A Generic and Modular Internet Charging System for the Cumulus Pricing Scheme

Burkhard Stiller, Peter Reichl[1], Jan Gerke, Placi Flury

Computer Engineering and Networks Laboratory TIK, Swiss Federal Institute of Technology, ETH Zürich, Switzerland
E-Mail: [stiller | gerke | reichl | flury]@tik.ee.ethz.ch

1. Peter Reichl is now with FTW, Wien, Austria.

Abstract

Pricing and charging are important management functions which future commercial networks will have to offer. Since the Internet is on the move to provide differentiated services, suitable and scalable management mechanisms are required for the backbone based on the Differentiated Services (DiffServ) Architecture. Since network management following actual market forces has to implement a price-driven and charge-driven approach for managing the usage of services, this paper introduces as a first contribution the design of a generic and modular Internet Charging System (ICS), offering a service-independent architecture and integrating economically controlled network management functions into present Internet technology. Moreover, a central part of this ICS, i.e. the Internet Charge Calculation and Accounting System (ICCAS), is presented in more detail and offers necessary user support functions. The second main contribution of the paper focusses on the newly developed Cumulus Pricing Scheme (CPS). CPS is unique to define clear relations between different time-scales for accounting, monitoring, and charging activities. Charges in this schemes are based on flat fees and, hence, are predictable and transparent, whereas, unlike pure flat rate schemes, the incorporated feedback mechanism nevertheless allows to take market forces into account. Finally, it is demonstrated how the ICCAS is well-suited to integrate both DiffServ technology and CPS.

Keywords: *Charging Architecture, Internet Pricing Model, Differentiated Services, Service Level Agreements.*

Short Title (Run-in Header): *Internet Charging with Cumulus Pricing*

1 Introduction

Despite the huge amount of bandwidth provided by today's Internet, bottlenecks have not disappeared, as dropped packets or long delays experienced by certain packets easily demonstrate. 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 appropriate mechanisms for managing, providing, and supervising a specified set of QoS parameters. Therefore, pricing and charging mechanisms will become highly relevant for offering more than a single service class, *i.e.*, a best-effort and a statistically guaranteed class, since financial incentives are the most effective stimuli for triggering user behavior according to users' real communication demands. These mechanisms are considered to be an essential part of the business and technology management as well as services management, as defined in the management framework [19]. However, the selection of a suitable pricing scheme for a single-service network is easier than for a multi-service network [12]. Therefore, the issue becomes even more complicated for multiple QoS levels.

Since Differentiated Services (DiffServ) define an emerging technology for a scalable, multi-services Internet backbone [28], an appropriate solution for pricing and charging these services is essential within a commercialized Internet services market. Internet Service Providers (ISP) are the key players in the communication market and they will offer differentiated Internet services in the near future. Flat fee pricing schemes seem to be favored by users, whereas usage-based charging mechanisms seem to be only viable, if they can be implemented efficiently [10], [34]. Therefore, a suitable combination of basic flat fees and time-dependent or volume-dependent charges as well as feedbacks from utilization of differentiated services could offer an appropriate mixture of both worlds. Such schemes are considered essential management mechanisms on different time-scales and they directly affect change management, operations management, and performance management [24] and need to be supported by an efficient Internet charging system. Although a variety of pricing schemes for Internet access exists, this paper establishes a new perspective on them by considering the flow of market management information itself to be as important as their content. Therefore, three main types of pricing schemes are 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 depend on the network status, depending indirectly on the behavior of a single user, i.e. through his externalities [15]. Flat rate schemes are generally independent of metering information. With respect to DiffServ, the future Internet backbone will consist of many DiffServ domains, each of them operating under a different administration. Therefore, two principally distinct locations for applying pricing and charging mechanisms exist. These are (a) the interface between a user and its DiffServ provider and (b) the interface between two DiffServ providers [34], [3].

To overcome the lack of suitable solutions, this work contributes (1) a generic and modular charging system for a multi-service Internet, specifically targeted towards DiffServ service classes, and (2) a better than flat fee pricing scheme for multiple service classes termed Cumulus Pricing Scheme (CPS). The Internet Charge Calculation and Accounting System (ICCAS) offers a service-independent architecture and functions, which can be instantiated according to particular ISP requirements. CPS is the only approach known so far, which defines a clear relationship between different time-scales of accounting periods, measurement periods, and charging periods.

The remainder of this paper is organized as follows. Based on related work, terminology, and an introduction into DiffServ, the developed ICCAS is presented in Section 2. Afterwards, the proposed CPS is defined and discussed in Section 3. While Section 4 maps CPS onto the ICCAS in a DiffServ environment, Section 5 summarizes the work, draws conclusions, and discusses future work.

1.1 Related Work

The number of projects concerned with pricing, charging, and accounting tasks in the Internet have increased quite significantly over the last few years. Therefore, only a few recent systems and modeling approaches are summarized. While pricing model overviews can be found in [21], [36], many projects dealing with charging and accounting functionality on the network level try to achieve a high independence from such pricing models [37]. However, it has been argued that pricing in general and usage-based pricing in particular can impose a high overhead on telecommunication systems [25], [31]. Any form of usage-based pricing for various telecommunication services is interesting, because underlying resources (such as satellites, frequencies, cables, routers/switches, and most notably operating personnel) are scarce and very costly. Internet pricing has been criticized constantly in the past years for its economic drawbacks of not being incentive-compatible [5], [6], [18], and [31]. Furthermore, 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 that provides a similar service to all customers. Therefore, the multi-service paradigm needs to be investigated with respect to the heterogeneous networking infrastructures and technologies of the Internet [30].

An early per-flow billing system for TCP (Transmission Control Protocol) flows and initial ideas on a billing service design are presented in [9] and [33] respectively. Advanced per-flow charging and accounting approaches based on reservations have been tackled in [11] and [22]. The objectives of the Swiss National Science Foundation project CATI (Charging and Accounting Technology for the Internet) [34] included the design, implementation, and evaluation of charging and accounting mechanisms for Internet services and Virtual Private Networks (VPN) [17], mainly focussed on flow-based approaches. This covered the technology support for open, Internet-based E-Commerce platforms in terms of usage-based, per-flow transport service charging [8], [11] as well as advanced and flexible configurations for VPNs. The main assumption of “Lightweight Policing and Charging” is that a multi-service packet network may be achieved by adding classification and scheduling to routers, but not policing [3]. Another highly important question concerns the issue of user acceptance of pricing. In 1998 the INDEX project (Internet Demand Experiment) started in order to investigate user reaction when exposed to various pricing schemes for different qualities of Internet access. It turned out that the users were not averse to flexible pricing models. Moreover, the widespread flat rate model, at least in its pure form, proved to tend toward a waste of resources, unfairness among users, and revenue losses for ISPs [10]. The project M3I (Market-Managed Multi-service Internet) [27] started recently and aims to design and implement a next-generation system enabling Internet resource management through market forces, specifically by enabling differential charging for multiple service levels.

1.2 Terminology

Based on related work and the necessity to obtain an uncontradicted terminology, the following definitions are applied.

- **Accounting** defines the summarized information (accounting records) in relation to a customer’s service utilization. It is expressed in metered resource consumption, *e.g.*, for the end-system, applications, calls, or any type of connections.
- **Billing** defines the collection of charging records, summarizing their charging content, and delivering a bill or 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 within this work as an overall term, depicting at a higher layer of abstraction 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 of charging.
- **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. Prices may be calculated on a cost/profit base or on the current market situation.
- **Tariff** defines the algorithm used to determine a charge for a service usage. It is applied in the charge calculation for a given customer for the service he utilizes. Tariffs may contain, *e.g.*, discount strategies, rebate schemes, or marketing information.

1.3 Differentiated Services in the Internet

In the early days of the Internet transport of data was designed as a best-effort service. With increasing QoS requirements of applications, several proposals have been submitted to overcome this heritage. The important ones include the Integrated Services Architecture (IntServ) [2] and the Differentiated Services Architecture (DiffServ) [28]. In DiffServ single application flows are assigned and accumulated to Behavior Aggregates (BA) at edges of the DiffServ domain. Packets belonging to a specific BA are identified by DiffServ Code Points (DSCP). The association and enforcement of the BA's QoS is fulfilled by specific forwarding policies, the Per Hop Behaviors (PHB). Services in terms of network delivery characteristics can be described by BAs and their associated PHBs. Figure 1 depicts a situation, where four different flows (flow 1 to 4) are accumulated into two different BAs. Since the DiffServ Border Router and the network offers only a limited amount of resources, flow set up has to be requested and accepted flows have to comply to specified traffic characteristics. The access request, implying a commonly known admission control, and the conditioning of the traffic, necessarily according to a contract, demand signaling and appropriate contract negotiations.

Figure 1

DiffServ quantifies services to a small number - at least compared to IntServ - of classes. Applications need to meet their QoS requirements by choosing from this pool of service classes. These service classes either result from the combination of BAs and their PHBs or are simply best-effort services. The relationship between the customer, i.e. service user, and the ISP, i.e. service provider, is required to be expressed by one or more contracts called Service Level Agreements (SLA) [1]. An SLA comprises technical, *e.g.*, expected bandwidth requirements or certain QoS parameters, and legal and financial aspect, *e.g.*, tariff structure or usage charges, for service delivery and in case of service rejection for possible compensations. The specific form of SLAs applicable to the Internet is still rather open and subject to discussion. Traditional telecommunication SLAs have been utilized in telephone networks for years, but they lack characteristics of Internet traffic.

From an SLA, which is defined and set up by human parties, Service Level Specifications (SLS) are deduced, representing the commitment for resource reservations [16], [38]. Each SLS contains technical details for service provision. SLSs have the advantage that they can be negotiated and set up automatically between communication parties without any user intervention, which is especially advantageous for dial-up and mobile customers. From an SLS Traffic Conditioning Specifications (TCS) are derived which specifies classifier, metering, marking, shaping and dropping rules which are to be applied to the selected BA. A typical SLA setup and enforcement requires, that a customer contacts his ISP for specification of his requirements. In this primary user point of view, the customer estimates the QoS he needs for his applications. According to his knowledge parameters can either be very precise, *e.g.*, max delay and jitter, or high-level, *e.g.*, 'support of a web server'. At this stage it is up to the ISP to map a customer's QoS formulation to its supported DiffServ classes and to its available bandwidths. This mapping is transparent to the customer, nevertheless an SLA determines the result of a dialog, which customer and ISP have to agree upon. Besides technical aspects financial and legal ones have to be discussed and affirmed as well. The customer-ISP relation will be defined for the developed pricing approach formally in Section 3.1.4 and applied to the ICCAS later in Section 4.1.

2 Design of a Generic and Modular Charging System

In existing charging systems of today's providers, setting prices, the function of charge calculation, and billing itself are integrated; and the maintenance of service classes, user profiles, customer data, identities, and banking account data are included.

Although these tasks can be distinguished clearly (cf. terminology in Section 1.2), for existing systems they are almost completely centralized within a single system. However, future charging systems need to integrate a variety of different charging and accounting records, Call Detail Records (CDR) and Internet Protocol Detail Records (IPDR) [7], from different communication providers or content providers, since customer's demand is determined by the so-called "one-stop billing" approach [37]. This suggests strongly the need to divide existing monolithic charging systems into several components with open interfaces, thus, enabling the exchange of individual components and allowing for the integration of components supporting different technologies without having to adapt the entire system. Therefore, the approach developed provides a generic and modular charging system in support of various pricing schemes applicable to different communication technologies [14], [35].

2.1 Tasks of an Internet Charging System

Task descriptions for a generic and modular charging system provide the basis for design guidelines of its internal components and external interfaces. Charging systems are supposed to support the following tasks:

- **Perform transport, service, and content charging**: The optimal design for an Internet Charging System includes a combined approach of different levels of charging. Transport charging, termed network charging or network access charging as well, forms the basis for providing a system to deal with the transfer of data, based on a packet-based network infrastructure, such as the Internet. The service charging located above this level allows for the clear distinction of different services including different QoS requirements and resource consumptions. Finally, content charging resides a the top.
- **Perform accounting tasks according to transport and multi-service definitions**: Data gathered from the physical infrastructure and mediated due to policies, need to be accounted for. This requires the knowledge of "sessions", "durations", or "flows". Mainly, this information is derived from data metered, such as "begin-of-session" or "end-of-flow". If such information cannot be determined explicitly, heuristics need to be applied for session or flow detection purposes. In any case, the "length" of a communication will be recorded, if any usage-based charging approaches are to be supported.
- **Support different levels of security for charging information**: All data, which can be mapped to monetary equivalents, have a certain degree of sensitivity. Based on the dedicated level of interest, a single accounting record, a single metered routing data, or a charging record may not be a security problem, since their lifetime and validity, and therefore value, are short. But combinations of aggregated data, *e.g.*, flow-related information in terms of usage information, duration, and customer identification, form critical information.
- **Support different pricing and enterprise policies**: As network service provisioning forms a part of an open market, different business models, pricing schemes, and enterprise policies need to be flexibly supported by the Internet charging system.

2.2 Charging System Design

The design of a charging system requires a clear separation of tasks. As depicted in Figure 2a, components are based on IP routers, which are part of the ISP's network. Therefore, the networking architecture itself (with or without charging support) remains unchanged, only metering needs to be integrated. For advanced feedback control loops between service usage and the user, a QoS component inside the IP router is essential to perform QoS management tasks. Therefore, the components located within the market management area of Figure 2a are inherently part of the network management architecture for economic-driven network operation and support systems.

2.2.1 Components and Operation Basics

Every provider maintains a network consisting of routers and network links between them, metering functions, mediation systems, accounting systems, and a billing system. Metering functions can be implemented in independent devices (passive metering) or integrated into routers (active metering). In either case, they generate usage data. Due to the fact that the amount of usage data is usually too large to be sensibly processed, it must first go through mediation where it is transformed into a form suitable for further processing. The resulting accounting information (base accounting records) are gathered and accumulated in accounting systems. These accounting systems, in turn, forward all accumulated and perhaps abstracted accounting information through a charge calculation function towards the billing system interface. These raw usage data are not yet associated with a specific user or customer. Therefore, it is necessary to identify a customer responsible for every sequence of data units sent. This is performed within mediation, since the amount of data leaving the mediation components can be much smaller due to the aggregation of usage data. The charge calculation, which receives pricing information from the pricing component, translates the accounting information into charging records, hence, it maps the resource-oriented information from the accounting systems into monetary values. Within the charge calculation, discounting strategies, marketing-driven pricing schemes, or simply fixed prices can be applied. All pricing and services-related information are specified in a tariff. Finally, the billing system uses these values to prepare the bills to be sent to customers. These tasks are conceptually coordinated within an overall high-level architecture and need to inter-operate as depicted in Figure 2a.

Figure 2

2.2.2 Design Dimensions

While the components for the charging system have been identified, still it has to be determined how these components are implemented and deployed in possible scenarios with potentially several different ISPs. There exist three dimensions in which a charging system can vary according to the scenario (cf. Figure 2b). 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 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 other dimensions of location and replication as well as heavily on the component type itself. ***Time-scales*** define the important dynamic criteria for feedback handling as discussed for pricing schemes (cf. the design of CPS in Section 3.1 later).

Table 1

According to [19], overall management time scales are distinguished as follows: 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 potential feedback in a round trip-time or milliseconds. As summarized in Table 1, 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. The atomic (communication-relevant) time-scale involves sending packets, round-trip times, and managing feedback between sender(s) and receiver(s). The short-term (application-relevant) time-scale is concerned with the usual duration of applications like file-transfer, video-conferencing, or IP phone calls. The tasks of accounting and measuring are closely related to these activities. The medium-term (billing-oriented) time-scale for performing billing actions is arbitrary, since it is influenced strongly by the usual lifestyle habits of humans, which includes, *e.g.*, monthly payments of rents, phone charges, or newspapers. Finally, the long-term (contract-specific) time-scale in this context is the duration of contracts between customers and ISPs, which usually varies from several months to years

2.3 Charging System and ICCAS Architecture

Following the design dimensions and task descriptions, a modular and generic architecture with open interfaces and interaction paradigms has been developed, including an ICCAS subsystem (Internet Charge Calculation and Accounting System).

2.3.1 Charging System Provider Interactions

Refining the concept into components and their interactions and interfaces results in the overall architecture as depicted in Figure 3a, as it has been designed for the M3I project [20], [35], and [27]. While important tasks of these components are discussed later in more detail in Section 2.3.3, interactions between two neighboring providers take place at two levels. The first one is 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.

Since the transport of this data is not 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.

2.3.2 ICCAS External Components

For describing the charging system completely, an outside-first approach has been taken, which illustrates all components external to the ICCAS itself first. ICCAS details follow in Section 2.3.3.

As shown in Figure 3a, ***metering*** is integrated in the IP router. Alternatively it could be placed directly on the wire. Indeed, such a solution introduces supplementary expenditures, *e.g.*, an entity needs its own IP address or requires special protocols. Furthermore, it can only monitor the actual usage of the link and has no knowledge of usage of any critical resources relevant for congestion control within the router. Therefore, the interconnection of several metering units to reconstruct the current router status is not feasible. Because of this, it would be necessary, in spite of having metering units on the wire, to know the state of the router, so an explicit interaction of the charging system and the router would be required. 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. For calculating charges of transmitted data, prices are important. These prices are applied in the charge calculation component (see Section 2.3.3) but they are not set at this point. There are many different ways to set prices and, therefore, a separate

pricing component performs this task. It can make use of economic models or just use fixed prices set by hand. For dynamic pricing models most often an input from metering is needed, since the amount of data traffic influences prices. The functions of an existing ***billing*** system affect the charging system and the ICCAS in particular 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 as mentioned above.

Figure 3

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 covers and controls the ISP’s business strategy with respect to its implementation and configuration of the given networking equipment. 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. In this case, the host agent acts on behalf of the user. This can include the negotiation of services with the ICCAS, an automatic reaction to communicated charges, or even payment information. A host agent can also restrict user’s options, when the customer in control of these users wants to restrict the behavior of his users he pays for. In particular, this is the case for companies, where the company is in the role of the customer of an ISP and the employees act as users of the services offered. Finally, since in general a user tends to lack a complete understanding of QoS in technical terms, he will be unable to specify detailed requirements in a way that can be used as a direct input to the ***QoS component*** within the router. Instead the user has a higher, application level view of quality. This view must be translated into technical values, which can be used for setting parameters in QoS components and for charging according to technical usage data. This translation takes place in the ***service directory***.

2.3.3 ICCAS Architecture and Internal Components

After having discussed external components of the ICCAS, its internal entities consist of a charge calculation, an accounting, a customer support, and a user support component. The separation of the ICCAS into these components increases the required degree of flexibility, since these components can be physically distributed as discussed above. Embedding the ICCAS into the overall charging system is provided through eight distinct interfaces (cf. Section 2.4). Concerning the flow of data within the ICCAS internally, it has been divided in two logical paths (cf. Figure 3b). The Accounting Information Path (AIP), depicts the flow of pure charging-relevant data. The Control Policy Path (CPP) is used to manage and configure the ICCAS, especially all entities involved with the processing of the charging data. These two paths differ mainly in the order and direction they process data. The AIP starts from the bottom of the graph (taking raw technical data) by processing metered, mediated, and pricing data. It ends on the top of graph, where complete charging records are handed over to the billing system. In contrast, the CPP starts from the top of the graph (taking business-related data) by receiving enterprise policy control information and processes down to the bottom of the graph, resulting in QoS control data to be handed over to the underlying router and optionally an agent.

The ***accounting*** component receives all metered and mediated usage data and is responsible for storing these. It must provide these stored data to other ICCAS components and interfaces for further processing, feedback, or statistic evaluation. Accounting is the central usage data storage component. The ***charge calculation*** component processes the accounted for usage data. It calculates appropriate charges for resource usage applying a tariff (which is for instance defined by the newly developed Cumulus Pricing Scheme in Section 3), communicated by the pricing component. To be able to determine all charges fully, it needs input from the user support, *e.g.*, user identifications. An ISP may have many and a large number of different customers. Additionally, a customer is not the same as a user, *e.g.*, one customer might pay the bills of several users. A customer negotiates a contract with the ISP. The content of this contract, *e.g.*, number of users covered by the contract and their names and accounts, are managed within the ***customer support***. While the customer support component is responsible for keeping all contract information the ***user support*** component is responsible for making sure that those contracts are kept. On one hand, this means that he blocks any user requests that are not covered by the contract the customer has. On the other hand, he must make sure that a service requested by a user is delivered to him, if the contract allows it.

2.4 ICCAS Interfaces

In general, all interfaces between the components described are designed to act (1) as protocols, allowing for the 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. The following list summarizes interfaces required, identifies the Accounting Information Path (AIP) and Control Policy Path (CPP), and outlines important examples of AIP functionalities. Further details on these interfaces can be obtained from [14].

- **QoS Interface (CPP):** To provide services to the customer it is necessary to control the QoS component of routers. This interface can be used to set QoS parameters of routers, depending on the technology in place.
- **Mediation Interface I-MA (AIP):** The mediation interface is responsible for collecting data from several mediation entities, possibly even from mediation entities of different types. The usage data, which have been mediated after data gathering, need to be transferred to the ICCAS. The data exchanged across this interface (cf. Table 2) will include one of the following alternatives, which depend on the particular scenario: a simple hand-over of data gathered by metering or a hand-over of mediated data based on the particular inputs from the enterprise policy control. This may result in the dedicated specification of specialized data to be required for the ICCAS, some special aggregation of these data, or even the neglecting of data resulting from the gathering process.
- **Enterprise Policy Control Interface (CPP):** This is the interface for changing parameters of the ICCAS after the system has been deployed. By using this interface the enterprise policy control can install new services or request and receive charging or accounting data.
- **Service Interface (CPP):** This interface can be used to read service definitions out of the service directory.
- **Billing Interface (AIP):** This interface is responsible for sending calculated charging records to the billing system providing access to billing institutions. It supports facilities to identify customers and enterprises in financial terms and is responsible for accessing economic accounting tasks. Furthermore, it identifies the paid-for service. This is done by giving it an ID (corresponding to the one in the charging data base) and by attaching all the requested characteristics of the service. The service contract and the costs are also passed to the interface to legitimize and detail a list of costs.
- **Pricing Interface I-CP (AIP):** Pricing is responsible for setting prices used by the charge calculation component, therefore, this interface is used to send calculated prices to the charge calculation component. The ICCAS requires input from the pricing component, in particular details on prices and tariffs for services. A generic way to exchange this price information between different participants is defined, allowing for a large range of information and efficient transmission. Charge calculation-relevant details for this protocol encompass memory-efficient data structures for prices and tariffs, processing-efficient data structures (context-free) for prices and tariffs, and inclusion of customer and/or user identification information for prices and tariffs. Since the pricing reflects a well-defined means for communicating prices and tariffs across networks, a software-based interface between the “hosting” component and the pricing component are local matters only.

- **Feedback Interface I-FI (AIP):** To set suitable prices the pricing component uses price models with various input variables. Some price models need usage or charge information as input variables, hence, these can be communicated to the pricing component via this interface. The feedback interface provides the necessary input for the pricing component to adapt and change prices according to the current network status. The feedback represents the status of the network in terms of resource usage, i.e. quantitative usage given by the amount and characteristic of the accounting and charging records, as well as in terms of the current charging and accounting policy.
- **Host/Gateway Agent Interface (CPP):** This interface is responsible for the optional communication with the customer. Mainly, this includes the selection of services the customer can use or the transfer of a feedback signal from the service provider to a user. This interface is open for future enhancements.

3 Cumulus Pricing Scheme

Over the last couple of years, the issue of pricing and tariffing Internet services has become an important and wide-spread research issue. However, there is still a lack of definite solutions, since proposed schemes either have turned out to be too inflexible or too complex for a satisfying technical solution. Hence, there is still an urgent need for new pricing schemes that are able to deal with the various requirements, such as technical and economic ones, demanded of them.

3.1 CPS in Detail

The new Cumulus Pricing Scheme[2] (CPS) takes advantage of different existing schemes while avoiding most of their problems. Therefore, based on the identification of requirements the general idea is presented first. Afterwards, the new methodology of time-scale mappings in pricing is introduced and a formal representation of CPS is provided.

3.1.1 Requirements

CPS has been developed with respect to the following three main requirements as illustrated in Figure 4.

- **RT-1: Customer.** Over the last couple of years, there has been an extensive discussion on the preferences customers show towards dynamic tariff schemes for Internet services, *e.g.*, INDEX [10], CATI [34], and M3I [27]. It has turned out that in this context flat-rate pricing is still regarded as one of the most important approaches, if not in fact the most important one. All investigations performed within the work presented here have taken this fact into consideration. Therefore, the idea of CPS uses flat rate pricing as a conceptual starting point by enhancing it and bringing a degree of dynamicity into play.
- **RT-2: ISP - Economic.** As an economic entity an ISP is interested in maximizing either network utilization, total revenue, or similar objectives. To this end, charging and pricing schemes represent an important interface to the customer: prices may, *e.g.*, be used to indicate good behavior or misbehavior of the customer or to signal the congestion state of the network. The idea is to use market forces (which reveal themselves in terms of prices) for network management. But this is only possible, if there is a feedback mechanism, which allows the ISP to communicate the relationship between current customer behavior and overall system status.
- **RT-3: ISP - Technical.** The aforementioned feedback mechanism depends on the existence of respective mechanisms for technical accounting. This requirement opens a vast field of possibilities as to which detailed data on the network status are to be obtained. It appears almost impossible to commit ISPs to the availability of a certain minimal measurement requirement as the technical conditions vary enormously. Therefore, the approach proposed here pre-supposes the availability of

2. Due to naming conventions as defined in Section 1.2, the developed Cumulus Pricing Scheme reveals a to be in fact a Cumulus Tariff Scheme, but for historical reasons, the name has been kept as is.

measurements, but leaves the level of detail and granularity deliberately to the respective ISP. This openness will turn out to be one of the major advantages of the new scheme.

Figure 4

3.1.2 General Idea of Cumulus Pricing

Common Internet pricing schemes usually have trouble with at least one of the three requirements presented above. The fundamental decision between static and dynamic schemes touches immediately the 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. A prominent example is provided by Kelly's effective bandwidth approach for ATM pricing, where practical considerations forced the implementation to be based on bounding approximations to the effective bandwidth function instead of the latter function itself, cf. [32], p. 24f. In this situation, the Cumulus Pricing Scheme 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" indicating user behavior), and it allows a huge flexibility in terms of the technical prerequisites, especially concerning the measuring and accounting mechanisms of the required data records. 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*. As discussed in Section 2.2, these time-scales play the important role for today network management principles. With respect to CPS, the Cumulus Point Rule (cf. Section 3.1.4 below) refers to the medium-term time-scale, whereas the Reaction Rule belongs to the long-term scale. Moreover, measurements as described in Section 3.2 take place on the short-term scale, hence, in this respect CPS may be characterized as multi-dimensional mapping of the atomic time-scale onto the other three.

3.1.3 New Methodology

Moreover, these three requirement types may be assigned to appropriate time-scales. In doing so, it is noticeable that they are placed apparently at transition points between time-scales: transparency and predictability are relevant on medium to long-term scales, economic efficiency refers to short and medium-term scales, and the accounting technology deals with atomic and short-term events. Thus, the problem of balancing these requirements turns out to be an issue of reconciling these different time-scales. Therefore, the analysis of how these scales may be used for constructing a solution is important.

Usual flat rate schemes are basically intended to have no a priori time limitations, i.e. the rate is supposed to remain constant more or less forever. In order to introduce market forces, the first necessary step is to break with this supposition by introducing the possibility of price changes on the level of contracts (long time-scale). I.e. the tariff no longer guarantees the flat rate to be constant in the long run, but explicitly allows renegotiation of the contract (and here especially of the charges) in case of, *e.g.*, user misbehavior or fundamental system changes. The second important time-scale is the one connected with billing periods (medium time-scale). On this level, *e.g.*, in connection with their monthly bill, the user receives feedback about their current behavior and its consequences for the future validity of the contract. If the user has been detected by the ISP to strongly overuse capacities or to misbehave in some sense, they will usually receive some sort of warning that in case of unchanged behavior the contract may sooner or later be terminated by the ISP. On the other hand, if the user is underutilizing the service, this behavior may be rewarded by some sort of bonus system. This feedback system depends strongly on measurement activities that are supposed to take place on the application time-scale, but are deliberately left open to the ISP. Hence, one could think of (1) an ISP monitoring each packet or each connection in one extreme, (2) ISPs undertaking periodic monitoring, (3) ISPs measuring in statistical framework, or (4) at the other extreme, ISPs not measuring at all as a less interesting approach.

The basic communication process, however, still has to happen at the atomic level. As these activities have been identified as causing the essential feasibility problems, it is proposed to ignore them and instead express them by activities on the three remaining time-scales. Therefore, the resulting pricing scheme is characterized by mapping of the atomic scale onto the short, medium, and long-range scale. This paradigm shift has immediate consequences, especially with respect to the design process itself, which is reduced to finding useful combinations of possible mechanisms on the three larger time-scales.

3.1.4 Mathematical Description

Suppose that ISP *I* offers a service, and initially customer *C* has stated their expected resource requirement *x* according to a contract (the SLA), *e.g.*, in terms of bandwidth in Mbit/s, jitter in ms, or packet losses, whereupon ISP *I* has offered a flat rate tariff of *a* $/month for this service which customer *C* has accepted. In reality, the resources 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 their statement *x*, i.e.

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

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

Cumulus Points (CP) 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:

Cumulus Point 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 depending on* $\operatorname{sgn}\Delta_i$ *iff*

$$0 \le \theta_{c_i} \le \Delta_i < \theta_{c_i+1} \quad or \quad \theta_{c_i-1} < \Delta_i \le \theta_{c_i} \le 0\,. \quad (2)$$

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, then $|c|$ cumulus points are assigned, where c now is a negative number, hence the cumulus points are referred to as “green” ones, whereas for positive c the cumulus points are “red”.

Now the cumulus points c_i are accumulated over time according to

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

hence, Γ_n describes the total sum of cumulus points assigned since the start of the contract.

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

$$|\Gamma_n| \geq \Theta . \tag{4}$$

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

Note, the renegotiation of the contract may have different forms, as shown in Section 3.3 with respect to the example. Moreover, this approach can be extended to the case of any combination of resource requirements expressed by various QoS parameters $x_1, x_2, ..., x_q$ as mentioned above, such as delay or jitter, i.e. a QoS vector $x = (x_1, x_2, ..., x_q)$. To this end, let $\|.\|$ be a suitable q-dimensional metric, *e.g.*, based on maximum or euclidian metrics. Rewriting Equation (1) as

$$\Delta_i = \int_{t_{i-1}}^{t_i} \|(V(t) - x)\| dt \tag{5}$$

for q-dimensional *V* and *x* yields the desired generalization in a straightforward manner.

3.2 ISP Policies

The central question to be answered is how to deal with these CPs. The concept of Cumulus Points has a long tradition, especially in Switzerland and Germany. Usually, the green version of them is used to stimulate customers towards buying more often e.g. in supermarkets of one specific chain see, *e.g.*, the Swiss Migros "Cumulus Card". The red form of CPs may be found, *e.g.*, in the German system of dealing with illegal behavior in car traffic. Here, crossing red lights or driving cars in a not too sober state yields (if a policeman was around the corner) a certain small number of points to be noted in a central file located in Flensburg, a Northern city in Germany. As soon as in the course of two years 12 or more points have been accumulated there, you automatically lose your driving license. On the other hand, points may also be deleted, *e.g.*, in the case of good long-term behavior, of taking additional courses about correct traffic behavior.

These examples show that the rules mentioned in Section 3.1.4 of how to use the scheme with respect to communication services and bandwidth in fact is completely up to the ISP. Different ISP policies include especially the following ones:

- **Measurements:** It is almost impossible to find a standard way of network monitoring and accounting that is compulsory for all ISPs. With the CPS proposal it is up to the ISP, on which data measurements the distribution of CP is based. An interesting feature in this context in particular is the measurement policy enabling ISPs to provide incentives customers for shifting traffic from peak to off-peak times, simply by defining measurement points. *E.g.*, if it is agreed between ISP and customer that the distribution of measurement points is concentrated during peak times, the latter will probably receive the incentive to send traffic in times of low measurement frequency.
- **CP assignment:** Being assigned CPs depends on violating certain thresholds in terms of utilization or bandwidth. Fixing these thresholds is up to the ISP. As the CP assignment depends crucially on the accounting mechanism, thresholds should be set such that different measurement techniques applied to the same consumption pattern eventually do not differ by more than one CP. Moreover, setting $\theta_{\pm 1} \neq 0$ prevents smaller oscillations around *x* to result in superfluous CP assignments.

- **Accumulation:** Usually, CPs are supposed to be accumulated over subsequent billing periods. However, as the traffic sinner example shows, CPs may be allowed to expire, or red CPs may be charged up against green ones.
- **Contract renegotiation:** Another threshold to be set freely by the ISP concerns the point at which the contract with the customer is supposed to be renegotiated. The way of renegotiating is also open. Either the customer delivers a new statement about expected QoS requirements, and the provider offers a new flat rate, or the old contract remains valid, and the accumulated discrepancy is resolved, *e.g.,* by an extra payment. Note that offered flat rates will depend on the size of resource consumption, i.e. the higher the capacity requirement, the lower the price per MB due to usual discounting. Moreover, the size of the extra payment in principle corresponds to an extra contract "ex post" over the duration time of the original contract and the capacity discrepancy. Hence, the flat rate per MB of the additional contract is higher, and the customer's behavior is forced to be in accordance with the contract (*e.g.,* as shown in the example below, cf. Figure 5a during April).

3.3 Example

Figure 5a describes a typical example of how CPs are used. Customer *C* has stated their 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, their consumption falls below the expected value (one green CP in March), before it behaves exactly according to the contract in April (which is apparently the ideal situation). Later on, in May and June this value is exceeded again. The accumulation of the CPs as of end of June sums up to five red CPs and eventually requires a renegotiation of the original contract.

Continuing the example it can be safely assumed that it relies to the following ISP policy: *For each month the customer is awarded up to two CPs in each direction (one for slight, two for heavy discrepancy between expected and actually used bandwidth). Five red CPs form the threshold for renegotiation of the contract.*

Figure 5

Note that these five red CPs actually indicate that over a longer time period there has been an extensive mismatch between expected and consumed bandwidth. To correct this mismatch, one could think of several possibilities.

- **Case 1:** The customer believes that the mismatch has been only temporary and that their initial statement is still valid. Then the ISP may set an additional charge for each red CP, upon paying the charge the respective CP expires, and the entire system is basically reset to the initial condition (cf. Figure 5b).
- **Case 2:** Assume again that like in case 1 the customer does not want to change their commitment, but unlike in case 1 they do not want to make an extra payment in order to cancel the accumulated CPs. In this case, they could make a special additional contract (maybe also fixed in time, *e.g.,* over five months) over the average amount *x'* the bandwidth has been over-consumed over the last few months. If this additional contract is not used at all, it will solely produce green CPs month after month that may be charged up against the existing accumulation of red CPs until all of them have lost their validity. In this example, there are still three red CPs valid as of end of July.
- **Case 3:** Assume the customer now is convinced that the initial statement has not been correct. In this case a change of the contract would be necessary, now stating a higher expected bandwidth consumption *x"*. The accumulated sum of five red CPs could either be removed by an extra payment according to case 1, or by accidental under-utilization of the new statement *x"*, leading to a status of 2 red CPs altogether as of end of July.

4 Results and Evaluation

The integration of various pricing schemes into the developed ICCAS can be performed. This section discusses the integration of the pricing scheme CPS into the ICCAS based on a DiffServ networking environment, since both concepts are well suited to charge differentiated data transmissions in a DiffServ network. They are combined in a way which minimizes protocol overhead for collecting measurement information and which is feasible in small as well as in large networks.

4.1 CPS in the Context of DiffServ SLAs

CPS is tailored to complement DiffServ in the provision of QoS and in the appeasement of customers. According to the DiffServ philosophy, CPS is applied 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 SLAs 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 economic, since both apply on the same service, a similar granularity, and at identical locations in the network. Therefore, an SLA includes besides legal commitments financial commitments, *e.g.*, the sum of financial forfeits on contract violation, both of which are not concerned with the charging task directly.

The process to set up an SLA is sketched in a generic fashion. The customer and the ISP initiate a negotiation phase in which technical requirements and possibilities are assessed, warranties on QoS and performance are given, the processes to verify them are specified, and in which charges are set, i.e. for CPS the price is fixed on the foundation of the customer’s traffic volume estimation. The negotiation between the customer and the ISP is vulnerable to subtle inaccuracy, since it does not only handle technical topics, which are expressed in precise and reliable parameters or mathematical formulas. Thus it is indispensable for the customer as well as for the ISP to consult with a legal adviser. Concluding, it is up to the customer, the ISP, and a legal adviser to define the SLA.

Because SLAs do not already define technical parameters, ISPs have a lot of 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 the pricing, CPS, although they are specified in the same SLA. This is illustrated by a *specification layer*, which is separated into a *charging plane* and a *DiffServ plane.* Concerns of charging are located in the *charging plane*. Likewise, this generic representation allows one to locate and specify parameters of charging systems and pricing schemes, other than CPS. Parameters necessary to be specified within an SLS (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.

4.2 The Charging System in Support of CPS

Contracting between customers and a provider founds the basis for connectivity. In the case of a DiffServ networking environment, end-to-end services are supported by SLAs and their handling by the charging system. The DiffServ architecture concentrates network’s and system’s intelligence at network edges and supports an approach to meter those edges. 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 SLAs (cf. Section 1.3). 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. Note, this approach does not return to micro-flow orientation, but it supports reservations for BAs (aggregated super-flows containing many micro-flows) by bandwidth brokers of the DiffServ network. A settlement between different domains also does not take place on a micro-flow basis, but for BAs only. Between domains SLAs exist also, which are negotiated on the basis of estimations domain operators have of their traffic, which originally is caused by the traffic of end users. The specific settlement between two neighboring domains is also performed by applying the CPS scheme. Interfaces of the Accounting Information Path are considered at this stage to explain the use of CPS and DiffServ within the ICCAS.

- **Mediation-Accounting Interface I-MA**: This interface is small. It is used to collect metered usage data of each BA of each border router, where mediation components are located. This data can either be transmitted periodically or upon request from the ICCAS. To set parameters in mediation components, *e.g.*, the interval for automatic transmission of metering data, the ICCAS must also configure mediation components remotely.
- **Pricing Interface I-CP**: The price interface is not needed, since dynamic prices do not exist by applying the CPS scheme. Prices changes on a longer time-scale are reflected by appropriate changes of SLAs themselves. Therefore, prices for data transmissions can be queried easily from the customer support inside the ICCAS (cf. Section 2.3.3) where SLA information is stored in the contract control.
- **Feedback Interface I-FI**: This interface is also not required in case of DiffServ and CPS, since the only feedback a customer receives are cumulus points at the end of each billing period via a bill. Its possible to offer the user an option to query his current cumulus points' standing. Though, it appears to be more sensible for the user to measure his data traffic himself and, consequently, calculating his cumulus points using his SLA and comparing the result to the bill. This is one of the major advantages of the proposed integration of CPS into a DiffServ environment: The user can control his bills exactly and will immediately recognize differences with his provider. Measurement and calculation on the user's side will be performed automatically.
- **Billing interface**: The billing interface needs to receive charges for each customer to create a bill and all cumulus points for a current billing period. The numbers of cumulus points to be sent are calculated in the charge calculation component.

4.3 Exchanged Data of Interfaces and Protocols

Based on the description of interactions between all charging system components, overheads for charging information can be identified. Taking the Accounting Information Path (AIP) into closer consideration, interfaces are of local interest or data is exchanged over longer distances. The Mediation-Accounting interface (I-MA) serves as an example for the need of an exchange protocol. As depicted in Table 2, three families of data are distinguished. General data contains information such as identifications, originators, and date. Usage data includes duration and volume data. Finally, service-relevant data encompasses references on contracts and their handling.Configuration data to indicate the level of aggregation are transferred once to the mediation component. Data is requested periodically which results in the metering data being collected.

Table 2

Assuming a direct Internet Protocol (IP) encapsulation of I-MA interface interactions and protocol data units (similar to other signaling protocols such as RSVP), configuration messages contain, *e.g.*, per usage-based service to be monitored a 20 Byte IP header, its ID (4 Byte), the source and destination (2*32 Byte), and the date (4 Byte), in total 72 Byte. Periodic metering data include the IP header (20 Byte), the ID (4 Byte), the duration, if the flow has been finalized (4 Byte), and the volume (4 Byte), totalling to 32 Byte packets. Depending on the accounting period P_A and the number of services N_S to be mediated and accounted for, the total bandwidth $B_{I\text{-}MA}$ of I-MA for periodic data collected equals to:

$$B_{I\text{-}MA} \text{ [bit/s]} = 1/P_A * N_S * 32 * 8 \text{ [1/s * bit]}.$$

To continue with AIP interfaces, the Charge Calculation-Pricing interface (I-CP) contains a single price list only (cf. Table 3), which is distributed to the relevant ICCAS as often as the enterprise policy component decides on changes on prices per unit.

Table 3

For the CPS over DiffServ scenario, the Feedback interface (I-FI) is not required, however, it may contain a number of different families of data, such as accounting and charging templates or volumes for quantitative feedback information (cf. Table 4).

Table 4

The Billing interface (I-BI) transports data from the ICCAS to the billing system. The billing instantiation requires for the send operation service descriptions (4 Byte) and once its technical definition in terms of QoS parameters (sizeof(Q) Byte). Periodic billing records are transferred with the service ID (4 Byte), the charge (3 Byte), and the monetary unit of the charge (1 Byte), *e.g.*, CHF or US$. Worked on enhancements include the integration of authentication means. Depending on the billing period P_B, the total bandwidth $B_{I\text{-}BI}$ of I-BI for periodic billing record transmissions equals to:

$$B_{I\text{-}BI} \text{ [bit/s]} = 1/P_B * 8 * 8 \text{ [1/s * bit]}.$$

Table 5

4.4 Evaluation of CPS

Following the investigation of ICCAS interfaces and protocols, the most important properties of the Cumulus Pricing Scheme are summarized. They include the following properties by design:

- **Discrete:** Instead of reporting every small fluctuation of the user behavior, CPS allows the user behavior to fluctuate within certain boundaries. The feedback is given in a quantized form.
- **Cumulative:** Not single bursts, but only continuous change of promised behavior triggers reactions, *e.g.*, in terms of changing prices.
- **Early warning:** If some user starts to change their behavior, this will have no immediate consequences, but nevertheless they will know at an early stage that their pattern has changed, and that continuing the new usage habits will require sooner or later a renegotiation of the current contract.
- **Predictable and transparent:** As the entire scheme presents itself to the user as a variation of the fully accepted flat rate schemes, there will not be any trouble in terms of user acceptance. Charges for the user remain stable over long time-scales, and necessary changes are transparent to the user because of the early warning character of the feedback mechanism.
- **Market managed:** The contract between customer and ISP is based on information delivered by the customer about 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 their commitments, CPS introduces a slow penalizing mechanism in order to trigger well-behavior of the customer or change of the contract.

- **ISP policy dependent:** There is a strong influence of ISP policy on the realization of the scheme. The ISP may freely decide not only the measurement procedure, but also the threshold values for awarding Cumulus Points, the thresholds concerning their accumulation etc. On the other hand, the semantics of CP are sufficiently clear to allow the mapping of such thresholds and values between different ISPs.
- **Technically feasible:** Awarding Cumulus Points most probably should be justified by respective measurement results, but it is largely up to the ISP how to gain these data and how to interpret them. In the minimal case, a small number of very crude measurements or perhaps even estimations of user behavior may already be taken as sufficient by the ISP.

5 Summary and Conclusions

Existing network management functionalities have been extended within the work presented by a dedicated support of market-managed network management mechanisms, in particular the supply of charging and pricing functions for differentiated services. Their necessity is driven by a world of commercialization in networking and the demand for QoS-based services, but minimized technical effort for QoS provisioning and charging in networks. The development of a generic and modular Internet Charging System, in particular the Internet Charge Calculation and Accounting System (ICCAS) and a discrete, predictable, transparent, and technically feasible pricing scheme termed Cumulus Pricing Scheme (CPS) illustrates a feasible technical solution to the support of pricing differentiated services in the Internet.

The lessons learned from this work clearly indicate that the proposed pricing scheme would have avoided a recent failure of an all-you-can-eat offer as seen from the U.K.-based Internet Service Provider (ISP) Breathe [26]. Therefore, CPS as it stands today is claimed to offer a better scheme than traditional flat-rate schemes, as discussed in Section 4.4, however CPS currently is not proven yet to be economically optimal. But technical requirements of an ISP for performing traffic measurements are reduced, which is complemented by a simple set of measurement policies, negotiable between the customer and the ISP. They are sufficient for single user and enterprise contracts and can be supported effectively by the Internet charging system. Specifically, the ICCAS design is suited for handling an extensible variety of different services and pricing schemes, where particularly the CPS has been integrated without problems. In addition, all ICCAS interactions are functionally complete and sufficient for known pricing schemes today, ranging from usage-based to flat-fee approaches.

The work in progress includes the finalization of the implementation of the ICCAS. Future optimization of interfaces and protocols will follow based on iterative measurements taken, mainly according to the measured processing overhead for components. The simulation of the market-managed feature of CPS with multiple ISPs is under way and will be extended to inter-ISP scenarios. This includes the simulation of various ISP policies and their effects on prices and user behavior. In addition, the economic incentive compatibility of CPS will be studies in a broader range, including multi-provider markets. Since CPS is a first approach to combine consistently actual cost and charged prices to the customer, the different DiffServ aggregation levels are to be investigated. Finally, the optimization of target values as well as initializing CPS thresholds and prices are undertaken currently for the DiffServ scenario discussed above.

Acknowledgements

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 and partially in the ANAISOFT project funded by the Swiss

National Science Foundation, Bern. Special thanks go to M. Karsten, J. Schmitt, and O. Heckmann at the Technische Universität Darmstadt, Germany for the joint design of the networking architecture (Figure 3a) for M3I. Additionally, the authors like to acknowledge Hasan's contribution and many discussions with all of their project partners. The camera-ready version of this work has been supported partially by the Austrian Kplus program. Finally, many thanks to Lawrence and Lewis Tiffany for proofreading the final version.

References

[1] 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.

[2] R. Braden, D. Clark, S. Shenker: *Integrated Services in the Internet Architecture: On Overview;* Internet Engineering Task Force, RFC 1633, June 1994.

[3] 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.

[4] N. Brownlee: *Internet Pricing in Practice;* In: L. McKnight, J. Bailey (edts.) *Internet Economics;* MIT Press, Cambridge, Massachusetts, U.S.A, 1997, pp 77-90.

[5] D. Clark: *Combining Sender and Receiver Payments in the Internet*; Telecommunications Research Policy Conference, October 1996.

[6] 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.

[7] S.A. Cotton (edt.): *Network Data Management – Usage (NDM-U) for IP-Based Services;* IPDR Specification Version 1.1, June 2000.

[8] G. Dermler, M. Günter, T. Braun, B. Stiller: *Towards a Scalable System for Per-flow Charging in the Internet;* Applied Telecommunication Symposium, Washington D.C., U.S.A., April 17-19, 2000.

[9] R. Edell, N. McKeown, P. P. Varaiya: *Billing Users and Pricing for TCP;* IEEE Journal an 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, Bocca Raton, Florida, U.S.A., March 1998.

[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, 1999.

[13] P. Flury, P. Reichl, J. Gerke, B. Stiller: *Charging Considerations for Virtual Private DiffServ Networks;* TIK-Report No. 94, Computer Engineering and Networks Laboratory, ETH Zürich, Switzerland, August 2000.

[14] J. Gerke, P. Flury, B. Stiller: *The Design of a Charging and Accounting System for the Internet;* TIK-Report No. 91, Computer Engineering and Networks Laboratory, ETH Zürich, Switzerland, August 2000.

[15] R. Gibbens, F. Kelly: *Resource Pricing and the Evolution of Congestion Control;* Automatica, Vol. 35, 1999, pp 1969-1985.

[16] D. Grossman: *New Terminology for Diffserv*; Internet Draft, work in progress, August 2000.

[17] M. Günter, I. Khalil, T. Braun: *An Architecture for Managing QoS-enabled VPNs over the Internet;* 24 IEEE Annual Conference on Local Area Networks (LCN'99), Boston, Massachusetts, U.S.A., October 18-20, 1999, pp 122-131.

[18] 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.

[19] H.-G. Hegering, S. Abeck, B. Neumair: *Integrated Management of Networked Systems;* Morgan Kaufmann Publishers, San Francisco, California, U.S.A., 1999.

[20] M. Karsten: *Pricing Mechanisms Design;* Public M3I Project Deliverable 3, June 30, 2000. URL: http://www.m3i.org.

[21] M. Karsten, J. Schmitt, B. Stiller, L. Wolf: *Charging for Packet-switched Network Communications - Motivation and Overview;* Computer Communications, Vol. 23, No. 3, February 1, 2000, pp. 290-302.

[22] 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.

[23] S. Keshav: *An Engineering Approach to Computer Networking;* Addison-Wesley, Reading Massachusetts, U.S.A., 1997.

[24] C. Linnhoff-Popien, P. Reichl: *Netzmanagement;* Aachener Beiträge zur Informatik, Augustinus Verlag, Aachen, 1998.

[25] J. MacKie-Mason, H. Varian: *Pricing Congestible Network Resources;* IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, 1995, pp 1141 – 1149.

[26] N. McIntosh: *Heavy Surfers Pay the Price.* The Guardian, England, U.K., August 3, 2000.

[27] M3I: *Market Managed Multi-service Internet;* 5th Framework EU Project, IST Program, No. 11429, URL: http://www.m3i.org, 2000.

[28] K. Nichols, V. Jacobson, L. Zhang: A *Two-bit Differentiated Services Architecture for the Internet.* RFC 2638, July 1999.

[29] A. Odlyzko: *Paris Metro Pricing for the Internet.* AT&T Labs, Research, New Jersey, U.S.A., May 1997.

[30] P. Reichl: *Kelly's Bound, RUAA and the Pricing of Multiclass Traffic in Loss Networks;* U.K. Performance Engineering Workshop, Bristol, U.K., July 1999.

[31] 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.

[32] D. Songhurst (edt): *Charging Communication Networks - From Theory to Practice;* Elsevier, Amsterdam, The Netherlands, 1999.

[33] A. Sirbu: *Internet Billing Service Design and Prototype Implementation;* Carnegie Mellon University, Computer Science Department Technical Report, Pittsburgh, Pennsylvania, U.S.A., 1994.

[34] 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.

[35] 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, Computer Engineering and Networks Laboratory, ETH Zürich, Switzerland, September 2000.

[36] B. Stiller, P. Reichl, S. Leinen: *Pricing and Cost Recovery for Internet Services: Practical Review, Classification, and Application of Relevant Models;* to appear: NETNOMICS - Economic Research and Electronic Networking, Vol. 3, No. 2, 2001.

[37] 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, Track 3.

[38] B. Teitelbaum, P. Chimento: *QBone Bandwidth Broker Work Group.* URL: http://qbone.ctit.utwente.nl/BBroker/, August 2000.

6 Short Author Biographies

Prof. Dr. Burkhard Stiller received his diploma degree in computer science and doctoral degree from the University of Karlsruhe, Germany in 1990 and 1994, respectively. From 1991 until 1995 he has been a Research Assistant at the Institute of Telematics, University of Karlsruhe, being on leave in 1994/95 for an EC Research Fellowship at the University of Cambridge, England. From 1995 is with the Computer Engineering and Networks Laboratory, TIK at the Swiss Federal Institute of Technology, ETH Zürich, Switzerland as a Senior Researcher and Lecturer for multimedia communications and since 1999 as Assistant Professor for Communication Systems. His research interests include architectures for multimedia systems, charging and accounting, pricing schemes, Quality-of-Service models, and high-speed networking.

Dr. Peter Reichl has been studying Mathematics, Physics, and Philosophy in Munich, Germany and Cambridge, UK. He finished his Ph.D. degree with Aachen University of Technology and has been working at Bell Labs, Murray Hill, NJ, before he joined the Computer Engineering and Networks Laboratory (TIK) at ETH Zurich. Since January 2001, he is working as Senior Researcher at the Telecommunications Research Center Vienna (FTW). His current interests include pricing, QoS and traffic models in the future Internet.

Jan Gerke holds a diploma degree (master) in computer science from the University of Karlsruhe, Germany, where he did his thesis on pricing models for the Internet. Since April 2000 he is working towards his Ph.D. at the Computer Engineering and networks Laboratory TIK, ETH Zürich, Switzerland currently investigating economic aspects of QoS provisioning in the Internet.

Placi Flury received his master degree in electrical engineering at the ETH Zürich in March 2000, which was completed by his thesis work on remote administration of Linux hosts. He joined the TIK in May 2000 as a research assistant, where he works on network QoS, Internet economics, and Service Level Agreements.

7 Figures

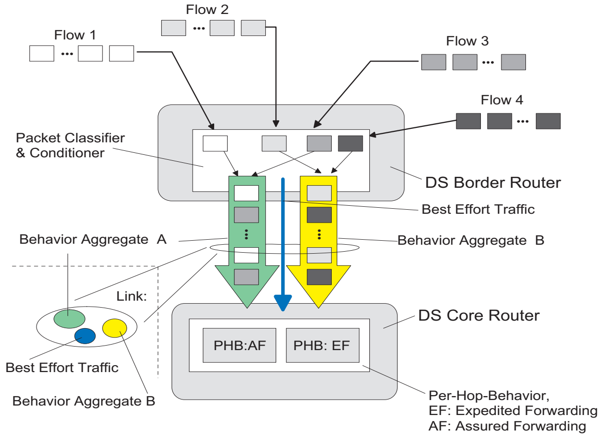

Figure 1: Flow Conditioning, Aggregation and Per-Hop Behavior

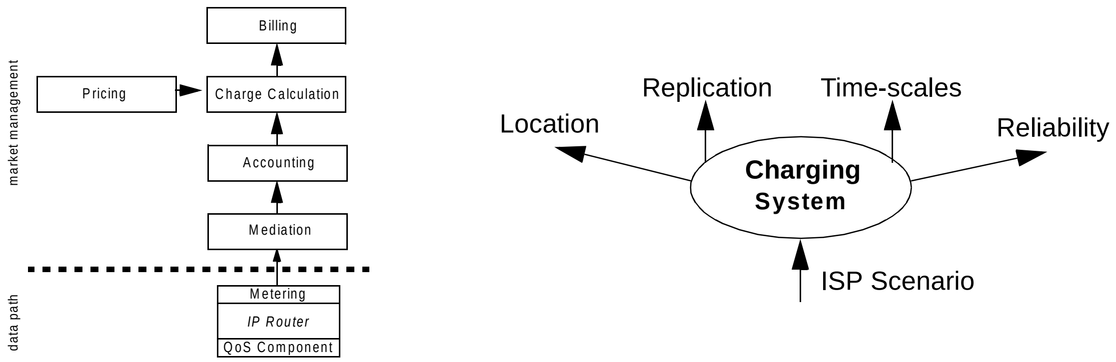

Figure 2: (a) Overall Concept of Charging and (b) Its Design Dimensions

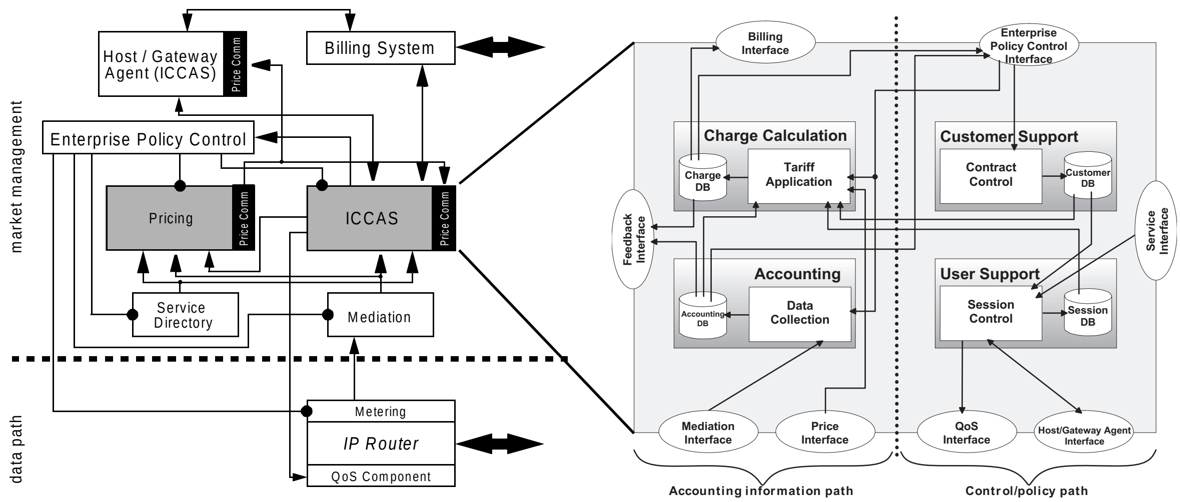

Figure 3: (a) Architecture of the Overall Charging System and (b) ICCAS Internal Architecture in Detail

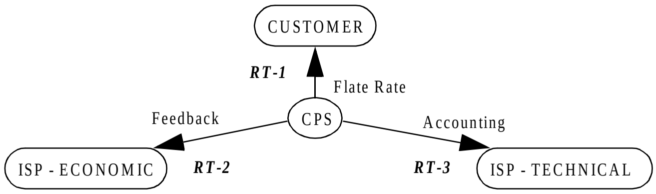

Figure 4: CPS Requirements

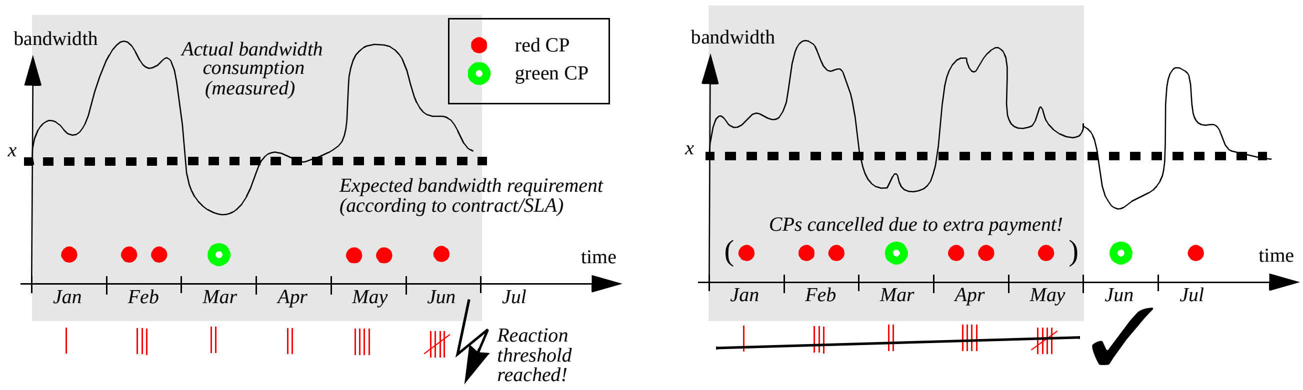

Figure 5: (a) Red and Green Cumulus Points and Their Accumulation over Time and (b) CPs Cancelled by an Extra Payment

8 Tables

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

Table 2: Message Content for I-MA

Family	Member	Size	Description
General	ID	4 Byte	Identification of customer and session (or service).
	Source/Destination	32 Byte	Originator and Consumer identification.
	Date	4 Byte	Date of service/session allocation.
Usage	Duration	4 Byte	Duration of service/session use.
	Volume	4 Byte	Effective bandwidth consumption.
Service	QoS Contract	sizeof(Q)	Agreements on jitter, loss rate, delay.
	Contract Change	1 Byte	Customer initiated renegotiation.
	Contract Break	1 Byte	Notification of network failures.

Table 3: Message Content for I-CP

Family	Member	Size	Description
Pricing	Price-list	sizeof(Pl)	List of price/unit, *e.g.*, per volume, per jitter, delay constraints. Prices may vary, *e.g.*, depending on time of day or congestion.

Table 4: Message Content for I-FI

Family	Description
Accounting templates	Information about the characteristic of the Accounting Records.
Charging templates	Information about the tariffs and the resulting Charging Records.
Charging and accounting volumes	Quantitative feedback about accounting and charging.

Table 5: Message Content for I-BI

Family	Member	Size	Description
Description	ID	4 Byte	General description of a service as seen by users.
Definition	QoS list	sizeof(Q)	The technical definition of a service in terms of QoS parameters.
Billing Data	Charge	3 Byte	Charging value of the service utilized.
	Monetary	1 Byte	Monetary charging unit.