 Overview and Objectives

The Workshop on “Economic Traffic Management (ETM)” had been jointly organized by the European research projects EC-GIN, EMANICS, and SmoohtIT. The common denominator of these three projects can be found in the topic of economic management, which includes the question, whether economics and economic theory is applicable in network management in general, in which way this will be beneficial compared to traditional network management approaches, and which players will benefit from such an approach.

Therefore, the main objectives of the workshop are:

- Exchange research ideas in the area of ETM;
- Exploration of economic management questions across project limits; and
- Discuss aspects of ETM, which are essential for their successful application in the Internet today/tomorrow.

To allow for a nicely structured set of presentations, 5 sessions have been organized on “Incentives and P2P”, “Congestion Control and Traffic Management”, “Bottleneck Detection and Distributed Capturing”, “Locality Mechanisms”, and “QoS Management and Traffic Optimization” as well as a single discussion on “Will Economic Traffic Management mechanisms be successful?”. Abstracts of all 13 presentations below show that the special problem addressed or the more general architecture problem tackled have a common basis in terms of considering decentralized and economically-driven characteristics. Therefore, ETM provides for the right incentives to ensure that all players are better off compared to traditional network and traffic management approaches.

Thus, in a nutshell, the ETM Workshop helped researchers to guide their respective area of work, mainly being influenced by practical application constraints as well as by seeing external effects and requirements, which could be considered useful to be integrated.

2 Incentives and P2P

The key concern in Peer-to-peer (P2P) networks and systems is driven by the fact that peers may consume more resources than offered by the same peer. Thus, the so-called free-riding problem had to be tackled by the right incentives, which ensure that upload and download of resources will get balanced. But furthermore, the role of an Internet Service Provider (ISP) is relevant for the transport of overlay traffic as well, since it may affect the planned traffic to be transported. Therefore, the ISP has to become a member of a collaborative game, which is driven by the overlay network provider

2.1 Private Shared History (Thomas Bocek, UZH)

This talk proposed a scheme termed Private Shared History (PSH), which is about combining a shared history, which is used to find transitive paths, and a private history to verify the correctness of this path. Finding such a path is important, if peers in the network have an asymmetry of interest. In such cases, a private history alone cannot be used as a basis for an incentive scheme and a shared history has to be used instead. However, both approaches have advantages and disadvantages. PSH exploits these advantages and minimizes disadvantages by combining both approaches in an efficient manner. The current implementation and its evaluation was presented. Finally, PSH extensions and its use in EMANICS, EC-GIN, and SmoothIT have been outlined.

2.2 SmoothIT Overlay Management Architecture (Peter Racz, UZH)

This talk provided a brief overview of the SmoothIT project and the architecture currently being developed. It presented the objectives of SmoothIT and discussed various incentives for all players to participate in ETM. Three main solution concepts (namely agreements, locality promotion, and QoS (Quality-of-Service)/QoE (Quality-of-Experience) differentiation) have been outlined and key requirements have been summarized. The TripleWin principle of an optimization of the cooperating roles of users, overlay provider, and underlay provider has been stated.

Finally, the SmoothIT Information Service (SIS) architecture and protocol have been presented, which serves in a client/server-based approach between Internet Service Providers (ISP) to optimize the peer/resource selection process of the overlay with a collaborative underlay.

2.3 Modeling of P2P-based Video Streaming (Tobias Hoßfeld, UniWue)

This talk showed how to model P2P-based video streaming and in particular to address the performance evaluation of Economic Traffic Management in the context of the SmoothIt project. As input for the video streaming model, some popular P2P-based video streaming have been characterized exemplarily. These measurements help to understand the system behavior and how to influence it, which in turn helps to derive further mechanisms for ETM. These measurements include a characterization of the formed overlay topology, the observed traffic characteristics, the applied edge-based intelligence (like bandwidth adaptation or re-routing on application layer), as well as a quantification of QoE depending on network disturbances.

In the second part of this presentation, it was shown how to model such a P2P-based video streaming system based on these measurements obtained. A crucial point was to determine the appropriate degree of abstraction, which is a trade-off between computational time, i.e. simulation efficiency, and accuracy of the model applied. This abstraction needs to allow for answering the desired performance questions. In the context of ETM and especially the SmoothIT project, it is necessary (a) to study the TripleWin situation when using ETM, *i.e.* to quantify the traffic optimization from different players' perspectives, and (b) to demonstrate the incentive to use the SmoothIT approach, *e.g.*, by showing the performance gain/loss when using/not using ETM.

3 Congestion Control and Traffic Management

Congestion control determines an important mechanism for managing traffic within a given network. Thus, standardization of respective mechanisms and metrics is essential for inter- and intra-domain scenarios. At the same time, the special case of overlay traffic appearing with a high percentage of the overall traffic in ISPs does need a careful handling to prevent unintended congestion as well as to maximize revenue for all types of traffic being carried. Interconnection issues on the physical and the business level have to be aligned. Finally, the effects of peer and locality awareness on traffic are investigated, while addressing the BitTorrent overlay system.

3.1 Current IRTF/IETF Congestion Control Work and How it Relates to P2P Systems (Michael Welzl, UIBK)

This talk gave an overview of current work related to congestion control in the IRTF (Internet Research Task Force) and IETF (Internet Engineering Task Force). This included an introduction to the scope of the Internet Congestion Control Research Group (CCRG) and a brief overview of recent discussions related to peer-to-peer traffic management. These discussions happened under the heading of TANA (Techniques for Advanced Network Applications), with a BoF (Birds-of-Feathers) session at a recent IETF-72 meeting in Dublin which might lead to the formation of a new working group.

3.2 Operator's Vision about Overlay Traffic Management (Juan Fernandez-Palacios, TID)

This presentation described the rationale behind the need of new traffic management mechanisms being able to promote the overlay traffic locality and provide required QoS for each application. Furthermore, a potential solution for traffic locality promotion and QoS differentiation was introduced. Such solution would be based on the combination of technical incentives and ETMs for ISPs and an overlay collaboration.

3.3 Insertion of ISP-owned peers and Locality Awareness in BitTorrent (Sergios Soursos, AUEB)

File-sharing overlay applications generate a large portion of the total traffic in the Internet. In this work, two approaches were investigated: How to modify the original BitTorrent protocol in order to achieve a more efficient use of the underlying network, and an evaluation run experimentally to study their impact both on the inter-domain traffic for the ISP and on the file download completion times for the end-users. In particular, a locality-aware mechanism was considered and applied to the tracker, based on which Autonomous System (AS) each peer of the swarm belongs to.

It was proposed to insert ISP-owned peers (IoPs) in the network as an alternative means to improve the download completion times. Experiments have been conducted of the aforementioned approaches using the ns-2 simulator and main results have been presented. The locality awareness achieves a reduction of inter-domain traffic, while the insertion of ISP-owned peers reduces further the amount of ingress traffic for the ISP that introduces the IoP. Furthermore, the introduction of an IoP improves the file download completion times. The combination of the two approaches is very effective also.

4 Will Economic Traffic Management mechanisms be successful? (Discussion Moderator: Burkhard Stiller, UZH)

To make the discussion short, the answer to this question was agreed upon to be: Yes. However, the constraints in which this "yes" will be true, have to be added as well. Thus, the following point of views have been expressed.

The decentralization of traffic and network management is a must to ensure scalability concerns of operators in an ever increasing world of new services, applications, and consequently traffic profiles. Thus, the traffic management as such needs to be efficient, but, at the same time, it has to reduce costs of the management tasks undertaken.

Since a direct inter-connection to billing systems may be way too costly, economic incentives should be integrated into the data signaling and data exchange process. This integration may happen at the edge of the network, however, it may not change existing charging schemes, such as the flat fee scheme for residential customers and the 95%-percentile scheme for interconnected ISPs.

Furthermore, the role of congestion control in that respect does play an important role, where traffic shaping as well as the support of fairness issues will be important. If such schemes can be integrated into today's Internet without the need to change protocols, the potential for reliable, secure, resilient, and efficient mechanisms is large.

However, it has to be taken into account that time scales of round-trip times are way shorter compared to overlay-to-underlay mapping feedback loops. Thus, the need to optimize these mapping loops becomes obvious, since a beyond packet-level approach may diminish the complexity, while at the same time, ease the traffic management.

Finally, the need to various traffic classes has been expressed — if it were 3, 4, or 16 or even more, has not been concluded unanimously. The problem of assigning an application or its traffic flow onto the respective class gives raise to further problems, but this cannot be neglected at all, since today overlay applications do — at least in principle — neglect the underlay as well. And this situation is characterized as being non efficient at all.

But the need for service differentiation as well as accompanying measures on the technical side as well the incentive aspect, thus, the economic relevance of the problem, has been stated clearly.

Therefore, the potential of economics being applied to traffic and network management is clearly seen and has to be supported in a variety of aspects, such as mapping functions, incentives for peer selection, pricing schemes for end-user traffic and ISP-to-ISP traffic, or service differentiation.

In which way the benefits of ETM can be quantified and proven is under heavy investigation in different projects and work packages. The success of ETM as such can be stated to be measurable, since revenues, cost reductions, and the minimization of maintenance efforts will form key dimensions and parameters to be used to show that ETM mechanisms are beneficial. This benefit will be visible for all players involved, including the customer, the overlay provider, and the underlay provider.

5 Bottleneck Detection and Distributed Capturing

The problem of shared bottlenecks in a given network shows that many flows may suffer a reduction of quality of this communication. Thus, a detection algorithm will benefit applications in a way, which will allow them to achieve their intended QoS metrics. Furthermore, the problem of network monitoring in a traditional manner sees a single mirroring device and a single or multiple analysis box in place. This approach fails to scale with respect to the data rates of the link, thus, a scalable and robust approach is essential for an efficient monitoring of traffic.

5.1 Shared Network Bottleneck Detection with SVD (Murtaza M. Yousaf, UIBK)

This talk presented a new mechanism for detecting shared bottlenecks between end-to-end paths in a network. This mechanism, which only needs one-way delays from endpoints as an input, is based on the well known linear algebraic approach SVD (Singular Value Decomposition). Clusters of flows, which share a bottleneck are extracted from SVD results by applying an outlier detection method. Simulations with varying topologies and different network conditions show the high accuracy of our technique.

5.2 DiCAP — An Architecture for Distributed Packet Capturing (Cristian Morariu, UZH)

IP (Internet Protocol) traffic measurements form the basis of several network management tasks, such as accounting, planning, intrusion detection, and charging. High-speed network links challenge traditional IP traffic analysis tools with their high amount of carried data that needs to be processed within a small amount of time. Centralized traffic measurements for high-speed links typically require high-performance capturing hardware that usually comes with a high cost. Software-based capturing solutions, such as libpcap or PFRING, cannot cope with those high data rates and experience high packet losses.

Thus, this presentation proposed a scalable architecture and its implementation for Distributed Packet Capturing (DiCAP) based on inexpensive off-the-shelf hardware running the Linux operating system. The prototype designed had been tested as an implementation and was evaluated against other Linux capturing tools. The evaluation showed that DiCAP can perform loss-less IP packet header capture at high-speed packet rates, when used alone, and that it can highly improve the performance of libpcap of PFRING when used in combination with those.

6 Locality Mechanisms

Locality determines typically the information within a given context, where a user, client, or provider is located geographically. This type of information may be of high importance depending on the application and service in use. Thus, the automated detection of jurisdiction forms a key part of an automated contract conclusion or a later claim handling procedure, for which location information in contracts is required. Furthermore, in overlay networks, the search for a resource can be driven by location information in a more efficient manner, especially overcoming the problems of server-based, central solutions. Finally, the Border Gateway Protocol (BGP) can be utilized to extract locality information, which can be applied in turn to rank peers in terms of their locality of their IP addresses.

6.1 Locality and Contracts (Martin Waldburger, UZH)

Location-related parameters such as a service provider's domicile, habitual residence, and establishment constitute key input parameters for private international law. Private international law procedures, also known as conflicts of laws, are relevant to international contracting, thus, to situations where a service provider and service user reside in different legal domains.

Driven by the motivation to automate contract formation in an international context, an attempt to formalize the Swiss federal private international law (IPRG) has been undertaken. The focus was set on determining jurisdiction in an international contract automatically. To that aim, IPRG-specific decision rules and needed input parameters — so-called connecting factors — have been presented in excerpts. This lead to visualizing those challenges faced when aiming at a direct law formalization. Consequently, a hypothesis-based approach to address these challenges was introduced.

6.2 Globase.KOM - A P2P Overlay for Fully Retrievable Location-based Search (Nicolas Liebau, Aleksandra Kovacevic, TUD)

Location-based services are becoming increasingly popular as devices that maintain a geographical position become more available to end users. The main problem of existing solutions to location-based search is keeping information updated, which typically requires the centralized maintenance at specific times. Therefore, retrieved results do not include all objects that exist in reality.

A P2P approach can overcome this issue, since peers are responsible for those information users are searching for. Unfortunately, current state-of-the-art overlays cannot fulfill these requirements for an efficient and fully retrievable location-based search. In this talk Globase.KOM was presented, a hierarchical tree-based P2P overlay that enables fully retrievable location-based overlay operations, which proved to be highly efficient and logarithmically scalable.

6.3 Peer Locality Information Using BGP (Amruth Juturu Kumar, UZH)

A P2P application constructs an overlay network for the purpose of efficient and scalable resource searching or sharing. The underlying network offers interconnected ISPs, either via peering or costly transit links. One key problem in this case is the high traffic load on transit links caused by a non-optimal selection of peers within the overlay network. A new approach to solve this problem is by applying ETM mechanisms based on incentives, where overlay nodes can query information from ISP-provided services in the underlay, to allow for an optimal selection of peers in the overlay. In this respect, information on whether a packet will be routed to a transit link, a peering link, or within the ISP, is highly beneficial to determine whether a remote peer is preferred or not by that ISP. Other useful metrics in these cases include locality and performance.

The current work did design and implement an infrastructure within a test-ISP that provides for respective information on locality of a peer relative to the querying peer. The locality information is based on the BGP routing table and also on the information from ISP-maintained databases.

7 QoS Management and Traffic Optimization

The QoS management driven by business indicators does show an approach, which can automate the policy-based management of commercial services. Thus, an integration of policy-based management approaches in a multi-domain case with business value becomes promising. Furthermore, for the optimal interconnection of ISPs the respective routes become crucial, if not selected carefully. The new routing algorithm proposed shows a concept, which may require a more difficult coordination between ISPs compared to ISPs at the benefit of a more detailed knowledge of QoS connectivity, resilience, and cost.

7.1 Issues Confronting Business-driven QoS DiffServ Management (Javier Rubio-Loyola, UPC)

Network and services policies have been proven to be an efficient vehicle to assess QoS DiffServ (Differentiated Services) management in intra-domain and inter-domain environments. Moreover, current frameworks that address this issue have been decoupled traditionally from the business value, even when the research community recognizes business profit as one of the main motivations for any management solution.

This talk gave an outline of the key aspects confronting business-driven QoS DiffServ management. It presented initially principles of the application domain of this research topic and provided an introduction to the technical approach that has been chosen to address it. Finally, it provides a scenario outlining the scope of this research and summarized the issues that are currently being addressed in this work.

7.2 Inter-domain Traffic Optimization in an Inter-carrier Environment (Miroslaw Kantor, AGH)

Due to the development of Next Generation Networks, which leads to a multiservice transport layer within a multi-domain environment, the importance of inter-domain traffic engineering issues keeps growing. As the telecommunications market is still increasing and the number of ISPs is growing, operators face different routing options with regard to service quality and cost. Also changes in pricing models and the explosive growth of traffic force carriers to deploy new routing models, since the business environment becomes very dynamic and routing changes are required in shorter time frames. Connections have to be routed according to the lowest cost paths to maximize operator's income. Inefficiencies in implementing interconnection strategies can decrease carriers' outcome and make them spending more time on network management.

Therefore, the need to develop algorithms supporting the choice of optimal interconnection routes becomes crucial. Least Cost Routing (LCR) algorithms to optimize the utilization of resources are proposed. By using the methodology proposed the best upstream/transit ISPs are selected. The chosen ISPs will assure low cost, good performance, and sufficient path diversity to protect against the network failures. By using the LCR algorithms proposed, the routing strategy can be more efficiently executed by incorporating the knowledge of cost with network conditions. The LCR algorithm can also decrease the time interval needed to analyze a huge number of alternatives and helps a carrier make decisions considering new agreements with other carriers within a dynamic framework.

8 Conclusions

The workshop has shown that Economic Traffic Management (ETM) mechanisms show a high potential, which has to be investigated and exploited in research and prototypes. While the advantages of a highly decentralized traffic management approach in the world of today's interconnected networks of the Internet is obvious — due to many providers and far more customers being interconnected — the need for a scalable management functionality in this world is emerging — mainly due to too many flows and applications to be supported. Thus, the application of incentives — either monetary ones or non-monetary ones — enables a high decentralization degree, which typically leads to economics, since fully decentralized markets show a significant number of commonalities, which a decentralized network and traffic management can exploit. Therefore, the economics are an important aspect of tomorrow's management approaches, since they combine the incentive metric with the traffic to be transported, monitored, and signalled. In conclusion, the ETM mechanisms addressed so far will play an important role in application areas and networks, where the benefits in terms of gains achieved, *e.g.*, in terms of revenue, cost savings, or smaller investments for providers, will be quantified and underlay providers, overlay providers, and customers will cooperate under determined strategies. This approach will lead to a TripleWin situation, where all cooperating parties will be better off, compared to traditional traffic management approaches.

A number of those areas tackled in the Workshop on "Economic Traffic Management" are being worked on in much more detail in a variety of national and European projects. The group of people at the workshop reflected, besides other project work, important views and goals, which are addressed in the Framework 6 Specific Targeted Research Project "Europe-China Grid InterNetworking" (FP6-2006-IST-045256-STREP), the Framework 6 Network-of-Excellence "EMANICS: European Network of Excellence for the Management of Internet Technologies and Complex Services" (FP6-2004-IST-026854-NoE), and the Framework 7 Specific Targeted Research Project "SmoothIT: Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies" (FP7-2008-ICT-216259-STREP).

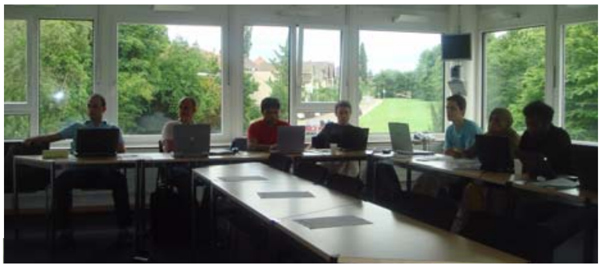
6th Framework Project FP6-2006-IST-045256-STREP
Europe-China Grid InterNetworking

6th Framework Project FP6-2004-IST-026854-NoE
Management of the Internet and Complex Services

7th Framework Project FP7-2008-ICT-216259-STREP
Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

9 Key Words

Incentives, SmoothIT Information Service, Modeling P2P Streaming Services, Congestion Control, Overlay Traffic Management, Locality Awareness, Shared Bottleneck Detection, Distributed Packet Capturing, Locality, Contracts, Business-driven Management, and Least Cost Routing.



The ETM Workshop Team during Presentations Given

10 People Involved

The team of highly motivated researchers and workshop participants are shown in the environment of the local host, the Communication Systems group CSG of the Department of Informatics IFI, at the University of Zürich UZH in the picture above and below.

The ETM Workshop Team during Presentations Given

Attendees: Thomas Bocek (UZH), Peter Racz (UZH), Tobias Hoßfeld (UniWue), Michael Welzl (UIBK), Juan Fdez-Palacios (TID), Sergios Soursos (AUEB), Murtaza Yousaf (UIBK), Cristian Morariu (UZH), Martin Waldburger (UZH), Nicolas Liebau (TUD), Amruth J. Kumar (UZH), Javier Rubio-Loyola (UPC), Miroslaw Kantor (AGH), Philipp Gschwandtner (UIBK), Humaira Ijaz (UIBK), and Burkhard Stiller (UZH).

Joint EC-GIN, EMANICS, and SmoothIT ETM Workshop on *"Economic Traffic Management"*

Welcome

Burkhard Stiller, UZH

Zurich
August 4-5, 2008

Welcome

- Joint EC-GIN, EMANICS, and SmoothIT workshop on "Economic Traffic Management"
  - Goals: To exchange technical information and discuss research topics of interest across project limits
  - Path: Technical presentations of 20 min max and following 10 min of Q&A, coffee breaks, discussion
- Brief introduction on a per-person basis:
  Who is who?
  Which project/topic are your working in/on?

Agenda Monday August 4, 2008

13.15 ***Welcome***

13.30 ***Session 1: "Incentives and P2P"***

T. Bocek: Private Shared History

P. Racz: SmoothIT Overlay Management Architecture

T. Hoßfeld: Modeling of P2P-based Video Streaming

15.00 *Coffee Break*

15.30 ***Session 2: "Congestion Control and Traffic Management"***

M. Welzl: Current IRTF/IETF Congestion Control Work and How it Relates to P2P Systems

J. Fdez-Palacios: Operator's Vision about Overlay Traffic Management

S. Soursos: Insertion of ISP-owned peers and Locality Awareness in BitTorrent

17.00 ***Discussion: Will Economic Traffic Management Mechanisms be successful?***

18.00 *End*

18.05 ***Take-off for Social Event***

Agenda Tuesday August 5, 2008

9.00 ***Session 3: "Bottleneck Detection and Distributed Capturing"***

M. M. Yousaf: Shared Network Bottleneck Detection with SVD

C. Morariu: DiCAP - An Architecture for Distributed Packet Capturing

10.00 ***Discussion: Will Economic Traffic Management Mechanisms be successful?***

10.30 *Coffee Break*

11.00 ***Session 4: "Locality Mechanisms"***

M. Waldburger: Locality and Contracts

N. Liebau, A. Kovacevic: Globase.KOM - A P2P Overlay for Fully Retrievable Location-based Search

A. Juturu Kumar: Locality Information Using BGP

12.30 *Lunch*

14.00 ***Session 5: "QoS Management and Traffic Optimization"***

J. Rubio-Loyola: Issues Confronting Business-driven QoS DiffServ Management

M. Kantor: Inter-domain Traffic Optimization in an Inter-carrier Environment

15.00 ***Closing:*** B. Stiller: Wrap-up Discussion

15.30 *End*

Workshop on "Economic Traffic Management",
August 4-5, 2008, Zurich, Switzerland

PSH: A Private and Shared History-based Incentive Mechanism

Thomas Bocek[1], Wang Kun[2], Fabio Victora Hecht[1], David Hausheer[1], and Burkhard Stiller[1,3]

[1] Department of Informatics IFI, Communication Systems Group CSG, University of Zurich
[2] Research Institute of Telecommunications Transmissions, CTTL, China
[3] associated with the D-ITET, ETH Zürich

Motivation and Background
Basic Design, PSH Algorithm
Implementation, Evaluation

Introduction

- P2P Systems have many advantages over centralized approaches
  - Scalability, load balancing, redundancy, no SPOF, ...
  - Popular applications: Skype, BitTorrent, Zattoo, Emule, ...

Problem Statement

- Problems in decentralized network
  - Free-riding, selfish behavior, malicious peers
    - Gnutella: 70% peers share no files
  - → Incentives schemes necessary!
    - Tit-for-Tat (TFT), Private Shared History Incentive Mechanism (PSH)
- TFT incentive mechanism for file-sharing (BitTorrent)
  - TFT incentive mechanism works well with symmetric interest
    - A downloads from B, B downloads from A chunks from same file
    - TFT incentive mechanism uses private history (aggregated information from local observation)
- TFT incentive mechanism enables fair sharing
  - Provide unused resources now, demand resources later

Motivation

- TFT fails for asymmetric resource interest
  - Example DHT: Peer A stores data on B, but B does not store anything on A
  - Example IPTV: Channel switching in IPTV, new channel may require building up private history

→ For asymmetric resource interest: Transitive TFT
  - Peer A consumes from B, B consumes from C, C consumes from A
  - Key question: How does C know about A and vice versa
    - Use shared history (observations from other peers)

Basic Design

- **P**rivate **S**hared **H**istory Incentive Mechanism (PSH)

- PSH combines private and shared histories

  1st step: Collect information using shared history
  2nd step: Evaluate and verify information using private history

Background

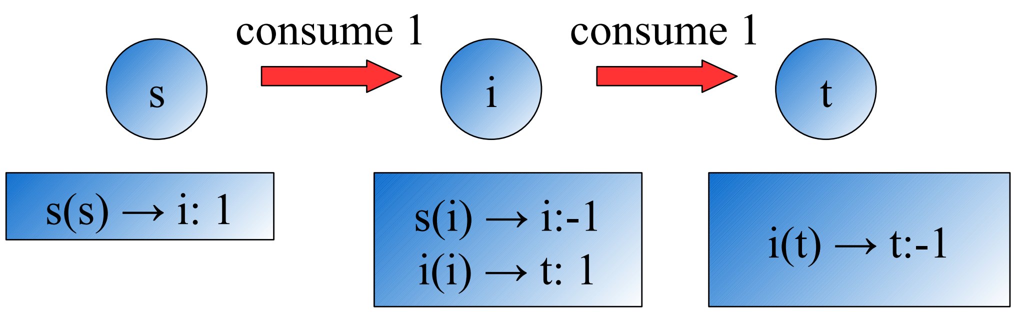
- Private history (TFT)
  - Advantages
    - Collusion resistant
    - No transaction information distribution overhead
  - Disadvantages
    - Works with symmetry of interest
- Shared history (Transitive TFT)
  - Advantages
    - Works with asymmetry of interest
  - Disadvantages
    - Not collusion resistant / not performing well
    - Transaction information distribution overhead

PSH Advantages / Disadvantages

- Private history (TFT)
  - Advantages
    - Collusion resistant
    - No transaction information distribution overhead
  - Disadvantages
    - Works with symmetry of interest
- Shared history (Transitive TFT)
  - Advantages
    - Works with asymmetry of interest
  - Disadvantages
    - Not collusion resistant / not performing well
    - Transaction information distribution overhead

PSH Algorithm (1)

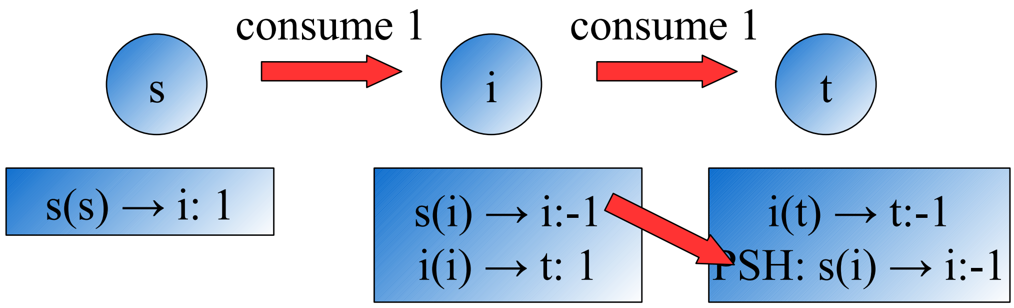
- PSH algorithm in 4 steps with examples, initial credit 1
  1. Distribute latest transaction information with every request/reply
     + no new connection required, - message size larger



Notation *v(w)→x:z* peer v consumes z resources from peer x, observed by peer w

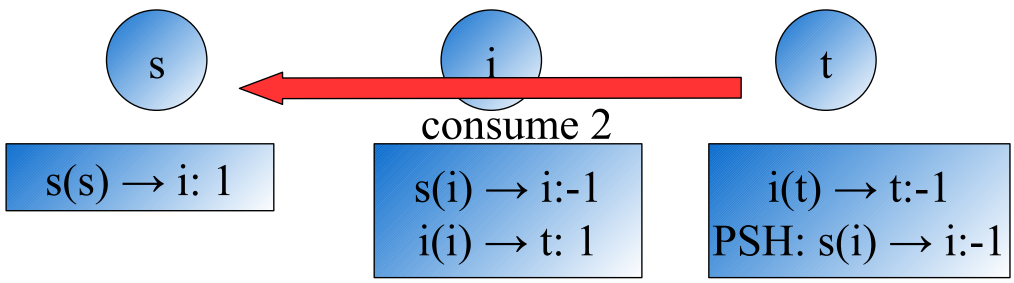
PSH Algorithm (1)

- PSH algorithm in 4 steps with examples, initial credit 1
  1. Distribute latest transaction information with every request/reply
     + no new connection required, - message size larger



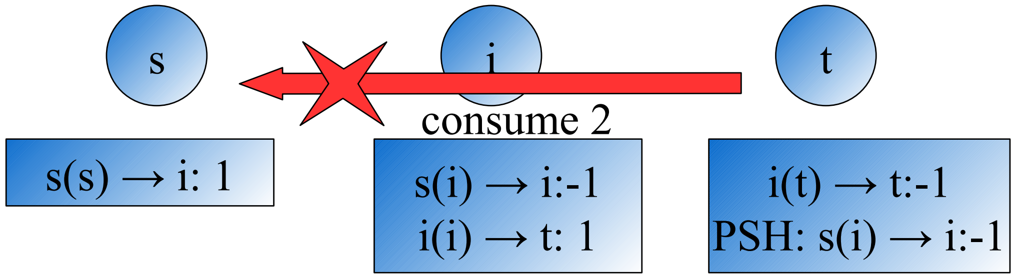
PSH Algorithm (2)

2. If TFT fails for transaction t, path of length l>2 is searched in shared history



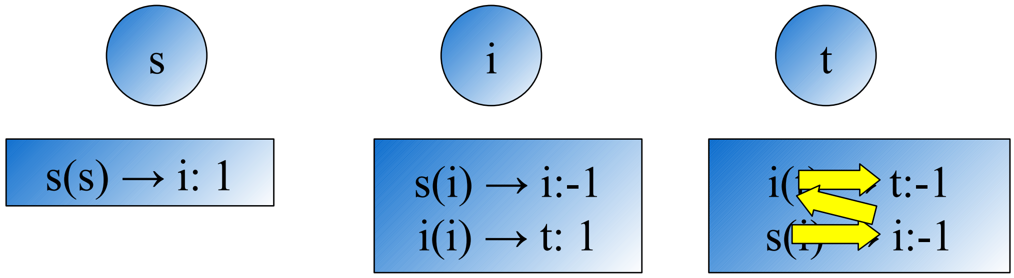
PSH Algorithm (2)

2. If TFT fails for transaction t, path of length l>2 is searched in shared history



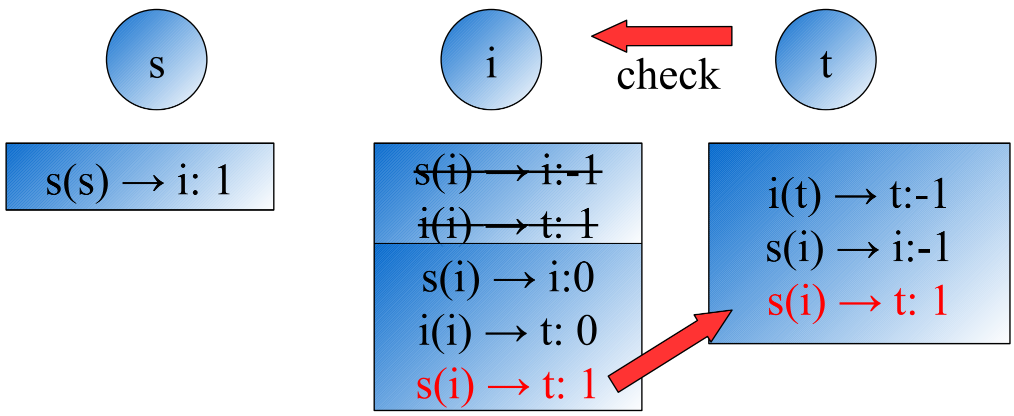
PSH Algorithm (3)

2. If TFT fails for transaction t, path of length l>2 is searched in shared history



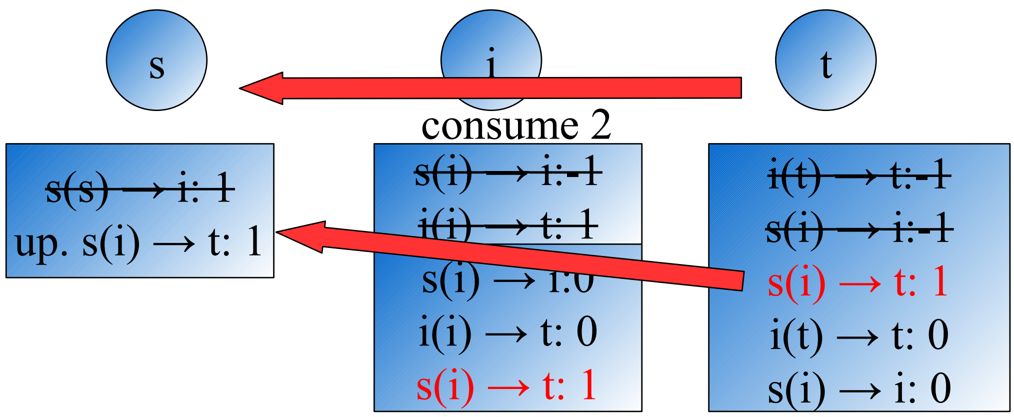
PSH Algorithm (4)

3. If path is found, issue check c
4. Apply check c on peers involved in transaction t and execute transaction t again



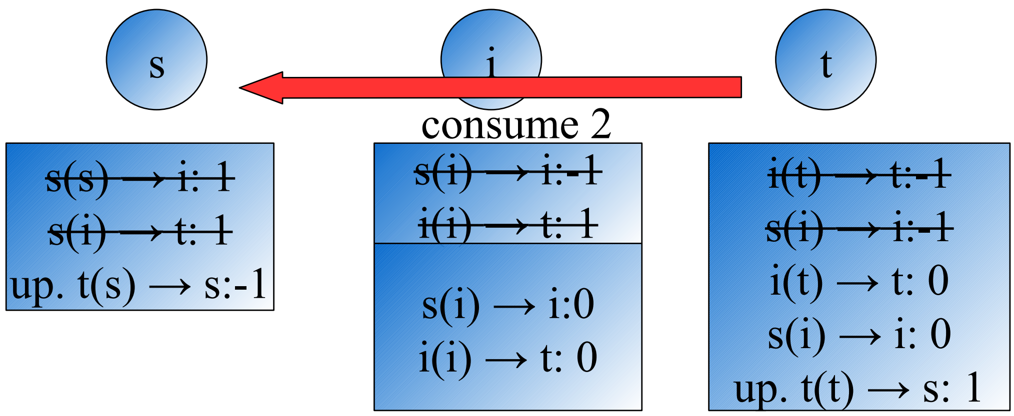
PSH Algorithm (4)

3. If path is found, issue check c
4. Apply check c on peers involved in transaction t and execute transaction t again



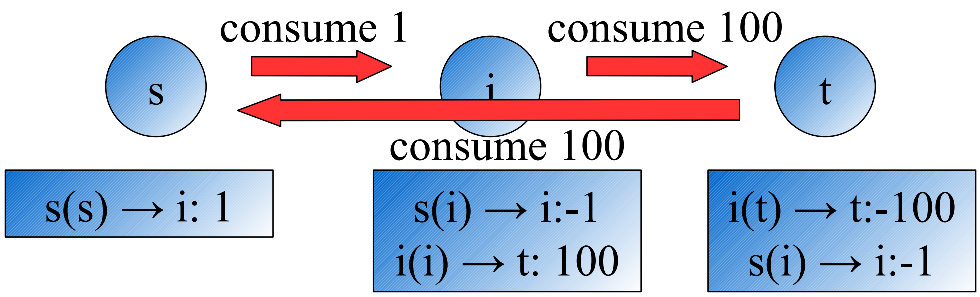
PSH Algorithm (4)

3. If path is found, issue check c
4. Apply check c on peers involved in transaction t and execute transaction t again



PSH Algorithm – Collusion-proof

- Maximum flow for path s → i → t is 1



- Missing updates/wrong checks/outdated information/ malicious peers
  - Peer s decides based on its private history
  - Balance on peer i remains the same
  - For peer t: amount on check may be lost

PSH Algorithm - PSH_r

- PSH sends with each request latest transaction information
- PSH_r reduces distribution overhead
  - No distribution of transaction information beforehand
  - If transaction fails, transaction information exchanged
    - + less overhead
    - + smaller messages
    - - only paths of length 3 can be found reliably (one intermediate node)

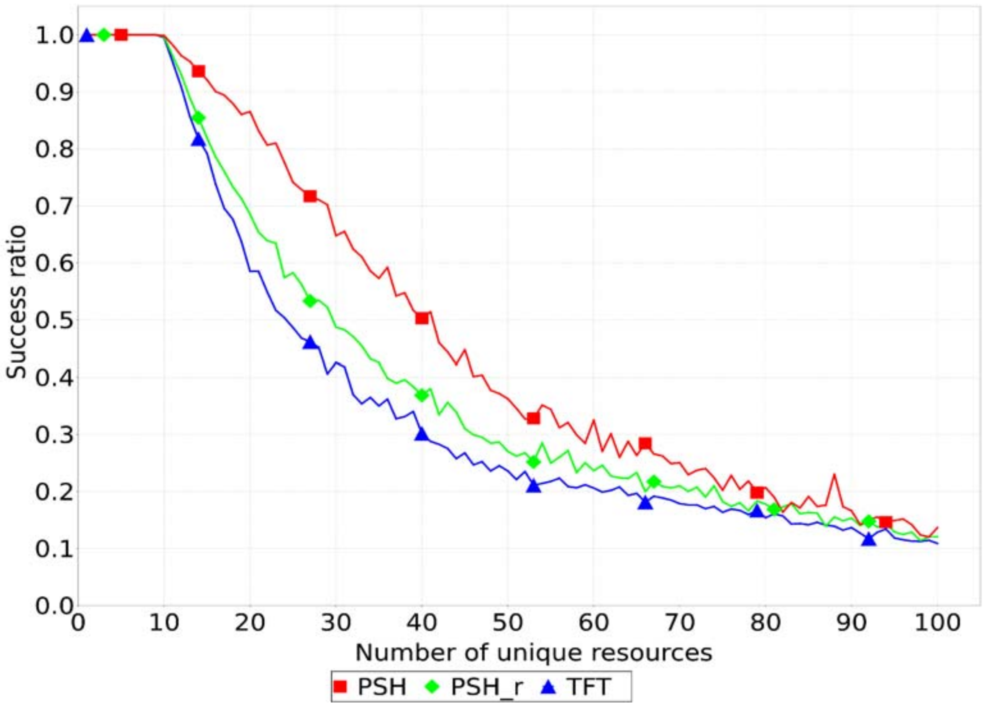
Implementation

- PSH implemented in Java
- PSH uses a TomP2P (DMM, XOR metric)
- Evaluated with n peers, n=[1..N], N=100
- Experiments run on one machine, operation executed sequentially
- Compared mechanisms:
  - PSH (max. 3 retries)
  - PSH_r (no retries)
  - TFT (no retries)

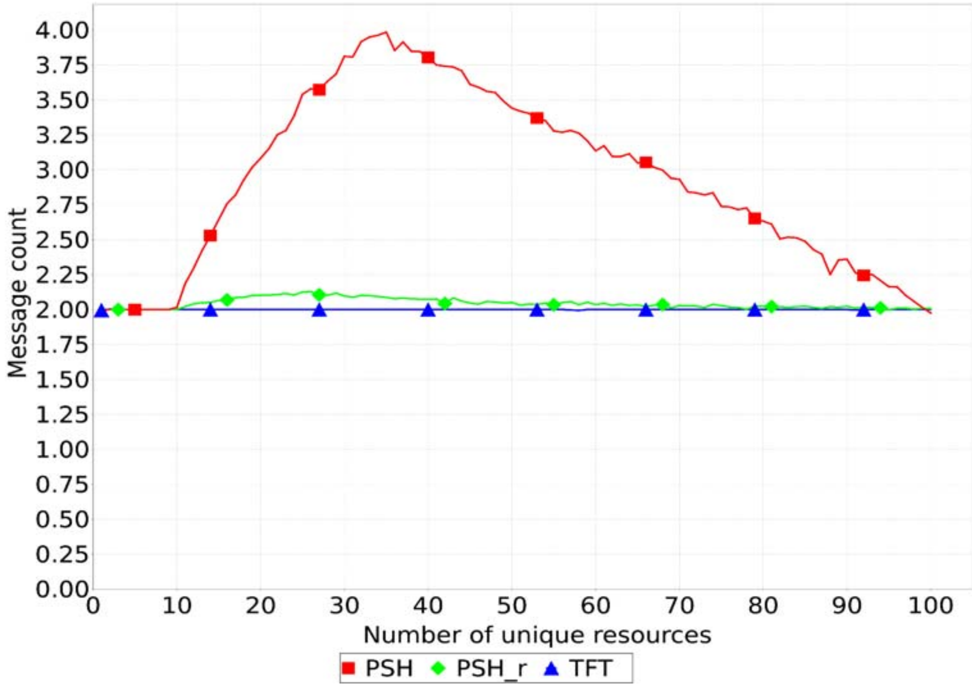
Evaluation (1)

- Parameter "number of unique resources" varied, p=[1..N], N/p peers share the same resource
  - p=1 → all peers shares the same resource (symmetric)
  - p=100 → all peers share different resource (asymmetric)
  - p=50 → 2 peers share same resource
- Plotted against:
  - Success ratio
    - With perfect knowledge success ratio is 1.0
  - Average message count
    - Number of messages per transaction
  - Message size
    - Total size for all messages

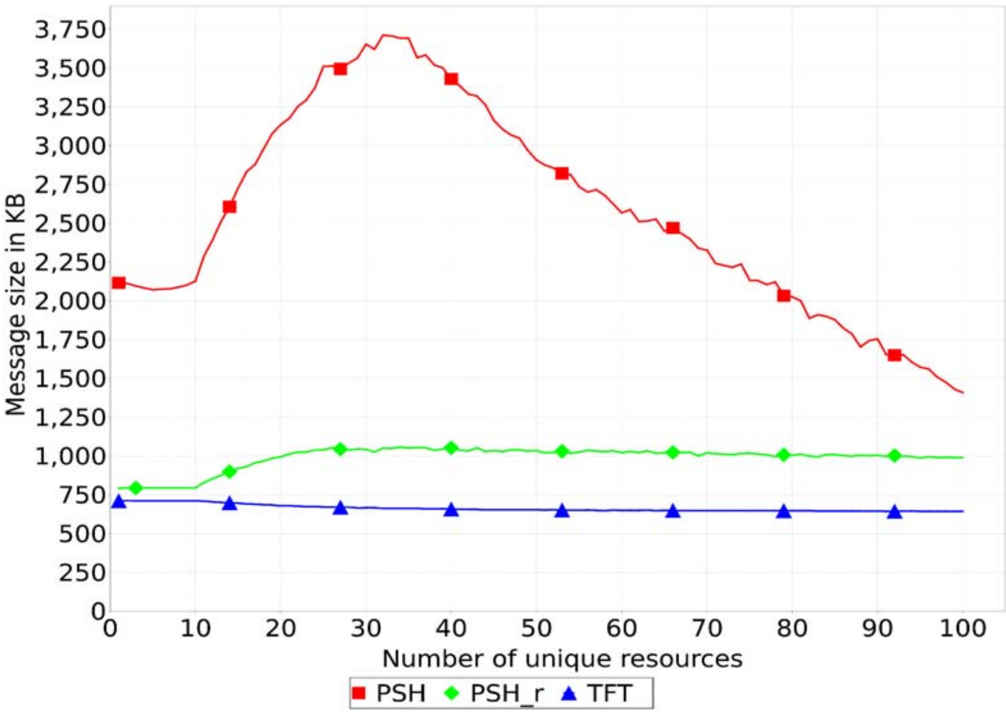
Evaluation (2) – Success Ratio



Evaluation (3) – Count



Evaluation (4) – Size



Summary and Conclusions

- PSH combines private and shared history
- PSH works up to 73% better than TFT incentive mechanism for asymmetric resource interest
- PSH is collusion resistant
- Distribution of transaction information introduces overhead

Future Work

- PSH2: Reduce message size, based on PSH_r, uses bloom filters to find intermediate node → supports only one intermediate node
- Distributed testing environments
  - PSH2 deployment on EMANICSLab
- Possible PSH2 applications
  - SmoothIT, P2P TV, demo application (channel zapping)
  - EMANICS, P2P collaboration, P2P voting
  - EC-GIN, PSH2 as basis for trust mechanism

Questions?

Thank you for your attention.

Reference:
T. Bocek, W. Kun, F. V. Hecht, D. Hausheer, B. Stiller: *PSH: A Private and Shared History based Incentive Mechanism.* 2nd International Conference on Autonomous Infrastructure, Management and Security Resilient Networks and Services (AIMS), July 2008

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

European Seventh Framework STREP FP7-2007-ICT-216259

SmoothIT Overlay Management Architecture

UZH, DoCoMo, TUD, AUEB, PrimeTel, AGH, ICOM, UniWue, TID

Peter Racz, UZH

ETM Workshop
Zurich
August 4-5, 2008

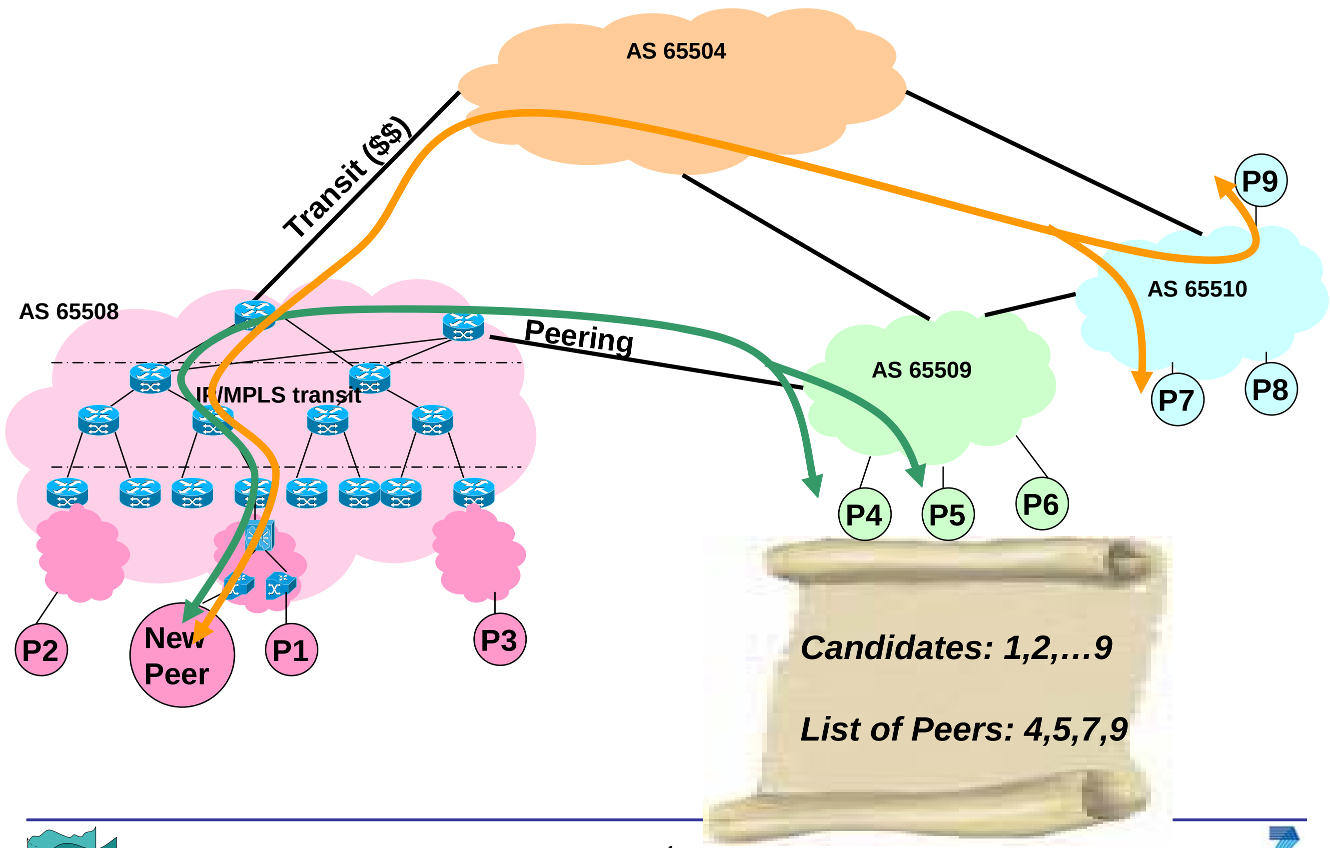
Outline

- Motivation and Example Scenario
- SmoothIT Objectives
- Incentives to Participate in ETM
- Solution Concepts
- Key Requirements
- SmoothIT Information Service
  - Architecture
  - Protocol
- Summary

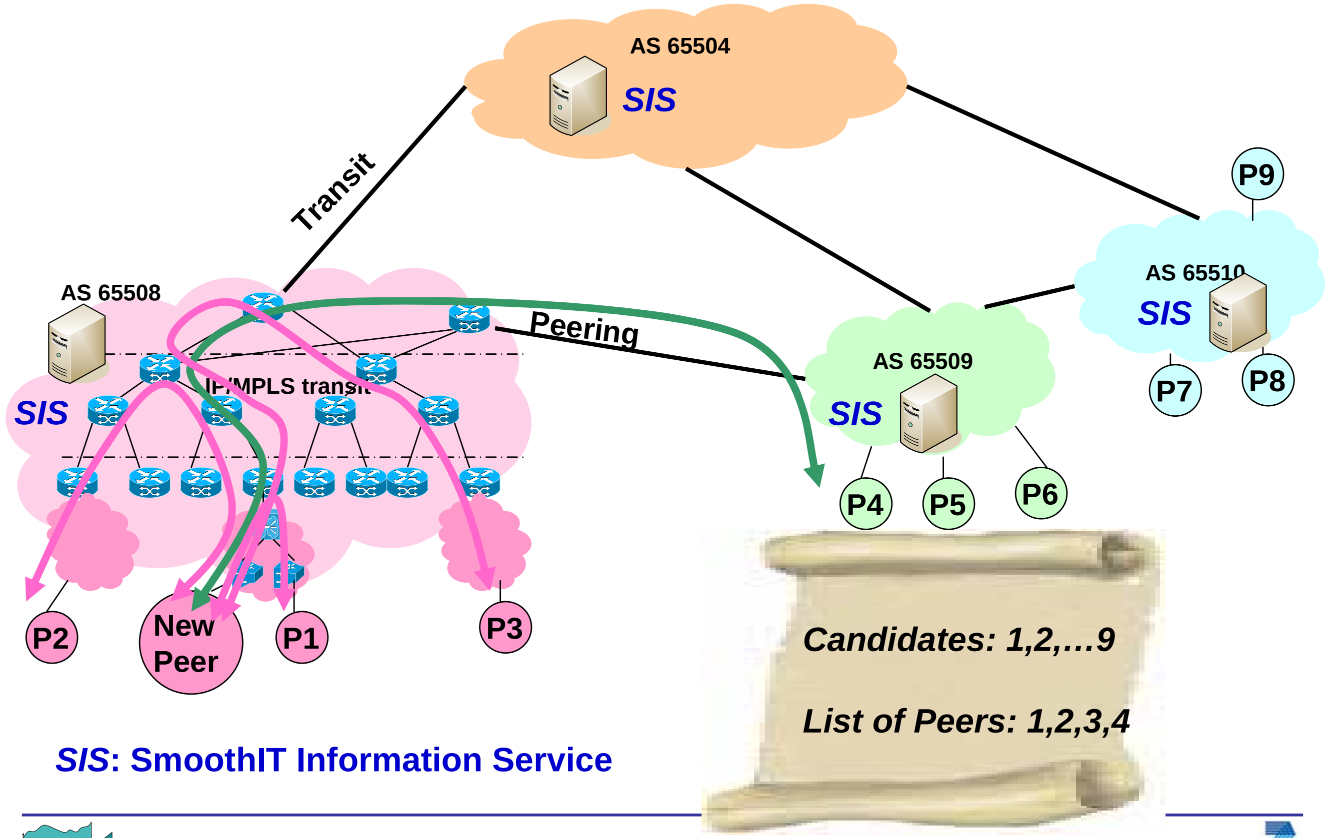
Motivation

- P2P applications and traffic
  - Significant and increasing amount of P2P traffic
  - Suboptimal peer selection due to information asymmetry
    - Underlay *topology*, incl. *routing metrics and values*, unknown to overlay
    - Overlay *requirements,* incl. traffic characteristics, unknown to underlay
- Consequence
  - Non-optimized overlay traffic in the underlay
    - Higher costs in underlay
    - Lower QoS in overlay
  - Conventional traffic management techniques not suitable
- Goal of the SmoothIT project
  - Bridge overlay with underlay
  - Apply Economic Traffic Management (ETM)
  - Optimize traffic and achieve win-win situation for all parties

Example: Locality-unaware Overlay



Example: Locality-aware Overlay



***SIS*: SmoothIT Information Service**

SmoothIT Objectives

- **Structure Internet-based overlay networks** to be efficient and optimal for **users, overlay providers**, and **ISPs** **leading to the "*win-win-win" (triple win)* situation**
  - Investigate, design, and apply specialized **economic theory** for **decentralized network-efficient Internet-based overlay services in multi-domain scenarios,** including wireless access
  - **Develop** an optimized **incentive-driven signaling approach** for defining (theory) and delivering (technology) economic signals in support of cooperating and competing providers

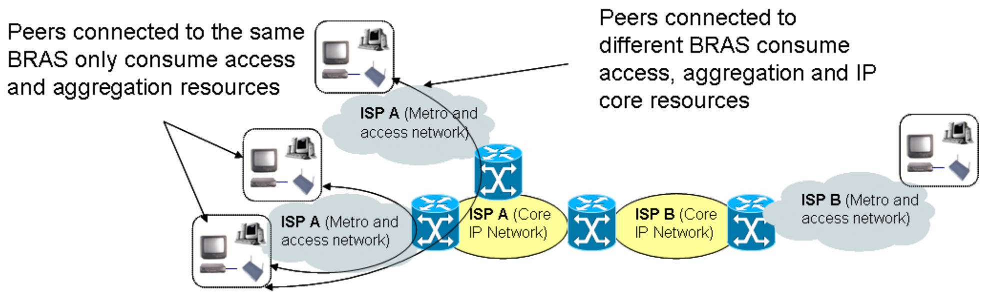
- **Operator-orientation:** demonstrating key results through a strong focus on **ISP and telecom requirements** (*e.g.*, NN)
- **Implementation-orientation**: design, prototype, and validate the **networking infrastructure** (real-life test-bed)

Triple Win in Detail

- Management of **overlay networks** based on a collaboration between the overlay provider and the network (underlay) provider in support of the user (information is the key)
  – Cost and investment recovery for operators
- Incentives for operators
  – Reduce overlay traffic and inter-domain traffic, reduce costs
  – Keep overlay services (boost flat rate tariffs; keep customers)
  – Avoid to be on an overlay block list and “make money” with transport `http://www.azureuswiki.com/index.php/Bad_ISPs`
- Incentives for overlay providers
  – Active role in traffic management increases service quality
  – Increased user base due to better performing services
- Incentives for user
  – Increased service quality, *e.g.*, in terms of reliability, RTT, bandwidth

Solution Concepts

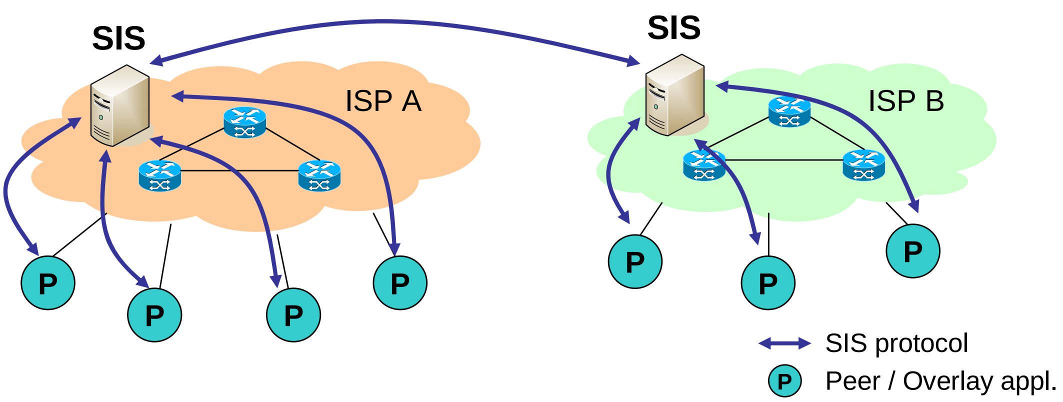
- Agreements between overlay provider and operator
  – *E.g.*, active caching:
  the operator provides explicit local caches for overlay content
- Locality promotion
  – Operator provides information about how to achieve best quality in overlay, *e.g.*, operator prioritizes alternative peer interconnections
- QoS/QoE differentiation (application-awareness)
  – Operator knows overlay application traffic (labels, deep packet inspection) and applies application-aware traffic management



Key Requirements

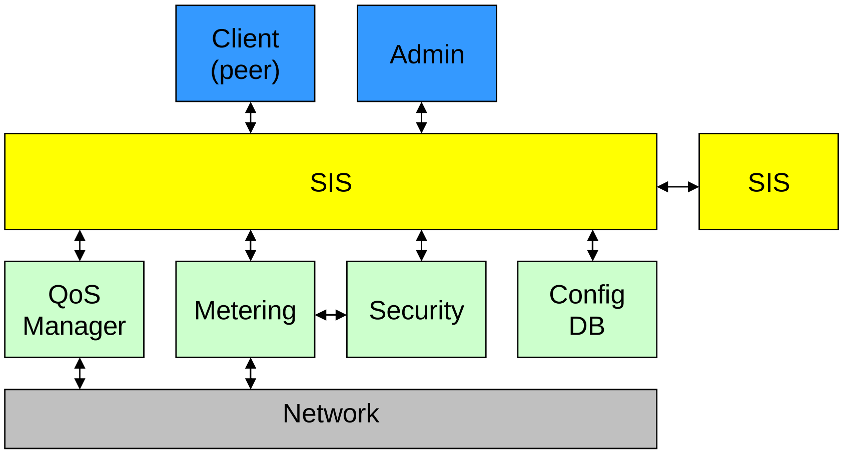
- Incentive-compatibility and traffic optimization
  - Provide incentives for all parties to achieve triple win
- Support of different overlay applications over a common interface
- Interface supporting various optimization schemes
  - E.g., different kinds of applications, high throughput, low delay, free/premium service
- Inter-domain support
- QoS support for network services
- Mobile network support
  - Node mobility, shared medium, heterogeneity of node and link capacities
- Easy deployment
  - In overlay applications and in ISPs' networks
- Extensibility
  - New applications, new metrics
- Scalability, efficiency, and robustness
- Security
- Standard compliance

SmoothIT Information Service (SIS)



- Deployment of SIS components in the ISPs' network
  - To convey information between overlay and underlay
- Client-Server architecture
- Overlay applications interact with SIS in order to select „better" peers
  - Reducing costs of ISPs
  - Improving QoE of users

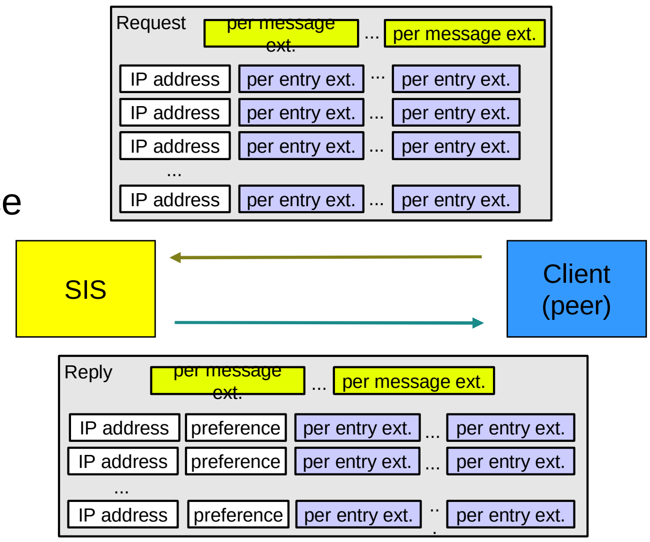
SIS Architecture



- SIS
  - Contains ETM logic
  - Aggregates information and calculates preference values
- Metering
  - Collects information from the network, e.g., BGP routing, topology
- QoS Manager
  - Performs QoS provisioning
  - Support of QoE schemes
- Security
  - Authentication and authorization
- Config DB
  - Stores various information about the network, e.g., topology, capacity

SIS Protocol

- Between SIS and overlay appl.
- Stateless request-response interaction scheme
- Application-independent
- Basic preference information service
  - Request: list of identifiers/peers (IP addresses)
  - Reply: list with preference values
- Optional further parameters
  - Per message or per parameter e.g., application type, desired QoS, capacity, locality, pricing information, peer availability



Possible Implementation

- Protocol selection criteria
  – Platform independence
  – Language independence
  – Standard compliance
  – Maturity and stability
  – Availability of parser libraries in different languages
  – Simplicity, ease of use for overlay applications
  – Efficiency, low protocol overhead
  – Human-readability
  – Formal and verifiable definition of message format
- Candidates
  – Web Service (SOAP + XML)
  – REST
  – JSON
  – YAML
  – Google protocol buffers
  – HTTP-based custom protocol

Summary

- Management of overlay traffic is necessary
  – Large amount of traffic
  – High costs for ISPs
- SmoothIT Information Service
  – Deployed in the network of ISPs
  – Provide information to overlay applications
  – Optimize traffic and achieve win-win situation

Thank you for your attention!

Thanks to SmoothIT's project partners:

UZH, DOCOMO, TUD, AUEB, PrimeTel, AGH, ICOM, UniWue, TID

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

European Seventh Framework STREP FP7-2007-ICT-216259

Modeling of P2P-based Video Streaming

UZH, DoCoMo, TUD, AUEB, PrimeTel, AGH, ICOM, UniWue, TID

Tobias Hossfeld, UniWue

Workshop on ETM, Zürich
August 4-5, 2008

P2P-based VoD: Joost Measurements

- Test PC in Würzburg: mainly connections to peers in Europe
- P2P downloads video from peers world-wide and resembles video stream out of received pieces: **multi source download**

Why is ETM suitable for P2P?

- Conventional traffic management not sufficient
  - peers are “wildly” connected in overlay → e2e links over several domains, independent of underlay network → different providers of sub-networks → e2e-TM difficult
  - many connections to provide one service, e.g. multi-source download for video streaming → QoE is composed of QoS of individual flows
  - dynamics of P2P systems, changing network topologies
- Economic traffic management

  - provide incentives for win^x for x players to do traffic management; e.g. overlay selects peers, underlay provides information, …

  ➔ traffic optimized inherently

What is necessary for ETM and its performance evaluation ?

- Characterization of P2P application
  - measurements to understand system behavior and how to influence it
  - quantification of QoS disturbances on QoE

  → helps to derive mechanisms for economic traffic management / optimization potential
- Modeling of P2P application
  - required for performance evaluation via emulation, simulation, analysis
  - determine appropriate abstraction level to evaluate desired goals:
    - study win^x when using ETM → quantify traffic optimization
    - performance loss when not using ETM → demonstrate incentive

Agenda

- Characterization of P2P video streaming
  - overlay topology
  - observed traffic characteristics
  - edge-based intelligence
  - Quality of Experience

- Modeling of P2P video-on-demand
  - degree of abstraction
  - relevant overlay mechanisms
  - SmoothIT simulation

P2P Live TV: Zattoo

- license agreements, user groups according to content (channel, country regulation, language)
- cluster may be taken into account by ETM

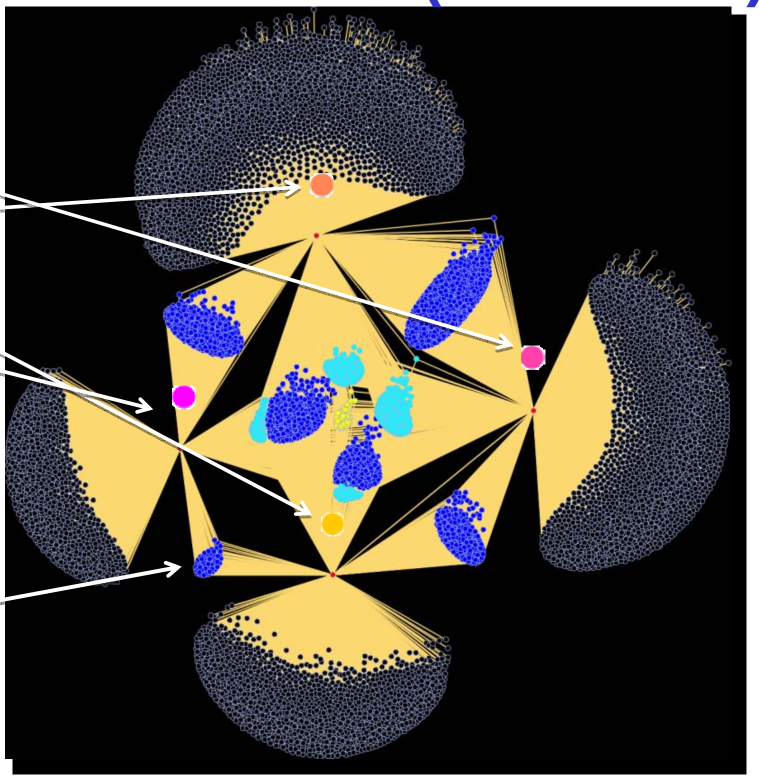
SopCast: Spain vs. Russia (Semi-Final)

Test machines, #connections

- Bartholomeus, 4267
- Castor, 5429
- Arminius, 2783
- Home, 2145

Peers known by *n* test machines

- 4
- 3
- 2
- 1

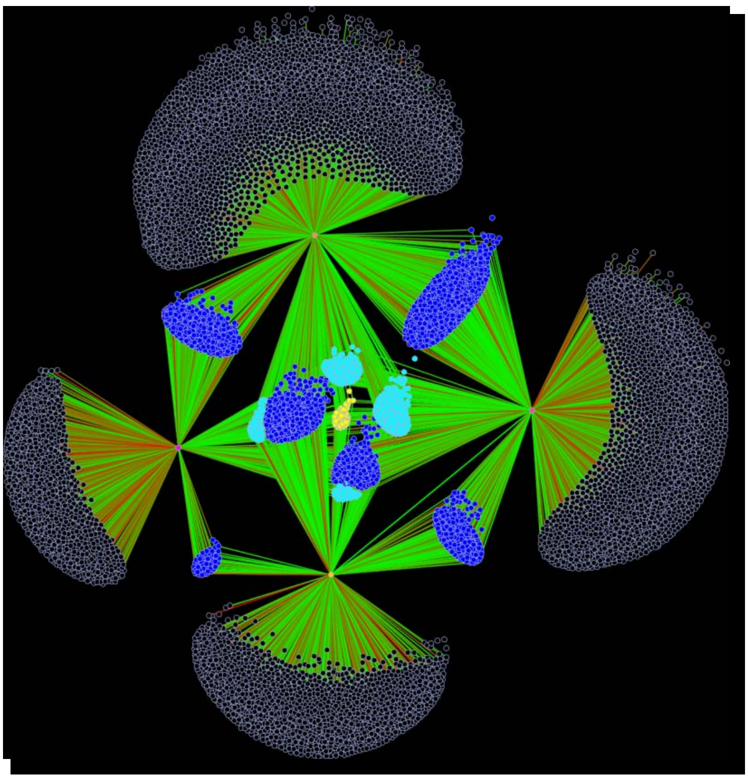


not yet published, paper in preparation by SmoothIT partners UniWue and UZH

End-to-End Delay between Peers

Ping (in ms):

0.0
500.0
1000.0
1500.0
2000.0
2500.0
3000.0
3500.0
4000.0

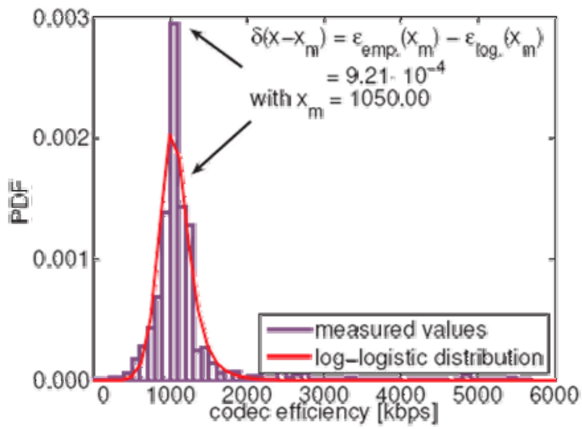


Agenda

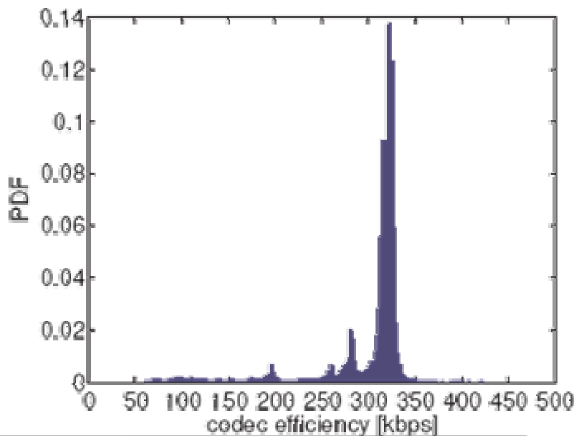
- Characterization of P2P video streaming
  - overlay topology
  - observed traffic characteristics
  - edge-based intelligence
  - Quality of Experience
- Modeling of P2P video-on-demand
  - degree of abstraction
  - relevant overlay mechanisms
  - SmoothIT simulation

Video Files and Streams

- OnlineTVRecorder.com



- YouTube video streams



		mean	std	CoV	skewness	kurtosis	median	min	max
OTR 11563 samples	duration [min]	[illegible]	[illegible]	[illegible]	[illegible]	.42	45	1	195
	size [MB]	[illegible]	[illegible]	[illegible]	[illegible]	.31	305.87	0.06	1236.87
	effici[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.05	1038.42	0.71	16310
YouTube 21014 samples	duration [s]	[illegible]	[illegible]	[illegible]	[illegible]	.64	252	5	10233
	size [MB]	[illegible]	[illegible]	[illegible]	[illegible]	.25	9.41	0.07	274.59
	efficiency [kbps	[illegible]	[illegible]	[illegible]	[illegible]	.81	318.54	1.12	1040.52

detailed models for video contents

A Qualitative Measurement Survey on Popular Internet –based IPTV Systems
HUT-ICCE, June 2008

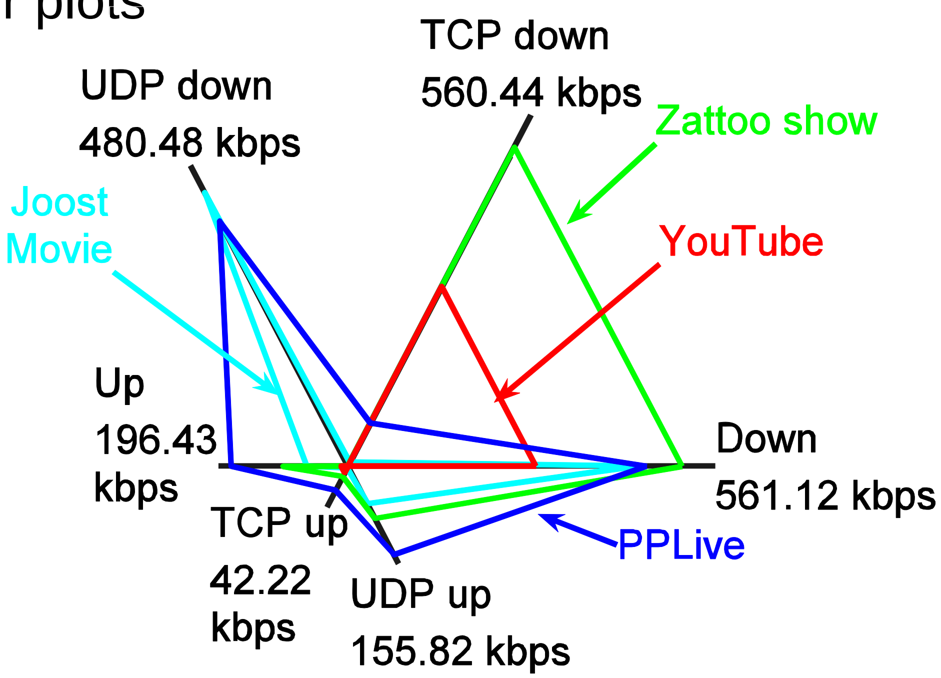
Traffic Characteristics

	TCP	UDP	up	down	TCP(up)	TCP(down)	UDP(up)	UDP(down)
Joost (short)	86.12	391.01	91.46	356.14	9.08	77.04	88.99	302.15
Joost Movie	9.85	546.75	69.24	487.08	3.02	6.82	66.32	480.48
Joost (Japan)	9.07	522.78	12.88	516.33	1.06	8.01	11.84	508.77
Zattoo (short)	285.15	104.27	28.68	359.51	11.44	273.71	19.06	86.83
Zattoo show	578.11	92.18	108.85	561.12	17.67	560.44	91.47	0.68
PPLive (short)	209.29	479.17	94.26	582.08	44.68	164.63	50.90	428.44
PPLive	117.95	586.80	196.43	502.29	42.22	75.75	155.82	430.95
PPLive (Japan)	159.73	547.79	196.69	509.95	42.15	117.59	154.58	392.73
YouTube	326.63	0.21	11.19	315.49	11.17	315.47	0.02	0.02

- Lot of numbers …
- … different kind of visualization !

Comparison of Traffic Characteristics

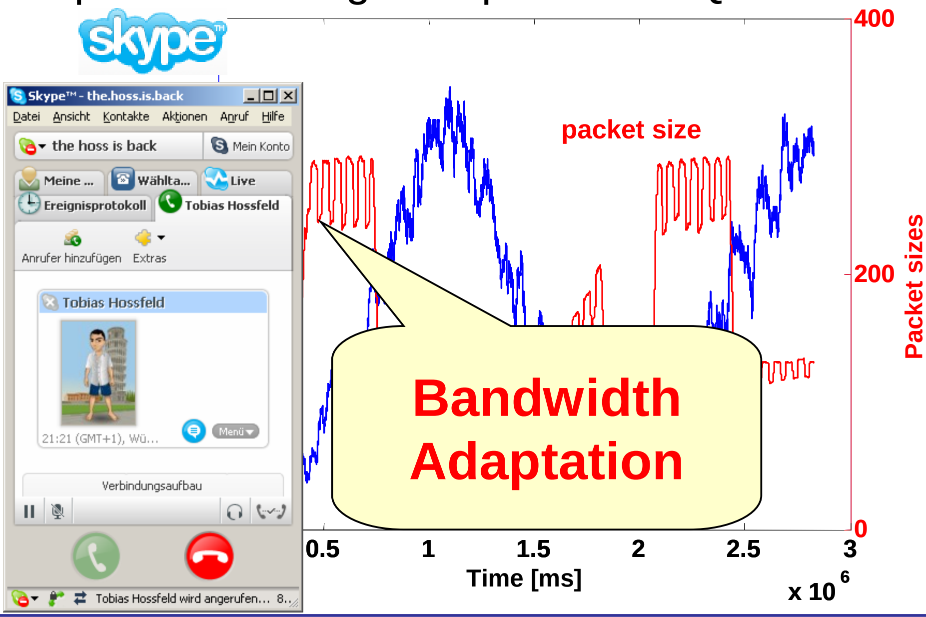
- Application classification via observed shapes in spider plots



Edge-based Intelligence: Bandwidth Adaptation

- Emulate network conditions to see system‘s reaction
- *Example:* Skype and time-varying packet loss
- Packet size adapted according to experienced QoS

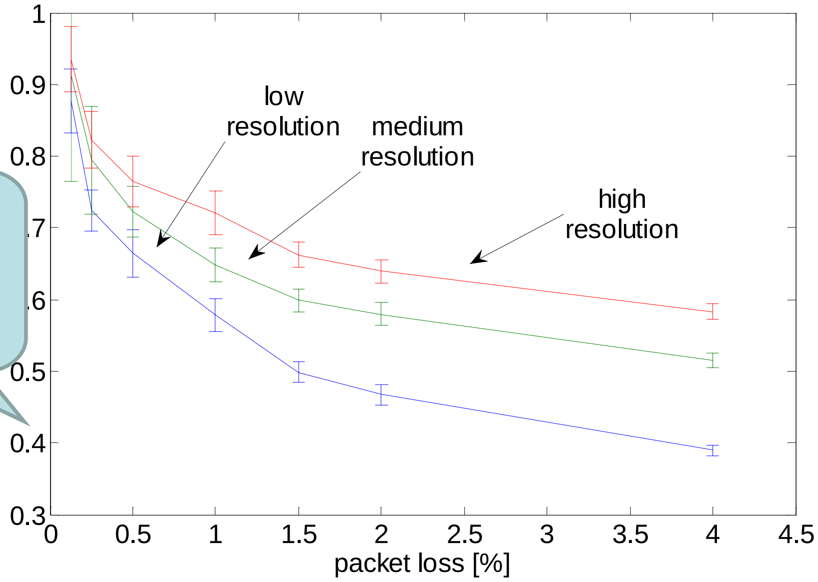
work in progress for video streaming



Quality of Experience for Video Streaming

- determines real user behavior
- quantification of QoS disturbances on QoE
- non-[illegible] ıctors

work in progress for video streaming



Agenda

- Characterization of P2P video streaming
  - overlay topology
  - observed traffic characteristics
  - edge-based intelligence
  - Quality of Experience
- Modeling of P2P video-on-demand
  - degree of abstraction
  - relevant overlay mechanisms
  - SmoothIT simulation

Performance Evaluation in SmoothIt

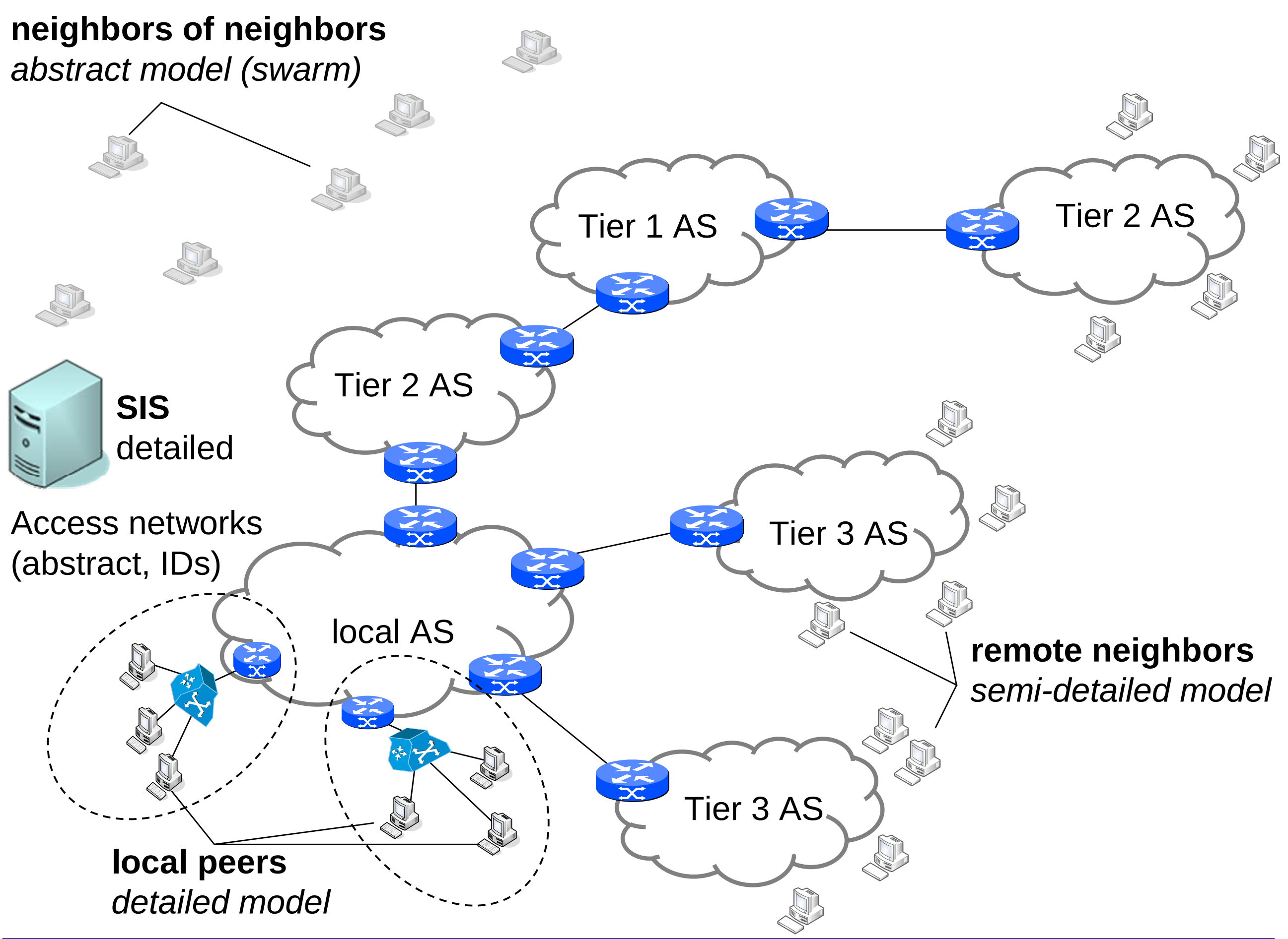
- Simulation
  - evaluate ETM concept and TripleWin situation
  - show benefit using SIS / drawback without SIS
  - scalability in large-scale environment
  - (needs realistic input values)
- Internal trial
  - feasibility of technical solution
  - quantify QoS (and even QoE)
  - (needs traffic characterization for emulation)
- External trial
  - test user incentives, real user behavior

Divide and Conquer

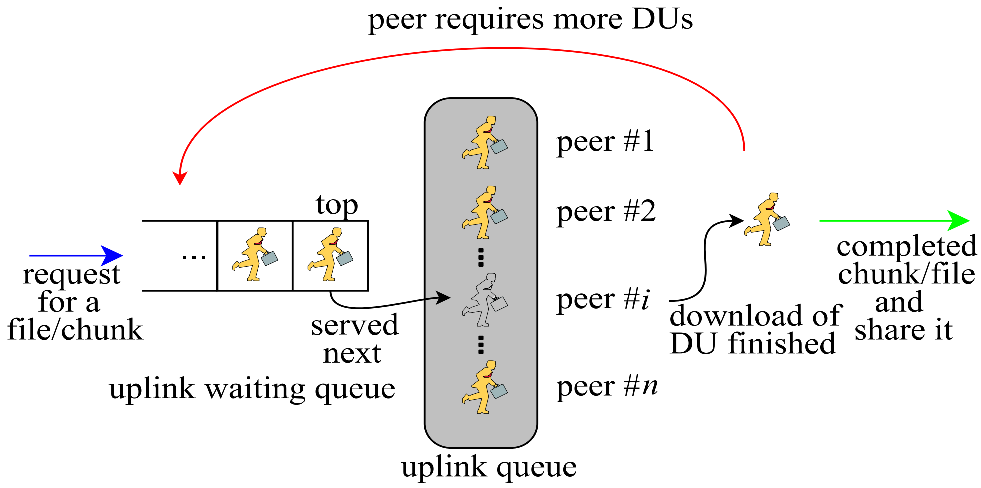
- Separate investigations if possible
- QoE is approximated by QoS parameters (WP1)
- Dimensioning of SIS (analytical): architecture discussion (WP3), theory and modeling (WP2)
- feasibility of SIS, response times etc.: engineering (WP3) and operation of internal trial (WP4)

Simulation Model for Video Streaming

- Key metrics for TripleWin: costs for ISPs, quality for users, performance of overlay
- Determine abstraction level
  - relevant mechanisms (i.e. ETM and overlay application, e.g. key functionality of SIS or modified chunk/peer selection strategies)
  - large number of peers, underlying ISP topology
  - user behavior and capabilities (e.g. churn) and reaction of players on incentives
- Detailed simulation of peers and SIS within ISP
- Approximate external peers using measurements
- Models for background traffic

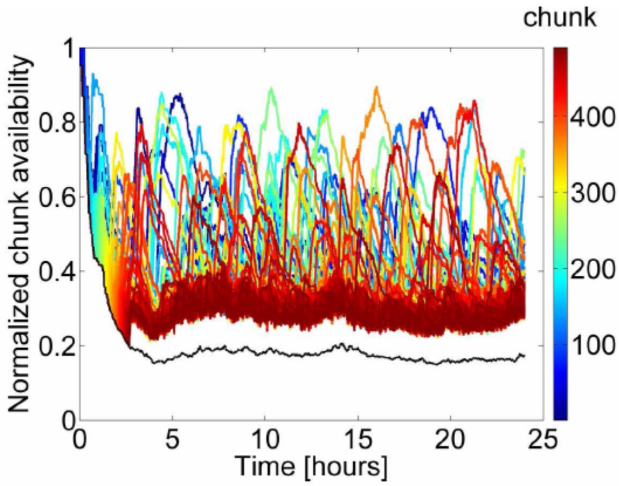


Peer and Chunk Selection @ Local Peers

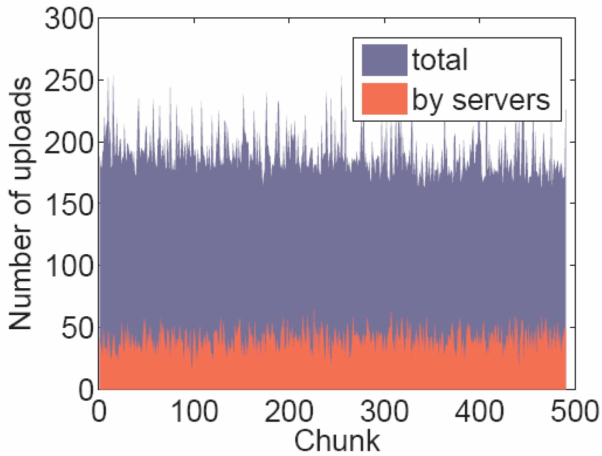


- Chunk selection of downloading peer
  - for video, not yet downloaded chunks which can be played back (deadline)
- Peer selection of uploading peer
  - based on incentives, tit-for-tat not useful for video → give-to-get

Least-Shared-First for Video Streaming



(a) Chunk availability

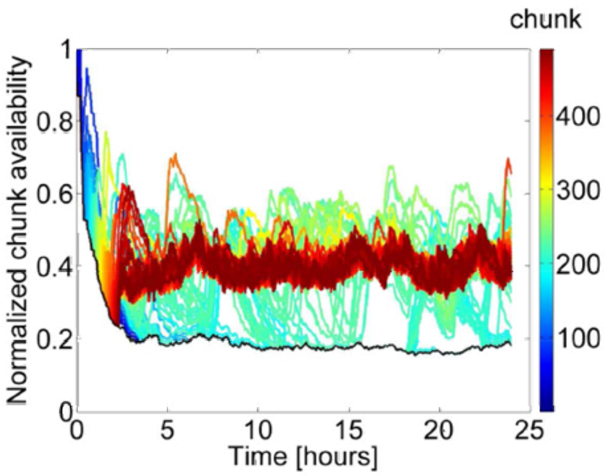


(b) Chunk uploads

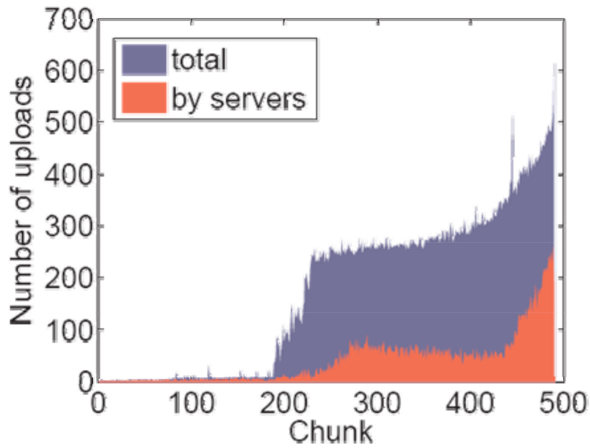
- deadlines of blocks not considered → bad QoE
- adapt peer and chunk selection strategies

T. Hoßfeld, S. Oechsner, F. Lehrieder,
C. Bergner, P. Tran-Gia
From File Sharing to Video Streaming: Impact of
Chunk Selection Strategies
submitted, July 2008

Taking into Account Deadlines



(a) Chunk availability



(b) Chunk uploads

- Strategies required beyond state of the art
  - peer selection, chunk selection, incentives
  - support by **economic traffic management**

Simulation Platform in SmoothIT

- Large scale environment requires simulation on flow level, abstraction of background traffic, …
- Parts of simulation is implemented by different partners → well documented API
- Implementation of SIS and sophisticated P2P mechanisms from the scratch, using existing knowledge on BitTorrent simulations
- Possibility for re-utilization of simulation code for internal trial

- ➔ Taskforce "Simulation" decides for ProtoPeer

Conclusions

- Characterization of P2P video streaming
  - to understand complex system dynamics and behavior
  - to get realistic values for models
  - allows to derive new solutions

- Modeling of P2P video streaming
  - needs appropriate degree of abstraction
  - to evaluate desired performance questions

- In SmoothIT, we have to take care of
  - overlay mechanism, key functionality of SIS, locality of users
  - to evaluate TripleWin

Current IRTF/IETF Congestion Control Work and How it Relates to P2P Systems

Michael Welzl http://www.welzl.at
Institute of Computer Science
University of Innsbruck, Austria

Economic Traff. Mgmt Workshop
University of Zurich, Switzerland
4 August 2008

Outline

- IRTF ICCRG
- IETF Transport Area
- P2P app considerations (TANA)

IRTF Internet Congestion Control Research Group (ICCRG)

- Standard TCP shows its limits (high bandwidth, long delay, etc)
  - Yet, flows should be TCP-friendly (not send more than TCP would)
  - HighSpeed TCP (RFC 3649): be more aggressive than standard TCP in high bandwidth environments with little loss only
  - Previously, either TCP-friendly or better-than-TCP; no combinations!
  - Now, plenty of proposals: CUBIC, BIC, FAST, Compound TCP, HTCP, XCP...
  - Can we agree on one of them?

- Note: scope of ICCRG is wider
  - Interactions with QoS mechanisms, traffic engineering, lower-layer technologies such as optical-burst-switching
  - DoS attacks vs. mechanisms against them vs. congestion control
  - How to be fair, how to define fairness
  - Open issues: currently being collected in a draft

Deployment of high speed TCPs

- High-speed TCP proposals have been on the table for quite a while
  - IETF did nothing: conservative about changing TCP
  - So people started using experimental mechanisms themselves

- Many mechanisms have long been available in Linux (pluggable CC)
  - pluggable CC now also available in FreeBSD

- After major press release (Slashdot: “BIC-TCP 6000 times quicker than DSL“), BIC became default TCP CC. in Linux in mid-2004
  - Now replaced with CUBIC

- Compound-TCP (CTCP) = default TCP CC. in Windows Vista Beta
  - For testing purposes; disabled by default in standard release

- Will this lead to an arms race?

The role of the IRTF / IETF

- The IETF wants interoperable mechanisms, specified in RFCs
  - so, authors of TCP proposals should be asked to specify their mechanisms
- Process devised: proposals will be pre-evaluated by IRTF (ICCRG)
  - Evaluation guidelines: RFC 5033, Transport Models Research Group (TMRG)
  - CTCP and CUBIC proposals currently on the table (October 2007)
  - See: http://www.irtf.org/charter?gtype=rg&group=iccrg for more details
- Procedure
  1. Write a draft
  2. Get reviews in the IRTF ICCRG; reviewers should check:
     - Does the proposal have a conflict with draft-floyd-tsvwg-cc-alt?
     - Were the TMRG metrics used in performance evaluations?
  3. Then go to the IETF, where reviews should be taken into account
- Currently happening for CTCP, CUBIC, HTCP

Some open issues of interest to ICCRG

- Reaction to corruption (*DCCP spec asking*)
  - Note: corruption and congestion can be heavily correlated on short time-scales, and links can have strange properties (e.g. HSDPA, 802.11B)
- TCP over IETF mobility / ad hoc protocols
  - Can we show that the problem space is equal to another one, e.g. load changing on a single path?
- Evaluation of (implicit and explicit) feedback signals
  - Interactions with QoS, Traffic Engineering (real-time), IPSec, lower layers, congestion = f(bytes or packets?)
- Pseudowires
  - E.g., some consume bandwidth independent of the payload *(Pseudowire WG charter mentions CC, but drafts and RFCs restrict use to dedicated paths because proper CC unknown)*

Some open issues of interest to ICCRG /2

- Precedence for elastic traffic (related to MLPP docs)
- Misbehavior of senders and receivers *(TCPM discussions)*, Denial-of-Service
- What is effective for media streams *(RTP profiles)*
- UDP based application layer protocols (*IRIS, SYSLOG – Sally Floyd's congestion control recommendation RFC is too unspecific for these groups*)
- Congestion control at the application layer (*SIP overload, ETSI GOCAP*)

ICCRG Slow Start Design Team

- Slow Start not as well founded in theory as AIMD part of TCP
  - Slow Start Design Team looks at issues and possible improvements
  - Led by Dirceu Cavendish
- Issues
  1. High-speed networks: overshoot can be quite large
     - fixes: increase by less than 1 when cwnd is large
       - RFC: Limited Slow Start
       - Proposal: Conservative Slow Start
  2. Initial phase may be too conservative
     - bad because many connections spend their whole life in SS
     - Why not increase cwnd by 2 for every ACK when cwnd is small? why not by 3?)
     - SS after periods of quiescence: sender may have some idea about the path in use
       - Proposal for faster-restart in TFRC (DCCP CCID)

IETF: Transport Area WGs related to CC

- DCCP Datagram Congestion Control Protocol
  - protocol for unreliable yet congestion controlled data transfer
  - “framework“ for CC mechanisms
- pcn Congestion and Pre-Congestion Notification
  - making DiffServ more dynamic
- rmt Reliable Multicast Transport
  - multicast CC
- tcpm TCP Maintenance and Minor Extensions
  - the name says it all :-)
- tsvwg Transport Area Working Group
  - misc (e.g. SCTP)

TANA BoF - IETF-72, Dublin, Ireland

- TANA = Techniques for Advanced Networking Applications
  - Evolved from workshop on P2P infrastructure (P2PI)
- “TANA is a transport-area BoF that will focus on broadly applicable techniques that allow large amounts of data to be consistently transmitted without substantially affecting the delays experienced by other users and applications.“
  - Main concern: P2P apps uploading over thin home uplinks
- BoF explored the following potential work items:
  - A cc. algorithm for less-than-best-effort "background" transmissions
  - A document that clarifies the current practices of app design and reasons behind them and discusses the tradeoffs surrounding the use of many concurrent transport connections to one peer and/or to different peers

TANA technical challenges

- Less-than best-effort service: solutions exist
  - end-to-end (seems to be primary interest here for TANA): delay-based CC proposals (I will probably write a survey) vs. router supported (RFC for less-than-best-effort DiffServ PHB)
  - Significant interest in “scavenger service“
    - P2P, Grid, ..
    - example at TANA BOF: VLBI (Very Long Baseline Interferometry): building a huge telescope out of many smaller ones
    - Note: mismatch with load based pricing schemes
- Tradeoffs of using multiple concurrent transport connections
  - requires characterizing multiple flows

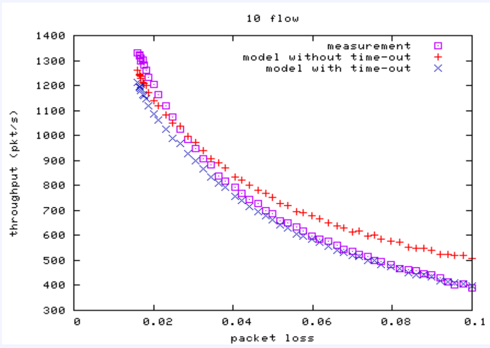
We have done some work on this... (UIBK, Dragana Damjanovic)

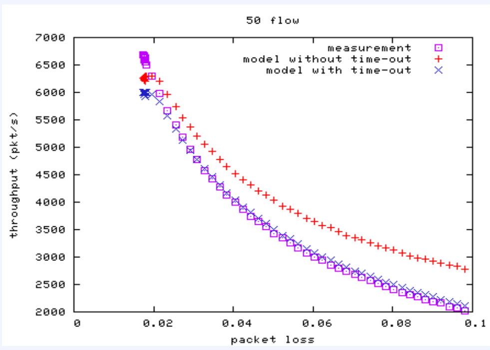
$$E[X] = \frac{2j^2pb - npbj + \sqrt{(n^2p^2b^2j^2 - 4np^2b^2j^3 + 4j^4p^2b^2 + 24n^2pbj}}{6n^2p}$$

$$E[W] = \frac{nE[X]}{2b} + \frac{3n^2E[X]}{2bj}$$

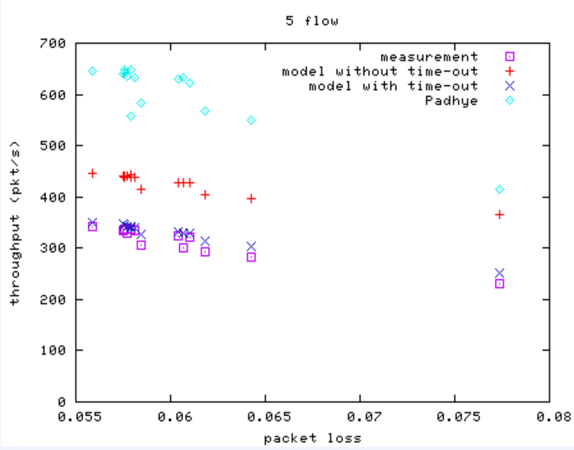
$$nTO = (j\ n/E[W])\ \frac{RTO\ (1 - 32p^2)}{X\ \ RTT}$$

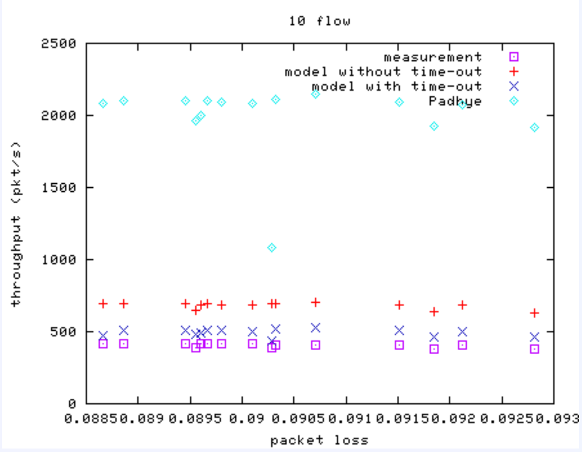
$$B = \frac{1}{p\ E[X]\ RTT}\left(1 - \frac{nTO}{n}\right) + \frac{nTO}{RTO\ (\ 1 + 32\ p^2\ )}$$



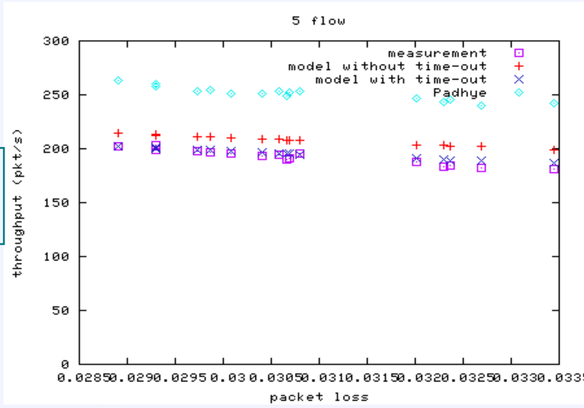


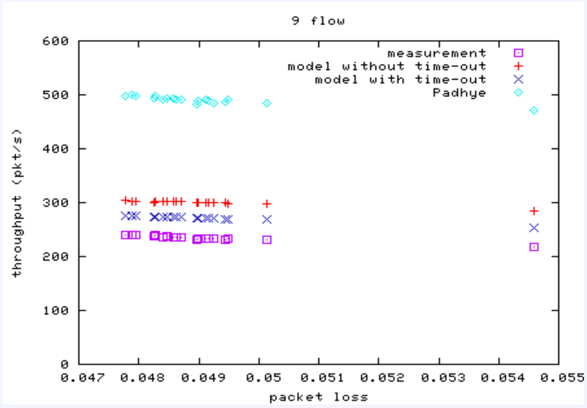
IBK - Ireland





IBK - Texas





Conclusion: what now?

- Plenty of interesting and relevant CC work to be done
- IETF: TANA
  - Meeting minutes not yet posted, but significant interest, unanimous agreement to start a WG
  - Decision now with IESG
  - Discussions currently happening in TSVArea mailing list
- IRTF: ICCRG
  - currently rather quiet
  - activity very welcome
  - volunteers for reviewing high-speed TCP drafts extremely welcome :-)

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

European Seventh Framework STREP FP7-2007-ICT-216259

Operator's vision about overlay traffic management

Juan Fernández-Palacios (jpfpg@tid.es)

Telefónica I+D (TID)

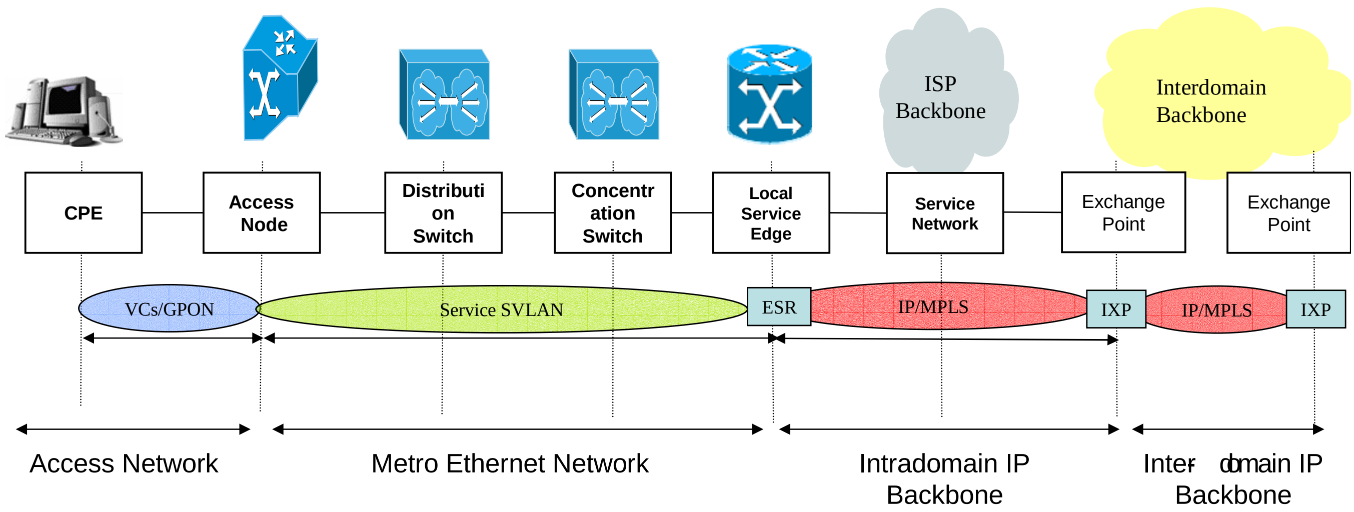
August, 2008

Operator's vision about overlay traffic management

- …The different networks, technologies and players involved in the overlay traffic transport
- …The current existing options for ISP interconnection
  - Peering and transit
  - SLAs
  - Charging schemes
- …What are the main issues in overlay traffic management
  - QoS differentiation
  - Locality
- … What technical incentives may be offered by the network operator in order to promote the traffic locality

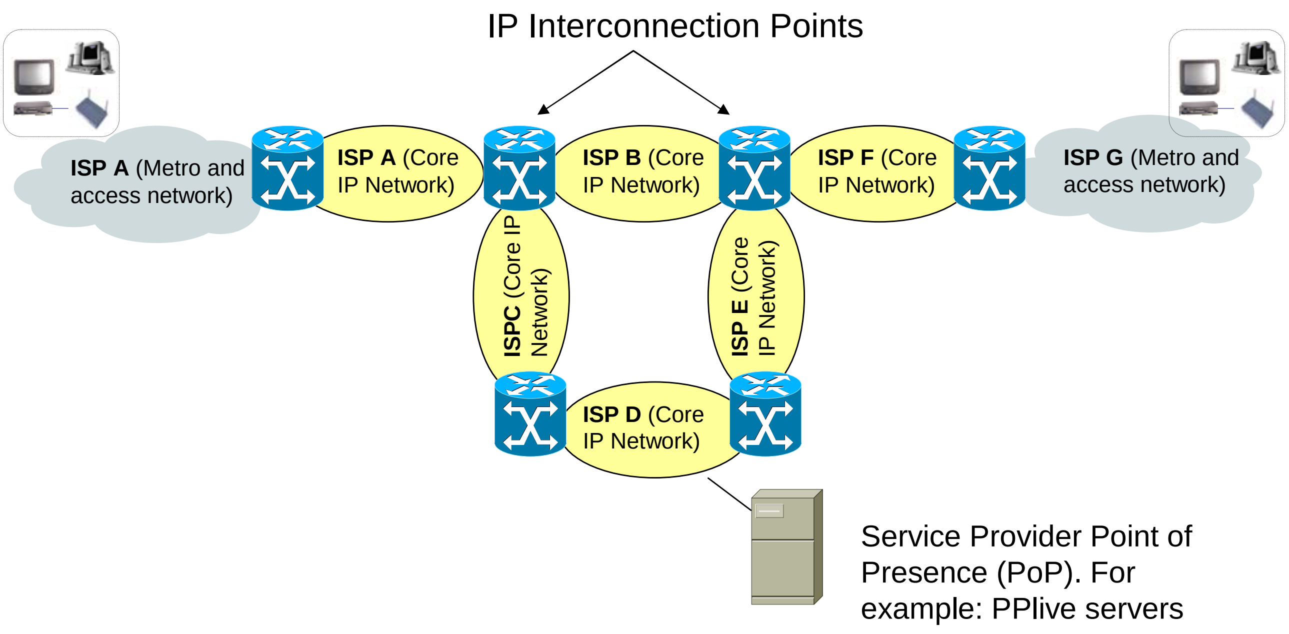
Network Topology

E2E Reference Network

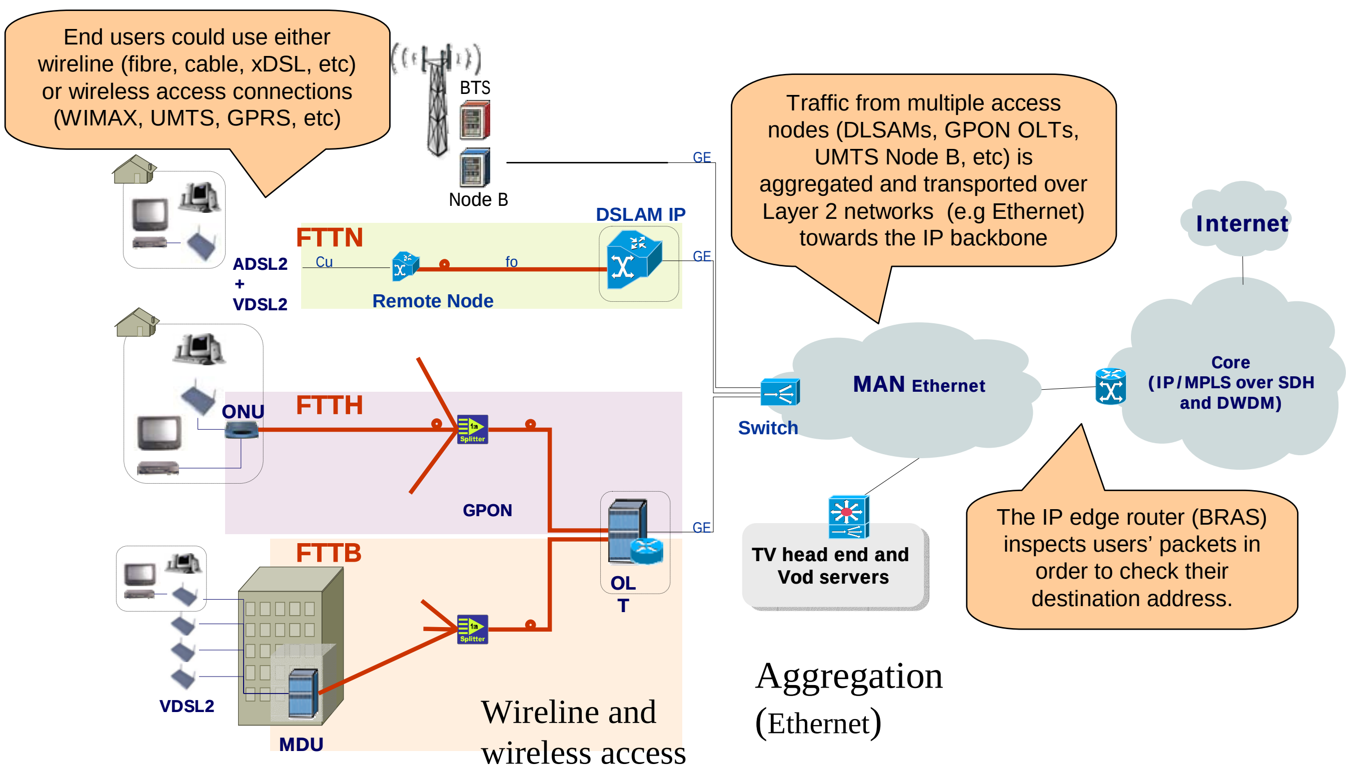


End to End Architecture

- Traffic between two peers could pass through multiple networks

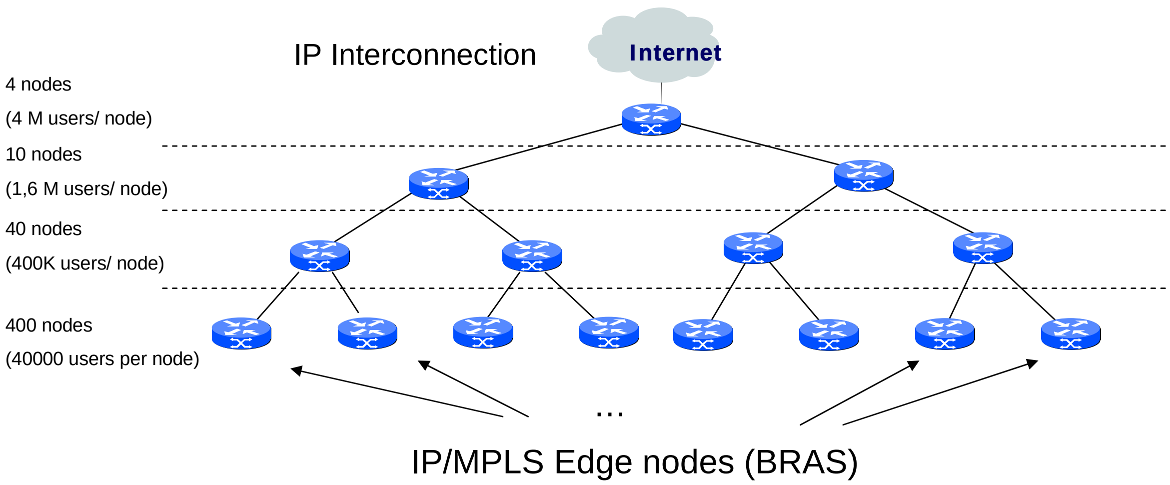


Metro and Access Network



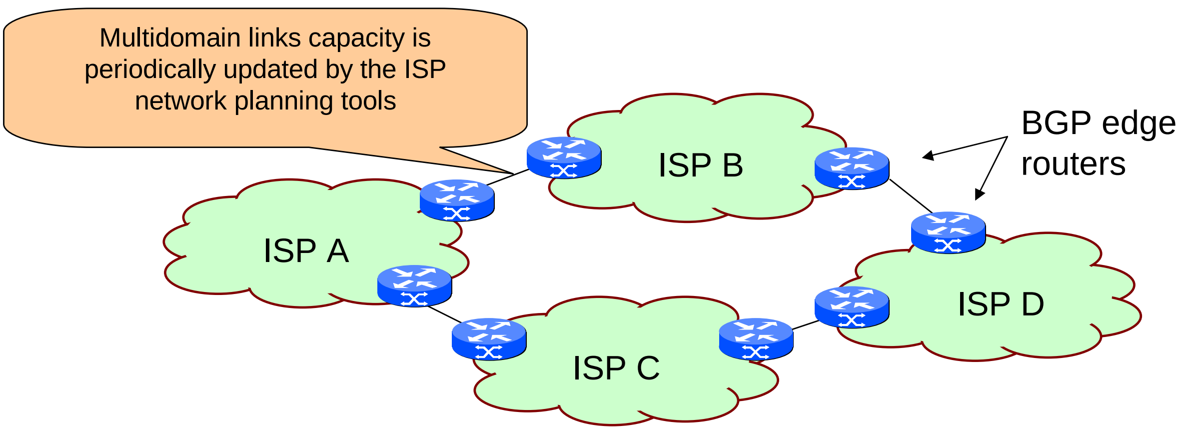
IP intradomain Backbone

Typical delay per hop (due to queuing, propagation, processing and switching)	3 ms
Propagation delay	1ms/ 200 Km
QoS control	DiffServ Different MPLS LSPs per CoS
OAM mechanisms	BFD and LSP Ping



ISP Interconnections-Routing

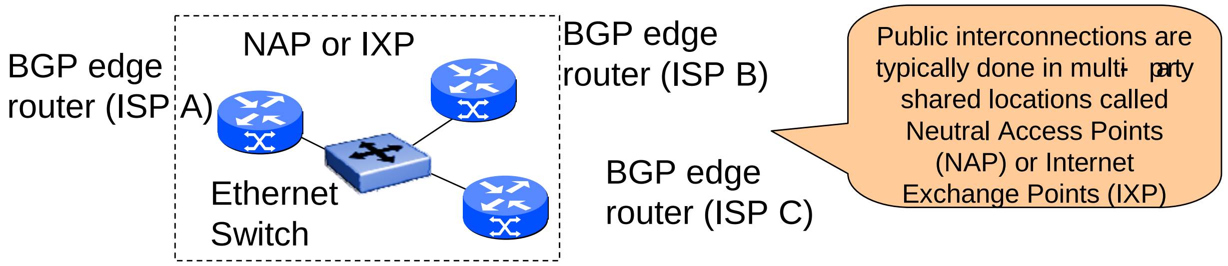
- ISPs’ networks are interconnected as autonomous routing domains
- Global routing is based on BGP (Border Gateway Protocol)
  - BGP works by maintaining a table of IP networks or 'prefixes' which designate network reachability among autonomous systems (AS)
  - BGP neighbors, or peers, are established by manual configuration between routers



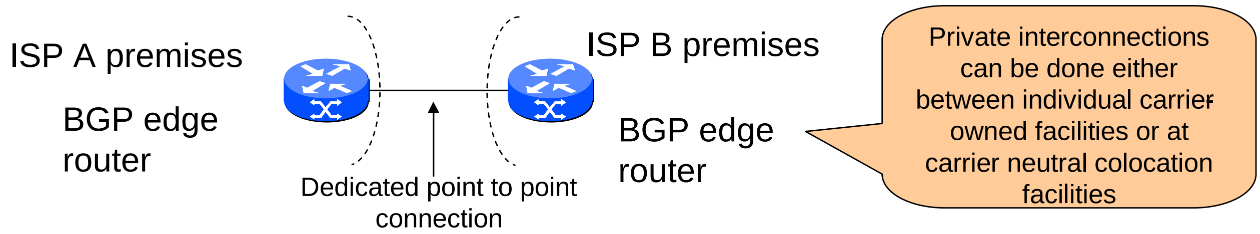
- Routers that sit on the boundary of one AS, and exchange information with another AS, are called border or edge routers.

ISP Interconnections-Physical connectivity

- The physical ISP interconnections are categorized into two types:
  - **Public interconnection**: Interconnection utilizing a multi-party shared switch fabric such as an Ethernet switch.

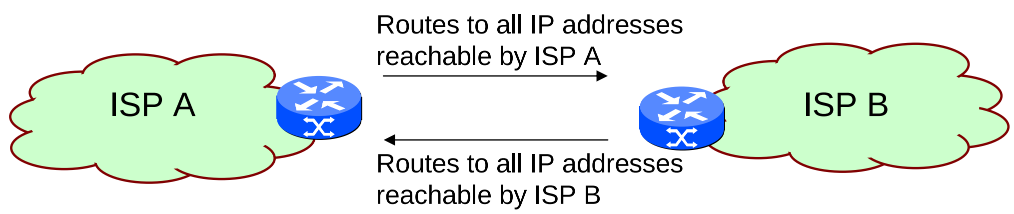


  - **Private interconnection**: Interconnection utilizing a point-to-point interconnection such as a patch-cable or dark fiber between two parties.



ISP Interconnections-Business relations

- The ISP interconnection consists of the advertisement by an ISP of
  - …routes to its customer's IP addresses to the other ISPs (i.e soliciting inbound traffic)
  - …a set of routes to all of the reachable destinations by the other ISP, to the ISP's customer (i.e soliciting outbound traffic)



- Such exchange of reachability information and traffic could be done freely or not depending on the ISPs business relationship

- The relationships between ISPs are generally described by one of the following categories:
  - Peer: Two networks exchange traffic between each other's customers freely
  - Transit: An ISP pays to another ISP for the traffic exchange

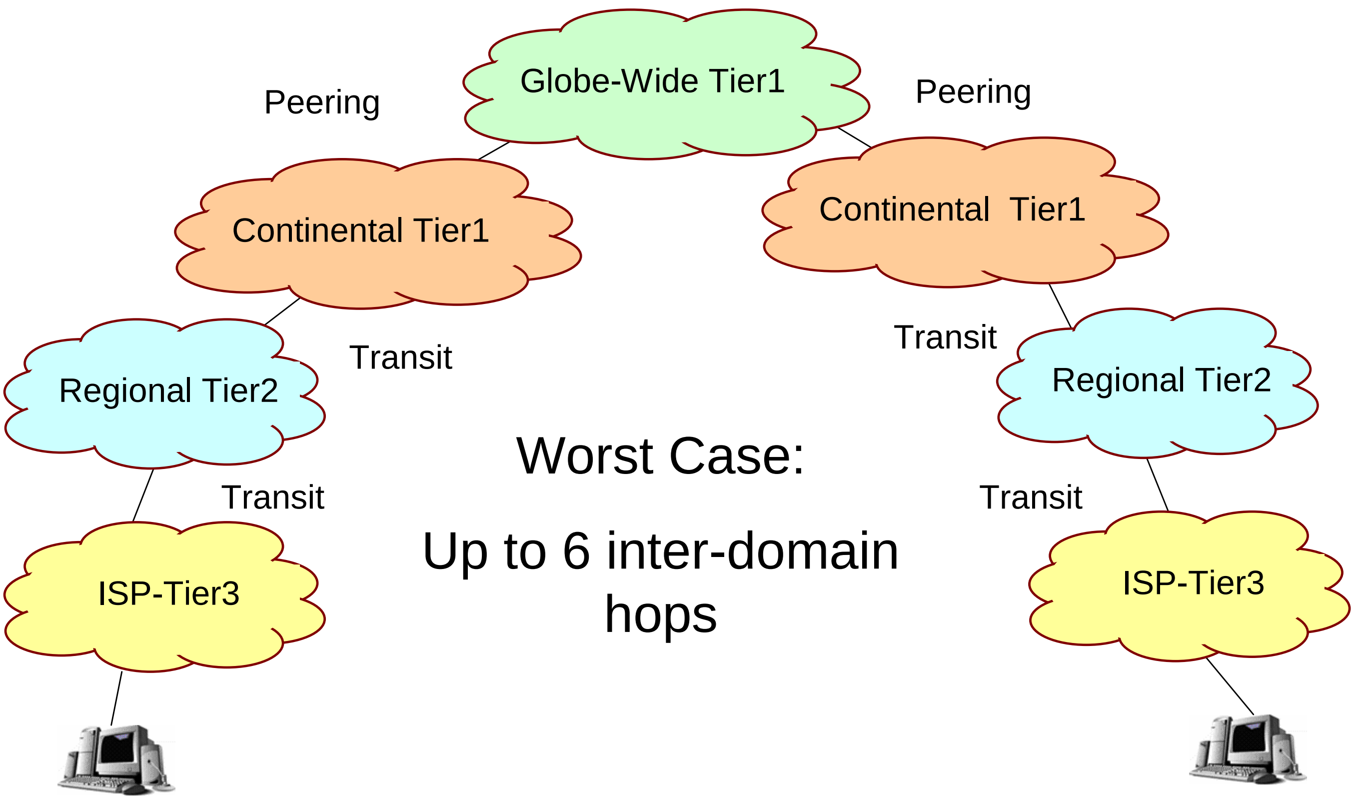
ISP Interconnections- Charging

- In the IP transit model the purchaser has to pay the difference between outbound and inbound traffic
- Pricing is typically offered on a Mbps/Month basis and requires the purchaser to commit to a minimum volume of bandwidth
- For example a common charging model for IP transit is based on **95th percentile method:**
  - The difference between the average inbound and outbound traffic is measured every 5 minutes and recorded in a log file
  - At the end of the month, the top 5% of data is thrown away, and that next measurement becomes the billable utilization for the month

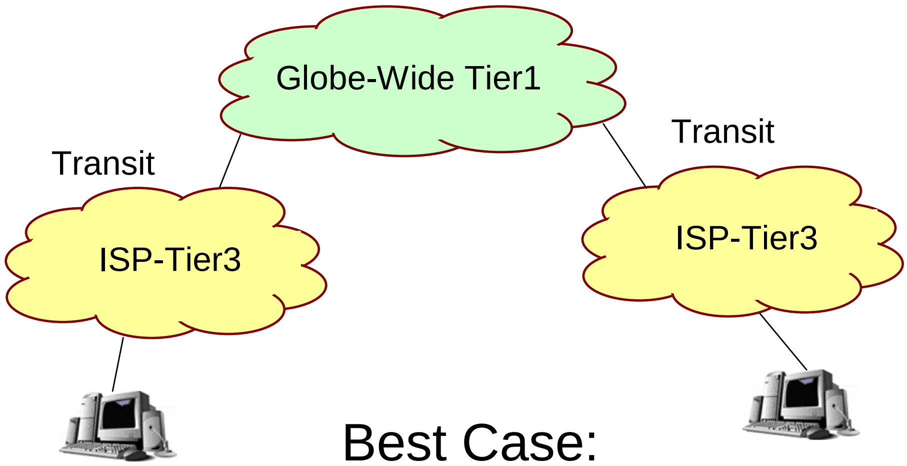
$$C_{month} = P_{95} \{ \text{Average(outbond-inbound)} \} \times \text{Price/Mbps}$$

C_{month} : Cost per month; **P_{95}**: 95th Percentile; **Average(outbond-inbound):** Average traffic samples measured every five minutes during one month; **Price/Mbps**: Price per Mbps agreed between the ISPs

IP Intra-domain Backbone



IP Intra-domain Backbone

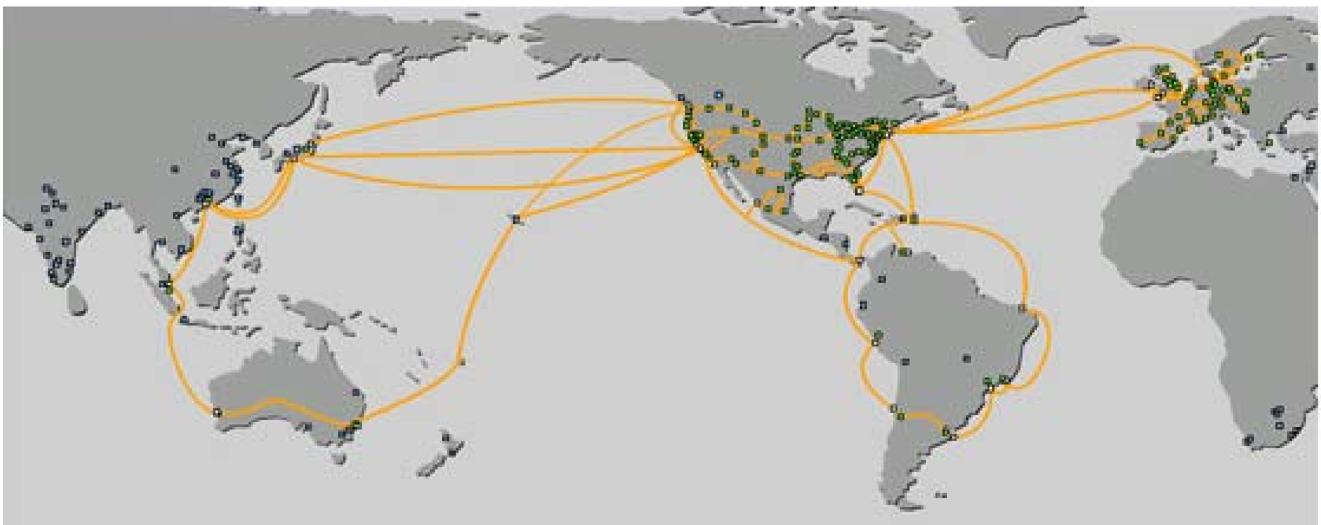


Best Case:

Only 2 inter-domain hops

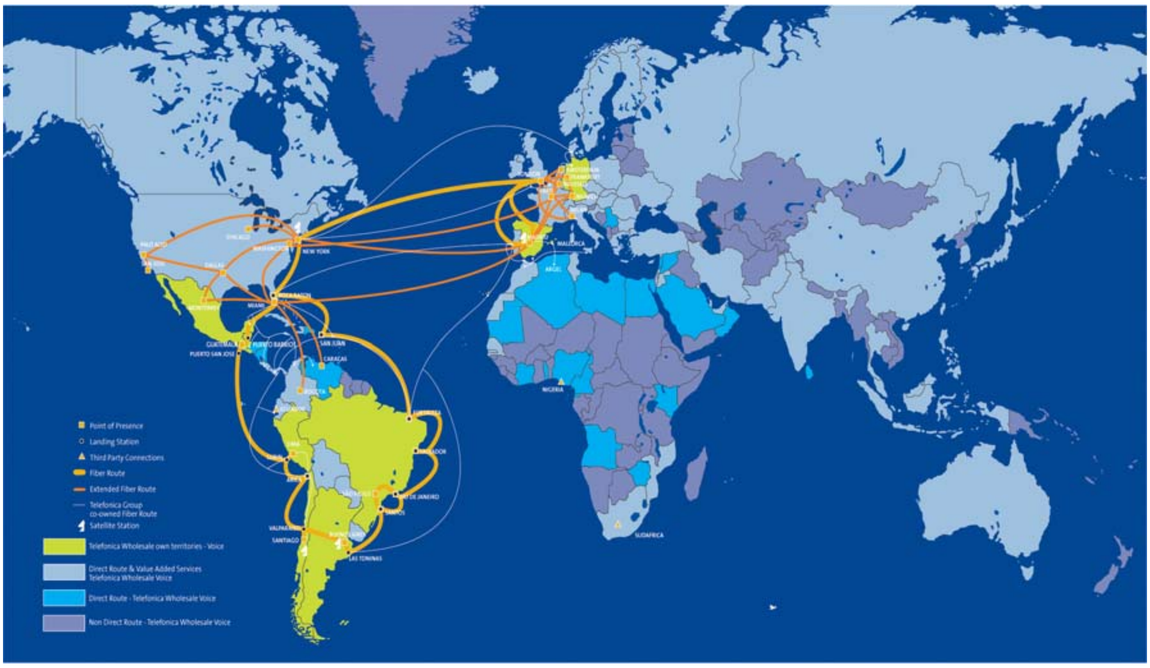
Global-Wide Tier 1

Global Crossing Network is covering large areas around the world



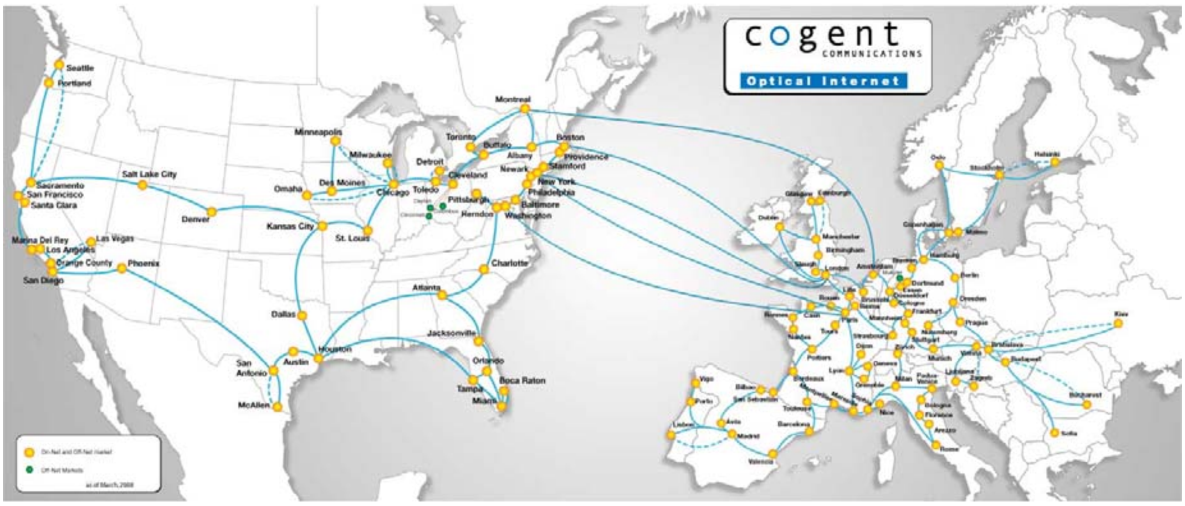
Regional Tier 1

TIWS is covering large areas of Europe, EEUU and Latin America



Regional Tier 2

Cogent covers Europe and EEUU



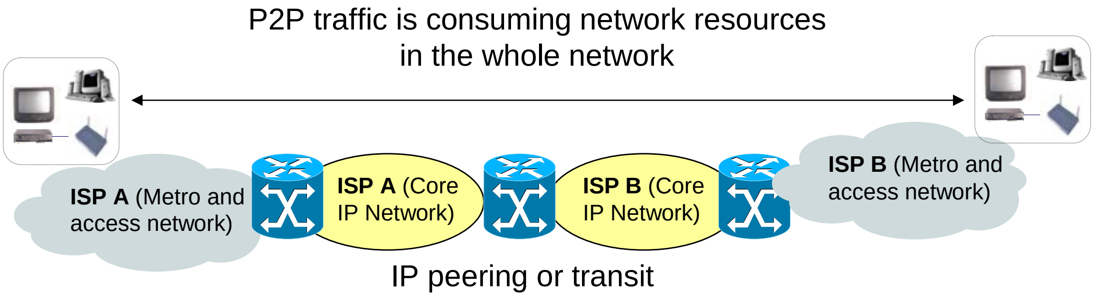
Main issues in overlay traffic management

Main issues in overlay traffic management- QoS differentiation

- Currently, Internet traffic (e.g http, overlay, etc) is transported according to a "best effort" approach
- However, some overlay applications such as IP-TV, VoD, VoIP, videoconference or gaming present strict requirements in terms of delay and/packet loss
- The introduction of application-aware transport services able to provide the required QoS for each application would improve the QoE perceived by the end user
- Why might operators be interested in increasing the QoE of overlay applications?
  - To increase the broadband customers fidelity and reduce the churn rate
  - To sell new broadband connectivity services specially adapted to Internet real-time and streaming applications

Main issues in overlay traffic management - Network planning

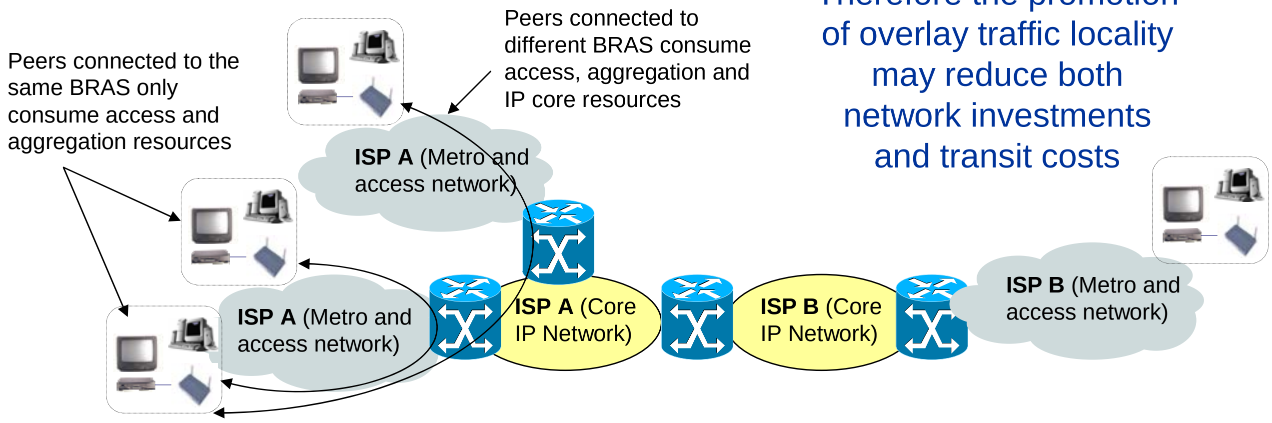
- Currently, a very high percentage of Internet traffic comes from overlay applications
  - For example, around an 80% of total Internet traffic in Spain is generated by P2P applications
- The amount and distribution of overlay traffic strongly impacts total network costs (CAPEX and OPEX).
  - For example:If and ISP customer is exchanging P2P traffic with a customer of another ISP then such traffic is consuming resources in the whole ISP network: access, aggregation, IP "national" core and IP interconnection (peering or transit)



Traffic Locality

- As higher the percentage of "multidomain" traffic as higher the network resources consumption and total costs:
  - Multidomain traffic passes through the whole network. Therefore, it consumes more transmission and switching resources than internal traffic
  - In case of having an IP transit agreement then multidomain traffic should be paid to another ISP
- On the other hand, internal P2P traffic doesn't consume interconnection bandwidth

Therefore the promotion of overlay traffic locality may reduce both network investments and transit costs



Potential incentives for File-sharing applications

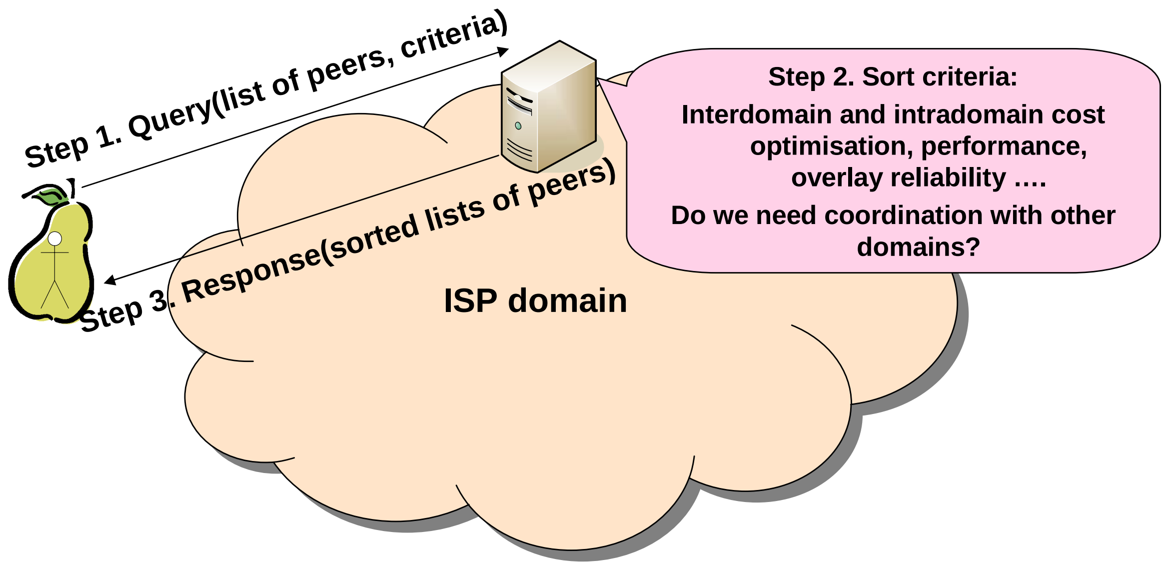
Main objective of ISP and P2P collaboration

ISP and P2P collaboration should be profitable for both ISP and end user

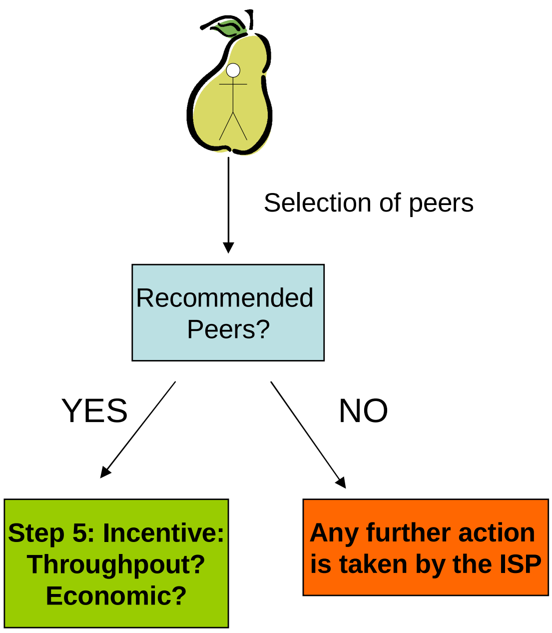
- **ISP benefits**: Increase the percentage of P2P intra-domain traffic
  → Cost optimisation (especial attention to interconnection costs).

- **User's benefits**
  - Faster downloads (↑ Throughput)
  - Potential economic incentives if different charging schemes than flat rate are used (e.g Charging schemes per traffic volume)

Steps 1-3



Steps 4: Selection of peers



Sorted list of peers provided by the ISP

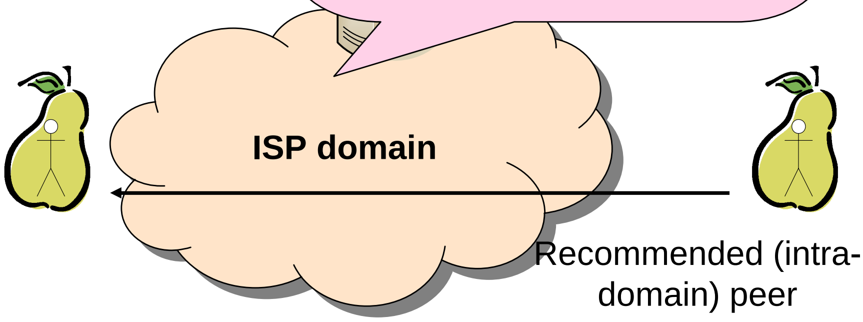
Peer 1	**ISP Recommended**
Peer 2	**ISP Recommended**
…	**…**
Peer N	**Not recommended**

Step 5: Technical incentive

Sorted list of peers provided by the ISP

Peer 1	ISP Recommended: High Throughput
Peer 2	ISP Recommended: High Throughput
...	...
Peer N	Not recommended

The ISP could automatically increase the recommended peer upstream capacity. However, the key technical challenge is how to assure that such bandwidth increase will not be also used for inter domain flows. Furthermore P2P applications use to limit the upstream capacity.



SUMMING UP: The introduction of technical incentives for file-sharing applications might be very complex

Step 5 (bis): Economic incentive

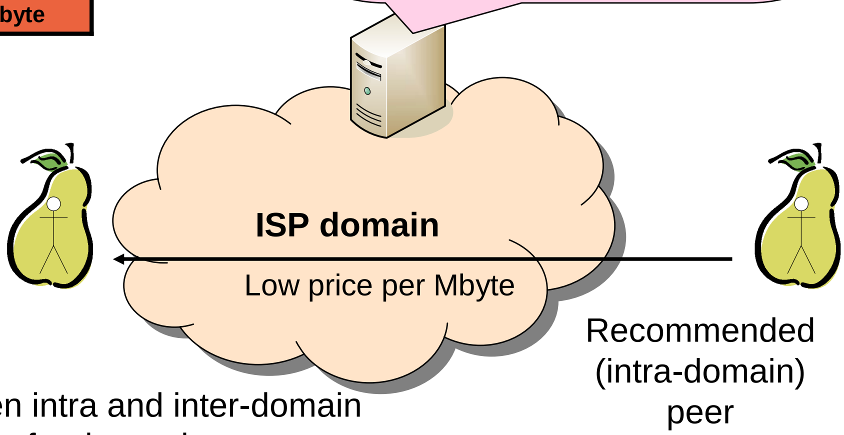
Sorted list of peers provided by the ISP

Peer 1	ISP Recommended: Low price per Mbyte
Peer 2	ISP Recommended: Low price per Mbyte
...	...
Peer N	Not recommended: Standard price per Mbyte

In charging models based on the bandwidth consumption, the total price at the end of the month depends on the amount of information transported over the Internet connection (Cmonth = Mbytes x Price/Mb). This charging model is often used in wireless (3G, 2G) Internet connections. Such charging models might be updated in order to reduce the Price/Mbyte of intradomain flows.

Important: privacy requirement should be considered.

Repositories of P2P users might generate legal problems



Billing differentiation between intra and inter-domain traffic might not be very clear for the end user

Potential incentives for P2P real-time Applications

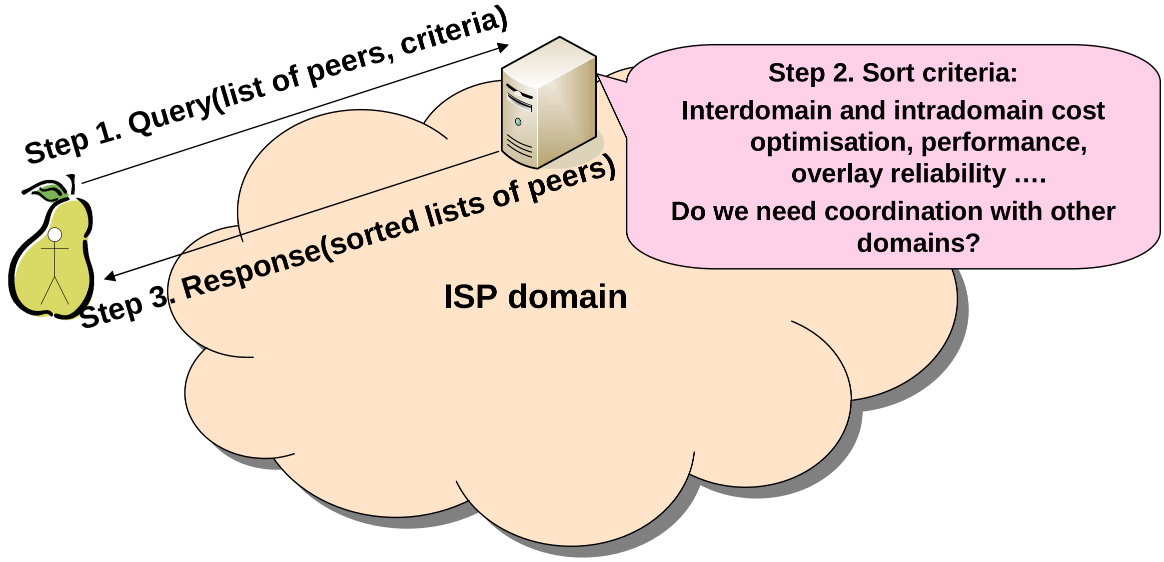
Main objective of ISP and P2P collaboration

ISP and P2P collaboration should be profitable for both ISP and end user

- **ISP benefits**: Increase the percentage of P2P intra-domain traffic
  - → To increase the broadband customers fidelity and reduce the churn rate
  - → To sell new broadband connectivity services specially adapted to real-time and streaming applications
  - → Cost optimisation (this benefit is more significant in file-sharing applications)
- **User's benefits**
  - Better QoE (lower delays and packet loss rates)

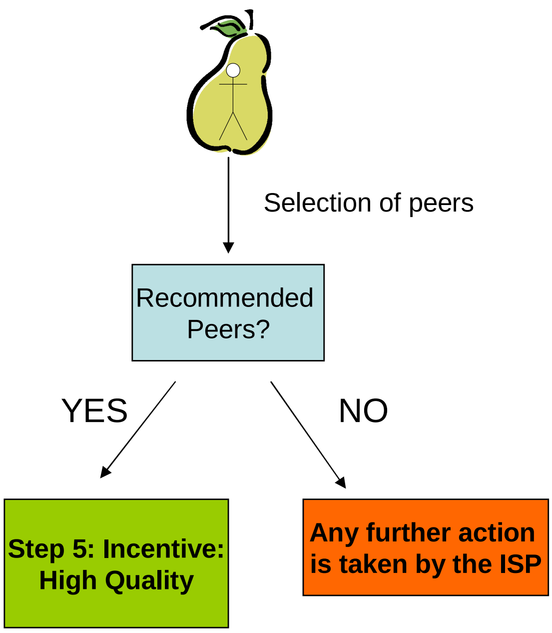
User's potential benefits are different in this case

Steps 1-3



The sorted list sent back to the users would be organised according to QoS criteria

Steps 4: Selection of peers



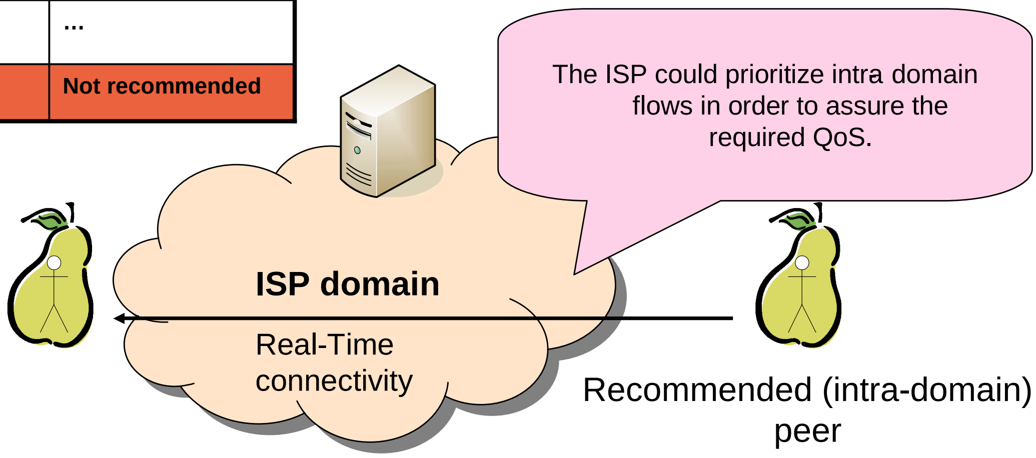
Sorted list of peers provided by the ISP

Peer 1	**ISP Recommended High Quality**
Peer 2	**ISP Recommended High Quality**
…	**…**
Peer N	**Not recommended**

Step 5: Technical incentive

Sorted list of peers provided by the ISP

Peer 1	ISP Recommended: High Throughpout
Peer 2	ISP Recommended: High Throughpout
...	...
Peer N	Not recommended



SUMMING UP: The provision of QoS incentives is technically feasible

Conclusions

- There are two key traffic management measures that would strongly impact on both the operator's network planning and the QoE perceived by the end users

1) Overlay traffic locality
   - Under an operator's perspective traffic locality promotion may reduce both network investments and transit costs
   - Under an end user's point of view locality will imply technical incentives (e.g faster downloads)
2) Overlay traffic QoS differentiation
   - Under an operator's perspective QoS differentiation would allow:
     - To increase the broadband customers fidelity and reduce the churn rate
     - To sell new broadband connectivity services specially adapted to real time and streaming applications
   - Under an end user's point QoS differentiation will imply better QoE in real time and streaming overlay applications

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

European Seventh Framework STREP FP7-2007-ICT-216259

ISP-owned peer & Locality Awareness in BitTorrent

UZH, DoCoMo, TUD, AUEB, PrimeTel, AGH, ICOM, UniWue, TID

Sergios Soursos, AUEB

Workshop on ETM, Zürich
August 4-5, 2008

Outline

- BitTorrent Optimization Potential
- Simulations
- Results
- Conclusions & Future work

BitTorrent Optimization Potential

Topology Awareness

- Issue: Tracker replies a ***random*** list of peers to each peer's request!
  - Inefficient use of the underlay
  - Affects also the performance of the overlay
- Idea: Alternative peer selection at the tracker
  - Apply a proximity criterion
- Potential proximity criteria: Autonomous System , number of hops, Round-Trip-Time, congestion, price
- Topology information must be available to the tracker
  - Information provided by the underlay
  - Or by the peers themselves – Incentives to be truthful?

Locality Awareness

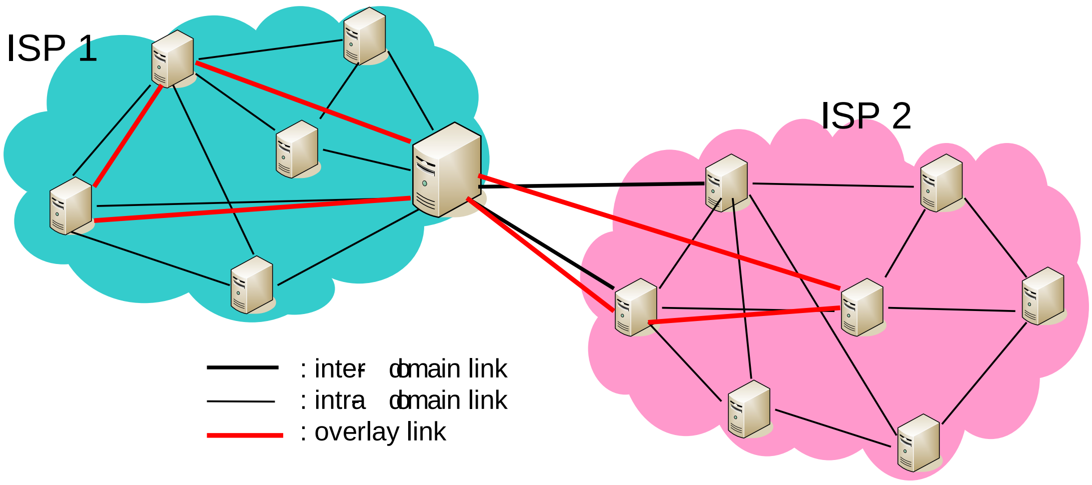
- Proximity criterion: ***Autonomous System***
- Tracker's reply list comprises of (Bindal *et al.**):
  - *K* out of *N* peers selected within the same AS with the requesting peer
  - *N*-*K* selected from other ASes
- Important reduction of ingress inter-domain traffic is achieved
- No improvements on peer's completion times are observed
  - For some peers the completion time is increased!

* R. Bindal, P. Cao, W. Chan, J. Medval, G. Suwala, T. Bates, A. Zhang, **"Improving Traffic Locality in BitTorrent via Biased Neighbor Selection"**, 26th IEEE International Conference on Distributed Computing Systems, p. 66, 2006

Insertion of ISP-owned Peers (I)

- Insertion of ISP-owned peers (IoPs) with high upload capacity is expected to have performance improvements both for the ISP and the end-user
- Motivation: The introduction of IoPs is expected to achieve reduction of ingress inter-domain traffic
  - This peer pre-fetches chunks by participating in the BitTorrent swarm (worst case!)
  - No redundant copies of these chunks are downloaded to the ISP from external peers
- But IoPs will also be selected from external peers to download from because of high upload rates & tit-for-tat mechanism
  - As a result extra egress inter-domain traffic is generated
- Tradeoff: reduction of ingress inter-domain traffic vs. increase of intra-domain & egress inter-domain traffic

Insertion of ISP-owned Peers (II)



1. In a pure BT network: Interaction with *'tit-for-tat'* mechanism
2. In a locality aware BT network: IoP is highly likely to be selected by the peers within the same AS

Simulations

bittorrent.patch* for the *ns*-2 simulator

- BitTorrent-like protocol, functions simplified
- Four classes implemented: Application, Tracker, Connection, Message
- BitTorrent implementation is modular; e.g. peer and piece selection algorithms can be replaced by alternatives
- Network model:
  - FullTCP: bidirectional data transfers
  - Uplink is assumed to be the bottleneck in the whole network
  - Downlink is neglected

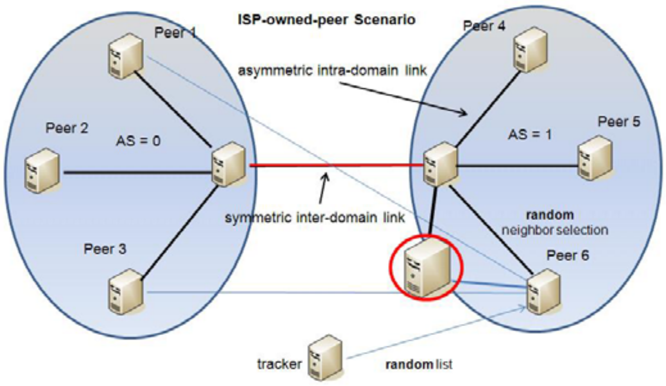
* K. Eger, T. Hoßfeld, A. Binzenhöfer, G. Kunzmann, **"Efficient Simulation of Large-Scale P2P Networks: Packet-level vs. Flow-level Simulations"**, 2nd Workshop on the Use of P2P, GRID and Agents for the Development of Content Networks (UPGRADE-CN'07) in conjunction with IEEE HPDC, Monterey Bay, USA, June 2007

Scenarios (I)

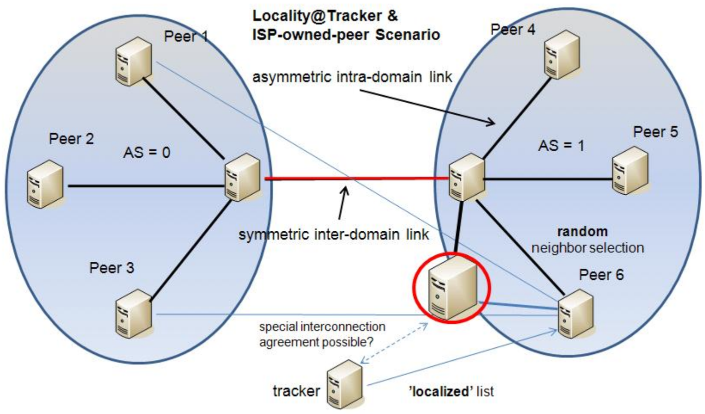
1. Pure BitTorrent
2. BitTorrent with Locality Awareness
3. BitTorrent with Insertion of ISP-owned peer
4. BitTorrent with Insertion of ISP-owned peer and Locality Awareness

- Symmetric or Asymmetric ASes
  - Symmetric: 25 peer per AS, e.g. 2 Tier-4 ISPs
  - Asymmetric: 35 and 15 peers in each AS, e.g. Tier-3 and Tier-4 ISPs respectively
- All-together or Split
  - All-together: Joining time of all peers ~U(0,10)
  - Split: Joining time of 5 peers in each AS ~U(150,300), whereas joining time of the rest of the peers in each AS and the ISP-owned peer ~U(0,10)

Scenarios (II)



Insertion of IoP in BitTorrent without locality awareness



Insertion of IoP in BitTorrent combined with locality awareness

Simulation parameters

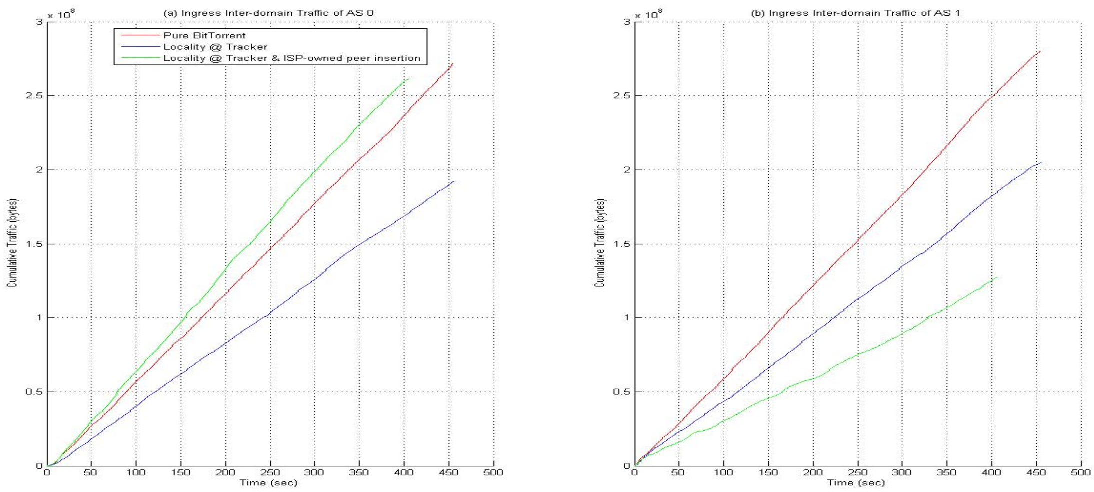
Description	Value
Number of peers	50
Number of seeds	1
Number of ASes	2
Number of peers per AS	(25,25), (35,15)
Upload capacity of regular peers	512K
Download capacity of regular peers	4096K
File size	20M
Number of peers requested from tracker (Size of tracker's list)	25
Number of local peers replied by tracker	20
Number of connections	20
Choking interval	10
Number of unchoked connections permitted per peer	4, 10 (in case of IoP)
Number of ISP-owned peers	1
Upload/download capacity of ISP-owned peers	40960K

@ AS 0 (Number of seeds)

@ AS 1 (Number of ISP-owned peers)

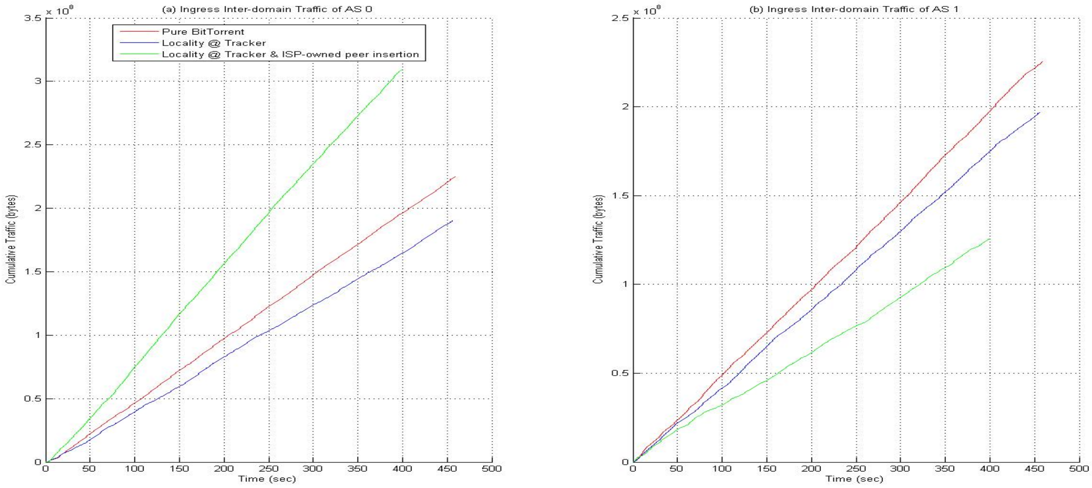
Results

Symmetric case: Ingress inter-domain traffic



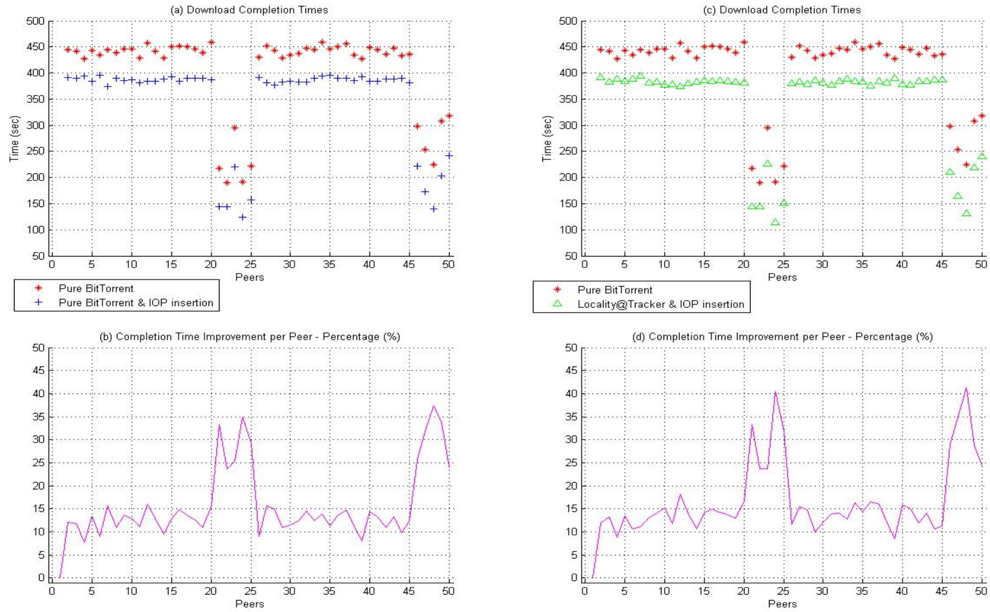
- Locality awareness achieves important improvements of ingress inter-domain traffic of both ASes – up to 30%
- Insertion of IoP combined with locality awareness further reduces the ingress inter-domain traffic of AS 1 – up to 60%
- As expected the ingress inter-domain traffic of the AS 0 is increased, however it is still less than the pure BitTorrent case

Asymmetric case: Ingress inter-domain traffic



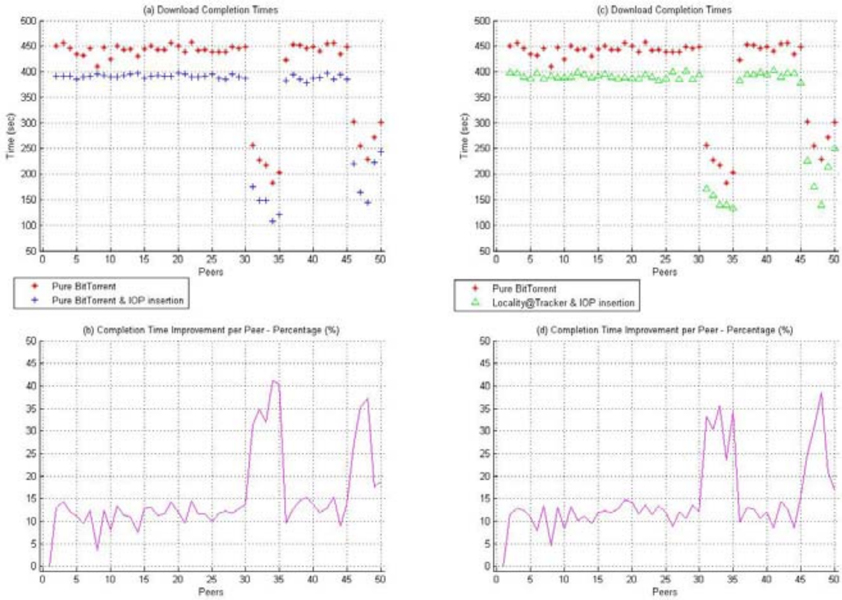
- Respective results hold also for the asymmetric case
- Locality awareness improves ingress inter-domain traffic for both ASes up to 15%
- Insertion of IoP combined with locality awareness further reduces ingress inter-domain traffic for AS 1 up to 45%
- However, the ingress inter-domain traffic for AS 0 is increased up to 30% compared to pure BitTorrent!

Completion Times – Symmetric case



- Insertion of IoPs improves ***all*** peers' completion times: 10-15%
- Especially for peers that join the swarm later than the IoP, the completion times are further improved up to 36-40% for the symmetric case

Completion Times – Asymmetric case



- Similar results as in previous case

Conclusions & Future work

Conclusions

- ISP deploying only locality awareness risk to lose customers because of completion time deterioration
- The insertion of IoP improves both ingress inter-domain traffic and peers' completion times for the AS that deploys the extra peer
- Consequently:
  - Interconnection agreement may be modified in favor of the AS that deploys the extra peer
  - The AS does not risk losing customers, on the contrary it may attract new ones

* I. Papafili, S. Soursos, G.D. Stamoulis, **"The Impact of Insertion of ISP-owned Peers and Locality-Awareness in BitTorrent"**, submitted to the 3rd International Workshop on Self-Organizing Systems (IWSOS '08), Vienna, Austria, December 10-12, 2008

Future work

- Asymmetric case & IoP → increased ingress inter-domain traffic for AS 0
  - Is there an incentive for AS 0 to introduce an IoP as well?
  - Study how the results will be affected.
- So far: One swarm & one IoP
- Variations:
  - One swarm and multiple IoPs
    - Trade-off: extra performance improvement vs. extra resources
    - How many IoPs?
    - How do their total resources scale?
  - Multiple swarms and multiple IoPs
- Study how interconnections agreements are affected!
  - E.g. under the 95th percentile charging scheme
- What about content promotion vs. locality promotion?

Thank you!

Questions?

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

European Seventh Framework STREP FP7-2007-ICT-216259

Investigation of BitTorrent Swarm Sizes

UZH, DoCoMo, TUD, AUEB, PrimeTel, AGH, ICOM, UniWue, TID

Workshop on ETM, Zürich
August 4-5, 2008

Tobias Hoßfeld, UniWue

Goal of investigation

- Is it possible to utilize locality information in a BitTorrent swarm?
- Is it possible to use ISP-owned peers?
- How to model the population sizes, request process, user behavior within a single swarm?

→ number of swarms (for popular contents)
→ number of seeders and leechers per swarm
→ time-dynamics within a single swarm

Measurement Scenario

- Downloaded .torrent files from thepiratebay.org
  - 63867 swarms measured, offering video contents, movies, TV series, documentary → swarm sizes obtained every 23 minutes (roughly) per swarm
  - most popular movies (w.r.t. #seeders and w.r.t. #leechers) traced individually

Top Ten

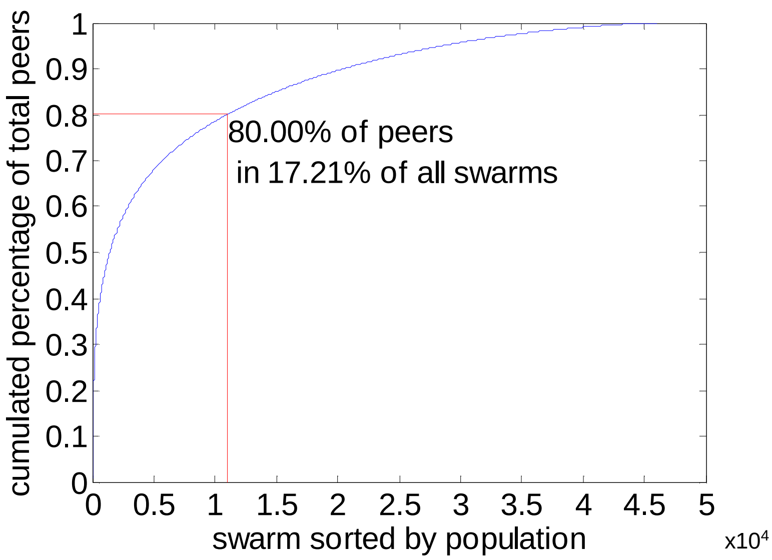
- Same video content (maybe different encoding), but offered in different swarms

- The_Chronicles_of_Narnia_Prince_Caspian_2008_Eng_TS_DivX_LTT
- Sex_and_the_City__2008__TS_Occor_avi
- Street_Kings_2008_DvDrip_Eng__FXG
- Kung_Fu_Panda_TS_XViD_mVs
- Futurama_The_Beast_With_A_Billion_Backs_DVDRip_XViD_HooKah
- Kung_Fu_Panda__2008_Eng__TS_DivX_LTT
- The_Incredible_Hulk_2008_CAM_SUBBED_XViD_nDn__MFD__avi
- Be_Kind_Rewind_2008_DvDrip_aXXo
- Charlie_Bartlett_2007_DvDrip_Eng__FXG
- Jumper_2008_DvDrip_AC3_aXXo

Pareto Principle for Swarm Sizes

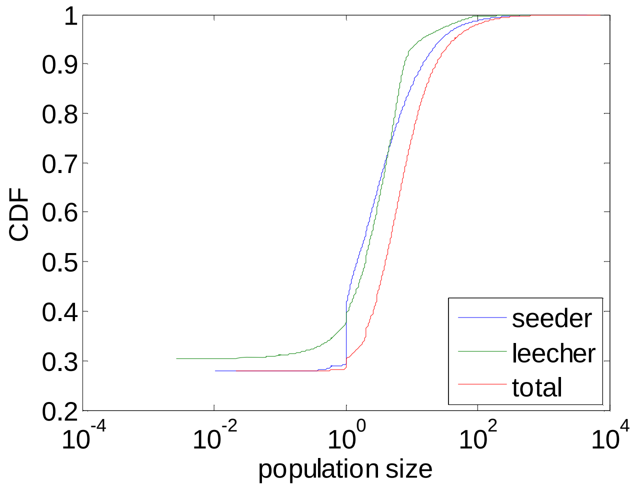
- 80-20 rule valid for BitTorrent swarm sizes
- 17861≡ 27.97% of swarms don't have any seeders; 1496+17861 swarms don't have any leechers

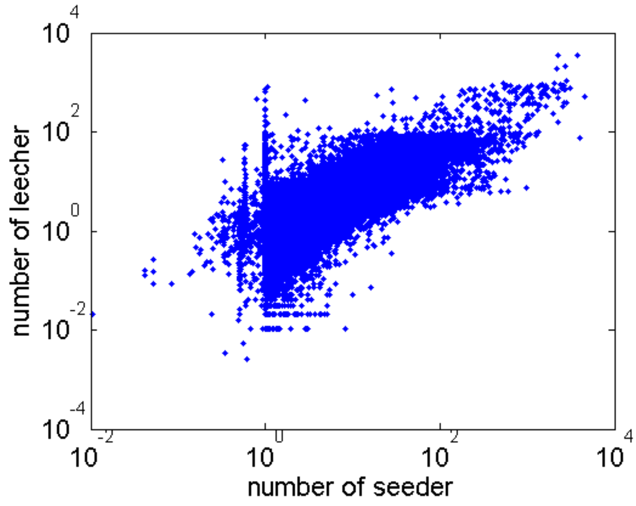
➔ Sufficient to insert ISP-owned peers in top swarms



Correlation of Seeder and Leecher

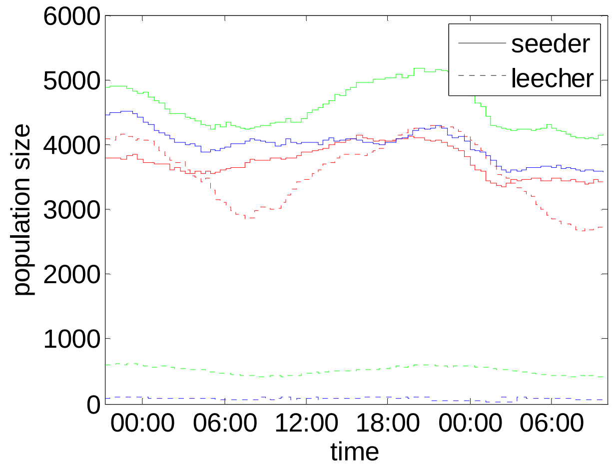
- correlation(#seeder,#leecher) = 0.70

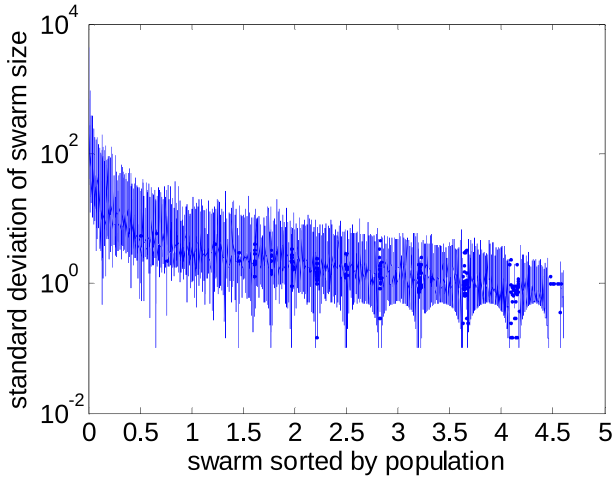




Time Dynamics

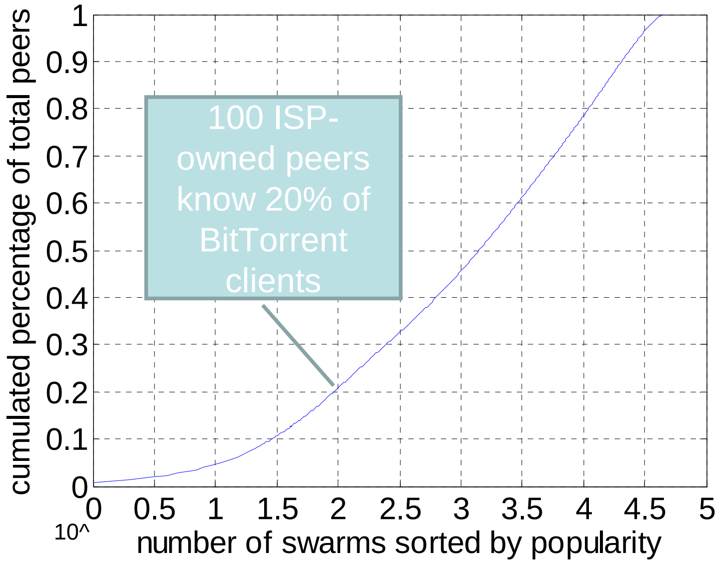
- Evolution of population sizes for some popular movies
- Standard deviation of swarm sizes during measurements





Conclusions

- ISP-owned peers possible



Thank you!

Questions?

Discussion:
“Will Economic Traffic Management Mechanisms be Successful?”

Burkhard Stiller, UZH (Moderator)

Zurich
August 4-5, 2008

Economic Traffic Management (ETM)

- The complete and integrated set of
  – issues in theory (economics, viability, efficiency …),
  – its design (signaling protocol, traffic management, …) and
  – its evaluations (simulation, proof-of-concept in a prototypical implementation) and understanding of inter-relations

  is exactly what ETM addresses.
- The solution proposed by ETM involves
  – decentralized decision making
  – guided by incentives and prices.
- **This approach is termed “Economic Traffic Management (ETM)”.**

Successful

- Webster: Resulting in success
- Thus, a measurable degree of succeeding!

- "User – Overlay (OL) Provider – ISP"
  - Win-win-win means "TripleWin"
  - Thus, a success is determined by a TripleWin!

- User is better off compared to pure OL application usage, better QoS/QoE
- OL provider better off compared to "old" applications without any service differentiation
- ISP better off due to optimized traffic handled

Dimensions to be Considered

- Technical overhead?
- New signaling protocols?
- Standardized mechanisms?
- Application adaptations?

- Viability measured in gains of costs

- Optimization in terms of bandwidth utilized
- Performance calculated in "good-put"

To be continued ...

Open Issues?

- How does an incentive-driven signaling approach (theory) and its respective protocol (technology) deliver economic signals across domain boundaries?

- Which schemes will be economically viable?
- Which schemes will be implementable?
- Which standardizations necessary?
- Which data protection acts to be considered?
- Which regulations, net neutrality issues dominate?

To be continued ...

SVD Based Shared Bottleneck Detection

Muhammad Murtaza Yousaf

Workshop on ‘Economic Traffic Management’ - Zurich
August 4-5, 2008

Shared Bottleneck Detection (Algorithm)

- What it does:
- Applications:
  - Network-aware Grid scheduling*
  - Precise File Transfer delay prediction*
  - Efficient file transfer in overlay networks *
  - Replica management in p2p like systems
  - Coordinated congestion management

** Directly related to EC-GIN*

Shared Bottleneck Detection(Algorithm – I)

One-way delay measurements

$D := (d_{i;j})_{\ i=1,\ldots,m;\ j=1,\ldots,n}$

One-Way Delays for window size 'w' *

Paths

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & \ldots\ldots\ldots\ldots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \ldots\ldots\ldots\ldots & d_{2n} \\ & & & & \\ d_{m1} & d_{m2} & d_{m3} & \ldots\ldots\ldots\ldots & d_{mn} \end{bmatrix}$$

** Optimal value of 'w' is found to be 300 ms*

Shared Bottleneck Detection(Algorithm – II)

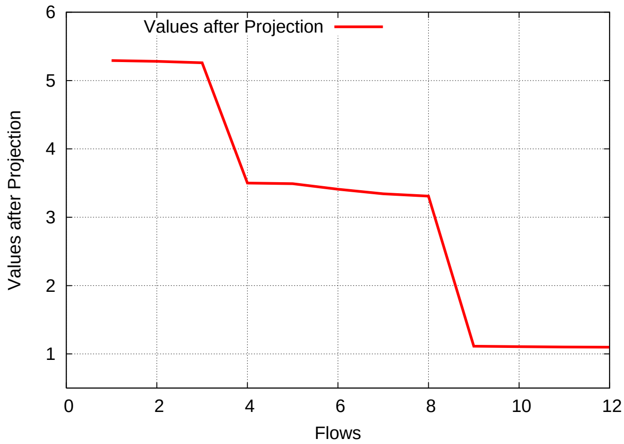
Applying SVD and getting Projection

$U\Sigma V^T = svd(D)$

U – Left Singular Matrix [Association among the paths]
Σ – Diagonal Matrix [Degree of association]
V – Right Singular Matrix

Getting Projection

$X_{mx1} = U_{mxr}\ X\ \Sigma_{rxr}\ (.;\ 1)$

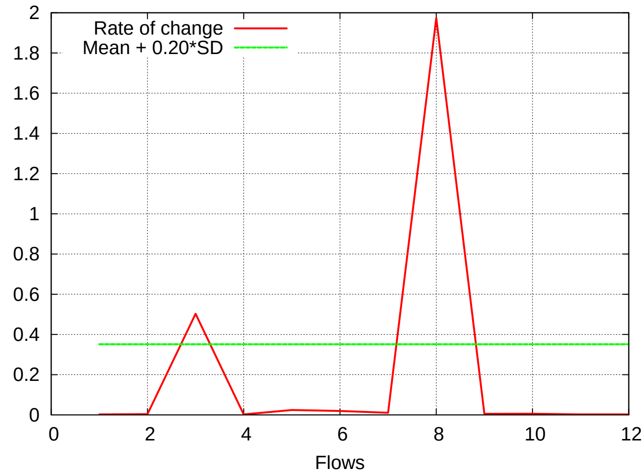


Shared Bottleneck Detection(Algorithm – III)

Outlier Detection

$$roc = \frac{x_i - x_{i+1}}{x_{i+1}}$$

$$\tau = \bar{z} + 0.2\sigma$$



Shared Bottleneck Detection (Algorithm) Analysis

- Suitable in real time
  - Time Complexity *O(m log m)*
  - *m – no. of paths*
- Performance Evaluation
  - *ns* simulations
  - *Emulab* experiments
  - *Implementation*: at a single server, based on collected data

Shared Bottleneck Detection (Algorithm) Status

- Ready and submitted in ToN
- Technical report is uploaded with complete algorithm details and documentation
  - *http://welzl.at/research/publications/sbdTechnicalReport.pdf*
- Development of real-life code has begun
  - Hopefully done by the end of this year
  - Then integrated with distributed *traceroute* to yield a full "peer awareness" system

Thank You

Questions ?

Workshop on "Economic Traffic Management" ,
August 4-5, 2008, Zürich, Switzerland

DiCAP - An Architecture for Distributed Packet Capturing

Cristian Morariu

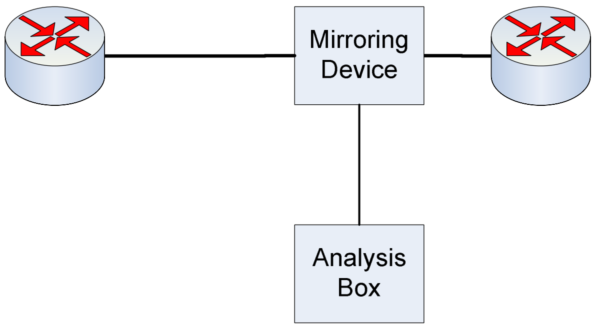
Department of Informatics IFI, Communication Systems Group CSG, University of Zürich

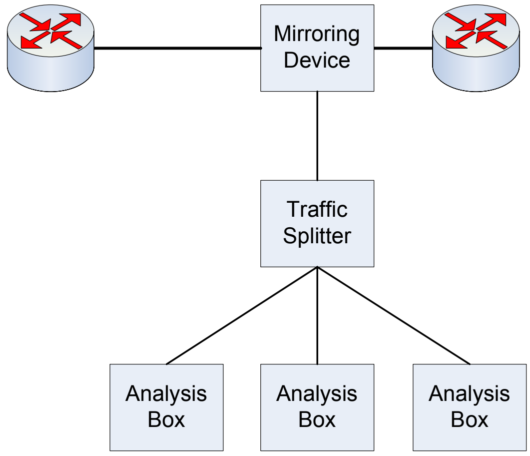
Motivation
Design and Implementation
Evaluation
Concluding Remarks

Overview

- Most network monitoring tasks require packet inspection
- Network monitoring
  - *Live*
    - packets are inspected in real-time
    - data is dropped after inspection
    - *e.g.* traffic accounting, IDS
- Before feeding data to monitoring applications, packets need to be captured.
- Solutions for high packet rates:
  - packet sampling
    - decreased measurement accuracy
  - dedicated hardware
    - Expensive

Traditional Architecture for Traffic Analysis





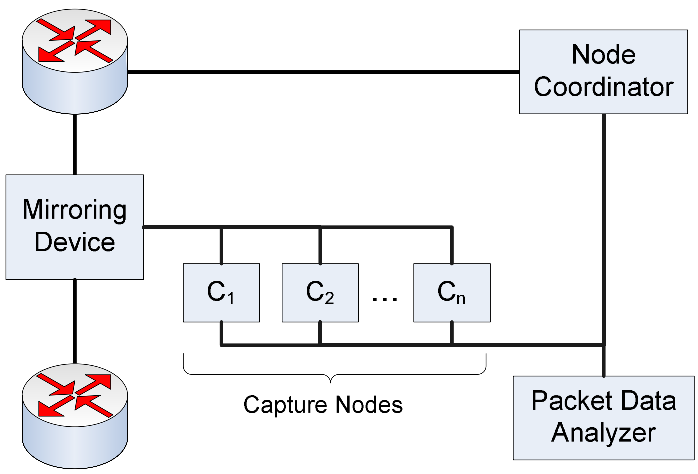
- Simple
- Easy to deploy
- Not scalable due to a single analysis box

- More complex
- The splitter is typically expensive
- More scalable with respect to performance
- The single point of failure still present

Motivation

- Build a distributed architecture for IP traffic capturing
  - Avoid a single point of failure
  - Avoid dedicated hardware
  - Based on stock, inexpensive PCs
  - Allows the increase of packets that can be captured

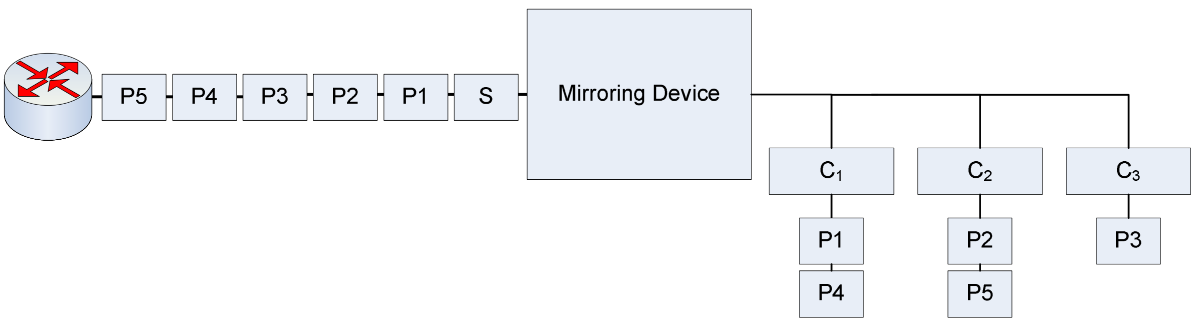
DiCAP Architecture



- Each node receives all packets
- Node coordinator decides the capture policy on each capture node
- The capture nodes selects which packets to capture packets so that no packet is captured by two nodes
- Packet data is sent to the packet data analyzer for further analysis

Selection Policies (1)

- Round robin selection
  - Node coordinator introduces control packets in the mirrored traffic
  - Capture nodes are logically organized in a chain
  - After each packet, capture responsibility shifts to the next node in the chain
  - Node coordinator configures the chain

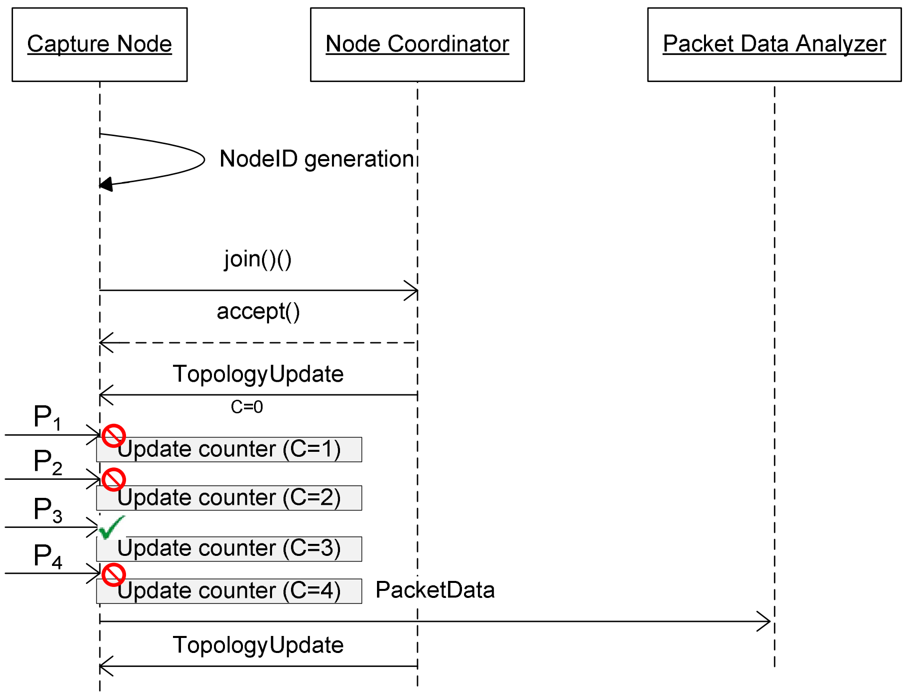


Selection Policies (2)

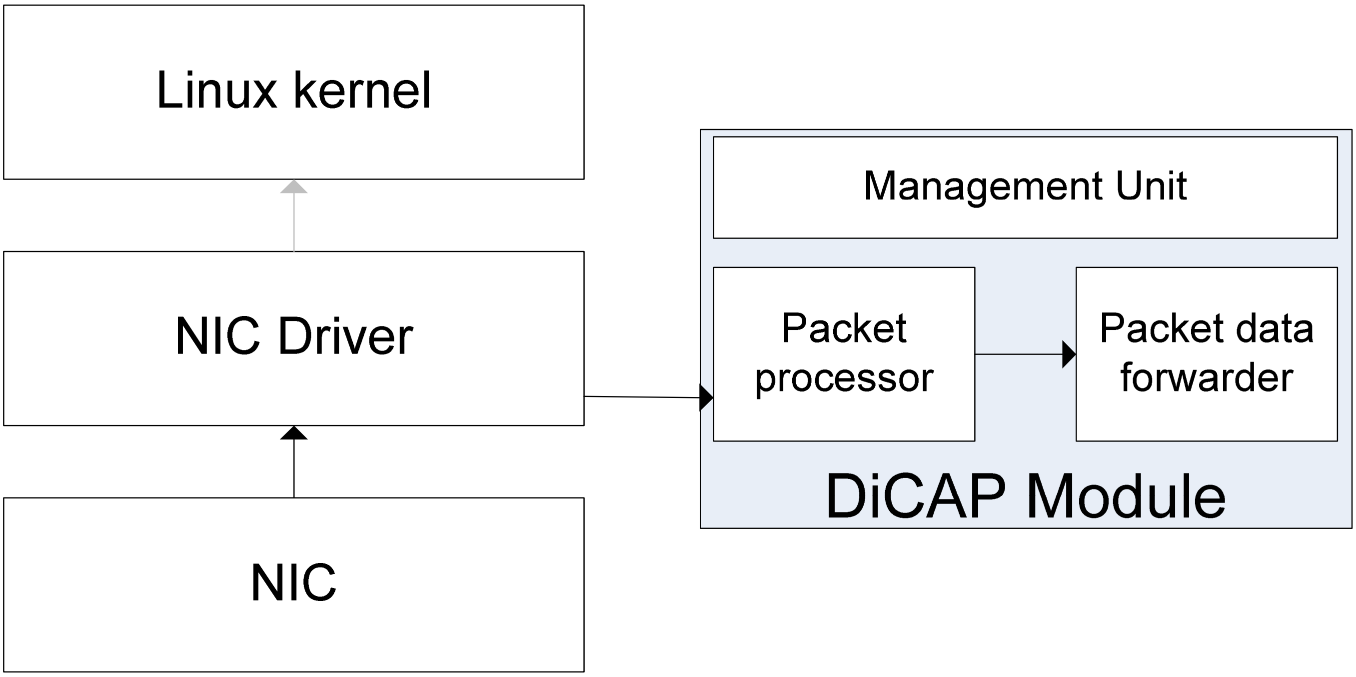
- Hash-based selection
  - A hash function is applied on packet headers
    - Current implementation uses IP identification field
  - Each node is responsible with a particular range of hash values
  - Each packet is captured by the node responsible with the respective range of hash values

- Advantages:
  - Easier synchronization
  - No need of control packet injection
- Disadvantages:
  - More computation needed for hash calculation

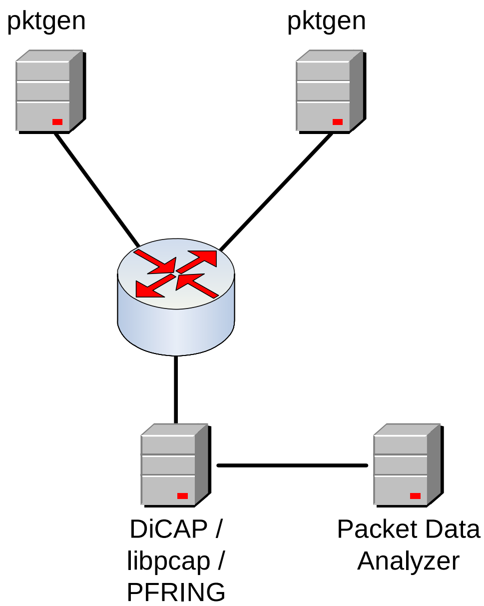
Capture Example in Round-Robin Selection



DiCAP Implementation



Evaluation Testbed



- Two nodes used for traffic injection using Linux pktgen
- One DiCAP node for testing capture performance of a single node
  - 1G Ethernet Broadcom NetXtreme BCM5721 card
- Packet data is sent to the packet data analyzer for further analysis
- One packet data analyzer

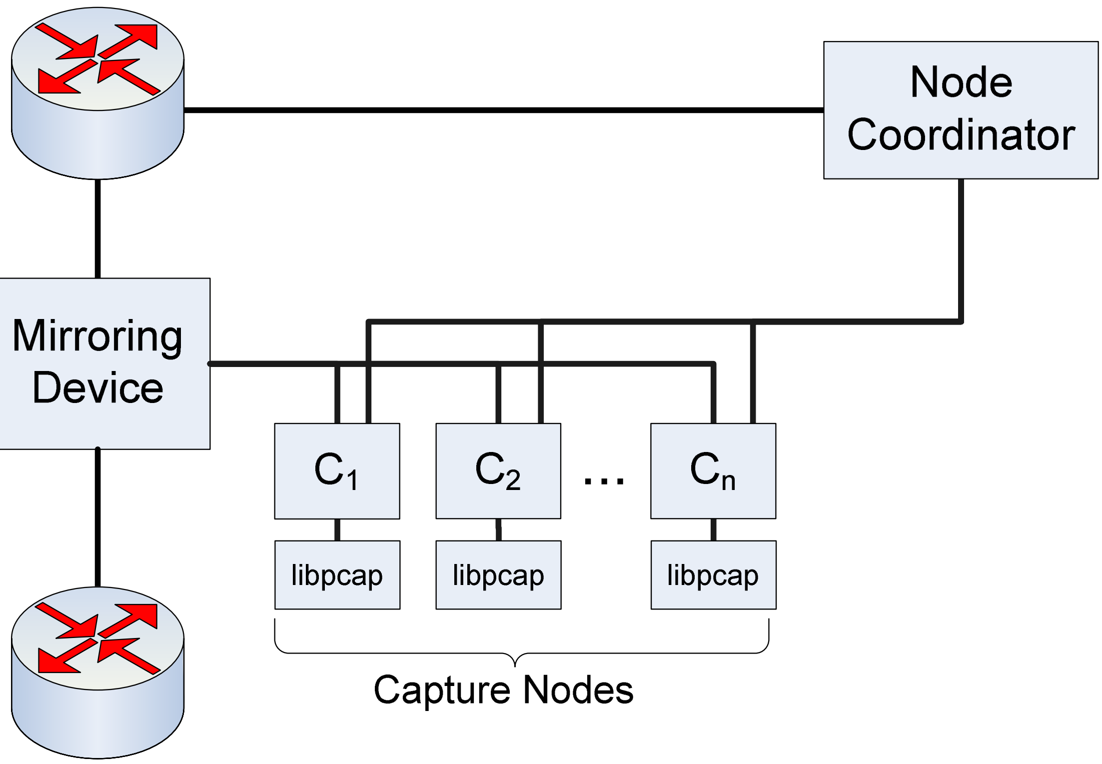
DiCAP Capture Evaluation

Packet Rate	libpcap Loss	PFRING Loss	DiCAP Loss
119 Kpps	0%	0%	0%
232 Kpps	10%	1%	0%
380 Kpps	75%	28%	0%
492 Kpps	90%	83%	0%
620 Kpps	93%	96%	0%

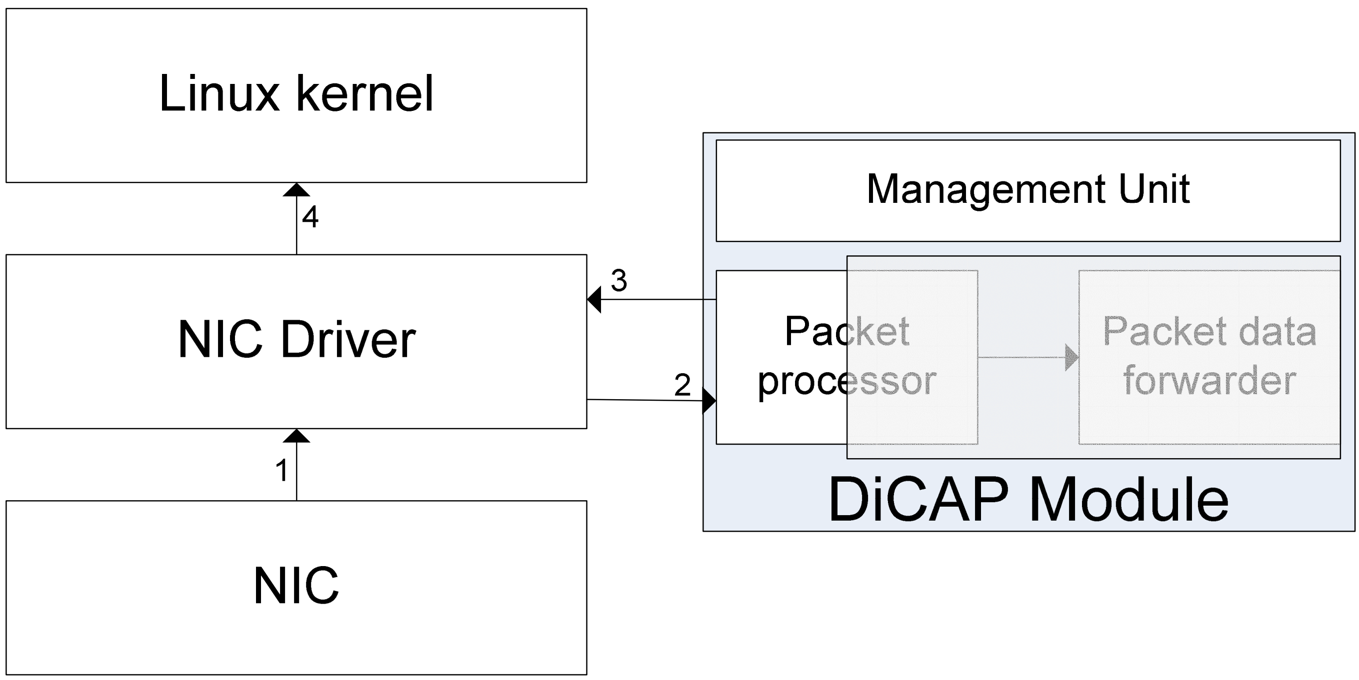
Operation Modes

- Capture Mode
  - Packet headers are captured and forwarded to a packet analyzer
  - No possibility to analyze payload of packets
- Distribution Mode
  - DiCAP does not capture anything
  - Decides which packets are forwarded to the higher levels
  - Capture takes place in user space applications
  - Allows parallel use of *libpcap*

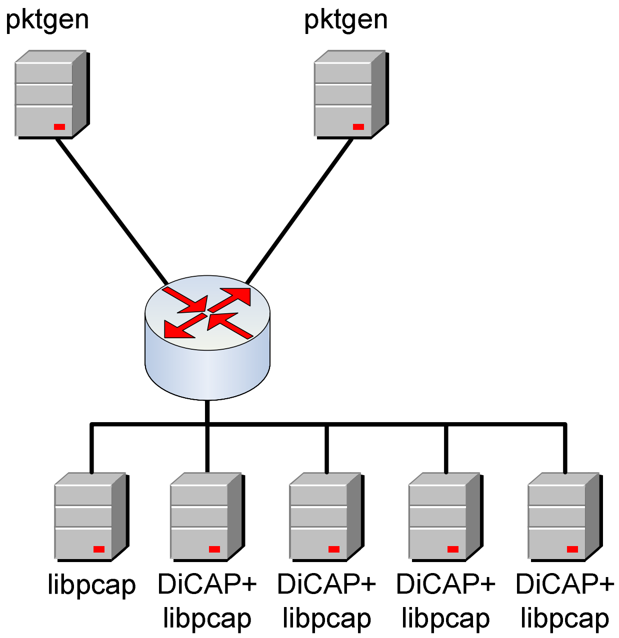
DiCAP in Distribution Mode



DiCAP Implementation – Distribution Mode

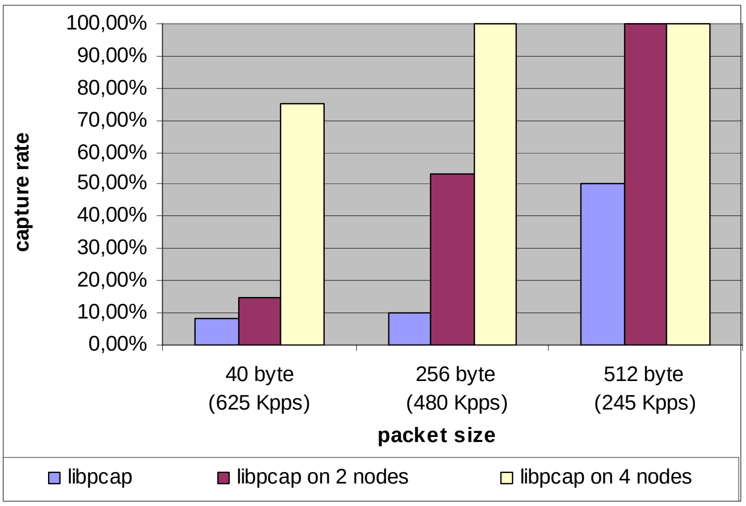


Evaluation Testbed



- Two nodes used for traffic injection using Linux pktgen
- One libpcap node for testing
- Up to 4 DiCAP nodes in distribution mode

Distribution Mode Evaluation



- 3 different tests were performed

1. traditional libpcap on a single PC
2. DiCAP in distribution mode on 2 PCs
3. DiCAP in distribution mode on 4 PCs

- DiCAP can improve the capture performance up to 700% when 4 PCs are used in parallel

Concluding Remarks

- DiCAP can be used to distributedly capture packets on a high-speed link
- It may be used to allow a distributed deployment of libpcap-based applications running on common, inexpensive PCs
- It may significantly improve libpcap performance
- Very simple design, easy to implement in hardware.
- Further investigation on using other hash functions may lead to performance improvement and better load ballance

Locality and Contracts

Martin Waldburger

Department of Informatics IFI, Communication Systems Group CSG, University of Zürich

Locality in International Contracts
Private International Law Procedure
Direct IPRG Formalization
Challenges and Solution
Summary and Conclusions

Locality in International Contracts

...

Contract parties (identity, applicable geographic location, contact details)

Language (contract and communications)

Validity (offer, contract duration)

Offerings (goods, services, main characteristics)

Price and terms (price, delivery, costs, payment)

Warranty, after-sales service, termination, remedies

Restrictions (conditions of purchase, geographic/time restrictions)

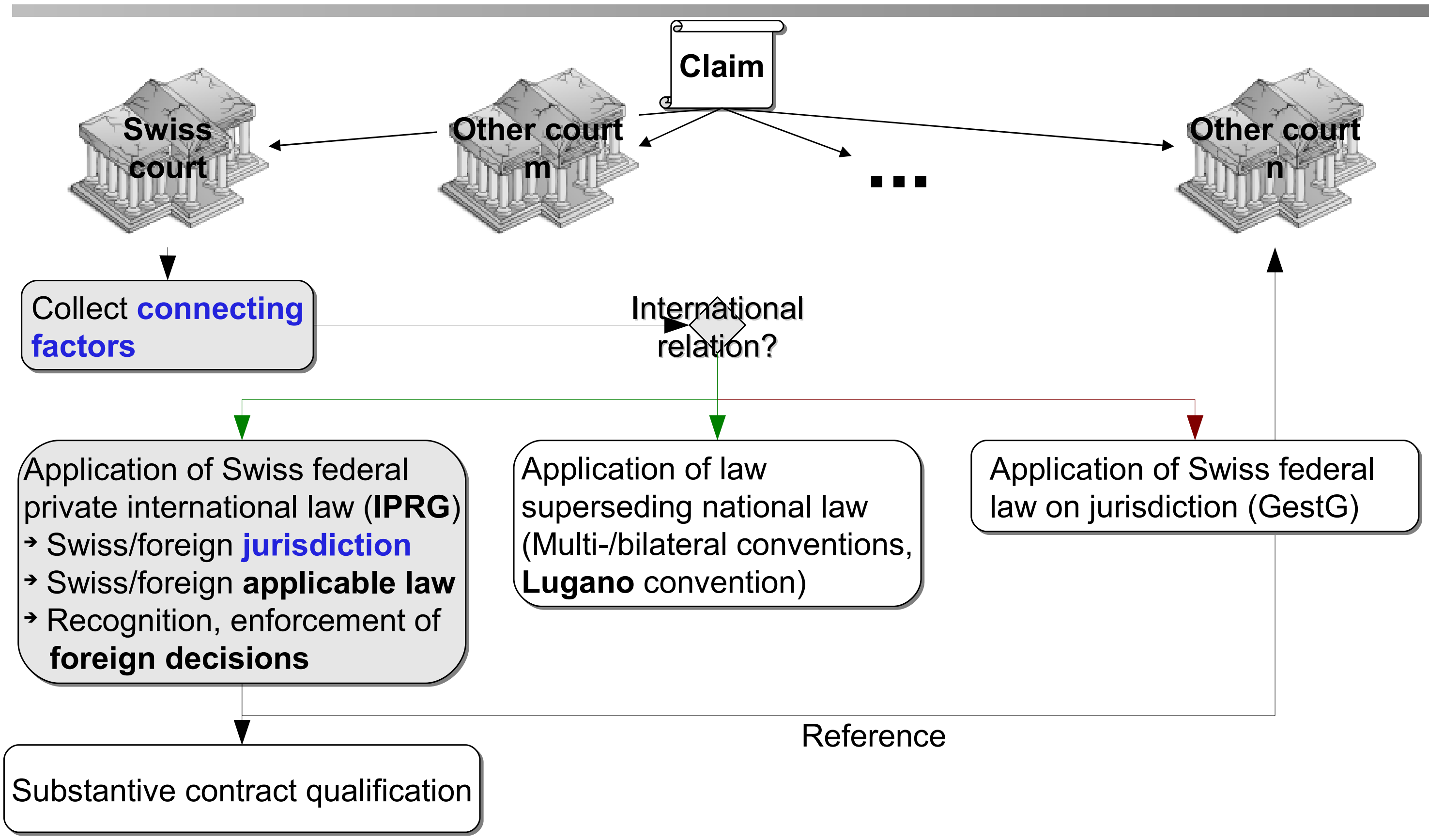
Dispute resolution (applicable law, jurisdiction, alternative dispute resolution)

...

adapted from International Chamber of Commerce guide for eContracting (http://www.iccwbo.org/policy/law/id3670/index.html)

Focus:
Determine jurisdiction automatically
Consequence:
• Formalize private international law procedures
• Collect connecting factors
Approach:
Example of Swiss law (IPRG) given

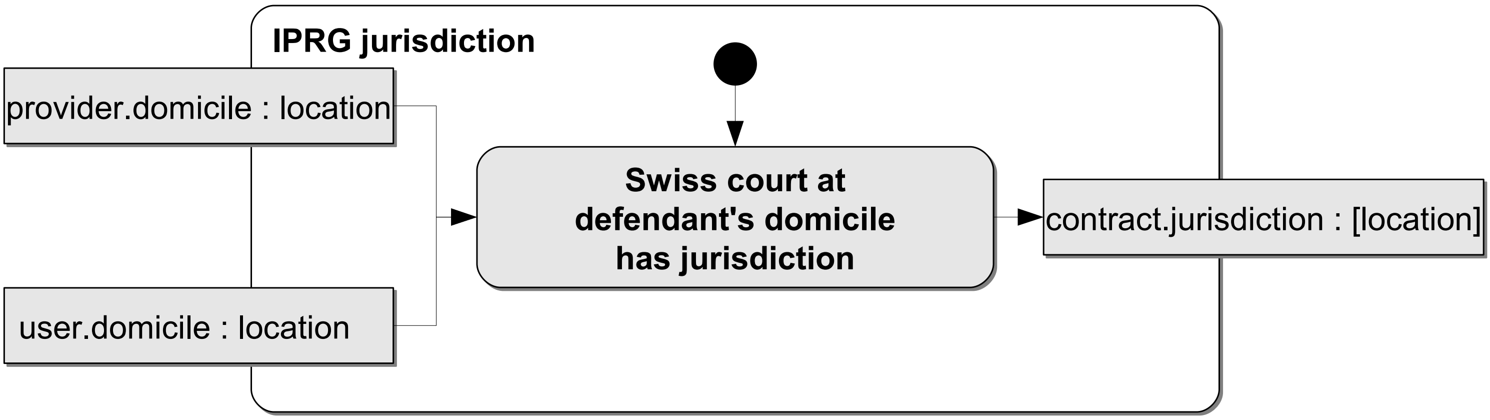
Private International Law Procedure



Direct IPRG Formalization

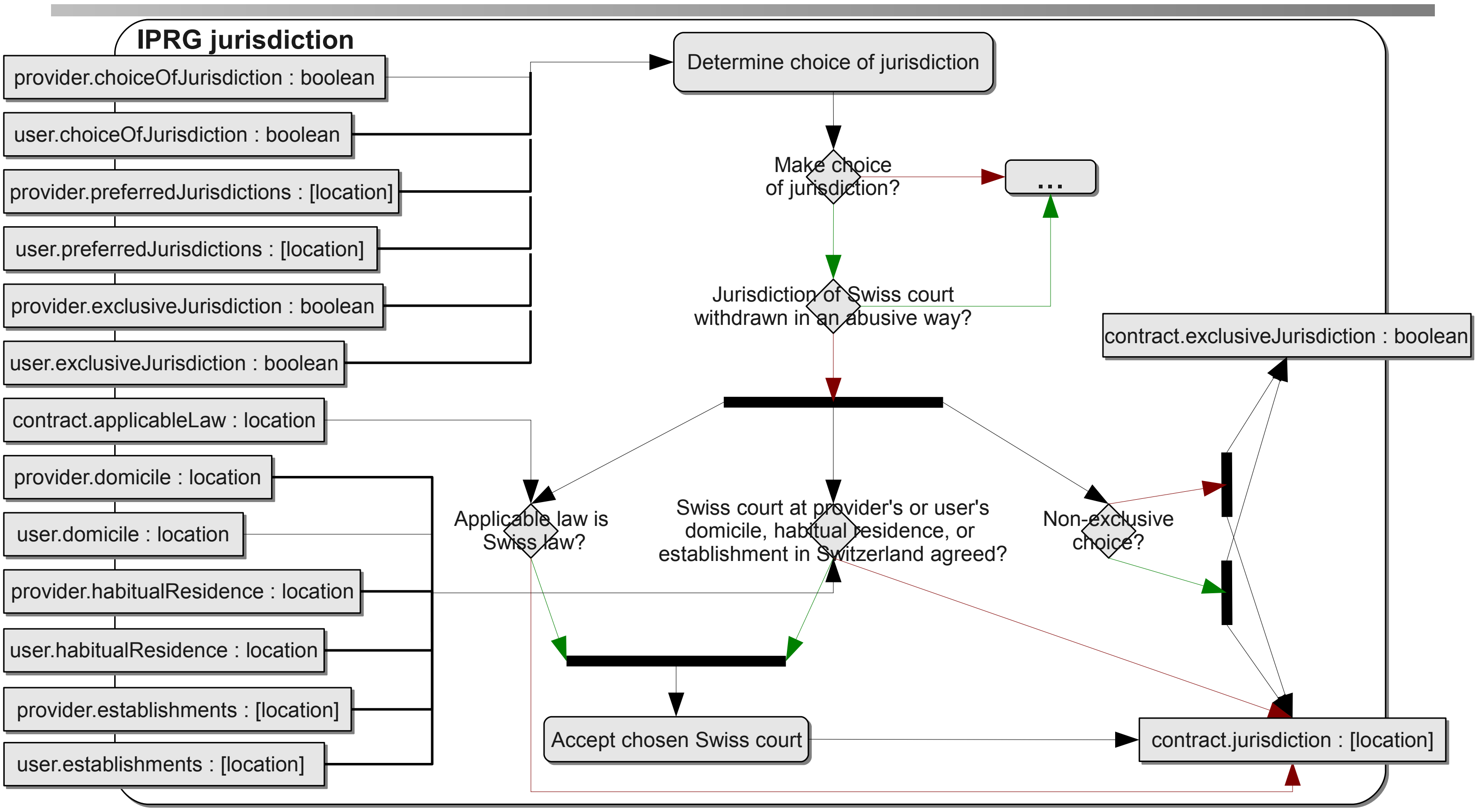
(Attempt)

General Jurisdiction Rule (IPRG)

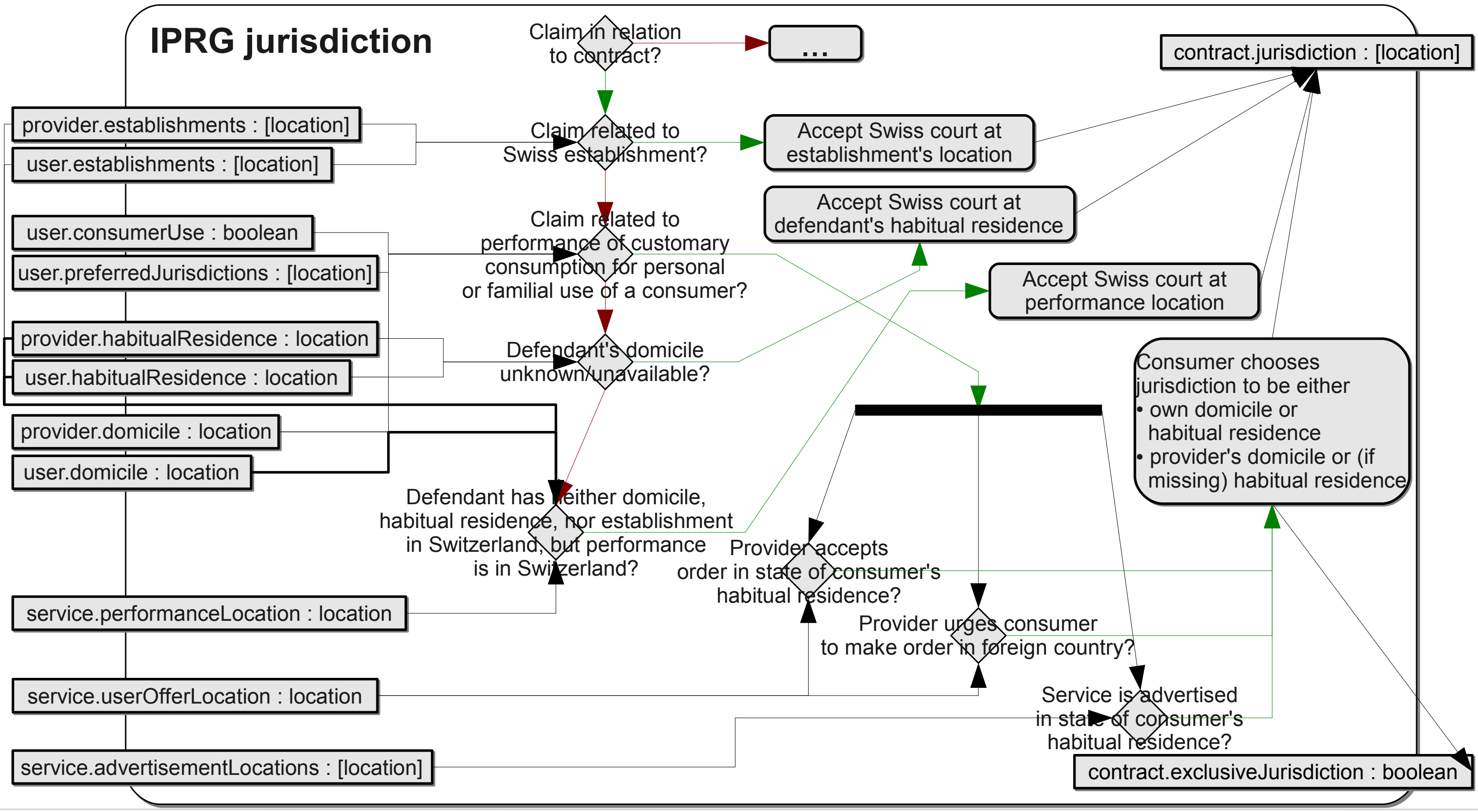


- Fallback application
- Claim-driven ↔ ex ante knowledge
- Location granularity
  - Territory
  - State

Choice of Jurisdiction (IPRG)



Claims of Contract (IPRG)



Connecting Factors (IPRG)

Provider
isNaturalPerson : boolean
domicile : location
habitualResidence : location
establishments : [location]
registeredOffice : location
effectiveAdministration : location
choiceOfJurisdiction : boolean
preferredJurisdictions : [location]
acceptableJurisdictions : [location]
exclusiveJurisdiction : boolean

User
Same as provider's connecting factors
+
consumerUse : boolean

Service
advertisementLocations : [location]
userOfferLocation : location
performanceLocation : location

Contract
exclusiveJurisdiction : boolean
jurisdiction : [location]
applicableLaw : location

- **27 connecting factors**
- **Jurisdiction-specific**
- **IPRG-specific**

Challenges and Solution

Claim-driven determinations

Case-based human interpretation

Incomplete information, ambiguity

Inconsistencies

Direct formalization...

... does not reduce complexity

... does not scale

... does not substantially improve legal certainty

... still requires human interference

... does not yield a single jurisdiction, applicable law

Connecting factors

Exception rules, interdictions

Timely order, logic interrelations

Not a single jurisdiction

→ Hypothesis-based Approach Adopted

Summary and Conclusions

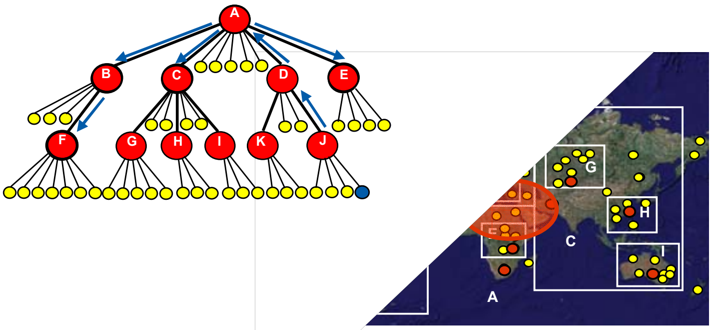
Location of key importance to international contracting
→ large set of location-related connecting factors

Direct formalization of private international law challenging
→ not reducing complexity, nor improving legal certainty

Hypothesis working approach
→ Determine relevant connecting factor settings
→ Determine manageable set of hypotheses
→ Estimate probabilities

Peer-to-Peer Approach to Location-based Search

QuaP2P-Vortragsreihe
April 16th, 2008



Aleksandra Kovačević, Nic as C. Liebau
sandra@KOM.tu-darmstadt.de, liebau@kom.tu-darmstadt.de
Tel.+49 6151 166886

TECHNISCHE UNIVERSITÄT DARMSTADT

QuaP2P

DFG Research Group QuaP2P
Technische Universität Darmstadt
Prof. Alejandro Buchmann, Ph.D.
Prof. Dr. rer. nat. Claudia Eckert
Dr.-Ing. Oliver Heckmann
Prof. Dr. Jussi Kangasharju
Prof. Dr. rer. nat. Max Mühlhäuser
Prof. Dr. rer. nat. Andreas Schürr
Prof. Dr.-Ing. Ralf Steinmetz (Speaker)
http://www.quap2p.tu-darmstadt.de/

Hungry for Čevapčići in Ireland (True Story…Almost)

TECHNISCHE UNIVERSITÄT DARMSTADT

Improving the QoS in Distributed Multimedia Communications

TECHNISCHE UNIVERSITÄT DARMSTADT

Argues that geographical location awareness can greatly help:

- enabling highly personalized services
- increasing the quality of multimedia content delivery

QoS parameters on the network layer:

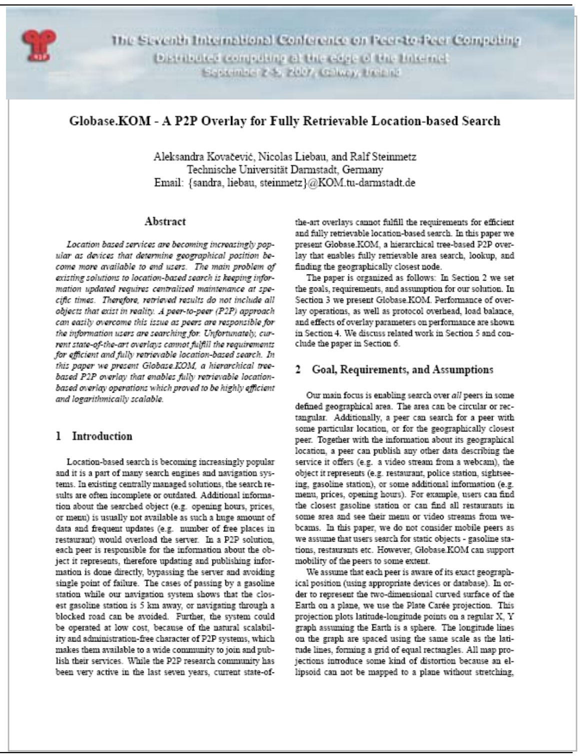
1. Loss
   - Today rarely critical
2. Throughput
   - Mainly depends on the available bandwidth
   - Available bandwidth ~ doubling every 10–14 mon
     [*A. M. Odlyzko, " Internet traffic growth: Sources and implications"*]
3. **Delay**
   - Queuing delay
     - Insignificant in comparison to other delay componen
   - Processing delay
   - **Transmission delay** $d_t = c \cdot l$

Presents Globase as an example

Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Location-aware Services**
- **Motivation for Peer-to-Peer Approach to Location-aware Services**
- **Overview of Existing P2P Solutions**
- **Our Approach: Globase.KOM**
  - Overview
  - Forming the Zones
  - Query Example: Distant Area Search
  - Area Search
  - Find the Closest Node
  - Interconnections
  - Evaluation
  - Proof of Concept: CamNet
  - Applications
- **Globase in QuaP2P**



The Seventh International Conference on Peer-to-Peer Computing
Distributed computing at the edge of the Internet
September 2-5, 2007, Galway, Ireland

Globase.KOM - A P2P Overlay for Fully Retrievable Location-based Search

Aleksandra Kovačević, Nicolas Liebau, and Ralf Steinmetz
Technische Universität Darmstadt, Germany
Email: {sandra, liebau, steinmetz}@KOM.tu-darmstadt.de

Abstract

Location based services are becoming increasingly popular as devices that determine geographical position become more available to end users. The main problem of existing solutions to location-based search is keeping information updated requires centralized maintenance at specific times. Therefore, retrieved results do not include all objects that exist in reality. A peer-to-peer (P2P) approach can easily overcome this issue as peers are responsible for the information users are searching for. Unfortunately, current state-of-the-art overlays cannot fulfill the requirements for efficient and fully retrievable location-based search. In this paper we present Globase.KOM, a hierarchical tree-based P2P overlay that enables fully retrievable location-based overlay operations which proved to be highly efficient and logarithmically scalable.

1 Introduction

Location-based search is becoming increasingly popular and it is a part of many search engines and navigation systems. In existing centrally managed solutions, the search results are often incomplete or outdated. Additional information about the searched object (e.g. opening hours, prices, or menu) is usually not available as such a huge amount of data and frequent updates (e.g. number of free places in restaurant) would overload the server. In a P2P solution, each peer is responsible for the information about the object it represents, therefore updating and publishing information is done directly, bypassing the server and avoiding single point of failure. The cases of passing by a gasoline station while our navigation system shows that the closest gasoline station is 5 km away, or navigating through a blocked road can be avoided. Further, the system could be operated at low cost, because of the natural scalability and administration-free character of P2P systems, which makes them available to a wide community to join and publish their services. While the P2P research community has been very active in the last seven years, current state-of-the-art overlays cannot fulfill the requirements for efficient and fully retrievable location-based search. In this paper we present Globase.KOM, a hierarchical tree-based P2P overlay that enables fully retrievable area search, lookup, and finding the geographically closest node.

The paper is organized as follows: In Section 2 we set the goals, requirements, and assumption for our solution. In Section 3 we present Globase.KOM. Performance of overlay operations, as well as protocol overhead, load balance, and effects of overlay parameters on performance are shown in Section 4. We discuss related work in Section 5 and conclude the paper in Section 6.

2 Goal, Requirements, and Assumptions

Our main focus is enabling search over *all* peers in some defined geographical area. The area can be circular or rectangular. Additionally, a peer can search for a peer with some particular location, or for the geographically closest peer. Together with the information about its geographical location, a peer can publish any other data describing the service it offers (e.g. a video stream from a webcam), the object it represents (e.g. restaurant, police station, sightseeing, gasoline station), or some additional information (e.g. menu, prices, opening hours). For example, users can find the closest gasoline station or can find all restaurants in some area and see their menu or video streams from webcams. In this paper, we do not consider mobile peers as we assume that users search for static objects - gasoline stations, restaurants etc. However, Globase.KOM can support mobility of the peers to some extent.

We assume that each peer is aware of its exact geographical position (using appropriate devices or database). In order to represent the two-dimensional curved surface of the Earth on a plane, we use the Plate Carée projection. This projection plots latitude-longitude points on a regular X, Y graph assuming the Earth is a sphere. The longitude lines on the graph are spaced using the same scale as the latitude lines, forming a grid of equal rectangles. All map projections introduce some kind of distortion because an ellipsoid can not be mapped to a plane without stretching,

Location-aware Services

Answering on questions:

- "*Where I am?*" = Locating
- "*What is near by? Where is …*" = Searching
- "*How can I go to?*" = Navigating

Location-aware services…

- are highly attractive for end-users and providers
- they can offer highly personalized services based on the user's location

Example:

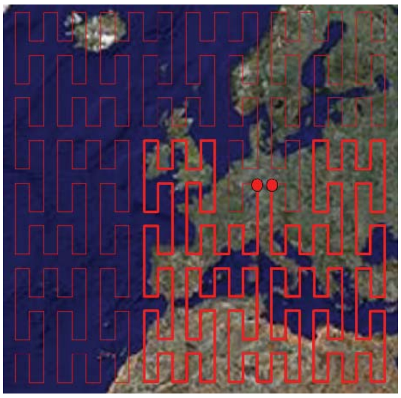
- List Italian restaurants within walking distance

Closest Italian restaurant?

Overview of Existing Peer-to-Peer Solutions

Re-using existing structured overlays using space filling curves [Chawathe05][Zhou03]

- Inefficient – geographically close nodes are not always close in overlay
- Not fully retrievable

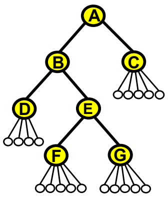


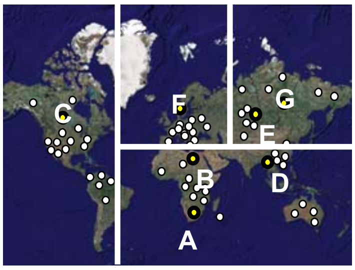
Mapping of the physical space into the CAN [Wang05]

- Lookup complexity
- Performance bottleneck at the higher level nodes

RectNet - binary distributed space partitioning tree [Heutelbeck05]

- Binary structure significantly reduces the search performance



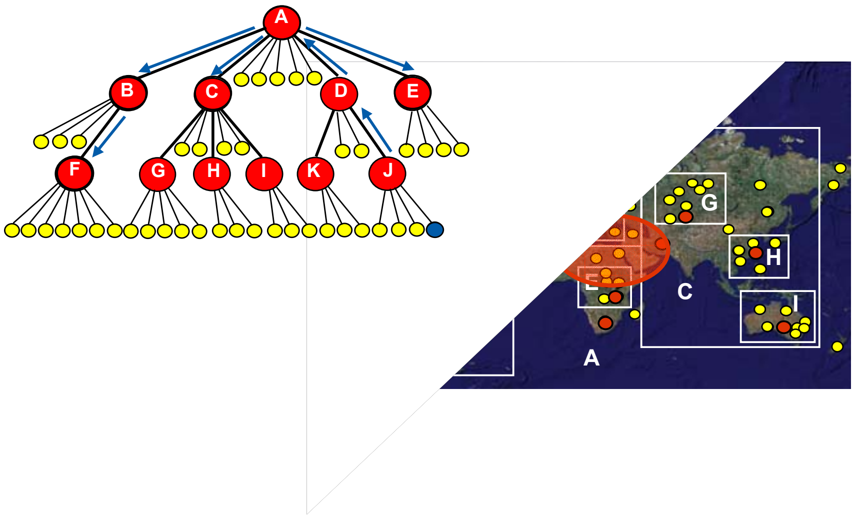


Our Approach: Globase.KOM

TECHNISCHE UNIVERSITÄT DARMSTADT

Globase.KOM: Design

- **Overlay Structure**
  - Overview
  - PeerID
  - Forming the Zones
  - Interconnections
- **Overlay Operations**
  - (Lookup)
  - Area Search
  - Find the Closest Node
  - (Join)
  - (Leave)
- **(Failure Recovery)**



Evaluation

Proof of Concept: CamNet

Overview

TECHNISCHE UNIVERSITÄT DARMSTADT

Globase.KOM
(**G**eographical **LO**cation-**BA**sed **SE**arch):

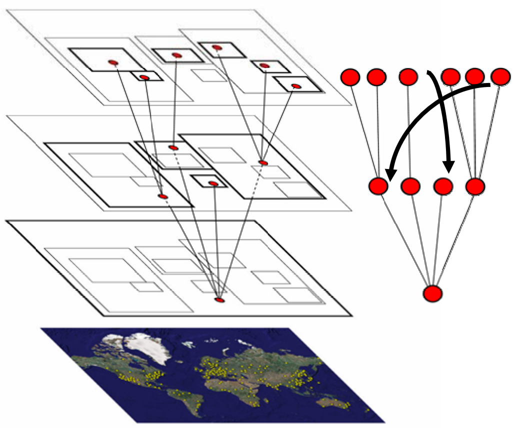
Superpeer-based overlay:

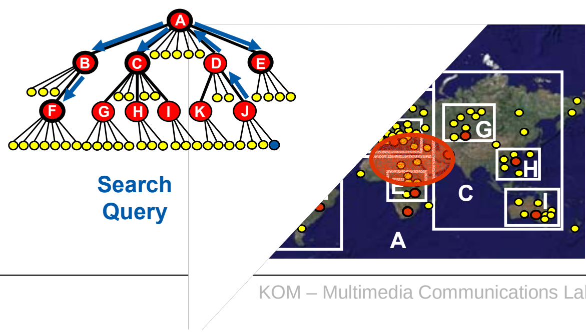
- Tree enhanced with interconnections
- Each zone assigned to a superpeer in the zone

World divided into rectangular, not overlapping zones

Failure recovery and search efficiency → interconnections:

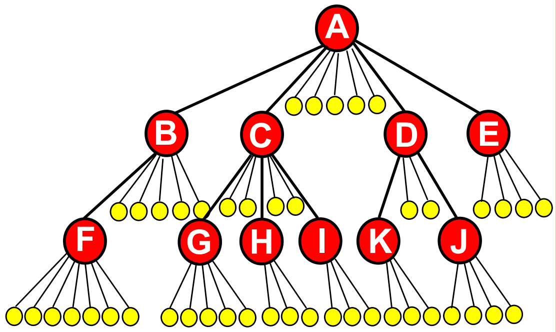
- Learning from received messages
- Enabling tree awareness

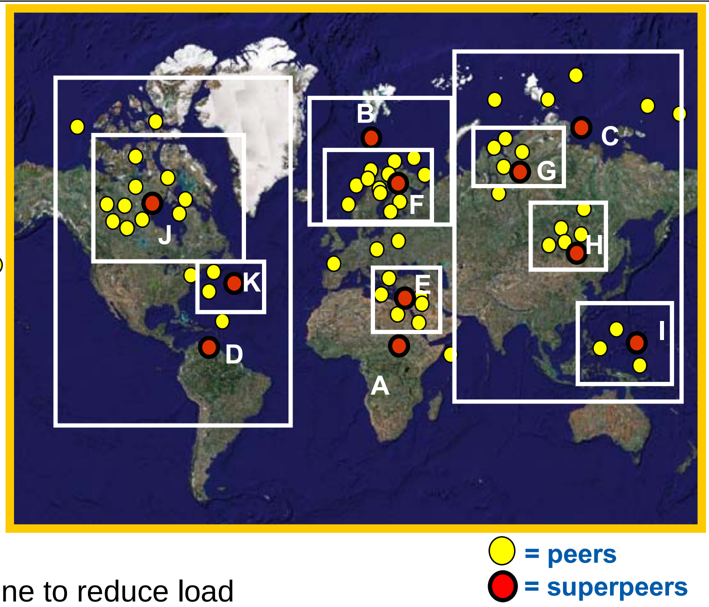


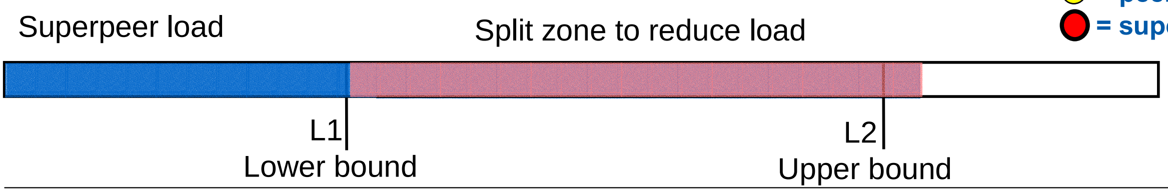


Forming the Zones

TECHNISCHE UNIVERSITÄT DARMSTADT

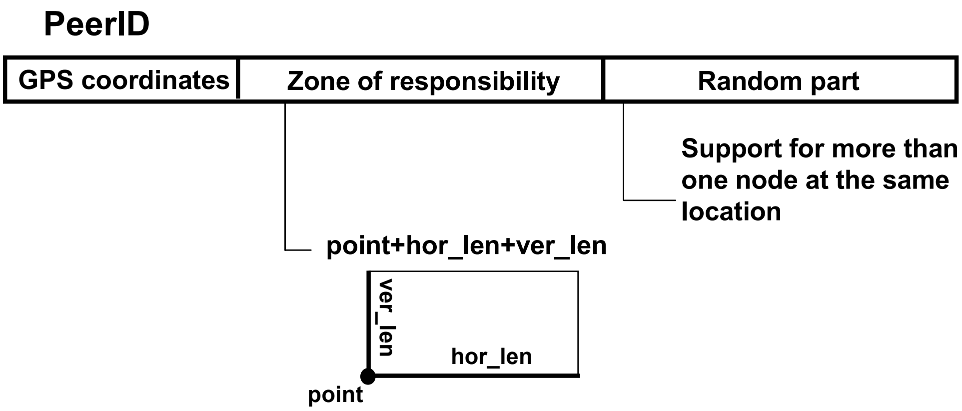






PeerID

TECHNISCHE UNIVERSITÄT DARMSTADT

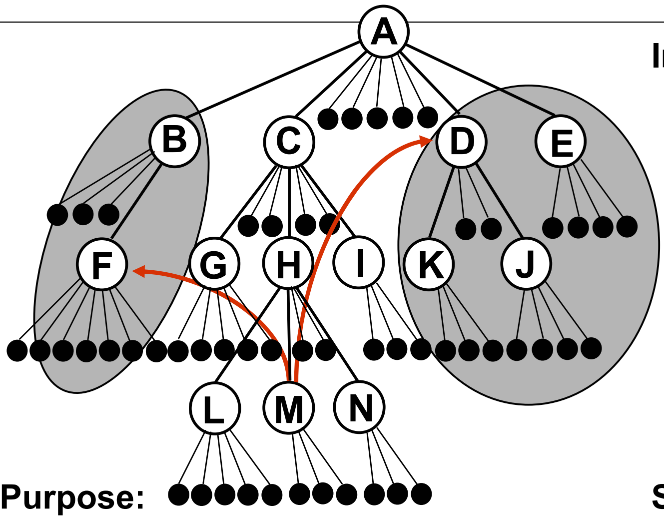


Advantages:

- Determining superpeer responsible for a given point without contacting it
- Controlled peer ID assigning process

Interconnections

TECHNISCHE UNIVERSITÄT DARMSTADT



Interconnections

- Additional routing tables

Purpose:

- Failure recovery
- More efficient search

Setting-up interconnections

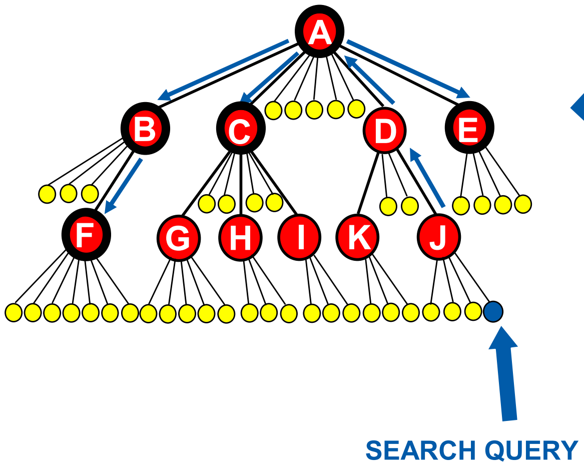
- Learning from received messages
- Obtaining from neighbours
- Enabling tree awareness

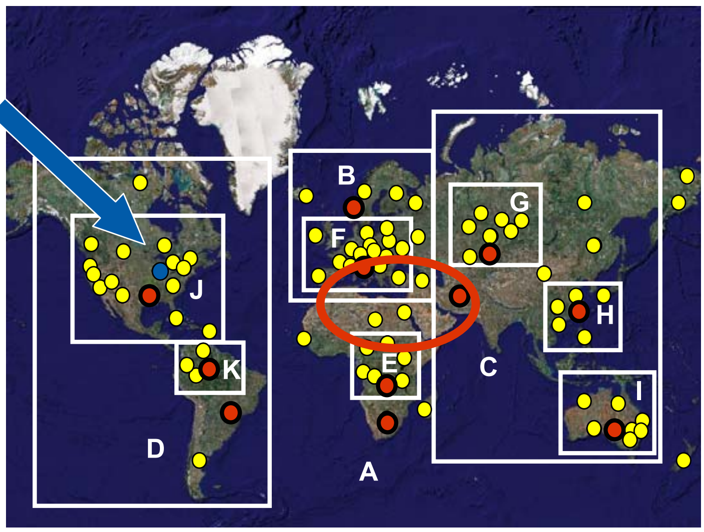
Theoretically:

- From tree → to expander graph

Query Example: Distant Area Search

TECHNISCHE UNIVERSITÄT DARMSTADT



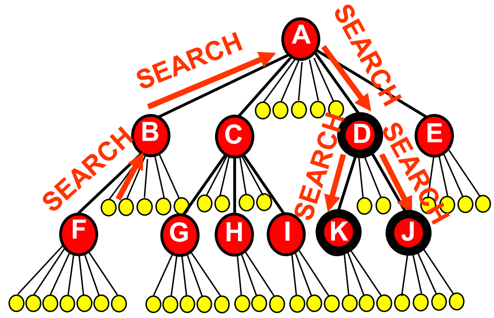


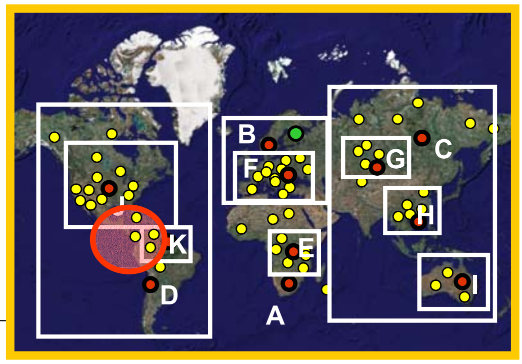
Steps for resolving area search:

- Query started at node J
- Towards responses from nodes A, B, C, E, and F

Area Search

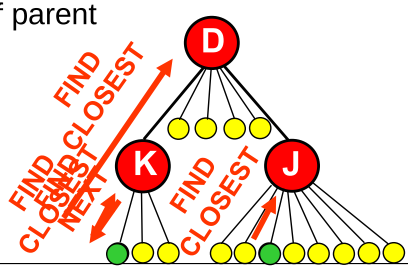
- **SEARCH:**
  - Center and radius
  - Query initiator + resp. superpeer
  - Sequence number
- **Superpeer calculates the searched ellipse onto the map projection**
  - Inside of my zone?
    - Me or children?
  - Inside of "interconnected zones"?
  - Forward message or send SEARCH_RESULT
- **Optimal value of a *timeout* is 2 sec (simulation results)**
- **For each received message interconnections are kept**

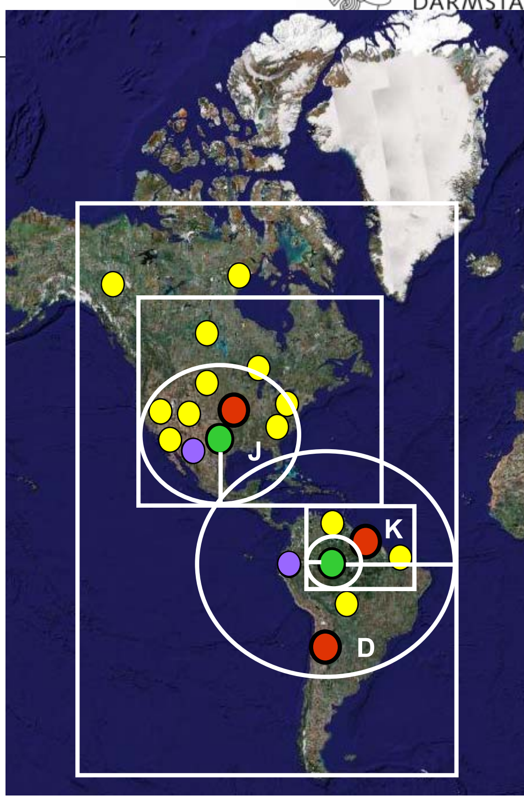




Find the Closest Node

1. **Calculate the closest border of the superpeer's zone (using parent's ID)**
2. **FIND_CLOSEST**
   - Calculated distance
3. **FIND_CLOSEST_RESULT**
4. **FIND_CLOSEST_NEXT**
   - Address of parent





Failure Recovery

TECHNISCHE UNIVERSITÄT DARMSTADT

Hybrid overlay structure requires:

- Fast superpeer failure detection
- Fast superpeer failure recovery

Failure Detection:

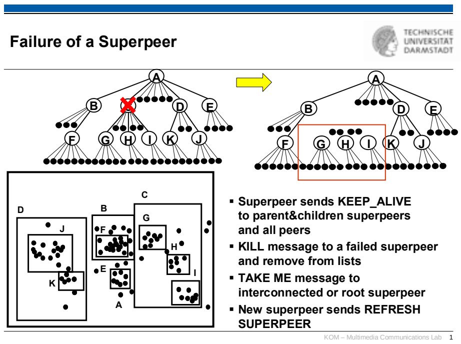
- **Active for superpeers:**
  - Periodical keep-alive messages (2 sec)
- **Passive for peers:**
  - Liveness information „on-demand“

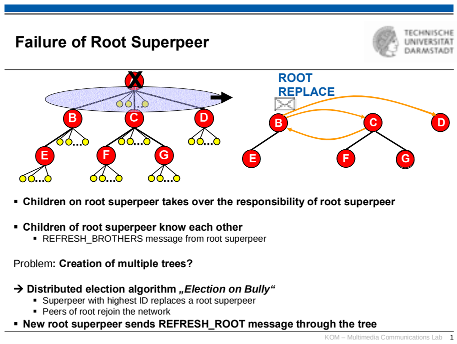
Failure Detection:

Best case

- **Using appropriate interconnections**
- **Contacting root superpeer**
- **Rejoining**

Worst case





Evaluation

TECHNISCHE UNIVERSITÄT DARMSTADT

Steps for a Performance Evaluation Study

[Jain *„The Art of Computer Systems Performance Analysis*]

1. **Evaluation Goals**
2. **List system services and possible outcomes**
3. **Metrics**
4. **Parameters**
5. **Factors**
6. **Techniques**
7. **Model**
8. **Design Experiments**
9. **Result Analysis and Interpretation**
10. **Result Presentation**

Graphic source: [Heckmann, Steinmetz et.al. 2006 a]

Evaluation Setup

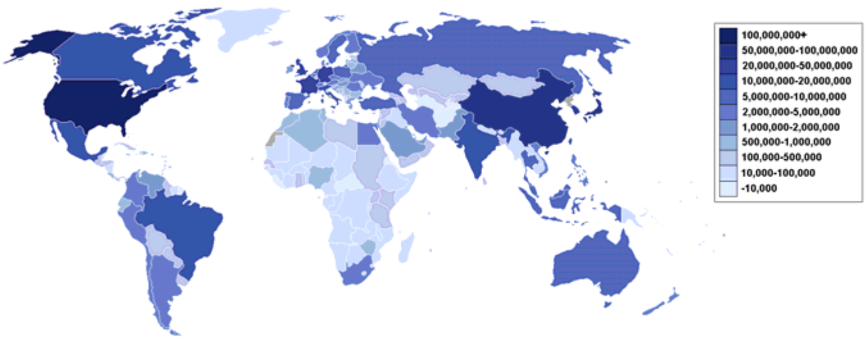
- **Metrics:**
  - Number of hops
  - Operation duration
  - Relative delay penalty (RDP)
  - Number of received messages
  - Depth and breadth of the tree
- **Parameters:**
  - L1 and L2
  - KEEPALIVE intervals and TIMEOUT
- **Factors:**
  - Number of users
  - Local vs. Distant Area Search
- **Techniques:**
  - Analytical
  - Measurements
  - Simulation

$$RDP = \frac{\sum OverlayDelays}{UnderlayDelay}$$

Model

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Implementation in PeerfactSim.KOM**
- **User Model:**
  - Peer distribution



  - User behavior
    - Relationship between location and online/offline time in day/week
    - User requests - Zipf distribution

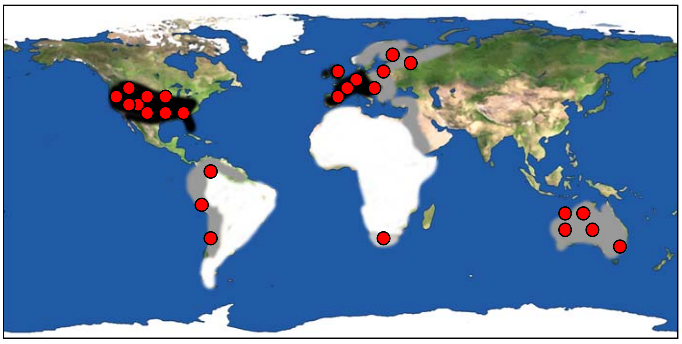
Source: http://upload.wikimedia.org/wikipedia/commons/7/7f/Internet_users_by_country_world_map.PNG

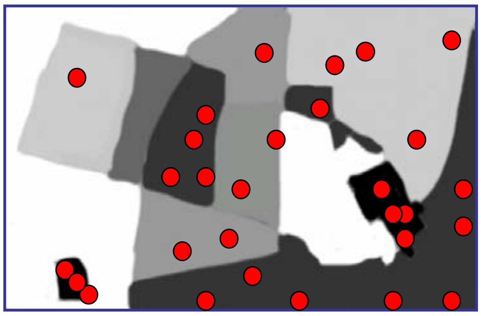
Bitmap-based Peer/Churn Distribution

TECHNISCHE UNIVERSITÄT DARMSTADT

Distribution of peers

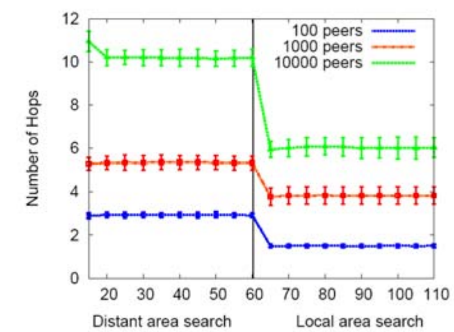
- Black = High density
- White = Low density

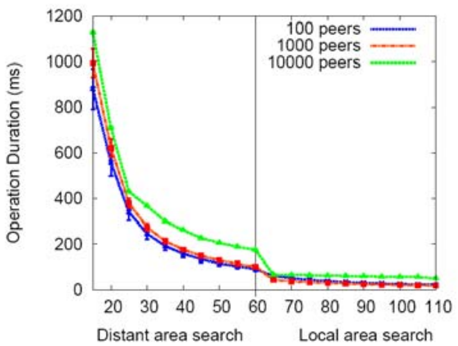




Area Search: Distant vs. Local

TECHNISCHE UNIVERSITÄT DARMSTADT





Globase.KOM	**L1 = 10, 25, 55** **L2 = 20, 50, 110**
Scenario	**Ideal + Churn MixLogNormal**
Experiment size	**100, 1.000, 10.000 peers**

First 60 ms start-up + distant area search
Next 50 ms local area search

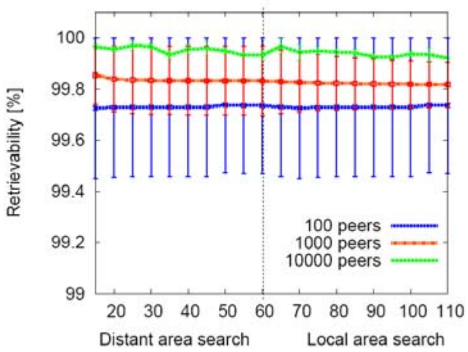
Number of hops:

- Less number of superpeers contacted

Operation duration

- Distance area search after start up phase similar to local area search
- Significance of interconnections

Area Search: Retrievability

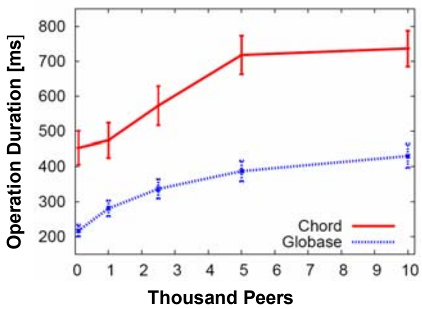


Globase.KOM	**L1 = 10, 25, 55** **L2 = 20, 50, 110**
Scenario	**Ideal + Churn MixLogNormal**
Experiment size	**100, 1.000, 10.000 peers**

Retrievability:

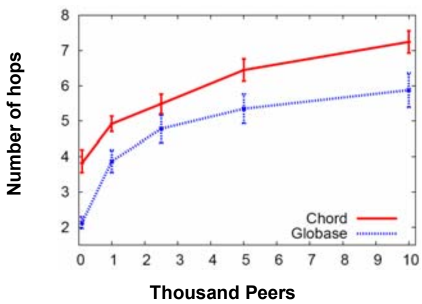
- In worst case only 0.3% results not delivered
- Missing results because of Search-Timeout
- No difference between distant and local area search
- No effects of interconnections

Scalability



Globase.KOM	**L1 = 10, 25, 31, 41, 55** **L2 = 20, 50, 62, 84, 110**
Chord	**10 successors,** **650 ms stabilization interval**
Experiment size	**100, 1.000, 2.500, 5.000, 10.000 peers**

Globase.KOM and Chord scale logarithmically

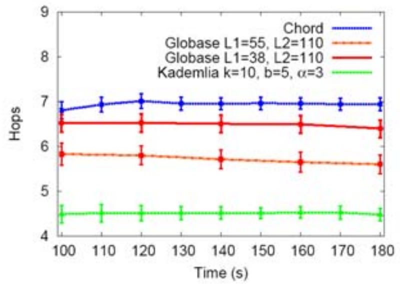


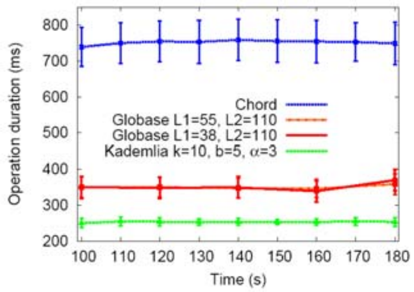
Globase.KOM vs. Chord

- Less operation duration (~ 70%)
- Less hops (~ 20%)

Lookup: Performance

TECHNISCHE UNIVERSITÄT DARMSTADT





Chord	**10 successors, 650 ms stabilization interval**
Scenario	**Ideal + Churn MixLogNormal**
Experiment size	**10.000 peers**

Not main focus of Globase.KOM!

Number of hops

- Chord needs on average 22.8% more hops than Globase.KOM [L1 = 55, L2 = 110]
- Kademlia performs 21% better than Globase.KOM due to big contact lists

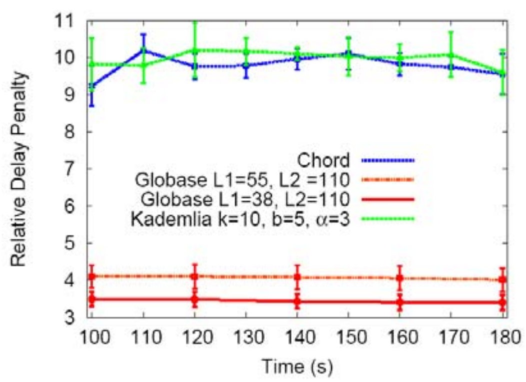
Duration of operations

- Globase.KOM [L1 = 55, L2 = 110] needs 38.4% more time than Kademlia but 53.5% less time than Chord

→Reason for shorter operation duration - Higher degree of underlay-awareness of Globase.KOM

Lookup: Costs - Relative Delay Penalty

TECHNISCHE UNIVERSITÄT DARMSTADT



Chord	**10 successors, 650 ms stabilization interval**
Scenario	**Ideal + Churn MixLogNormal**
Experiment size	**10.000 peers**

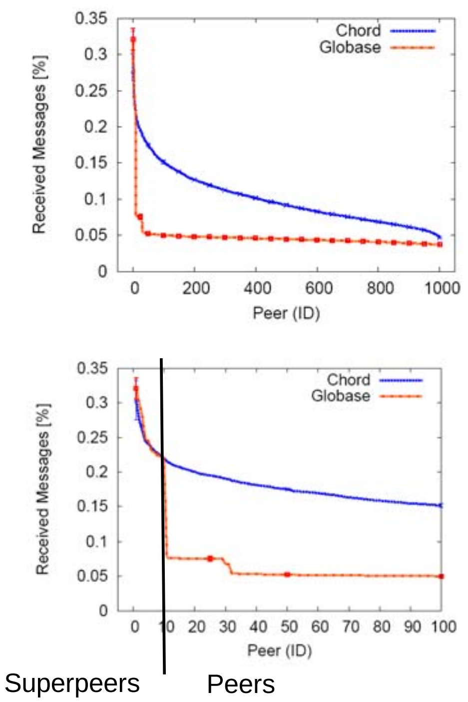
Globase.KOM

- RDP ~ 60% less than Chord and Kademlia

→Building of overlay more underlay aware

Lookup Costs: Load Balancing (1)

TECHNISCHE UNIVERSITÄT DARMSTADT



Chord	10 successors, 650 ms stabilization interval
Globase.KOM	L1 = 25, L2 = 50
Scenario	Ideal + Churn MixLogNormal
Experiment size	2000 peers

Chord

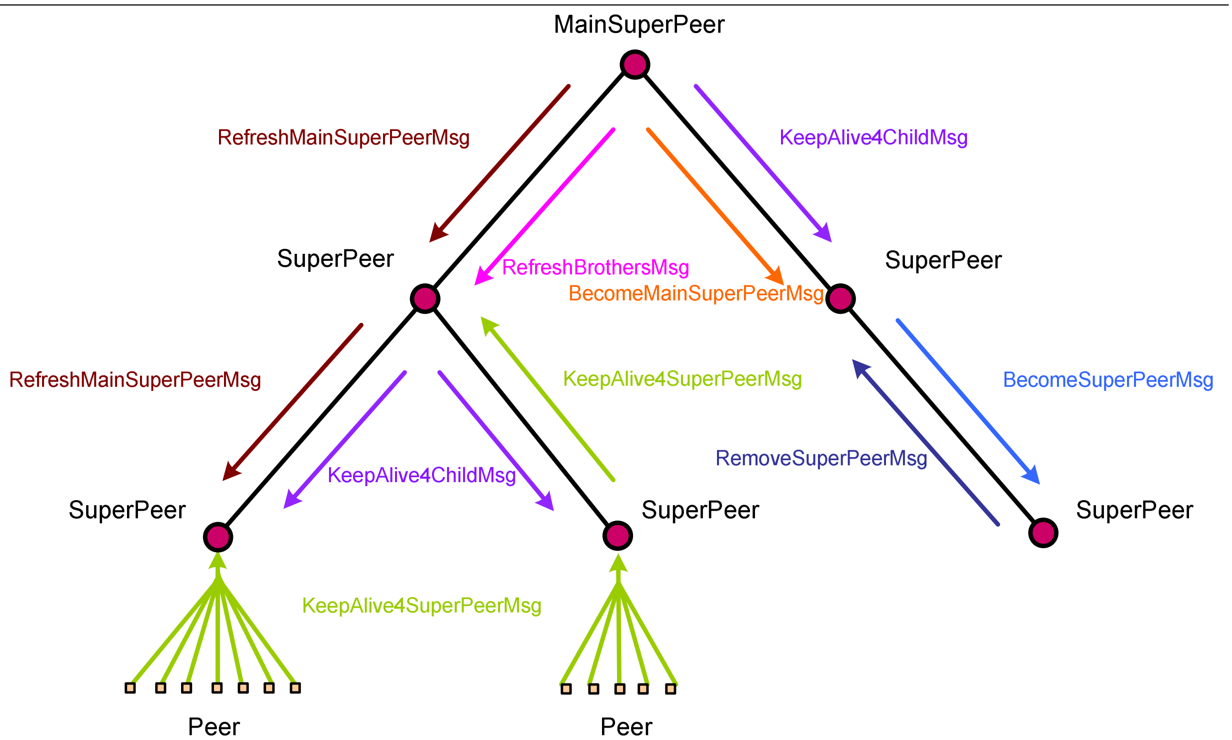
- Received messages per peer: 0,05 – 0,31%

Globase.KOM

- Received messages per peer: 0,04 – 0,32%
  - Peers: 0,04 – 0,07%
  - Superpeers: 0,21 – 0,32%
  - Root superpeer: 0,32%
- The higher a superpeer in hierarchy, the higher the load

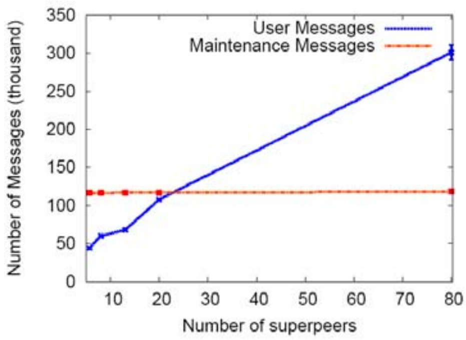
Globase.KOM: Protocol Overhead

TECHNISCHE UNIVERSITÄT DARMSTADT



Protocol Overhead

TECHNISCHE UNIVERSITÄT DARMSTADT



Globase.KOM	L1 =25, 50, 75, 100, 125, 150, 175 L2 = 50, 100, 150, 200, 250, 300, 350
Experiment size	2.000 peers

More superpeers → smaller zones → more superpeers involved in resolving the queries

Number of maintenance messages stays constant

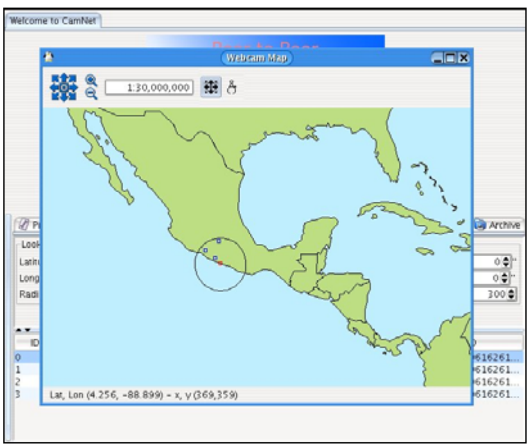
- More than 97.3% of the sent maintenance messages between peers and their responsible superpeer

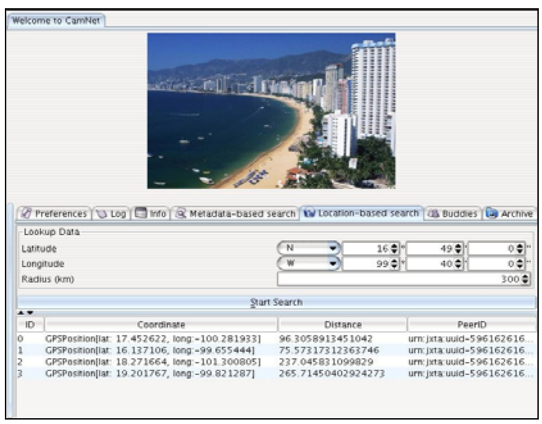
In optimal case (L2=100), number of maintenance messages - up to 50% smaller than the number of user messages

Proof of Concept: CamNet

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Peer-to-Peer network of webcams**
- **Efficient P2P location-based area search:**
  - "Find all webcams in a certain area"

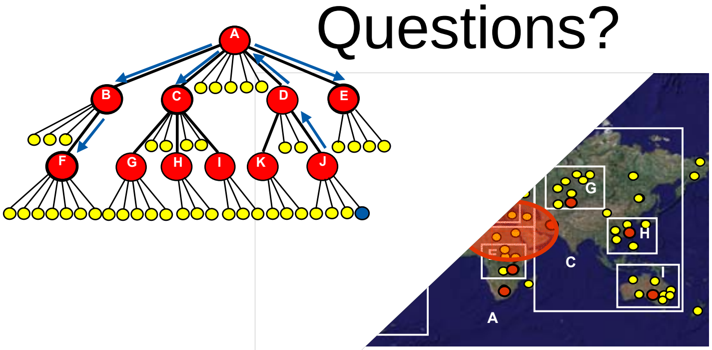




Thank You for Your Attention

Peer-to-Peer Approach to
Location-based Search
QuaP2P Vortragsreihe
April 16th, 2008

Questions?



Aleksandra Kovačević, Nicolas C. Liebau
sandra@KOM.tu-darmstadt.de, liebau@kom.tu-darmstadt.de
Tel.+49 6151 166886

DFG Research Group QuaP2P
Technische Universität Darmstadt
Prof. Alejandro Buchmann, Ph.D.
Prof. Dr. rer. nat. Claudia Eckert
Dr.-Ing. Oliver Heckmann
Prof. Dr. Jussi Kangasharju
Prof. Dr. rer. nat. Max Mühlhäuser
Prof. Dr. rer. nat. Andreas Schürr
Prof. Dr.-Ing. Ralf Steinmetz (Speaker)
http://www.quap2p.tu-darmstadt.de/

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

European Seventh Framework STREP FP7-2007-ICT-216259

Peer Locality Information using BGP

UZH, DoCoMo, TUD, AUEB, PrimeTel, AGH, ICOM, UniWue, TID

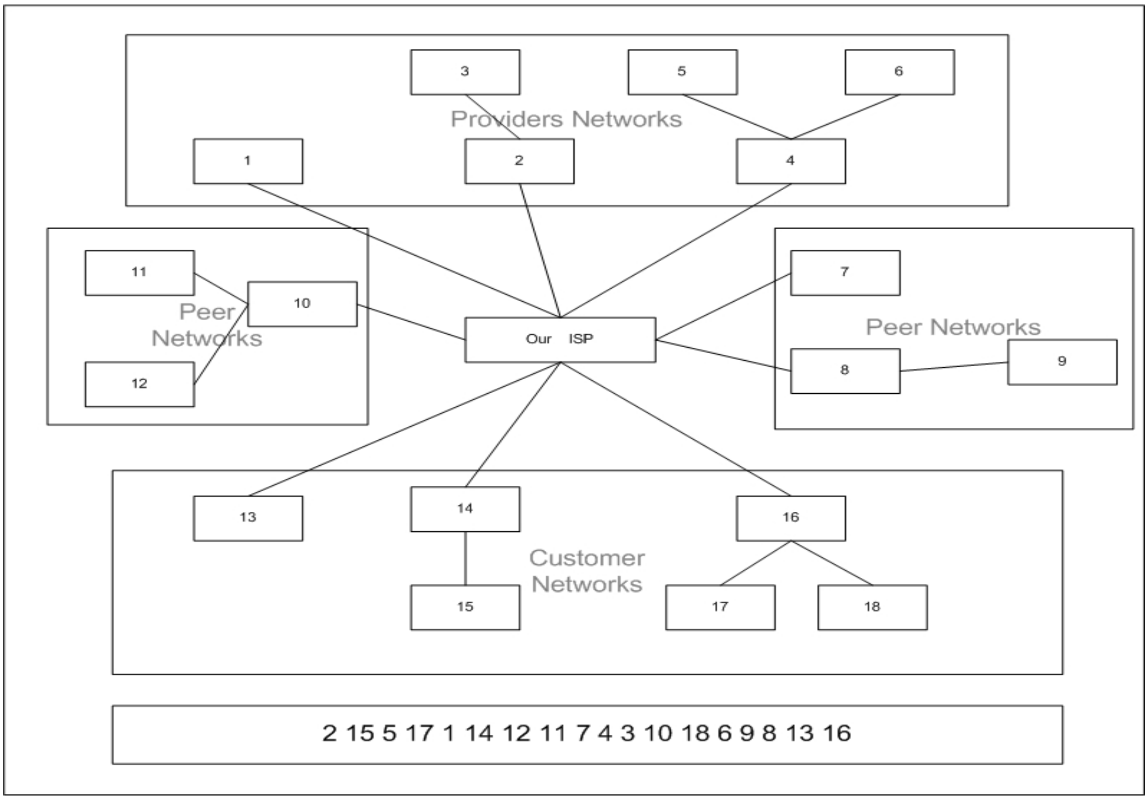
Amruth Kumar Juturu, Peter Racz, Fabio Hecht, Hasan, Burkhard Stiller

Workshop on Economic Traffic Management
Zurich
August 04 - 05, 2008

Outline

- Problem Statement
- Objective
- Overview of BGP
- Architecture
- Peer Prioritizing algorithm
- Softwares Used
- Implementation
- Evaluation
- Performance

Problem Statement



Objective

- To provide locality information of peers with respective to querying peer.
- Border Gateway Protocol (BGP) is used to extract information on locality of peers.
- ISP maintained Databases are also used as sources of information.

Overview of BGP

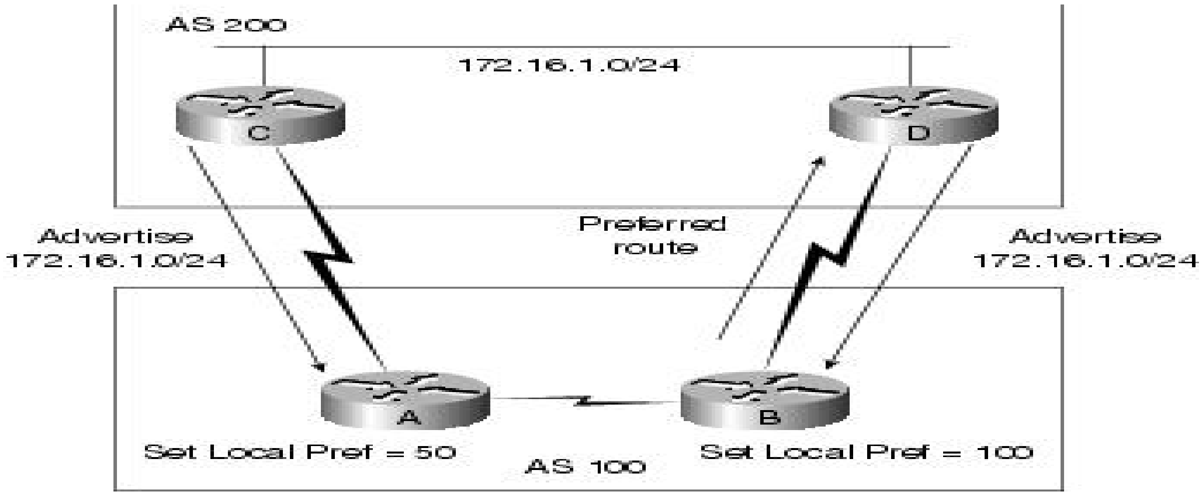
- Inter-Autonomous System (AS) routing protocol.
- Autonomous System: A network or group of networks under a common administration and with **common routing policies**.
- ISPs use BGP to exchange customer and ISP routes with other ISPs.
- When BGP is used between ASes, the protocol is referred to as External BGP (EBGP).
- If BGP is used to exchange routing information between routers in same AS, it is called Interior BGP (IBGP).

BGP route attributes

- Routes learned via BGP have associated properties called attributes.
- BGP selects best route to a destination from multiple routes.
- Selection is influenced by these attributes. They are
  - Weight
  - Local Preference
  - AS Path
  - Origin Code
  - Multiple Exit Discriminator (MED)

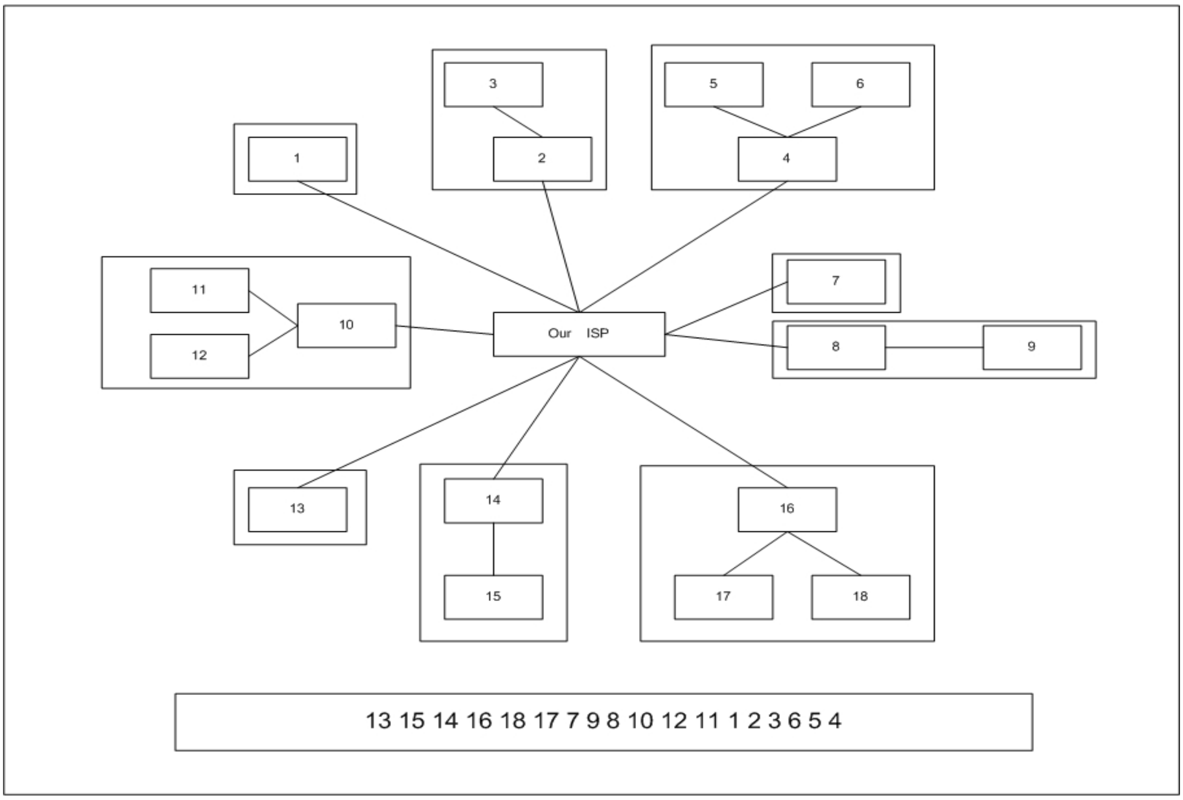
Local Preference

- Used to specify an exit point from within an AS to a destination.
- Route with highest Local Preference value is selected.



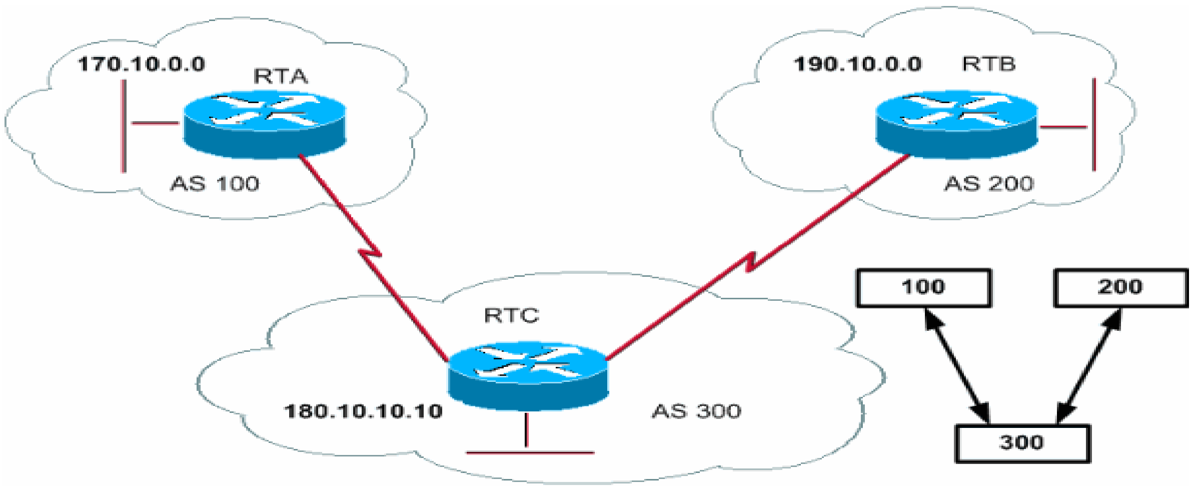
- Generally ISPs set Local Preference values reflecting the type of relations.
- For example
  - 50 - 99 might represent providers
  - 100 - 149 might represent peering agreements
  - 150 - 199 might represent customers

Example

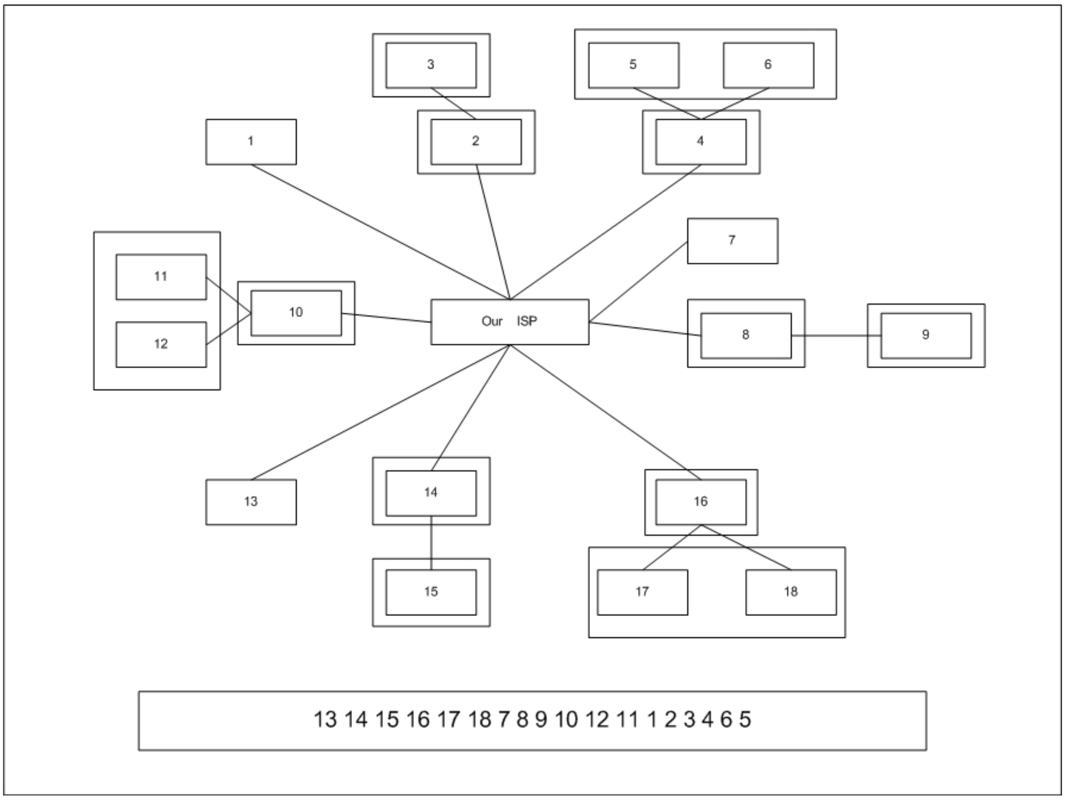


AS-PATH

- List of ASes traversed in the path to reach destination.
- Path with least number of AS hops selected.

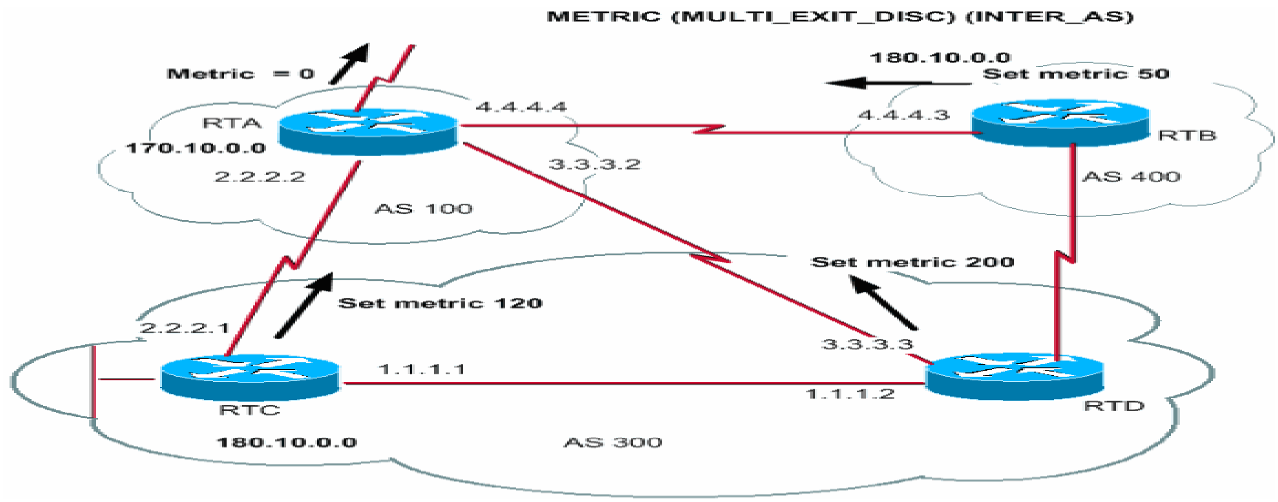


Example

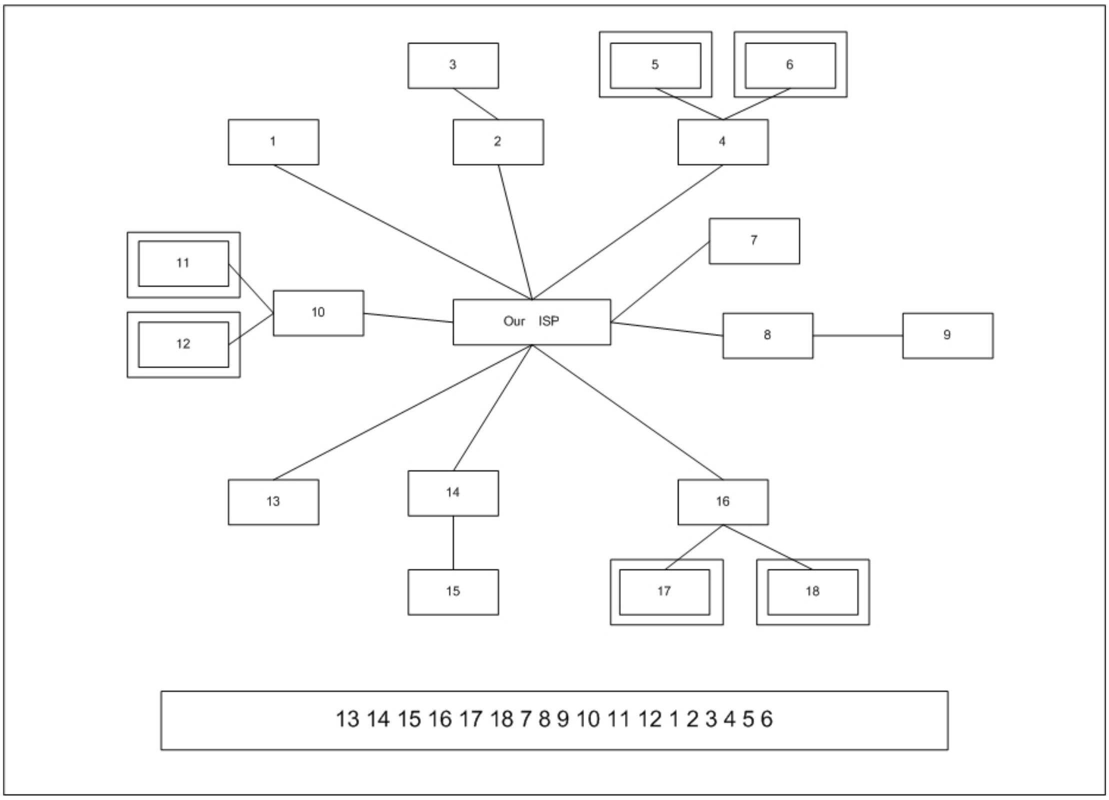


Multi Exit Discriminator (MED)

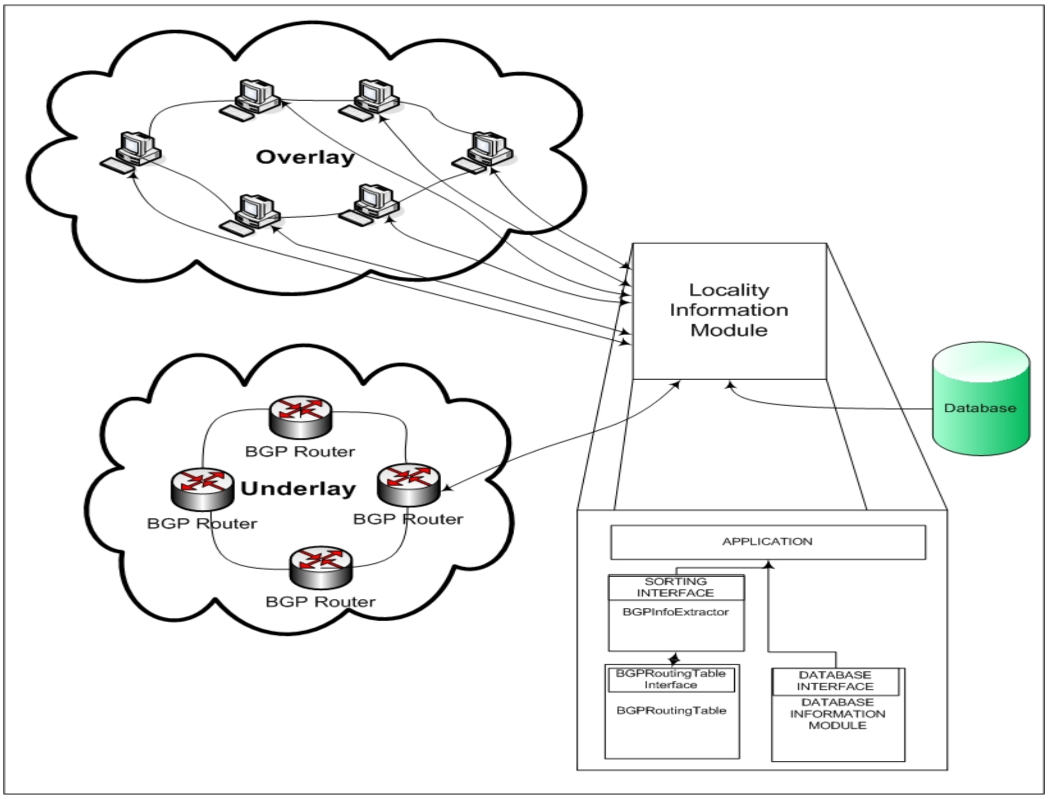
- Specifies a router's preference for an entry point into an autonomous system.
- A lower value is preferred.
- Also called Metric.



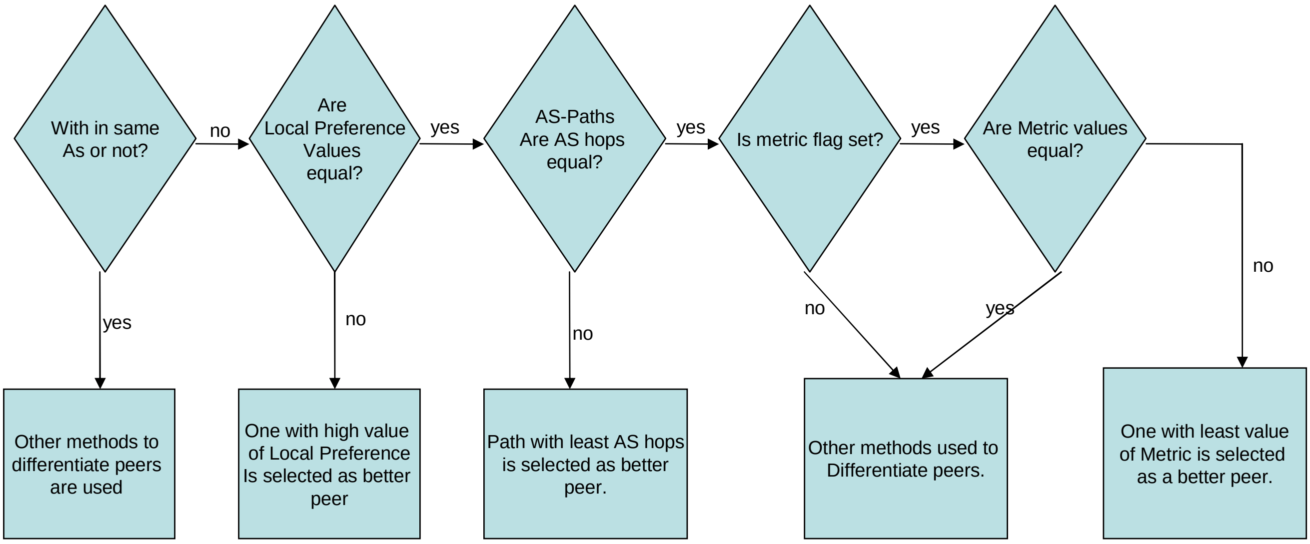
Example



Architecture



Peer Prioritizing Algorithm



Softwares Used

- Tools:
  - Quagga
  - Net-Snmp
  - MySql

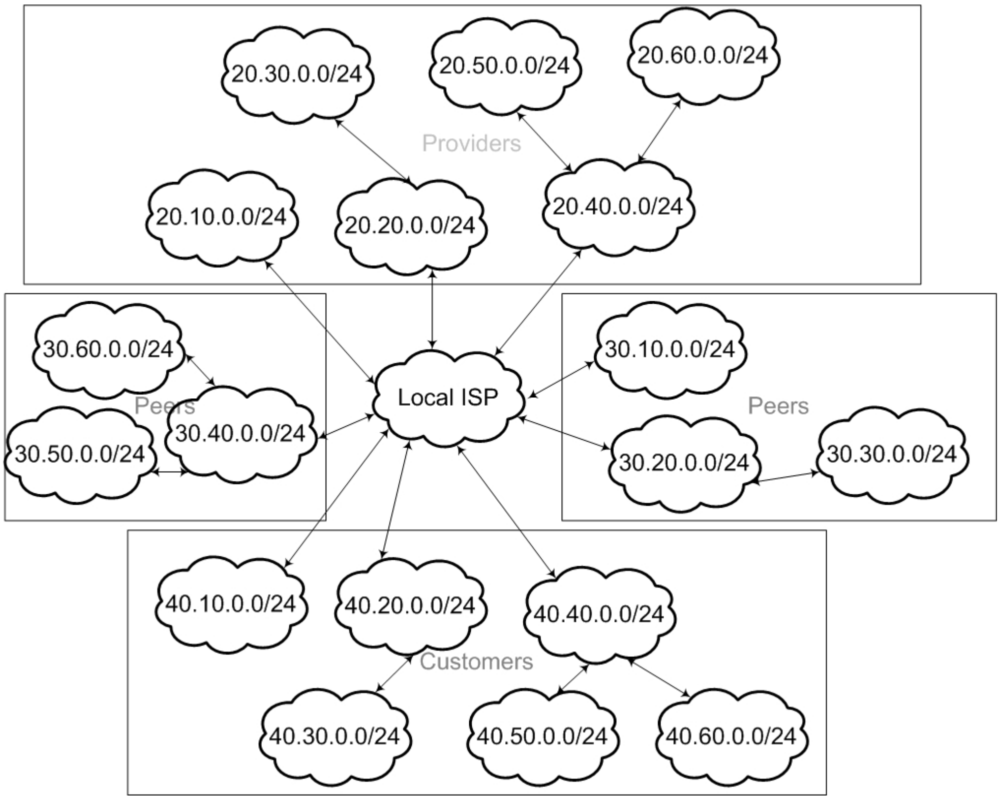
- Libraries
  - Snmp4j
  - JDBC

Implementation

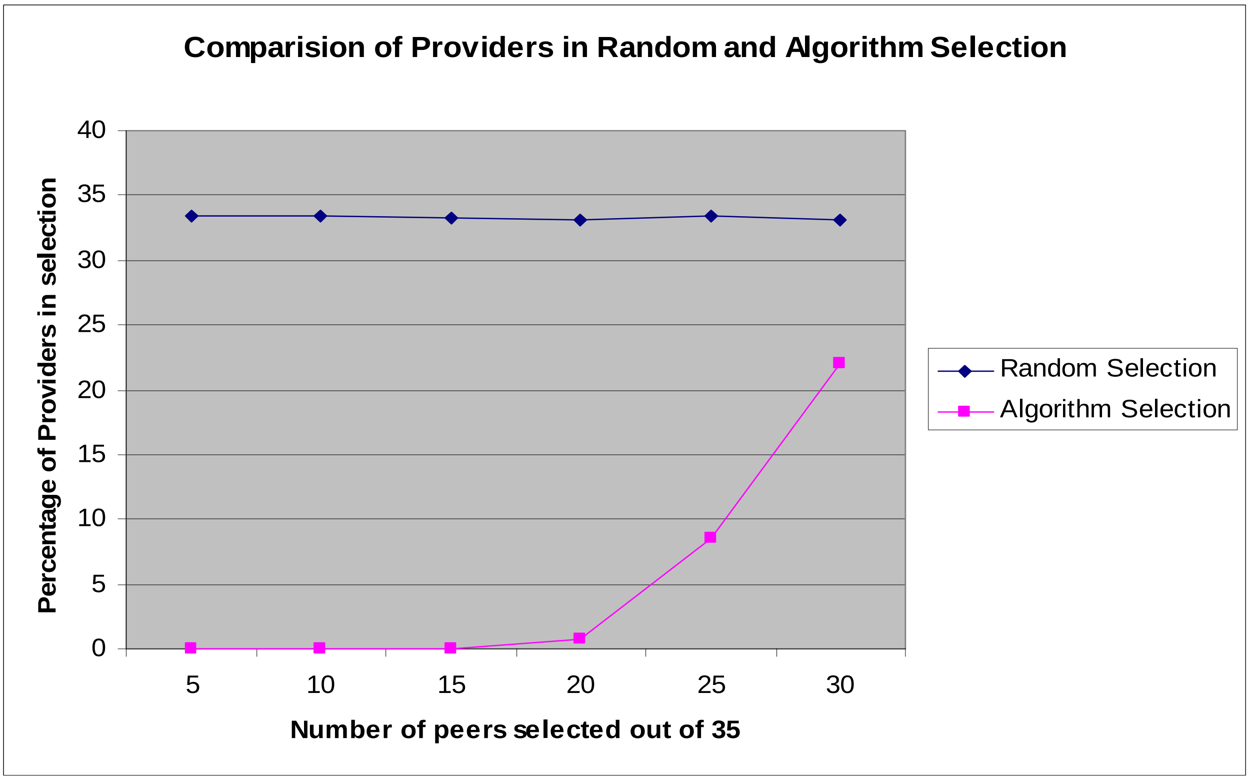
- BGPRoutingTableBuilder
  – Connects to BGP router through SNMP and reads subtree at OID 1.3.6.1.2.1.15 .
- BGPRoutingTable
  – BGP routing table read from the router.
- BGPRoutingTableEntry
  – Represents a BGP routing table entry
- BGPRoutingTableInterface
  – Interface for routing table that gives BGP routing table entry for a given IP address.

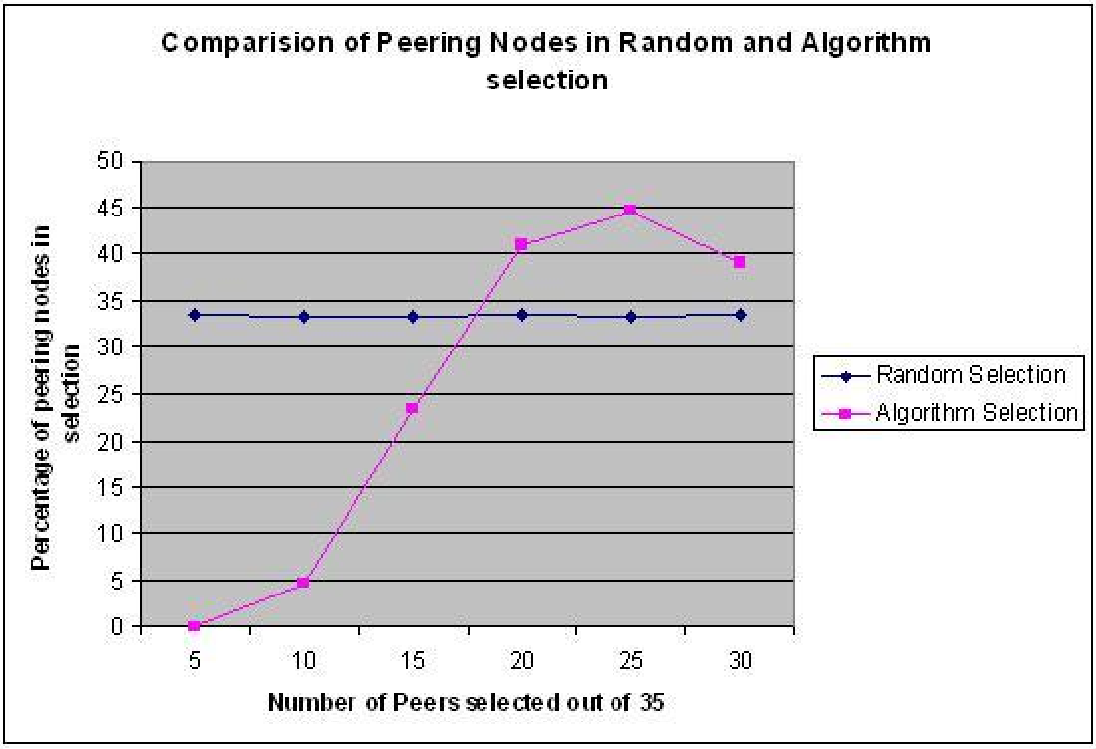
- BGPInfoInterface
  – Interface for getting individual BGP route attributes from routing table.

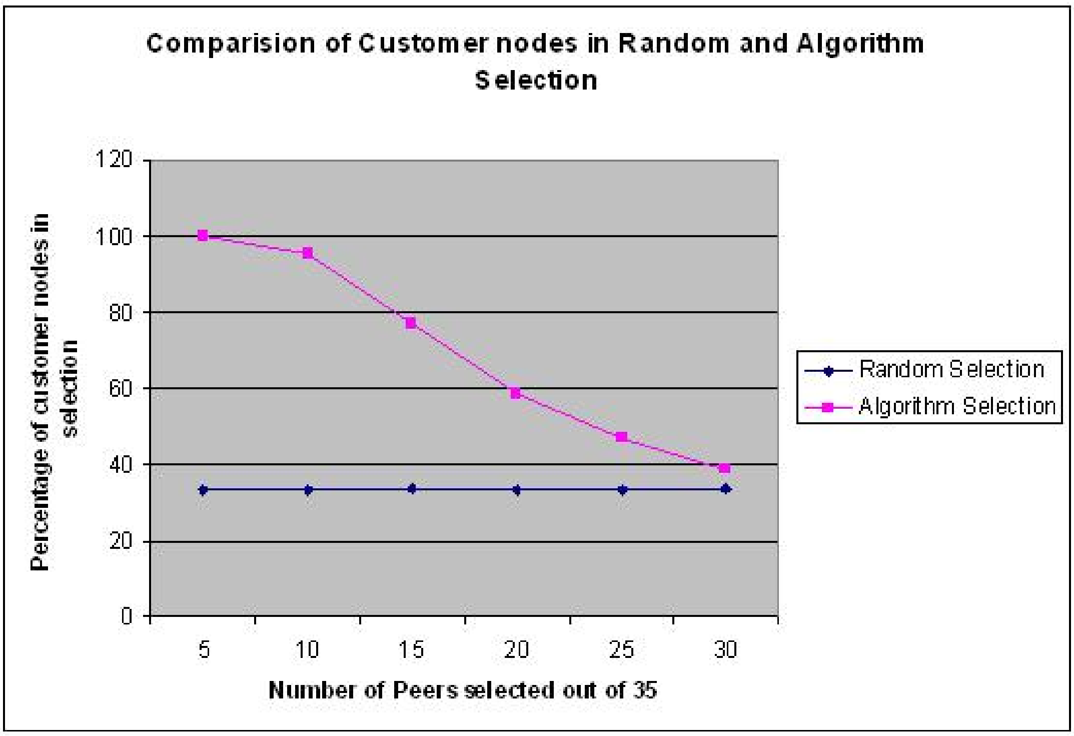
Sample Network Scenario



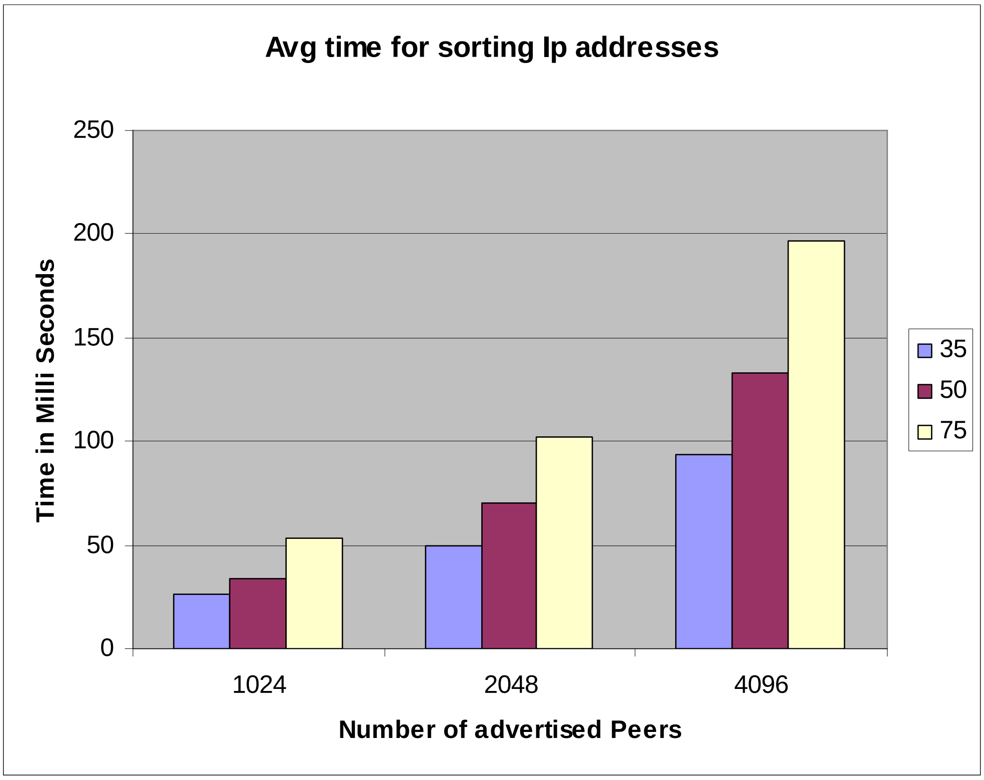
Evaluation Results




Comparision of Peering Nodes in Random and Algorithm selection
Percentage of peering nodes in selection
50
45
40
35
30
25
20
15
10
5
0
5
10
15
20
25
30
Number of Peers selected out of 35
Random Selection
Algorithm Selection


Comparision of Customer nodes in Random and Algorithm Selection
Percentage of customer nodes in selection
120
100
80
60
40
20
0
5
10
15
20
25
30
Number of Peers selected out of 35
Random Selection
Algorithm Selection

Performance

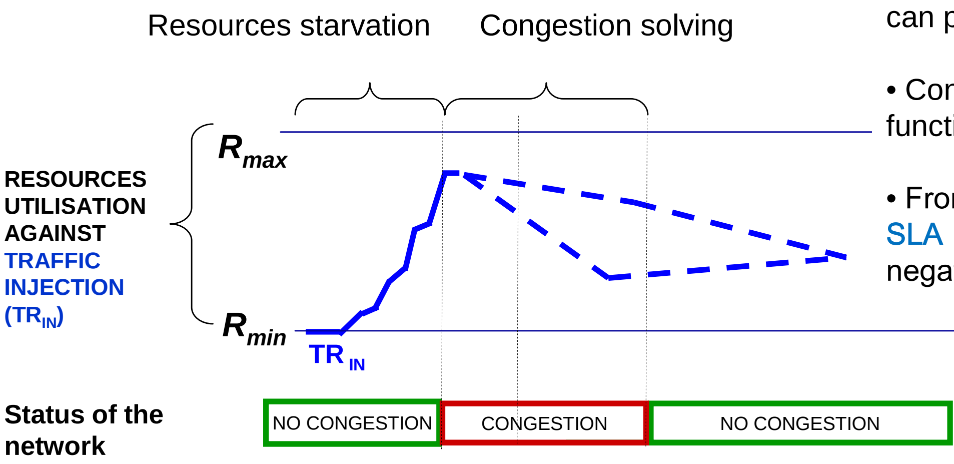


Thank you for your attention!

Issues Confronting Business-driven QoS DiffServ Management

Dr. Javier Rubio-Loyola

Introduction



- Traffic Injection without control can produce **congestion**
- Congestion **solving** is a function of user utilisation
- From a business viewpoint SLA violation could have negative impacts

Application Domain: the TEQUILA approach

Traffic Estimation Matrix
- Bounds of Traffic Demand
- Demand classified by QoS classes
- Demand for each traffc trunk

Service Management

Gereral Objectives
- Maximisation of traffic entering the network
- Commitment of Service Provider guarantees

Instruments
- Service Subscription Control (SLS-S):
  - Trade-off between subscriptions and resources
- Service Invocation Control (SLS-I):
  - Active services management
  - Quality of Service Commitment
- Traffic Forecast (TF):
  - Mutiplexing Factors for Estimations

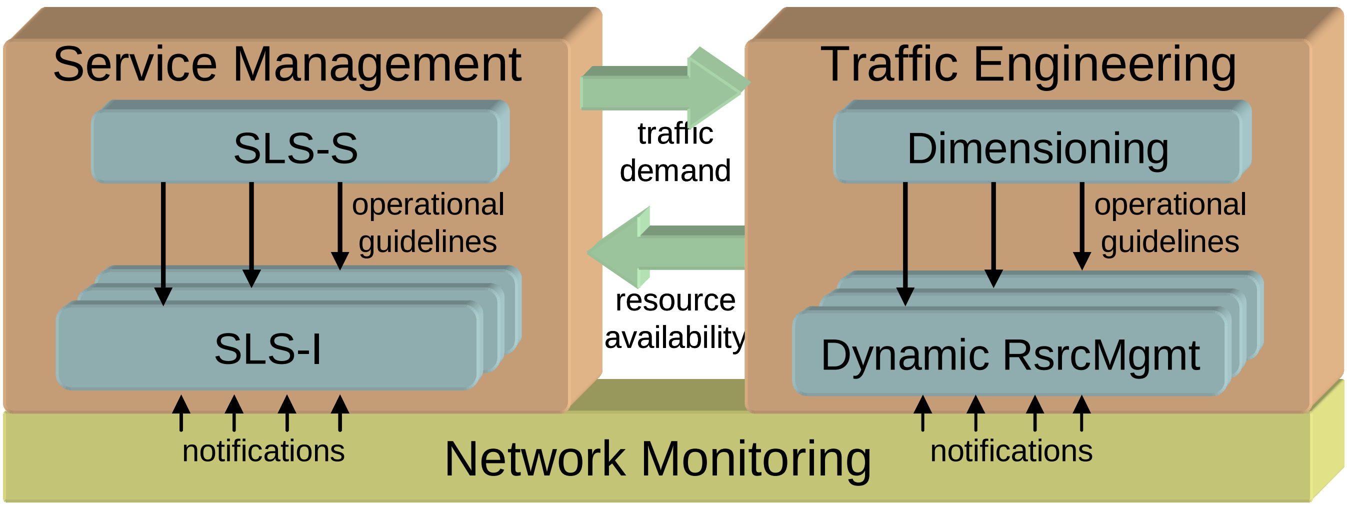
Traffic Engineering

Gereral Objectives
- Cost-effective physical resources management
- Adjust to statistical load fluctuations

Instruments
- Off-line Network Dimensioning (ND):
  - Accomodation of Predicted Load
- Dynamic management of resources (DRsM):
  - Load balancing between routes
  - Allocation of link resources between Per-Hop-Behaviors

Resource Availability Matrix
- Bounds of Available Resources
- Availability classified by QoS classes
- Availability for each traffic trunk

Application Domain: the TEQUILA approach



Policy-based Management as a Enabler of Programmability

- Control subscription volume
- Control service quality
- Control traffic injected to the network
- Control potential congestion built-up
- Control Network Dimensioning Directives
- Control Statistical Fluctuations of Traffic in the Core Network

EMANICS
Information Society Technologies

State of the Art

It is feasible to systematise the acquisition of the policy parameters (events, conditions, actions) needed to control QoS aspects:

- *Number of Subscriptions Controlled*
- *Traffic Injection Controlled*
- *QoS Degradation Prevented*
- *Traffic Demand Estimated*
- *Available Resources per-Traffic Trunk Calculated*
- *Dynamic Traffic Fluctuations Managed*

PROBLEM STATEMENT:

There does not exist a framework that allows controlling the life cycle of policies aligned to business directives, aware of QoS constraints

EMANICS

Information Society Technologies

Chosen Approach

SLA

Business Model

- **Money Profit**
- **Reputation**
- **Other**

Business Strategies

- **Consider business indicators (2)**
- **Evaluate the impact of incidents (3) occurring on the network affecting our business indicators**
- **Define priorities possibly by means of appropriate weights (4)**

Business- and QoS-oriented Management of Policies

Incidents affecting our business **(1)**

Network

- **Policy-controlled behaviour**

EMANICS

Information Society Technologies

Chosen Approach -ii

Business Strategies

- **Consider business indicators**
- **Evaluate the impact of incidents occurring on the network affecting our business indicators**
- **Define priorities possibly by means of appropriate weights**

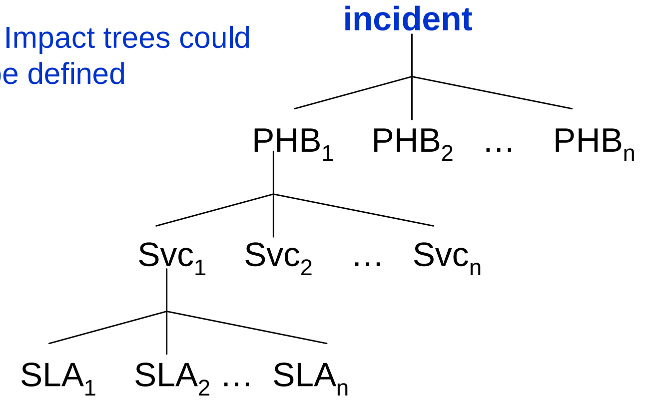
Incidents affecting our business

Network

- **Policy-controlled behaviour**

INCIDENT EXAMPLES:

- Network state; congestion, normal
- Traffic Injection Threshold Crossings
  - May have an impact on service level degradation and so forth
- Impact trees could be defined



Chosen Approach -iii

Business Strategies
- **Consider business indicators**
- **Evaluate the impact of incidents occurring on the network affecting our business indicators**
- **Define priorities possibly by means of appropriate weights**

BUSINESS INDICATOR EXAMPLE:

- Projected revenue loss due to violated SLAs

Incidents affecting our business

Network
- **Policy-controlled behaviour**

Chosen Approach -iii

Business Strategies
- **Consider business indicators**
- **Evaluate the impact of incidents occurring on the network affecting our business indicators**
- **Define priorities possibly by means of appropriate weights**

IMPACT:
Impact of an incident *i* on a business indicator *j*:
$I_j(i, t_i)$

Incidents affecting our business

Network
- **Policy-controlled behaviour**

EMANICS

Information Society Technologies

Chosen Approach -iv

Business Strategies
- **Consider business indicators**
- **Evaluate the impact of incidents occurring on the network affecting our business indicators**
- **Define priorities possibly by means of appropriate weights**

Weight ω_j is the degree of importance that business indicator *j* has for the service provider

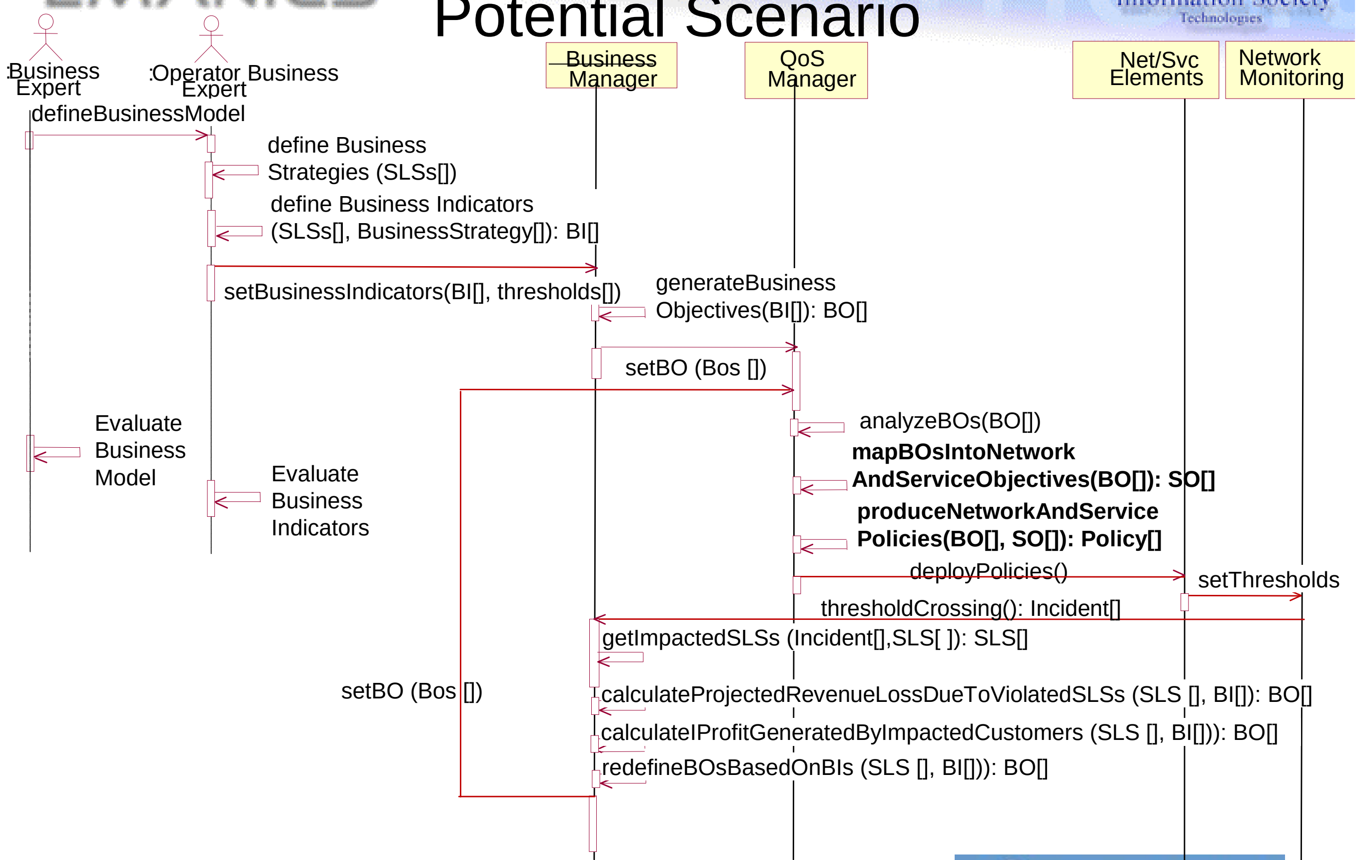
Incidents affecting our business

Network
- **Policy-controlled behaviour**

EMANICS

Information Society Technologies

Potential Scenario



Chosen Approach -v

$$\sum_{i=0}^{n} \sum_{j=0}^{m} \omega_j \, I_j\,(i, t_j) \quad \text{for} \quad \sum_{j=0}^{m} \omega_j = 1$$

m= number of business indicators
n= number of incidents affecting our business

ω_j is the degree of importance that business indicator *j* has for the service provider

Network state
- **Normal**
- **Congestion**

Service Rate Thresholds
- **0Mbps** .... **Rmax**

Internal BW alloc threshold crossings
- **0Mbps** .... **Rmax**

The TOTAL impact of an incident *i* on our business should be considered/measured

Appropriate business perspectives should result in corrective policy actions

Business- and QoS-oriented changes to minimise the negative impact of incidents

Chosen Approach -vi

Business Indicator Example:

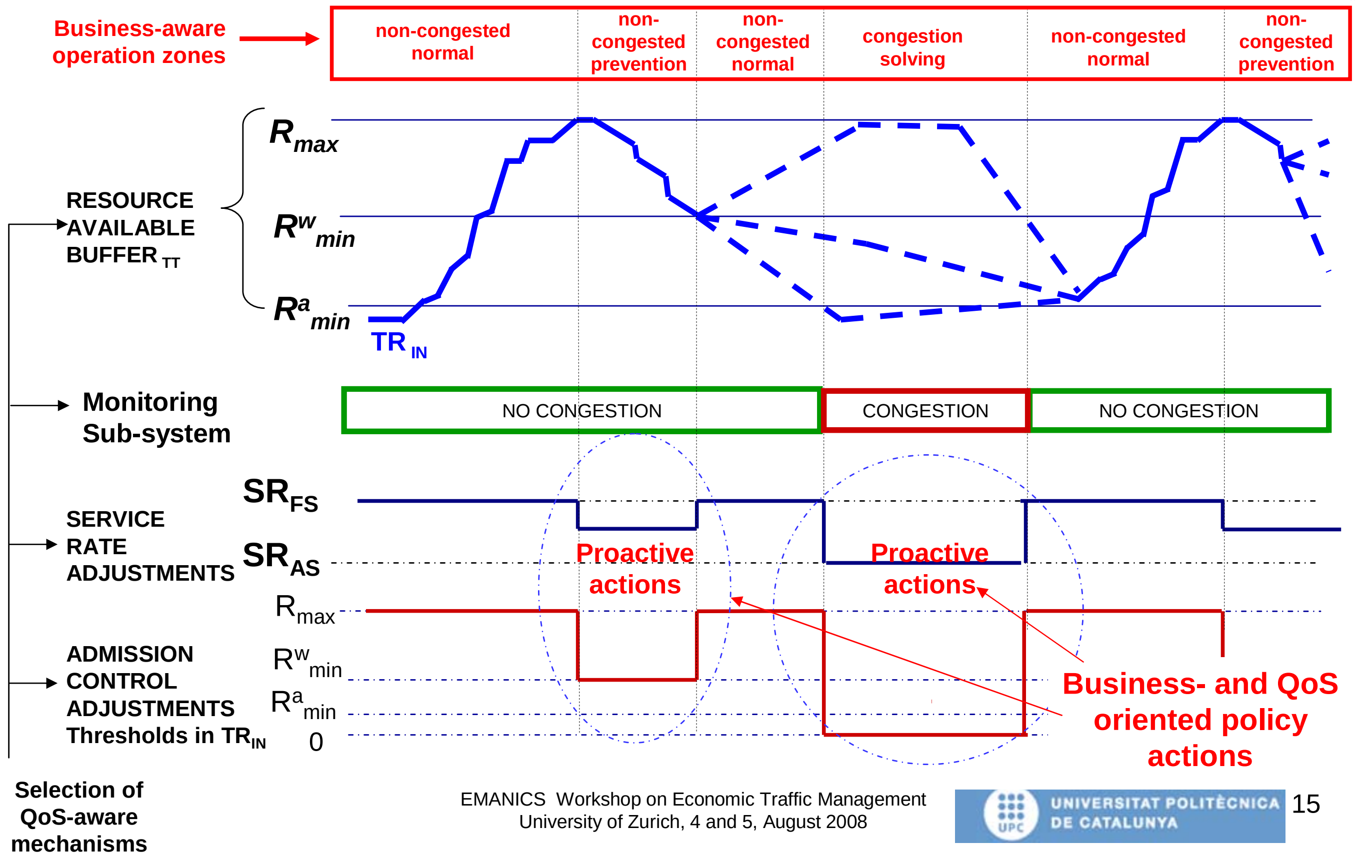
A Threshold on Projected revenue loss due to violated SLAs

Business Objective Example:

Maximize Profit without compromising the 80% of the QoS for VoIP Gold Services

Network and Service Policy Example:

```
inst oblig /Managers/moderateSubsAdmission {
on qtyLvlSet ( TT , SRamin , SRwmin , SRmax , serviceType ) ;
subject s = /Managers/SSM/BufferMO ;
target t = /Managers/SSM/SLSSPMA ;
do setSU(serviceType.PHB =EF and serviceType.PHB =AF4, TT.SUrab =
serviceType.Ramin + 0,8 * (serviceType.Rwmin - serviceType.Ramin)) ;}
```



Summary of Issues Confronting Business-driven QoS DiffServ Management

- We are dealing with Network Provider Aspects; business indicators for network providers still to be defined/standardised?
- A mapping and relationship between the business indicators and the business objectives has to be established – not a trivial issue
- There is no bidirectional framework for Business-oriented QoS DiffServ Management
  - Definition of the additional (if not existed) fields of a generic SLA to accomplish with our project objectives
    - Penalty/reward policies must be introduced
- No simulation platforms for policy-based management systems.
  - Intra-domain and inter-domain DiffServ/MPLS environment
- Advanced aspects for a control loop approach:
  - Stability and Convergence
  - Conflict detection and resolution

Future work

- A preliminar approach to address the above issues confronting Business-driven QoS DiffServ Management
- Finalisation and Implementation of a proof-of-concept prototype
- Simulations will be conducted to evaluate and to optimize the performance of diverse business directives under different patterns of service invocations and patterns of inter-domain traffic exchange between autonomous systems

Simple Economic Management Approaches of Overlay Traffic in Heterogeneous Internet Topologies

Workshop on "Economic Traffic Management"
Zurich, 4-5 August, 2008

Inter-domain Traffic Optimization in an Inter-carrier Environment

Miroslaw Kantor

Department of Telecommunications
AGH University of Science and Technology
Krakow, Poland

Outline

- Introduction
- Optimal routing elements
- General LCR model
- LCR optimization model
- Reliability optimisation model
- Conclusions

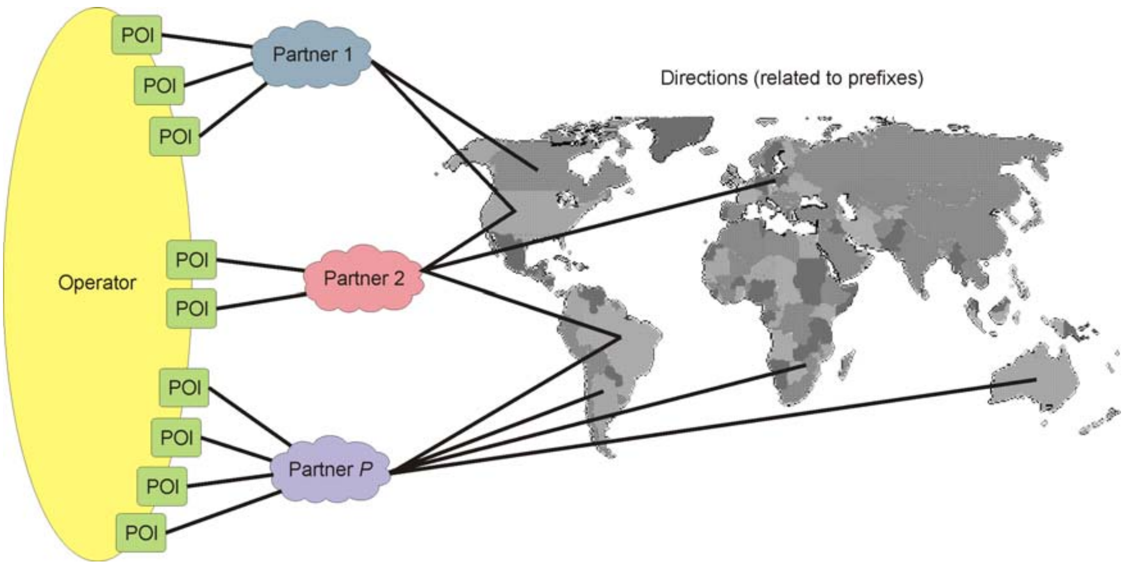
Introduction

- In today's competitive telecommunications environment, interconnect rates can be changed on a daily basis.
- Operators seek to minimise costs and enhance service quality.
- However, when terminating international voice and data traffic, carriers are faced with numerous routing, service and cost alternatives.
- To help for choosing the optimal routing solution the Least Cost Routing algorithm is proposed.
- The LCR algorithm will take into account carrier network engineering constraints including cost, quality, resilience, volumes, capacity and different agreement types.
- It will also support differential routing for different customer and traffic types.

Motivation and Objectives

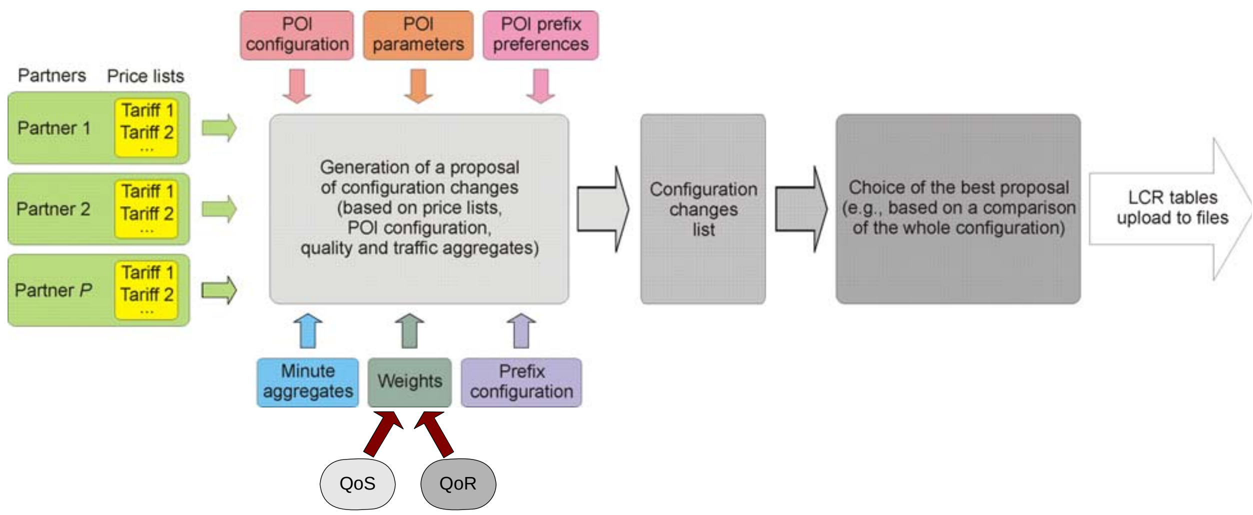
- searching optimal routing for effective interconnect cost reduction
- introduction of new services to build new revenue streams
- support for processes associated with entering into agreements with other operators

General connection model



POI – Point of Interconnection

General LCR model



POI – Point of Interconnection
QoS – Quality of Service
QoR – Quality of Resilience

Optimal routing elements

- LCR solution takes into account a number of parameters
  - tariff's based parameters, like terminating/transit charge
  - requirements regarding the quality of service (QoS)
  - requirements regarding the quality of resilience (QoR)
- These network parameters should be weighted according to the rule chosen by operator.
- As a result, the chosen route has to fulfill all the requirements with reference to both price and quality.
- Such route is called the optimal one.

Tariff's based parameters

- As a tariff we define a scheme of rates and regulations managing the charging of telecommunication services.
- A tariff model in most cases is composed of two elements:
  - a price: a monetary component;
  - a related tariff model, i.e., a calculation scheme, which provides a charging function enabling the calculation of costs considering charging variables (e.g., time of usage, volume transferred, allocated bandwidth) and charging coefficients (e.g., price per suitable unit).

Tariff's based parameters

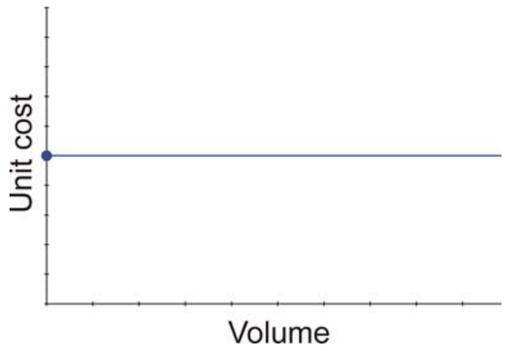
Numerous tariff models have been proposed for telecommunication services:

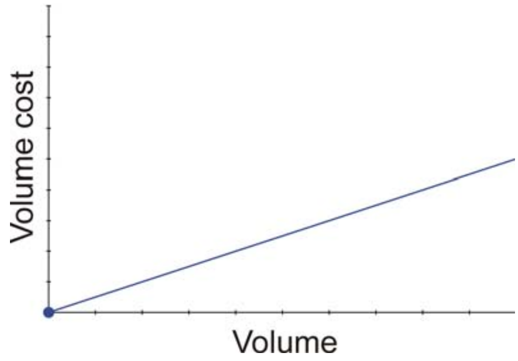
- linear tariff models - the price is a linear function of the number of served units
- non-linear tariff models- the price decreases with the volume (concave cost function)
- discounts - a special type of tariff model applied to decrease the total cost for the customer
  - they can be defined over total costs incurred for a special type of transaction or on total costs incurred for a certain period of time.
  - discounts are applied in combination with tariff models and result in an additional reduction of prices.

Tariff's based parameters

- In last years:
  - many operators had agreements where they paid each other months ahead, according to traffic measurements based on experience and qualified guesses,
  - interconnect and settlement was based on the size of the physical interconnect,
  - most common were long term agreements based on previous experience.
- Nowadays:
  - due to the significant change in nature of interprovider interconnect the big diversity of pricing structures exists within the Internet, often charging on a per-session, or call, basis,
  - the measurement of traffic parameters enables service providers to adopt new bandwidth pricing models based on what it is used for and how much of it is used,
  - some operators use pricing based on received volume, some on sent volume, some on a mix of sent and received volume, and some use pricing based on the access capacity irrespective of volume,
  - however, most operators charge for the amount of minutes routed for other operators.

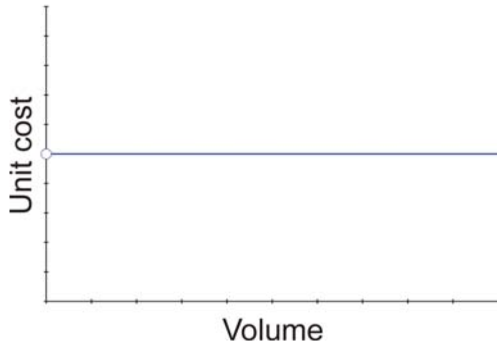
Cost structure in a tariff without beginning cost

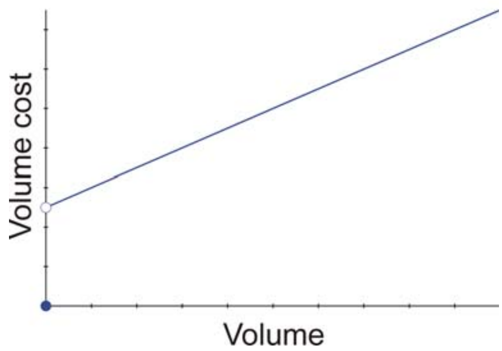




a) unit cost b) total cost

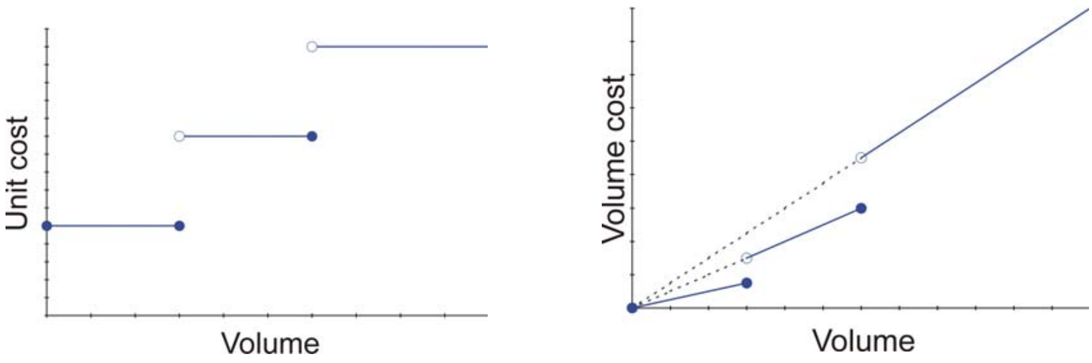
Cost structure in a tariff with beginning cost





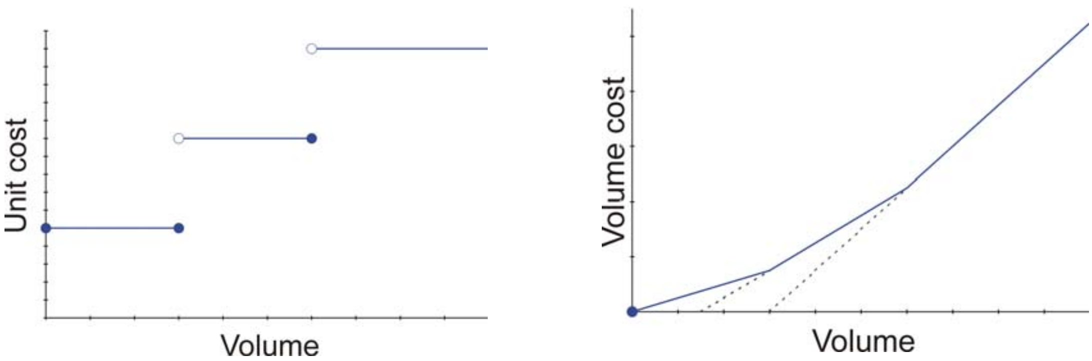
a) unit cost b) total cost

Cost structure in a non-linear tariff (1)



a) unit cost b) total cost

Cost structure in a non-linear tariff (2)



a) unit cost b) total cost

QoS parameters

- Quality of service in multidomain environment is an important factor in ensuring the establishment of effective interconnection and end-user services' satisfaction.
- In a multi-service network, very often rating is based on service type and Class of Service (CoS) related to QoS parameters.
- The QoS classes are defined taking into account following parameters: data and signalling traffic delay, hop-count, jitter, bit error rate (BER), packet loss ratio, throughput, traffic load, etc.
- These QoS parameters represent the requirements of the users regarding the service.

QoS parameters

- To provide end-to-end QoS over the interconnections of many autonomous domains, Service Level Agreements (SLAs) between operators are required.
- A key aspect of the SLA is definition of the QoS classes an provider can offer for its customers' traffic.
- Each operator then has to provision and configure its network resources so that traffic is forwarded in accordance with the agreed QoS levels.
- It has to be stressed, that every provider relies on every other provider to complete the end-user connectivity.
- As result of this dependent relationship, an individual provider's effort to provide significantly better service quality may have little overall impact on the total of client-delivered service quality.
- Service quality can be affected negatively by weak local engineering but cannot be uniformly improved beyond the quality provided by the partners along the end-to-end path connection.

Resilience issues

- Resilience is an important issue for network/service operators.
- The resilience in this context denotes protection against the failure of one or more interconnections.
- For example, if one of the interconnection partners, to which a huge amount of traffic is transmitted fails, it could happen that there is not enough free capacity available from the other interconnected transit providers to compensate the failure by rerouting the traffic destined for the failed provider.
- To protect the traffic against the failures, one of resilience schemes for interconnection diversity (e.g., connection protection) has to be applied.
- Connection diversity can also be used for load balancing, or to split a large flow into smaller parallel connections if there is no connections with enough bandwidth to support the whole flow.

Assumptions and goals of LCR optimisation model

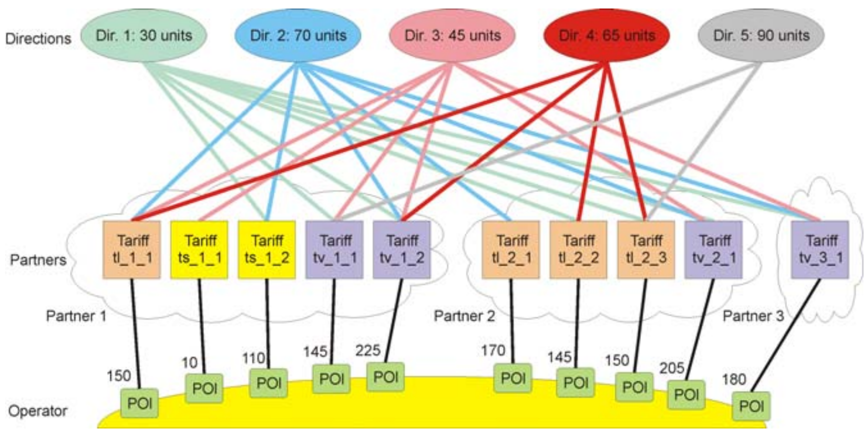
- The volume of the traffic to be carried to different directions is the main input for developed algorithm.
- Based on tariffs announced by operators, POI configurations, and quality the proposal of changes in running configuration is generated.
- The goal of the optimisation algorithm implemented in LCR solution is to find the distribution of the traffic which has to be send on required directions in the cheapest way, i.e. to choose the POI, partners and tariffs in order to pay the lowest total cost.

Basic LCR optimisation model BLCR

- *constants*
  - $h_{de} > 0$ - number of minutes that must be carried for direction d in timeband E
  - $z_{pte} > 0$ - maximum number of minutes that can be carried in tariff *t* of partner *p* (e.g., interface capacity constraints) in timeband *E*
- *continuous variables*
  - $x_{dpte} > 0$ - number of minutes carried by *p* in its tariff *t* for direction *d* in timeband *e*
  - $x_{pte} > 0$ - number of minutes transferred into all directions *d* which are carried by *p* in its tariff *t* in timeband *e*
  - $\delta_p > 0$ - total cost input related to partner *p*
- *goal function*
  - minimise the total cost for transmitted traffic

$$\min \sum_{i=0}^{P} \delta_i$$

Numerical example - constraints and assumptions



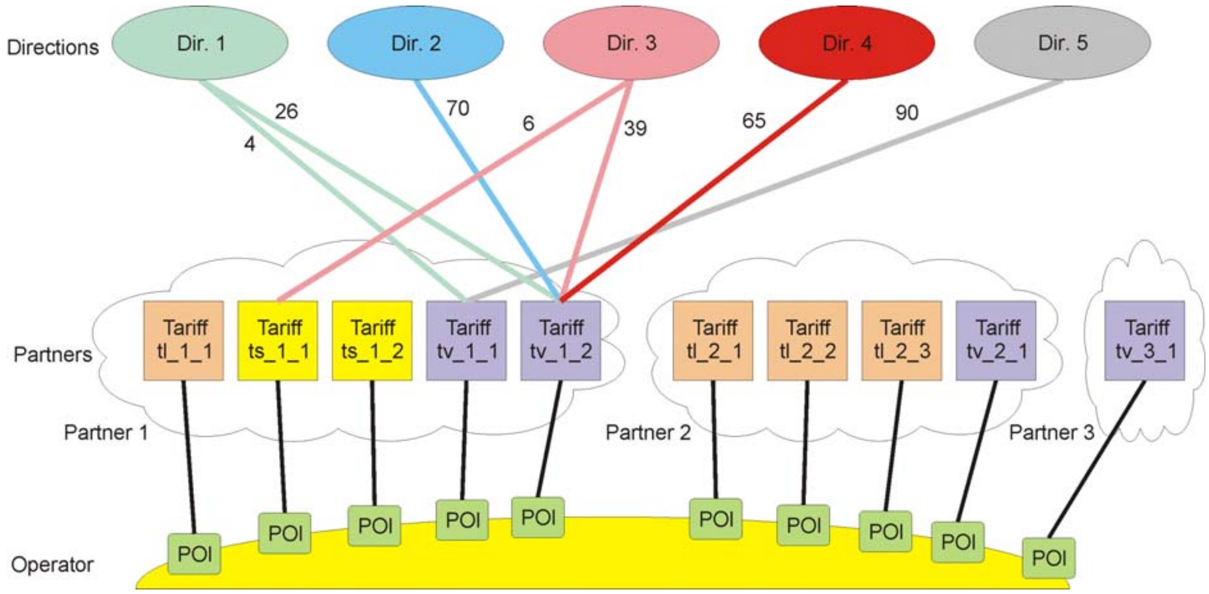
Global promotion

<table>
<tr><td>Partner</td><td>1</td><td>2</td><td>3</td></tr>
<tr><td>Threshold</td><td>250</td><td>200</td><td rowspan="2">no discount</td></tr>
<tr><td>Discount</td><td>30</td><td>20</td></tr>
</table>

Tariffs' structure

<table>
<tr><td colspan="2">tl_1_1</td><td colspan="2">ts_1_1</td><td colspan="2">ts_1_2</td><td colspan="2">tv_1_1</td><td colspan="2">tv_1_2</td><td colspan="2">tl_2_1</td><td colspan="2">tl_2_2</td><td colspan="2">tl_2_3</td><td colspan="2">tv_2_1</td><td colspan="2">tv_3_1</td></tr>
<tr><td rowspan="5">cost</td><td rowspan="5">8</td><td>cost 1</td><td>1</td><td>cost 1</td><td>6</td><td>cost 1</td><td>5</td><td>cost 1</td><td>3</td><td rowspan="5">cost</td><td rowspan="5">6</td><td rowspan="5">cost</td><td rowspan="5">8</td><td rowspan="5">cost</td><td rowspan="5">5</td><td>cost 1</td><td>2</td><td>cost 1</td><td>3</td></tr>
<tr><td>threshold 1</td><td>3</td><td>threshold 1</td><td>40</td><td>threshold 1</td><td>70</td><td>threshold 1</td><td>150</td><td>threshold 1</td><td>180</td><td>threshold 1</td><td>20</td></tr>
<tr><td>cost 2</td><td>2</td><td>cost 2</td><td>9</td><td>cost 2</td><td>9</td><td>cost 2</td><td>4</td><td>cost 2</td><td>9</td><td>cost 2</td><td>4</td></tr>
<tr><td>threshold 2</td><td>6</td><td>threshold 2</td><td>100</td><td>threshold 2</td><td>120</td><td>threshold 2</td><td>200</td><td>threshold 2</td><td>195</td><td>threshold 2</td><td>140</td></tr>
<tr><td>cost 3</td><td>5</td><td>cost 3</td><td>10</td><td>cost 3</td><td>10</td><td>cost 3</td><td>8</td><td>cost 3</td><td>10</td><td>cost 3</td><td>9</td></tr>
</table>

Numerical example - results



optimal goal function is 755.3

Reliability optimisation models

- The goal of the reliability optimisation algorithm implemented in LCR solution is to find the distribution of the traffic which has to be send on required directions to assure the required level of resilience.
- The formulation of LCR problem is based on Mixed-Integer Programming (MIP).
- With reliability in this context we mean protection against the failure of one or more interconnections.
- Basic optimisation model can be extended in several ways to include reliability.

Minimum Number of Partners optimisation model MNP

- MNP optimisation model is based on requirement for the interconnection with a minimum number of partners Pmin.
- Thanks to this multi-homing approach the dependency among partners (transit providers) is reduced.
- The proposed reliability model is simple and easy to apply.
- However, on the other hand, it does not provide a fine-grained control.

Minimum Free Capacity optimisation model MFC

- MFC optimisation model makes sure that there is a minimum amount of free transit capacity available at the POI interface, e.g. a percentage of the total traffic transmitting through that interface.
- This proposed reliability policy enables fine-grained control over the free capacity.
- However, it still has some drawbacks because if one interconnected provider who carries more than the fraction α of the traffic at the interface *i* fails, there will not be enough spare capacity to reroute the traffic.

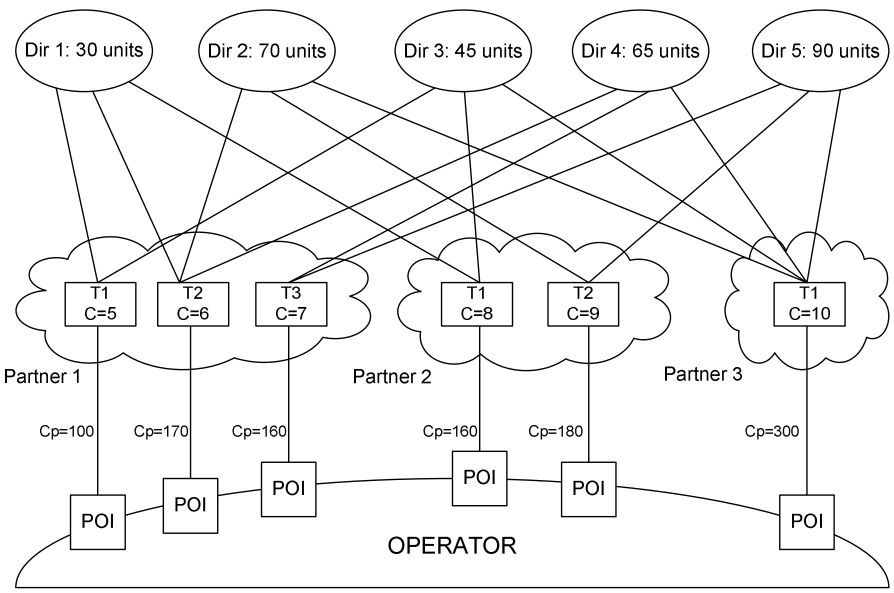
Single Partner Protection optimisation model SPP

- SPP optimisation model makes sure that there is enough spare transit capacity if a single partner fails completely.
- The following new constraint is added to the basic BLCR model: *constraint*

$$\sum_p f_p \geq \sum x_{dp}$$

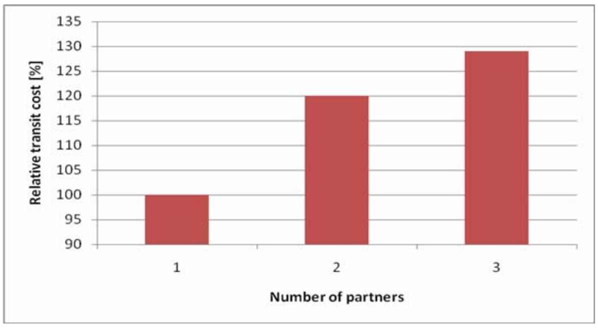
enforces the minimum amount of free capacity to protect against the failure of each provider *p*

Analysed scenario



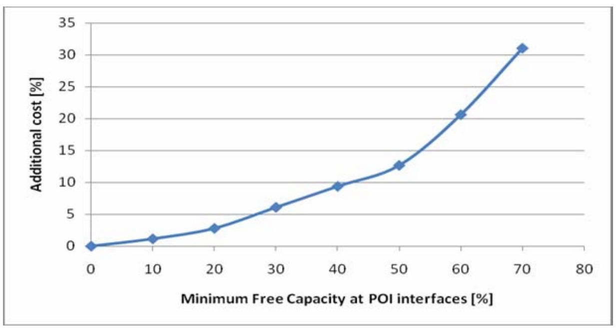
Results for Minimum Number of Partners MNP reliability policy

- The costs increase very quickly if the minimum number of transit providers is increased.
- It should be kept in mind that the reference cost (100%) is the cost of optimal solution (one partner with total cost equal to 1824).
- The MNP policy is easy for use, however, costs can grow up in uncontrolled manner, because the number of partners can be increased only in steps.



Results for Minimum Free Capacity MFC reliability policy

- The higher free capacity at the POI interface, the more probable that if – for example – one provider fails there will be enough spare capacity to send the traffic to required directions.
- However, at the same time the cost for sending the traffic increases while comparing with the optimal solution.
- In case of MFC reliability policy the operator has more control on trade off between level of reliability and cost.



Conclusion

- The concept of the least-cost routing in the inter-carrier context has been presented.
- The parameters influencing the choice of optimal route have been described.
- Simplified reliability optimization problems have been formulated and analysed.
- To increase the efficiency of operations within the interconnection environment the development of solution for finding the optimal routes becomes crucial.

Thank you very much