

Management of Differentiated Services Usage by the Cumulus Pricing Scheme and a Generic Internet Charging System

Burkhard Stiller, Jan Gerke, Peter Reichl, Placi Flury
ETH Zürich, Computer Engineering and Networks Laboratory TIK
Gloriastrasse 35, CH – 8092 Zürich, Switzerland
Phone: +41 1 632 7016, FAX: +41 1 632 1035
E-Mail: [stiller/gerke/reichl/flury]@tik.ee.ethz.ch

Abstract

Since the Internet is on the move to provide differentiated services, for the backbone based on the Differentiated Services Architecture (DiffServ), suitable management mechanisms are required. Scalable solutions for overload management are a must, such as economically-driven functions of charging and pricing. Future commercial networks need to offer as well a new view of pricing, considered as management information. Therefore, the Cumulus Pricing Scheme (CPS) explicitly proposed here for the DiffServ technology is the only approach known so far, which defines a clear relation between different time scales of measurement, accounting, and charging periods. The scheme is flexible enough to allow network management according to actual forces of the market. CPS's implementation is supported by a designed generic Internet Charging System offering a service-independent architecture and integrating interchangeably economically-controlled network management functions of charge calculation and pricing. In particular, it has been applied to DiffServ and the new CPS.

Keywords

Internet Pricing, Internet Charging System, Differentiated Services.

1. Introduction

Despite of the huge amount of bandwidth provided by the Internet today, bottlenecks are still far from having disappeared, *e.g.*, as dropped packets or the long delays experienced by certain packets easily demonstrate. Even a short-term overprovisioning will not solve the problem, since new application requirements will consume this capacity as well. But since a number of applications require at least a statistical guarantee of service characteristics, generally expressed in Quality-of-Service (QoS) parameters, the underlying networking technology needs to offer mechanisms for managing, providing, and supervising a specified set of QoS.

Therefore, pricing and charging will become highly relevant for offering more than a single service class, *e.g.*, a best-effort and a statistically guaranteed class, since financial incentives are the most effective stimulators for triggering user behavior accord-

ing to users' real communication demands. In addition, traditional congestion control methods may be operated within the network. However, economic functions are considered an essential part of business and technology management as well as services management, as defined in the management building [16]. Nevertheless, the selection of a pricing scheme for a single-service network is easier than for a multi-service one [12]. This complicates the issue for multiple QoS levels, but Internet Service Providers (ISP) are the key players in the communication market and will offer differentiated Internet services in the near future. Since Differentiated Services (DiffServ) defines a technology for ease-of-use of multiple services in the Internet's backbone, an appropriate solution for pricing and charging these services is essential within a commercialized Internet services market. Flat fee pricing schemes seem to be favored by users and usage-based charging mechanisms seem to be only viable, if they can be implemented efficiently. A variety of investigations and technology developments in these areas have taken place, such as INDEX [10] and CATI [26]. Therefore, a suitable combination of basic flat fees and time-depending or volume-dependent add-on charges and feedbacks for the utilization of differentiated services could offer an appropriate mixture of both worlds. Such schemes are considered essential management tasks on different time scales.

Even though a variety of pricing schemes for Internet access exists [28], this paper establishes a new view on them by considering the 'flow of management information' to be as important as their 'content'. In this sense, there are three main types of pricing schemes to be distinguished. Usage-sensitive or volume-based schemes rely on information on the user behavior that may be gathered at the interface between customer and ISP. Congestion-based pricing schemes, on the contrary, depend on the network status that depends only indirectly on the behavior of the single user, i.e. through her externalities. Flat rate schemes, finally, are generally independent of any metering information. With respect to the DiffServ approach, the future Internet backbone will consist of many different DiffServ domains, each of them operating under a different administration. Therefore, two principally distinct locations for applying the pricing and charging mechanisms exist: (a) the interface between a user and its DiffServ provider and (b) the interface between two DiffServ providers. For this reason, the view termed "Edge Pricing" by [7] has been extended to include all edges between neighboring providers. Initial ideas of these extensions have been expressed in [26] and [5]. Based on related work and a terminology description in Section 2, the proposed Cumulus Pricing Scheme (CPS) is discussed in Section 3. In support of this, Section 4 discusses a service-independent and generic Internet Charge Calculation and Accounting System (ICCAS), which was applied to CPS and DiffServ (Section 5). Finally, Section 6 summarizes the work and draws conclusions.

2. Related Work

The number of projects on pricing, charging, and accounting in the Internet has increased over the last few years. Therefore, only a number of recent charging-centric work of system's design and modeling is summarized below. Related pricing work is discussed in [28], [29], other overviews are to be found in [26].

Many projects dealing with charging and accounting functionality on the network level try to achieve a high independence from pricing models. However, it has been argued that pricing in general and usage-based pricing in particular can impose a high

overhead on telecommunication systems [19], [23]. Any form of usage-based pricing for various telecommunication services is interesting, because underlying resources (such as frequencies, cables, routers/switches, and most notable operating personnel) are scarce and very costly. The traditional, flat-fee Internet pricing model has been critiqued in the past for its economic draw-backs of not being incentive-compatible [23], [14]. Additionally, it is inflexible - *e.g.*, it does not allow for combined sender/receiver payments - and does not provide economic signals which are needed for network planning and expansion. But most importantly, the current model is based on the assumption of a single service best-effort network providing a similar service to all customers. Therefore, the multi-service paradigm needs to be investigated with respect to heterogeneous Internet networking technologies. An early per-flow billing system for Transmission Control Protocol flows and initial ideas on a billing service design is presented in [9] and [25]. Advanced per-flow charging and accounting approaches based on reservations have been tackled in [11] and [17].

The objectives of the project CATI (Charging and Accounting Technology for the Internet) [26] included the design, implementation, and evaluation of charging and accounting mechanisms for Internet services and Virtual Private Networks. This covered the enabling technology support for open, Internet-based Electronic Commerce platforms in terms of usage-based transport service charging as well as high-quality Internet transport services. The main assumption of the work on “Lightweight Policing and Charging” is that a multi-service packet network may be achieved by adding classification and scheduling to routers, but not policing [5]. Another important question covers the issue of user acceptance of pricing schemes. The INDEX project (Index Demand Experiment) investigated user reaction when exposed to various pricing schemes for different qualities of Internet access. It turned out that users were not disinclined to flexible pricing models [10]. The project M3I (Market-Managed Multi-service Internet) [20] aims at the design and implementation of a next-generation system that will enable Internet resource management through market forces, specifically by enabling differential charging for multiple levels of service.

2.1 Terminology

Based on the discussion of related work and the necessity to obtain an uncontradicted terminology, the following basic definitions will be applied within this work and are utilized for the task and component description.

- **Accounting** defines summarized information (accounting records) in relation to a customer’s service utilization. It is expressed in metered resource consumption, *e.g.*, for applications, calls, or any type of connections.
- **Billing** defines the collection of charging records, summarizing charging content, and delivering an invoice including an optional list of detailed charges to a user.
- **Charge Calculation** covers the complete calculation of a price for a given accounting record and its consolidation into a charging record, while mapping technical values into monetary units. Therefore, charge calculation applies a given tariff to the data accounted for.
- **Charging** is used as an overall term, depicting all tasks required to calculate the finalized content of a billing record. Sometimes in the literature, the term billing is utilized instead, however, it includes the full handling of invoices and customer data tasks, which are of second priority for the technological aspects only.

- **Mediation** is intended to filter, aggregate, and correlate raw technical data which in most cases has been collected by metering. Mediation transforms these data into a form which can be used for storing and further processing.
- **Metering** determines the particular usage of resources within end-systems (hosts) or intermediate systems (routers) on a technical level, including Quality-of-Service (QoS), management, and networking parameters.
- **Pricing** covers the specification and setting of prices for goods, specifically networking resources and services in an open market situation. This process may combine technical considerations, *e.g.*, resource consumption, and economical ones, *e.g.*, applying tariffing theory or marketing methods. Prices may be calculated on a cost/profit base or on the current market situation.
- **Tariffs** define algorithms used to determine a charge for a service usage. They are applied in the charge calculation for a given customer to the service he utilizes. For calculating charges the tariff may contain, *e.g.*, discount strategies, rebate schemes, or marketing information.

3. Cumulus Pricing Scheme — CPS

The issue of pricing and tariffing Internet services has become an important research issue over the last couple of years [28]. However, there is still a lack of definite solutions, as existing schemes either are too inflexible or complex for a satisfying technical solution. Hence, there is an urgent need for new pricing schemes that are able to deal with various essential economic, technical and user-specific requirements.

Moreover, pricing models can also be used as network management tools. This alternative point of view is based on eliciting user information about existing usage pattern. With this information, Internet Service Providers (ISP) are able to optimize various dimensions, *e.g.*, network configuration or admission control, regarding objectives like utilization or revenue. Each of these points of views is investigated with respect to one of their common denominators, *i.e.* their dependence on time-scales. This newly developed model identifies four different time scales that are equally relevant for network management as well as service tariffing. Afterwards a new Cumulus Pricing Scheme CPS is presented and investigated, based on compound activities on various time-scales. In addition, it is described how the pricing mechanism provided by CPS can be embellished differently using various policies. In this way, this approach provides for an easy introduction of basic economic theory, *e.g.*, on relevant cost models or marketing influences.

3.1 Time Scales

Time scales define the important dynamic criteria for feedback handling. According to [16], overall management time scales are distinguished: Short-term in minutes, medium-term in hours, and long-term in weeks or months. These scales are extended by an atomic scale for ultra-short times in seconds and below to allow for the definition of feedback in a round trip-time or milliseconds. Measurement intervals and units of measurement show the relevant timing data and information to be accounted for. The type of feedback is identified as well by describing its major content (*cf.* Table 1). Applying these time scales onto pricing-controlled activities results in (1) the atomic monitoring and control level, (2) the short-term intervention level, (3) the medium-term service deployment level, and (4) the long-term business/strategic level. More-

Table 1: Time Scales, Measurement, and Feedback Content

Time-scale Naming	Measurement Intervals	Measurement Units	Feedback Content
atomic	milliseconds, round trip-times	packets	communication-relevant data
short-term	minutes	flows/sessions	application data
medium-term	hours/days	billing periods	billing data
long-term	weeks/months	contract periods	contract data

over, measurements take place on a short-term time scale, hence CPS is characterized as a multi-dimensional mapping of the atomic time-scale onto the other three ones.

3.2 CPS: General Idea and Example

CPS has been developed with respect to three main requirement types:

- User- and customer-oriented: Simplicity, transparency, and predictability.
- Economic: Efficiency, market-orientation for the ISPs.
- Technical: Feasibility of accounting.

The fundamental decision between static and dynamic schemes touches immediately customer’s desires concerning price stability, *e.g.*, highly fluctuating auctions, whereas orienting a pricing scheme strictly according to the forces of the market readily induces technical infeasibility. In this situation, CPS is an approach to reconcile all three requirements. CPS is basically a flat rate scheme (but rates may vary over long time scales), it provides a feedback mechanism to bring market forces into play (where this feedback is not an immediate one, but requires the accumulation of a sufficient number of discrete “flags”, the so-called Cumulus Points (CP), indicating user behavior), and it allows a huge flexibility in terms of technical prerequisites, especially concerning measuring and accounting mechanisms of the required data records.

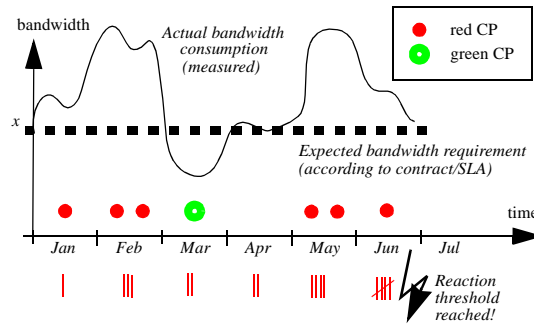


Figure 1: Red and Green Cumulus Points and Their Accumulation over Time

Figure 1 describes as an initial and typical example how Cumulus Points are used. Customer *C* has stated her expected bandwidth requirements to be x MB/s, but the actual bandwidth consumption exceeds the agreed upon one slightly in January and heavily in February. Accordingly, the consumer receives one red CP at the end of January and two additional red CPs at the end of February. Afterwards, her consumption falls below the expected value (one green CP in March), remains within the contract

in April (apparently the ideal situation), but exceeds this value in May and June. The accumulation of the CPs as of end of June sums up to five red CPs. The reaction threshold has been reached, hence a contract renegotiation has to be performed.

3.3 Mathematical Description of CPS

The key to the new solution proposed lies in building the contract between customer and ISP upon suitable information about the expected usage pattern of the service plus influencing the actual customer behavior by a new type of feedback mechanism that is specific in terms of its relation to different time scales. *Measurements* take place over a *short time scale* and allow evidence about *user behavior on a medium time scale*. This evidence is expressed in terms of discrete flags, yet not triggering some sort of *reaction* by themselves, but only as a result of their accumulation over a *long time scale*. Hence, if the user has been detected by the ISP to strongly overuse capacities or to misbehave in some sense, she will usually receive some sort of warning that in case of unchanged behavior the contract may sooner or later may be finished by the ISP. On the other hand, if the user is underutilizing her capacities, this behavior may be rewarded by some sort of bonus system. This feedback system strongly depends on measurement activities that are deliberately left open to ISPs. Hence, one could think of (1) ISPs monitoring each packet or each connection in the one extreme, (2) ISPs undertaking systematic monitoring, (3) ISPs measuring every now and then (maybe in a statistical framework), or (4) at the other extreme ISPs not measuring at all.

Mathematically, CPS may be described using two rules that refer to different time-scales, respectively. To this end, suppose that ISP *I* offers only one service, and initially customer *C* has stated her expected bandwidth requirements according to a contract (the SLA) to be x MB/s, whereupon ISP *I* has offered a flat rate tariff of a \$/month for this service which customer *C* has accepted. In reality, the volume consumed by *C* is described by a function $V(t)$ of time, which naturally may differ arbitrarily from the stated expected requirement x .

Let $\Delta_i = \Delta(t_i)$ describe the monthly over- or underutilize, respectively, of the customer with respect to her statement x , i.e.

$$\Delta_i = \int_{t_{i-1}}^{t_i} (V(t) - x) dt = \int_{t_{i-1}}^{t_i} V(t) dt - x(t_i - t_{i-1}) \quad , \quad (0)$$

where t_i describes the end of measurement period i , *e.g.*, the end of month, $i = 0, 1, 2, \dots$ (note that t_0 describes the start of the contract between ISP and customer).

Cumulus Points are assigned by the ISP *I* according to a rule (the so-called “CP Rule”) whose content is up to the ISP, but typically might look like the following:

CP Rule: Define θ_n , $n = -N, \dots, -1, 0, 1, 2, \dots, N$, to be the CP thresholds, $\theta_0 = 0$ and $\theta_{\pm(N+1)} = \pm\infty$ where N describes the maximal number of CPs that could possibly be assigned for one measurement period. Then for measurement period i , the customer is assigned c_i cumulus points iff

$$0 \leq \theta_{c_i} \leq \Delta_i < \theta_{c_i+1} \quad \text{or} \quad (1)$$

$$\theta_{c_i-1} < \Delta_i \leq \theta_{c_i} \leq 0 \quad , \quad (2)$$

the choice between (1) and (2) depending on $\text{sgn}\Delta_i$.

Hence, if Δ_i is positive (i.e. overuse in period i) and lies between thresholds θ_c and θ_{c+1} , then c cumulus points are assigned. If Δ_i is negative and between thresholds θ_{c-1} and θ_c , c cumulus points are assigned, where c now is a negative number, hence cumulus points are referred to as *green* ones, whereas for positive c cumulus points are *red*. Now cumulus points c_i are accumulated over time according to

$$\Gamma_n = \sum_{i=1} c_i, \quad (3)$$

hence, Γ_n describes the total sum of cumulus points assigned since the contract start. The reaction to CP accumulation is again basically up to the ISP and is the content of a second rule, the so-called “Reaction Rule”, typically looking like this:

Reaction Rule: Define Θ to be the reaction threshold. Then the contract between customer and ISP is in the state of imbalance and needs to be renegotiated after period n if

$$|\Gamma_n| \geq \Theta. \quad (4)$$

Depending on $\text{sgn}\Gamma_n$, there may as well be two different thresholds Θ^+ and Θ^- for red and green CPs, respectively.

3.4 ISP Policies

CPS as described is a highly variant pricing scheme. Its implementation may vary considerably according to different ISP policies that may include the following ones:

- **Measurements:** It is almost impossible to find a standard way of network monitoring and accounting that is compulsory for all ISPs. Hence, in the CPS approach it is up to the ISP, on which data measurements the distribution of CP is based.
- **CP assignment:** Being assigned one or more CPs depends on violating thresholds in terms of utilization or bandwidth. Fixing these thresholds is up to the ISP. Note that setting $\theta_{\pm 1} \neq 0$ prevents from smaller oscillations in x and result in a superfluous assignment of CPs.
- **Accumulation:** Usually, CPs are supposed to be accumulated over subsequent billing periods. However, it may be allowed for CPs to expire, or for red CPs to be charged up against green ones.
- **Contract renegotiation:** Another threshold to be set freely by ISPs concerns renegotiation. The way of renegotiating is also open. Either the customer delivers a new statement about expected QoS requirements, and the provider offers a new charge, or the old contract remains valid, and delta requirements plus the old red CPs are dealt with by a separate new contract, *e.g.*, by an extra payment.

4. A Generic and Modular Charging System

Following the introduction of the Internet networking technology DiffServ and one of its appropriate pricing schemes CPS, an economically operated network requires necessary functionality for performing the charging task. Based on the illustration of

major drawbacks of existing charging systems, the newly developed approach is introduced and its functionality is presented in detail.

In existing charging systems the setting of prices, the function of charge calculation, and billing itself are integrated, even additionally the maintenance of service classes, user profiles, customer data, identities, and banking account data are included. Although these tasks need to be distinguished for distributed and scalable systems as illustrated in [28] (in addition, cf. terminology as defined in Section 2.1), they are almost completely centralized within a single system. Therefore, future charging systems need to be able to integrate different charging and accounting records, *e.g.*, Internet Protocol Detail Records (IPDR) [8], since customer's demand is determined by the so-called "one-stop billing" approach [27]. By providing open interfaces it will become possible to exchange individual components and to integrate different components supporting different technologies without having to adapt the entire system. Therefore, charging systems are supposed to support the following tasks:

- Perform transport, service, and content charging.
- Perform accounting tasks according to transport and multi-service definitions.
- Support different levels of charging security.
- Support auditing.

4.1 Design and Implementation Dimensions

After identifying these components, it must be determined how they are implemented and deployed in possible scenarios with potentially several different ISPs. Four dimensions exist in which way a charging system can vary according to the scenario. The dimension of *location* defines, where components are located. In particular, an "inhouse" location refers to the fact that the ISP itself hosts this component and provides the according functionality internally. The dimension *replication* defines how many of these components considered exist in a given environment. Mainly the number of clients served by an ISP and the number of interconnection points with peering ISPs determine the number of replicated components required. The dimension *reliability* defines how reliable these components have to be. The needed degree of reliability depends only indirectly on the ISP type. It rather depends on the aforementioned dimensions of location and replication as well as heavily on the component type itself. Finally, the dimension *time scale* has been already introduced for CPS in Section 3.1 and is applied for the CAS similarly.

4.2 Charging System and ICCAS Architecture

An overall architecture for a charging system and its important subsystem, an Internet Charge Calculation and Accounting Systems (ICCAS), is based on the mapping of the conceptual view onto distinguished components. Refining this concept into components and their interactions and interfaces results in the overall architecture as depicted in Figure 2/left as it has been designed for the M3I project [20]. Interactions between two neighboring providers take place on two levels.

The first one is of course on the data path since providers must exchange data between their networks. Further inter-provider information exchange happens as part of the specific protocol processing as defined in the QoS model applied, *e.g.*, for resource reservation purposes such as using the Resource Reservation Protocol (RSVP) or

inter-Bandwidth-Broker communication, where messages are exchanged between border routers of neighboring providers. In these cases, a type of signaling or consolidation protocol has to take care of the distributed information scattered around in the network.

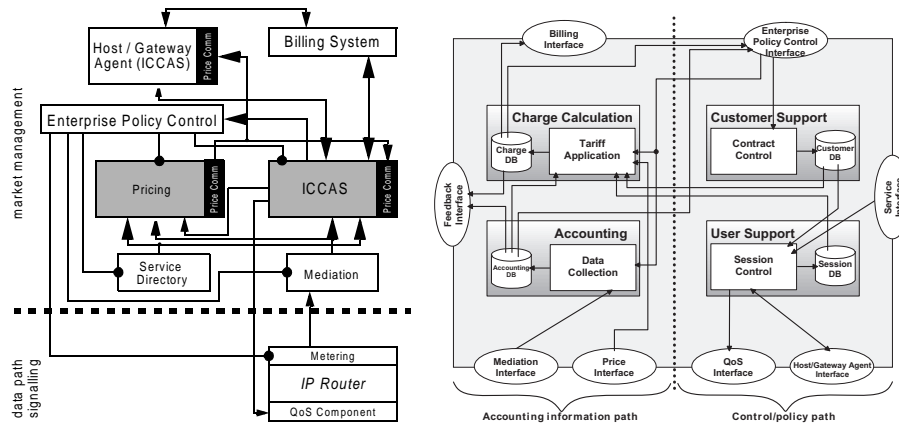


Figure 2: Architecture of the Overall Charging System and ICCAS Details

Since the transport of this data is not for free, ISPs will charge each other for the data transported. This leads directly to the second level of interaction. Each provider collects information of the amount of data transported and calculates a charge for it. He sends a bill through a billing system to the responsible neighboring ISP's entity. Either he bills another provider, if he provided a service for him or he sends the bill to one of his customers. So information exchange between providers occurs on the level of billing systems, where inter-provider invoices are exchanged. Instead of performing absolute billing between interconnected providers, they can also offset their claims against each other. A set of peering agreements and settlement schemes exist for today's ISPs, however, they are defined in a quite static manner and do not allow for immediate responses to bandwidth bottlenecks or further customer and user demands. In more specific detail, the tasks of each component of the ICCAS are defined within the following paragraphs.

Metering: As shown in Figure 2/left, metering is integrated in IP routers. Alternatively, it could be placed directly on the wire. This solution introduces supplementary expenditures, *e.g.*, an entity needs its own IP address or requires special protocols.

Mediation: In addition, the purpose of the mediation entity is to transform metered data (of each single meter), to merge data of different meters, and to reduce the amount of data metered. The mediation's effect of dramatically reducing the amount of data has a deep impact on the charging network design.

Pricing: For calculating charges of transmitted data, prices are important, which are applied in the charge calculation component, but they are not set at this point.

Billing: Major assumptions on functions of a billing system are to be made. It affects the charging system and the ICCAS in so far that interfaces need to be placed correctly and that plausible and complete scenarios can be created. It is used to describe inter-provider charging concerns.

Enterprise Policy Control: The enterprise policy control entity represents the ISP's interface for the management and supervision of all (except the billing system)

ICCAS related entities. It controls the ISP's business strategy with respect to its implementation and configuration of the given networking equipment. **Host/Gateway Agent:** The host/gateway agent performs two different functions. The first one is to communicate charges to hosts (users) and gateways to provide an optional feedback channel for their service usage. **Service Directory and QoS Component:** Since a user tends to lack a complete understanding of Quality-of-Service (QoS) in technical terms, he will be unable to specify detailed requirements in a way that can be used as an input to the QoS component within the router.

The ICCAS' internal entities consist of a charge calculation, an accounting, a customer support, and a user support component (cf. Figure 2/right). The separation of the ICCAS into these components increases the required degree of flexibility, since these components can be physically distributed. Concerning the flow of data within the ICCAS internally, it has been divided in two logical paths. On one hand, the Accounting Information Path (AIP), depicts the flow of pure charging-relevant data. On the other hand, the Control Policy Path (CPP) is used to manage and configure the ICCAS and entities involved with the processing of the charging data. These two paths differ mainly in the order and direction they process data. All interfaces between components are designed to act (1) as protocols, allowing for open communication between two remote entities of components, or (2) as software interfaces, reflecting the clear architectural decision, that the interaction between those components happens within a common address space. Further details can be obtained from [13].

5. Results and Evaluation

The integration of the pricing scheme CPS into the ICCAS based on a DiffServ networking environment has been performed. This combination does not lead to large protocol overheads and is feasible in small as well as in large networks.

5.1 CPS for DiffServ and the Charging System

CPS is tailored to complement DiffServ in the provisioning of QoS and in the transparency of customers. According to the DiffServ philosophy, CPS ties up at the edges of the network, i.e. at the border routers of any DiffServ domain. Consequently, CPS policies, which should be applied and enforced, need to be defined at and for edges of the network domain. Since DiffServ already uses Service Level Agreements (SLA) at the edges to comply with service provisioning tasks, CPS policy definitions and agreements preferably ought to be specified in the same SLAs. Merging service provisioning tasks with service charging tasks is appropriate and economical, since both apply on same service at a similar granularity and at identical locations in the network. Therefore, an SLA includes besides technical accordances, legal commitments and financial commitments, *e.g.*, the sum of financial forfeits on contract corruption, both of which not concerned with the charging task directly. The process to set up an SLA is sketched in a generic fashion in Figure 3.

Because SLAs do not already define technical parameters, ISPs have a some freedom to map the SLA contract to technical parameters. Furthermore, this freedom to comply with the contract allows to separate concerns of proper DiffServ technology from those of CPS, although they are specified in the same SLA. Figure 3 illustrates this fact by the *specification layer*, which is separated in a *charging plane* and a *DiffServ plane*. Concerns of charging are located in the *charging plane*. Likewise, this generic

representation allows to locate and specify parameters of charging systems and pricing schemes, other than CPS. Parameters necessary to be specified within a Service Level Specification include configuration and initialization parameters. In the case of CPS these cover, *e.g.*, threshold values for collecting cumulus points, prices associated to cumulus points, or refresh periods.

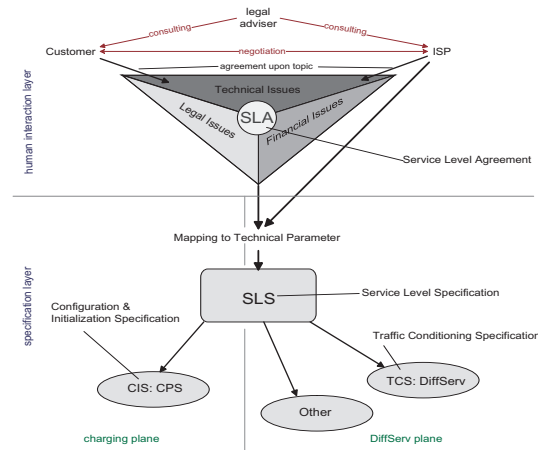


Figure 3: SLA of a DiffServ-CPS System

The main aspect of the DiffServ Architecture to concentrate intelligence at network edges suggests to meter at network edges only. Transmissions can be charged for according to their associated Behavior Aggregate (BA). However, an important question remains, how can end-to-end services be accounted and charged for? Each user has established a contract with his service provider. In a DiffServ world this contract is represented by an SLA. This SLA determines, which services the customer may use and how he will be charged for them using the applied pricing scheme, here CPS. However, the customer has only established an SLA with his provider, not with other domain operators, who play an important role in his data transmissions as well. Therefore, to guarantee the provision of a service not only for one domain, but for the entire transmission path, a flow establishment phase is needed. This setup can be performed by the User Support component of the ICCAS. This approach does not return to micro-flow orientation, it supports reservations for BAs, aggregated super-flows containing many micro-flows, by bandwidth brokers of the DiffServ network. A settlement between different domains does not take place on a micro-flow basis, but only for BAs. SLAs exist between domains, which are negotiated on the basis of traffic estimations of domain operators, which originally are caused by end user traffic. The specific settlement between neighboring domains is performed also by applying CPS.

5.2 Evaluation of CPS

The evaluation of the pricing scheme is based on the most important properties of the Cumulus Pricing Scheme, which include the following properties:

- **Discrete:** Instead of reporting every small fluctuation of the user behavior, CPS allows the user behavior to fluctuate within certain boundaries.

- **Cumulative:** Not single bursts, but only continuous change of promised behavior triggers reactions.
- **Early warning:** If some user starts to change her behavior, this will have no immediate consequences, but nevertheless she will know at an early stage that her pattern has changed, and that continuing the new usage habits will require a renegotiation of the current contract.
- **Predictable and transparent:** As the entire scheme presents itself to the user as a variation of fully accepted flat rate schemes, user acceptance will be a given. Charges for the user remain stable over long time-scales and necessary changes are transparent to the user.
- **Market managed:** The contract between customer and ISP is based on information delivered by the customer about her expected usage pattern. This information may be used by the ISP to determine the flat rate charge and moreover to optimize network utilization or revenue. As soon as the customer does not fulfill her commitments, CPS introduces a slow penalizing mechanism in order to trigger well-behavior of the customer.
- **ISP policy dependent:** There is a strong influence of ISP policy on the implementation of the scheme. The ISP may freely decide not only on the measurement procedure, but also on threshold values for awarding Cumulus Points or thresholds concerning their accumulation. On the other hand, semantics of CP allow the mapping of thresholds and values between different ISPs.
- **Technically feasible:** Awarding CPs most probably should be justified by respective measurement results, but it is largely up to the ISP of how to gain these data and how to interpret them. In the minimal case, a small number of very crude measurements of user behavior may already be considered as sufficient by the ISP.

6. Summary and Conclusions

This paper has presented the extension of current network management functionalities with charging and pricing functions for differentiated services, determining a business-oriented network management tasks. This extension is driven by a trend of commercialization in networking and the demand for QoS-based services, but minimized technical effort for QoS provisioning in networks. The development of a discrete, predictable, transparent, and technically feasible pricing scheme termed Cumulus Pricing Scheme (CPS) and the generic and modular ingredients of an Internet Charging System, in particular the Internet Charge Calculation and Accounting System (ICCAS) illustrates a feasible technical solution to the support of pricing differentiated services. Lessons learned clearly indicate that the proposed pricing scheme would have avoided recent failures of various all-you-can-eat offers, *e.g.*, as seen from a U.K.-based Internet Service Provider (ISP) [18]. In addition, a quite simple set of measurement policies, long-term negotiable between the customer and the ISP, are sufficient and can be supported effectively by the ICCAS. Specifically, ICCAS is suited for handling an extensible variety of different services, where particularly the CPS has been integrated. In addition, ICCAS interactions are functionally complete and sufficient for known pricing schemes, ranging from usage-based to flat-fee approaches.

The work in progress includes the ongoing implementation of the full ICCAS functionality in order to demonstrate the technical and economic feasibility. The optimization and tuning of interfaces and protocols will follow based on the measured

processing overhead for the ICCAS. The simulation of the market-managed feature of CPS with multiple ISPs is under its way and will be extended to inter-ISP scenarios. This includes the simulation of ISP policies, effects on prices, and user behavior. With respect to service deployment [15] and active networks it will be an interesting research topic to investigate the suitability of pricing models for injected code.

Acknowledgments

This work has been performed partially in the framework of the EU IST project Market Managed Multi-service Internet (M3I, IST-1999-11429), where ETH Zürich has been funded by the Swiss Bundesministerium für Bildung und Wissenschaft, Bern (No. 99.0536) and the ANAISOFI project funded by the Swiss National Science Foundation, Bern (No. 5003-057753/1). Special thanks go to M. Karsten, J. Schmitt, and O. Heckmann at the Technische Universität Darmstadt, Germany for the joint design of the overall architecture within M3I. Additionally, the authors like to acknowledge many discussions with all of their project partners.

References

- [1] F. Baumgartner, T. Braun, P. Habegger: *Differentiated Services: A New Approach for Quality of Service in the Internet*; In: H. van As (ed.): *High Performance Networking*, Kluwer, Dordrecht, The Netherlands, 1998, pp. 255-274.
- [2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss. *An Architecture for Differentiated Services*; Internet Engineering Task Force, RFC 2475, December 1998.
- [3] R. Braden, D. Clark, S. Shenker: *Integrated Services in the Internet Architecture: On Overview*; Internet Engineering Task Force, RFC 1633, June 1994.
- [4] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin: *Resource Reservation Protocol (RSVP)*; Internet Engineering Task Force, RFC 2205, September 1997.
- [5] B. Briscoe, M. Rizzo, J. Tassel, K. Damianakis: *Lightweight Policing and Charging for Packet Networks*; Third IEEE Conference on Open Architectures and Network Programming (OpenArch 2000), Tel Aviv, Israel, March 26-27, 2000.
- [6] N. Brownlee: *Traffic Flow Measurement: Experiences with NeTraMet*; Internet Engineering Task Force, RFC 2123, March 1997.
- [7] R. Cocchi, D. Estrin, S. Shenker, L. Zhang: *Pricing in Computer Networks: Motivation, Formulation and Example*; IEEE/ACM Transactions on Networking, Vol. 1, No. 6, December 1993, pp 614-627.
- [8] S.A. Cotton (ed.): *Network Data Management – Usage (NDM-U) for IP-Based Services*; IPDR Specification Version 1.1, June 2000.
- [9] R. Edell, N. McKeown, P. P. Varaiya: *Billing Users and Pricing for TCP*; IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, July 1995, pp 1162-1175.
- [10] R. Edell, P. P. Varaiya: *Providing Internet Access: What we learn from the INDEX trial*; Keynote Talk at Infocom'99 New York, INDEX Project Report #99-010W. URL: <http://www.index.berkeley.edu/99-010W>.
- [11] G. Fankhauser, B. Stiller, C. Vögtli, B. Plattner: *Reservation-based Charging in an Integrated Services Network*; 4th INFORMS Telecommunications Conference, Boca Raton, Florida, U.S.A., March 1998, Session MC-2.
- [12] D. Ferrari, L. Delgrossi: *Charging for QoS*; Invited Talk, 9th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV'99), London, U.K., 1999.

- [13] J. Gerke, P. Flury, B. Stiller: *The Design of a Charging and Accounting System for the Internet*; TIK-Report No. 91, ETH Zürich, Switzerland, August 2000.
- [14] A. Gupta, D. O. Stahl, A. B. Whinston: *Managing the Internet as an Economic System*; CISM, University of Texas at Austin, U.S.A., July 1994.
- [15] R. Haas, P. Droz, B. Stiller: *A Hierarchical Mechanism for the Scalable Deployment of Services over Large Programmable and Heterogeneous Networks*; IEEE International Conference on Communications (ICC'01), Helsinki, Finland, June 11-15, 2001.
- [16] H.-G. Hegering, S. Abeck, B. Neumair: *Integrated Management of Networked Systems*; Morgan Kaufmann Publishers, San Francisco, California, U.S.A., 1999.
- [17] M. Karsten, J. Schmitt, L. Wolf, R. Steinmetz: *An Embedded Charging Approach for RSVP*; 6th IEEE/IFIP International Workshop on QoS (IWQoS'98), Napa, California, U.S.A., May 18-20, 1998.
- [18] N. McIntosh: *Heavy Surfers Pay the Price*. The Guardian, England, August 3, 2000.
- [19] J. MacKie-Mason, H. Varian: *Pricing Congestible Network Resources*; IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, 1995, pp 1141 – 1149.
- [20] M3I: *Market Managed Multi-service Internet*; 5th Framework EU Project, IST Program, No. 11429, URL: <http://www.m3i.org>, January 2001.
- [21] K. Nichols, V. Jacobson, L. Zhang: *A Two-bit Differentiated Services Architecture for the Internet*. Internet Engineering Task Force, RFC 2638, July 1999.
- [22] A. Odlyzko: *Paris Metro Pricing: The Minimalist Differentiated Services Solution*; 7th International Workshop on QoS (IWQoS'99), London, U.K., June 1-4, 1999, pp 159-161.
- [23] S. Shenker, D. Clark, D. Estrin, S. Herzog: *Pricing in Computer Networks: Reshaping the Research Agenda*; ACM Computer Communication Review, Vol. 26, No. 2, April 1996, pp 19 – 43.
- [24] D. Songhurst (edt): *Charging Communication Networks - From Theory to Practice*; Elsevier, Amsterdam, The Netherlands, 1999.
- [25] A. Sirbu: *Internet Billing Service Design and Prototype Implementation*; Carnegie Mellon University, Computer Science Department Technical Report, Pittsburgh, Pennsylvania, U.S.A., 1994.
- [26] B. Stiller, T. Braun, M. Günter, B. Plattner: *The CATI Project: Charging and Accounting Technology for the Internet*; 5th European Conference on Multimedia Applications, Services, and Techniques (ECMAST'99), Madrid, Spain, May 26-28, 1999, LNCS, Springer Verlag, Heidelberg, Vol. 1629, pp 281-296.
- [27] B. Stiller, G. Fankhauser, N. Weiler, B. Plattner: *Charging and Accounting for Integrated Internet Services - State of the Art, Problems, and Trends*; The Internet Summit (INET'98), Geneva, Switzerland, July 21-24, 1998, Session Commerce and Finance.
- [28] B. Stiller, J. Gerke, P. Reichl, P. Flury: *The Cumulus Pricing Scheme and its Integration into a Generic and Modular Internet Charging System for Differentiated Services*; TIK Report No. 96, ETH Zürich, Switzerland, September 2000.
- [29] B. Stiller, P. Reichl, S. Leinen: *A Practical Review of Pricing and Cost Recovery for Internet Services*; to appear: NETNOMICS - Economic Research and Electronic Networking, Vol. 3, No. 1, March 2001.
- [30] I. Stoica, H. Zhang: *Providing Guaranteed Services Without Per-flow Management*; ACM Computer Communication Review (SIGCOMM'99), Vol. 29, No. 4, October 1999, pp 81-94.
- [31] B. Teitelbaum, P. Chimento: *QBone Bandwidth Broker Work Group*. URL: <http://qbone.internet2.edu/>, January 2001.