DYNAMIC CHARGING FOR INFORMATION SERVICES

Cliff Redmond and Vincent Wade *

1. Introduction

The last five years has seen an incredibly rapid growth in the number of subscribers to the Internet and in the number of organisations which now consider a presence on the world wide web as vital to their commercial interests. With such growth arises the possibilities of not just advertising services on the WWW, but actually offering, delivering and gaining payment for information services over the Internet. Such information service providers may be the Internet provider themselves, or represent value added service providers making use of a third party Internet service provider. This paper looks at the issues which are relevant to devising charging schemes based on information content and usage.

The paper first outlines some general approaches for charging for bandwidth and information services and then examines the properties of digital information which can effect the metering and charging of such information services. The paper presents the early ideas of a charging policy which discourages local (user) copies of digital information resources, suggesting two indexes which can be used to indicate the behaviour (usage) of the information. The paper also details the design and implementation of a generic (charging policy neutral) accounting management service based on the TINA C service architecture. This management service is being trialed on WWW based tele-educational service operating over a pan-European ATM based network.

2. Charging for bandwidth and information

There are many possible forms of service charging that can be used by service providers. e.g. flat-fee (usage insensitive charging) usage based (usage sensitive charging), responsive charging, information sensitive (content-aware) charging and information insensitive (content-blind) charging.

Flat-fee charging is perhaps the most basic charging mechanism that individuals and organisations use when paying for Internet access. This charging method has some incentives for users and service providers. For users it yields a predictable charge with little difficulty in understanding the bill, whereas for service providers it provides a predictable income with little accounting overhead. It must also be noted that flat-fee charging doesn't penalise information sharing; this sharing referred to as "public-spirited behavior", in [RFC1272], provides the ethos for Internet usage and accounting. Unfortunately this scheme typically leads to heavy network usage because there is no penalty (other than the resultant network congestion) for users consuming high levels of bandwidth.

* Cliff Redmond and Vincent Wade, Department of Computer Science, Trinity College, Dublin. {Cliff.Redmond,Vincent.Wade}@cs.tcd.ie.

To deal with congestion, usage-sensitive pricing schemes have been suggested, with common examples being statistical measures of effective bandwidth and charges for volumes of packets or cells [CANCAN]. MacKie-Mason and Varian [RPI] suggested a responsive usage-sensitive pricing mechanism that could actively deter congestion on networks. They proposed a 'smart market' where users indicate a maximum charge they would be willing to pay for bandwidth in a tight loop feedback mechanism. However, this scheme has had its detractors. For example Shenker et al. in [PCN] argue that the costs of implementing and policing such a policy might not actually cover the costs recovered. It is also questionable whether such a mechanism is actually implementable because the costs of congestion are inherently inaccessible. Instead, Shenker et al. suggest that "It is important to allow prices to be *based on* some approximation of congestion costs, but it is important to not force them to be *equal* to those congestion costs". Notably, they state that while support for pricing might be included, no pricing policy should be embedded in the network architecture. This would allow different network providers a choice in how they price a service.

While Internet Protocol packets cannot yet be aware of the nature of the traffic they are carrying, there are discussions about providing this functionality in future versions [IPv6]. Currently, ATM-based integrated services networks can support application-aware transport services, with quality of service guarantees [OPISN]. However, care is necessary if a network based on application-aware transport services is to flourish; the role of application-blindness in innovation and in the provision of *new* services should be considered [SACP]. A scheme allowing the network to be content-aware, while being application-blind might offer a solution; this papers outlines such a scheme.

The scheme proposed in this paper allows higher-layer service providers to pay for bandwidth requirements in an otherwise content-blind network, while users pay the providers directly for the services delivered. Previously such charging could only occur within inherently content-aware networks, such as on-line service providers. However, these on-line service providers are increasingly joining, or providing gateways to, the Internet, which is content blind, e.g. Prodigy, AOL and Minitel. These service providers recognise that the value of joining a network increases with the number of users already connected (termed "positive network externality" in [SACP]), but are increasingly only differentiating themselves by editorial services; the scheme presented in this paper would allow them to differentiate themselves on price as well.

2.1 Usage based charging for information services

When attempting to build any usage based charging scheme, it is necessary to have a mechanism to meter the use of a service's resources, store any data generated by this metering activity, and generate the charges based on appropriate tariffs. Once generated, these charges can be aggregated and used as a basis of a bill for the service user.

The Telecommunications Information Networking Architecture Consortium, TINA-C, has proposed a software based service architecture to support the rapid development and provisioning of a wide range of telecommunications services. It attempted to do this by defining a common approach for the design and management of services which allowed the reuse of service components. It also defined a number of common components for service accounting, subscription and session control, and proposed the use of a distributed computing environment to implement these components.

The TINA consortium has worked closely with the Open Management Group, OMG [OMG], and has adopted its Common Object Request Broker Architecture, CORBA, as its distributed processing environment. CORBA supports an abstract interface definition language (IDL) which allows components written on different platforms, in different programming languages, to inter-operate. This facilitates service component, and therefore software, integration and reuse.

The Prospect project has developed subscription and accounting management services based on the TINA consortium specifications on top of a CORBA component platform. Central to the TINA-C Architecture is the concept of a session. TINA sessions allow identification and authorisation of users (access sessions), service use, control and management (service sessions) and communication resource management (communication sessions). The project has concentrated on access and service session components because the Internet Protocol (IP) based services used do not need active communication resource management.

The implementation of common accounting components is described in section 5.2. The use of these components necessitated the mapping of the service session concept onto the inherently stateless protocol interactions of IP services. The service implementation section outlines how this was achieved for a WWW based information service, and describes the interfaces to the common accounting components. The following sections outline special considerations for the content that information services support, and describe a charging mechanism and policy suitable for such services.

3. Properties of information relevant to charging

There are several important considerations to take into account when charging for information in digital form. Some of the more important concern the nature of the information in this form. For example this information is:

- **medium-free**: this means that information exists essentially independently of its medium; it exhibits plasticity meaning its expression to the ultimate user can be transformed. For example, email discussion lists are regularly archived and made into an information resource, with threads, or links between emails used to enable the user to follow the discussion; still images can be taken from movies, and altered; recorded sound can be edited to change the order of speakers, their voices or background noises. The multimedia publishing industry is most immediately affected by this problem, and is faced with trying to licence information for uses for which it was never intended, often finding great difficulty even tracing the original producer. This has even lead for calls to set up a single ‘clearing house’ for multimedia publishers [ACM196], [Bangemann].
- **easily copied**: digital information is “... something that can be stolen without depriving the owner of it ...” [ElecPir] because instant and perfect copies are possible. Existing economic theories, based on pay-per-copy, are being stretched by this property [Cox]. Copyright - the traditional link between ideas, their expression and exploitation - is also suffering, with many newly framed laws that cover copies of digital information (even copies transient in nature), being criticised for upsetting the traditional balance between the owner of copyright and the public good that such law was originally designed to protect [ACM1294].

A solution to these problems might be to discourage users from copying information, other than in a transitory way. Thus multimedia producers could refer to the original form when transforming information, and copies of documents could refer back to their source when being read. There have been schemes that attempted to enforce this, including electronic document management systems [DigiBox].

Other important considerations that are relevant to charging for any form of information include:

- the **cost** of producing the information, including rewards for the primary producers (e.g. authors, film production company) and the publishers.
- the **nature** of the information; whether it is relatively dynamic (perhaps news normally found on television or in newspapers, or stock market quotes), or relatively static (perhaps reference material, e.g. a family encyclopaedia, or novels).
- the **user** using the information; traditionally done on a mass marketing basis, an interactive information system offers information producers much more feedback from their customers, on a per person basis, and will possibly lead to personalised information sources [Persapress], [PinaWeb] and much more niche marketing and production of information: “Think niche. It’s the net’s greatest strength” [O’Reilly].

The publishing industry has a number of roles in the “information society” [Toffler84], [Bangemann], including those of editing and selecting information, which are highlighted by another two properties of information. These properties are linked:

- information is generative; “If you use a piece of information, I can use it too...if we both use it, the chances are improved that we will produce more information...it is generative” [Toffler84]. This property is evidenced by the explosive growth of sites on the World Wide Web in the past five years [Netcraft] and their inter-referencing.
- too much information is, or can be, bad; “Up to a certain point more choice is better. [...] The user gets overwhelmed, and less is more.” [O’Reilly].

Thus publishing could continue overseeing both content and context, where providing context (or meta-information) could be as simple as providing a selection of links to favourite information sources, or as complicated as a sophisticated catologue, e.g.[Yahoo]. Many on-line services already offer a ‘best of the web’ service to their subscribers, for example Prodigy, AmericaOnLine and CompuServe. Such meta-information can be more valuable because their selection of other information resources, while still referencing the original source of information to allow the user to explore in more detail. This gives the original information producers an opportunity to gain in an usage based scheme.

The most common Web based information services finance their information publication by one of several indirect means: by subsidy, in the case of universities and research institutions; by advertising in the case of search engines and other Meta-information points; and by subscription, or a combination of subscription and advertising, in the case of on-line editions of newspapers or magazines (The Economist, The New York Times). As yet, very few (if any), sites charge for information accessed or used.

Two usage policies could be supported by the existing Web infrastructure. These policies depend on information in the protocol headers detailing the age of a document (the ‘Last-modified’ date), they are:

- Paying per access: this is equivalent to paying for a new copy of a document. Even though the user might use a local cached copy instead of a copy from the original server, the original server is still informed of its use by a conditional request for the document (an “If-Not-Modified” header).
- Paying per copy. This ensures a new copy of the document is downloaded each time it’s requested, which is a potential waste of bandwidth. This can be enforced by giving a document a “pre-expired” last-modified date, or none at all.

However, both of these charging schemes would tend to encourage copying of information; the user is effectively penalised for using the same information more than once. A policy outlined in the next section might be properly called pay-once-per-use, and could act as an encouragement not to copy.

4. A proposed dynamic charging policy based on information usage

At the moment the majority of service charges are based on fixed tariffs which include costs per session and a flat per use fee for each piece of information used within that session. This scheme has some in-built policy decisions, and doesn't discourage the copying of information as it is still a basic pay-per-access, or pay-per-copy scheme.

Some research has been performed on comparing usage of the service in different sessions, where a session could be defined as the application layer context in which requests to the service were made. This approach tried to account for users requesting some information more than once and not penalising them for this, since this could show they are still using the original source of the information, not a local copy. A single cumulative charge for the use of a piece of information over all sessions could be used, so that a user might actually be charged less for a piece of information if their usage profile fell within defined parameters. In order to compare a user's session, two basic indexes were devised, an inter-session repeat index, or ISRI, which gives the number of repeat requests for information in two consecutive sessions (where consecutive sessions were thought adequate on a brief examination of data) and an intra-session repeat index, or IaSRI, which monitors the number of repeat requests for a piece of information within the same session.

In order to have an adaptive pricing mechanism (where users might only pay once for a piece of information, and are rewarded by a decrease in the amount they pay if their profiles show they haven't copied information), it is necessary to have measures to estimate a typical user's repeat-request average (ISRI) and have a policing mechanism to stop users trying to take advantage of the system (IaSRI).

A policy based scheme which tries to discourage copying by referring to the original source of the information can only be effective if users are educated as to the adaptive nature of the pricing. Unfortunately there will always be the user that will try to 'beat the system' by requesting a document more than once within the same service session in the mistaken belief that the price of the piece of information will fall to its lowest level. The IaSRI is in place to attempt to police this needless waste of bandwidth.

4.1 Applying charging schemes in tele-education

The ISRI is a general index, and will probably vary widely depending on the nature of the information being served, i.e. how relatively dynamic it is. Figure 1. shows a graph of session overlap (ISRI) Vs. inter-session times for educational resources used in a web based educational course that was used by 1996/1997 final year database students. This graph shows a marked peak in the number of sessions with two to three repeat requests for a period of 24-48 hours after the previous session. One might expect that in an educational course the number of repeat requests for particular pages (information resources) would be small in consecutive sessions, but would be larger when separated by a period in the region of one week. That is to say students wouldn't need to refresh their memory on a daily basis, but would rather require reuse of information over a weekly period. As shown, the number of repeat requests peaks within a period of 48 hours after the previous access of the information. There also appears to be quite a low number of repeat requests in subsequent sessions (this would show as a pronounced diagonal ridge on the graph). This illustrates that the user usage pattern for information resources may not be intuitively predictable. Thus in order to define reasonable usage profiles/patterns, non-commercial trials must be used to generate acceptable patterns. A trial period is already common practice with many Internet based subscription services, to allow the subscribers' judge the product; by recording the usage patterns in such test periods the service provider could gain useful information too.

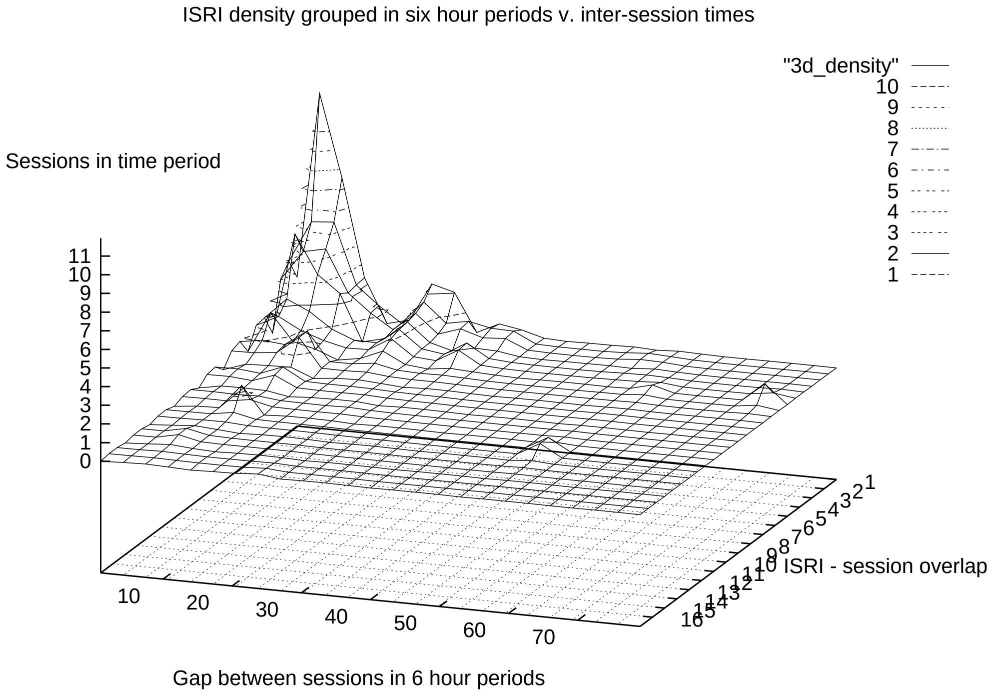

Figure 1.

Were the information available anything other than course materials, for example general reference material, quite a different ISRI profile might be expected. It must also be noted that most users used the course quite quickly, which could also explain the higher number of sessions with repeat requests within the first two days of use.

Given the ISRI for a pool of information, and any user's ISRI, a policy could be implemented that decreased the cost of information to that user if their ISRI fell within the expected range for the inter-session time. The presence of too high an IaSRI could be used to penalise a user for needlessly using a provider's bandwidth, it could be used as a basis for a charge for that user. Too high an ISRI might also be penalised, because this may be due to the usage of more than one user. The research on these indexes is ongoing, and further trials and experiments are planned.

5. Implementation of usage metering

The usage metering mechanism works on a per session basis, where a session is based on protocol state information in HTTP requests to a modified web server [Apache]. As mentioned, the accounting system is based on the TINA (Telecommunications Information Networking Architecture) Service Architecture, developed on top of a CORBA distributed processing environment.

In the TINA architecture a user is required to explicitly start a service session; this is achieved using Java/CORBA desktop components in the Prospect project, where a URL with a reference to the service specific MIME-type is returned as the service-specific reference (the MIME-type describes the content of the file requested). The service can be used independently of these desktop components once the user has a record of their service reference; in this case scripts can be used to start the accounting. Whatever mechanism is used, the service user can have fixed session starting point built into the first URL that they request from the server, which can be used to identify the user to the service. In this way the state information in the requests acts as a context in which requests are made, and therefore as an application layer session identifier.

5.1 Service implementation

Sessions are modelled using HTTP protocol state information, commonly called 'cookies'. This state information was originally introduced by one browser developer, but was implemented as a *de facto* standard by many web server developers as it offered web-sites a useful means of user tracking. It has subsequently come onto the Internet standards track [RFC 2109], and is now supported by most browser implementations.

The web server, or daemon, was modified to allow it to be configured to refuse to serve all but one service-specific MIME-type without the presence of the correct state information in the client's HTTP request. Without this state information, normally set by the server in the headers of the first response to the client, a user using a browser can only successfully request the service-specific MIME-type from the server. This modification included the addition of new configuration directives, acting on a per virtual server and per directory basis, that influence how the server handles requests

Given a request for the service specific MIME-type, the handler for this mime-type decodes the URL that was requested and uses the information contained in it to form and initiate the cookie using a 'Set-Cookie' HTTP response header.

If the user subsequently requests a file of any MIME-type from the server, the request will always include the state information associated with the first interaction. When a request arrives at the server, the requested file and user information, along with a code signifying the success or otherwise of the request, are forwarded to a standardised accounting subsystem via a CORBA one-way (non-blocking) call to a multithreaded user session manager (USM), if that part of the server's directory tree is so configured. (There is a description of the TINA based USM component in the next section). The addition of these CORBA calls does cause some overhead, with a small, but constant, delay of approximately 28 milli-seconds per request. This resulted in a reduction of server through-put from 1.57 MB/sec to 1.49 MB/sec for a 10k file.

The usage metering data is forwarded from the USM to an Usage Metering Data component (UMDataCO), which collects usage metering data for each session. When a session ends, the data for the user's session is stored in another accounting component, the Usage Metering Log component (UMLogCO), ready to be used as a basis for charging.

5.2 Component implementation

The components used to monitor and log usage, control tariffs and generate charges and bills are based on Computational Object specifications published by the Telecommunications Information Networking Architecture (TINA) Consortium in their 1995 Service Architecture.

The UMDataCO, or usage metering data computational object, is based on the X.742 Recommendation 'Usage Metering Function for Accounting Purposes'. A subset of the metering data information package outlined in the X.742 Recommendation is used. This consisted of usage data blocks to report session registration, requests and session completion (compete block). The implementation has concentrated on single user sessions, and the request block has come to signify a request made inside a session, rather than a request to join a session.

The UMDataCO accepts session registration and completion blocks from service independent parts of a service specific USM. As mentioned above, the USM receives usage information from the web server, via CORBA one-way calls to the Service dependent interface (Sdi). It can associate the information with a TINA service session and forward it to a UMDataCO for that user's session.

Figure 2. outlines the service dependent and independent parts of the accounting subsystem for the web based service. The notation used follows that used by Jacobson in [OOSE]. A detailed description of the interactions follows.

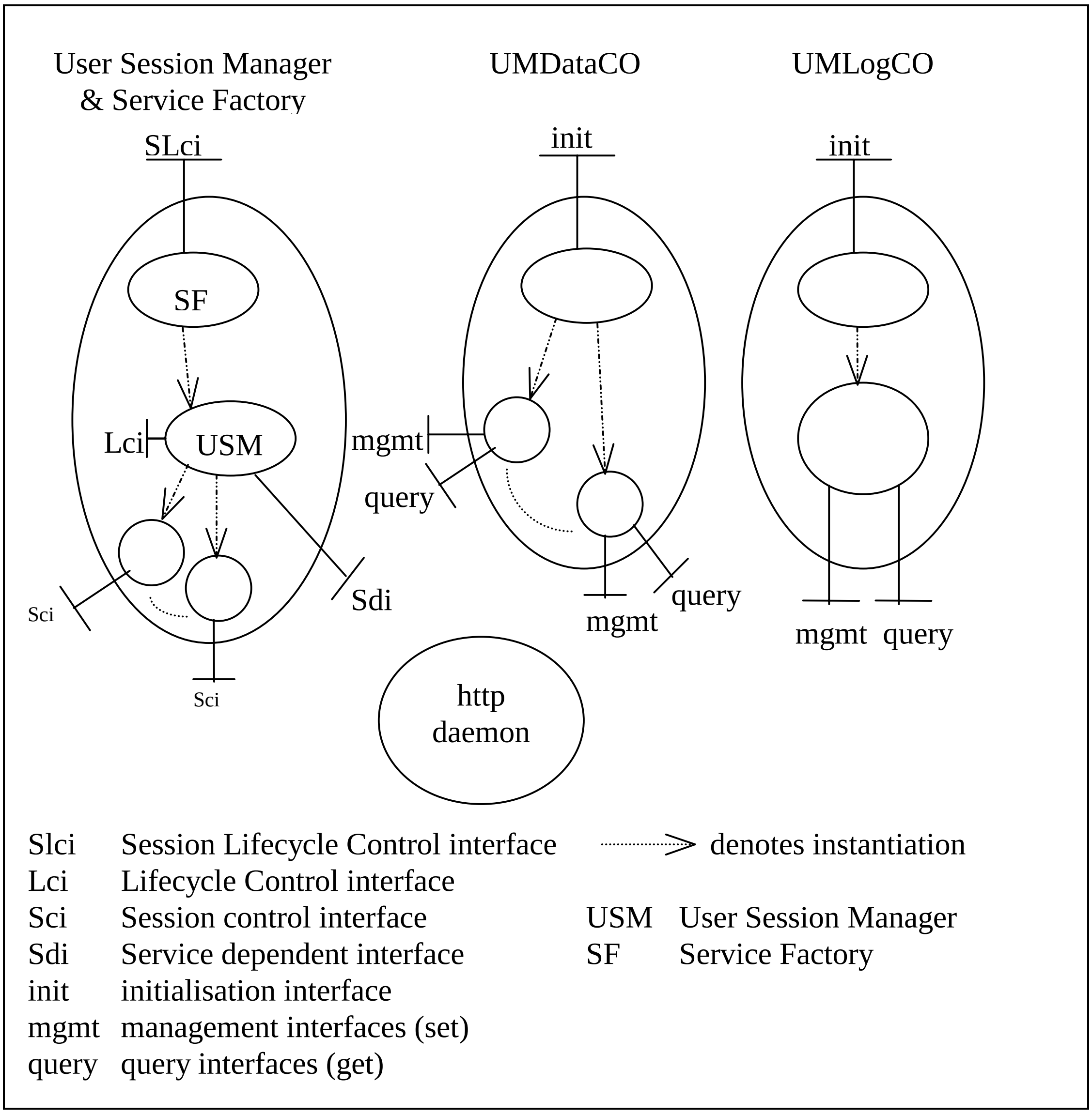

Figure 2.

There are two factory interfaces shown. The object that implements the init interface of the UMDataCO produces objects that meter a session's usage and the object that implements the USM's Lifecycle Control interface (Lci) produces Session control interface (Sci) objects. Each session control interface object has a corresponding object within the UMDataCO that meters that session's usage. When a session is started the Sci object forwards a registration block to the appropriate UMDataCO management interface (i.e. the management interface for that user), and when the session ends a complete block is forwarded.

Requests for resources are forwarded through client code embedded in the HTTP daemon and arrive at the Service dependent interface of the user session manager. The information in these calls signifies the session identifier of the requester and the information requested; the session identifier is used to find the corresponding management interface of the UMDataCO, and the request information is forwarded there. The implementation of the Sdi object takes advantage of multi-threading to ensure collection of all usage information; a session dependent interface object per user wasn't viable because of a need to keep the client code in the HTTP daemon simple and fast.

When a session ends, a cascade of interactions occurs. The object representing a session, it's Session control interface (Sci) object, is deleted by the Session Lifecycle control interface (SLci) object of the service factory (SF) via the USM's Lifecycle control interface (Lci). These interactions are internal to the SF/USM component. When the Sci object is deleted, this in turn deletes the corresponding object within the UMDataCO via the UMDataCO's init interface. When this happens the information associated with that session is stored using a call to the management interface of the UMLogCO component (the store operation). This information is essentially independent of any particular charging policy and can used by many different charging schemes, some of which have been outlined earlier.

6. Future work & conclusions

Several charging paradigms were outlined, including content-blind usage-sensitive charging and content-sensitive charging for information. Also outlined were the difficulties associated with information charging and an analysis of user usage patters for a web based course.

Current work is concentrating on incorporating some of the charging approaches discussed into the TINA accounting architecture, in which information services are not very well supported.

Future research will concentrate on re-requests of information from the point of view of the information being used, rather than the user using it; this would allow group usage schemes, which would be difficult to account for using the single user session focus of the analysis discussed above.

Future work will also accommodate HTTP caches into the charging scheme; users ought to be rewarded if they don't needlessly use a provider's bandwidth, so some discount might accrue when a browser makes a conditional (on the expirary date of the information) request. It might also be useful to account for badly administered or implemented caches that never verify information, by charging them a fixed price for information for all users in a group (or at a site) once one of them requests the information; here charging from the information's point of view, rather than the user's, shows its utility.

Acknowledgements

Thanks to David Lewis of University College London, and Alexander Richter of GMD-Fokus, Berlin for their comments and suggestions on an initial draft of this paper and the EU Commission for the partial funding of this research.

References

[ACM1294] Pamela Samuelson, *The NII Intellectual Property Report*, Communications of the ACM, Vol. 37, No. 12. (December 1994).

[ACM196] Pamela Samuelson, *Intellectual Property Rights and the Global Information Economy*, Communications of the ACM, Vol. 39, No. 1. (January 1996).

[CANCAN] CANCAN Project (AC014), *ATM Charging Schemes*, CANCAN Deliverable 5.

[COX] B. Cox, *Superdistribution*, Addison-Wesley, Reading MA, 1996.

[PCN] Shenker et al., *Pricing in computer networks: reshaping the research agenda*, Telecommunications Policy, Vol. 20, No. 3, Pergamon, 1996.

[SACP] J. MacKie-Mason et al., *Service Architecture and content provision*, Telecommunications Policy, Vol. 20, No. 3, Pergamon, 1996.

[DigiBox] Olin Subert et al., *The DigiBox: A Self Protecting Container for Information Commerce*, Proceedings of the First Usenix Workshop of Electronic Commerce, Usenix Association, Berkeley, CA, USA, 1995.

[ElecPir] John Chesterman and Andrew Lipman, *The electronic pirates : DIY crime of the century*, Routledge, London, 1988.

[OPISN] Q. Wang et al., *Optimal Pricing for Integrated Services Networks*, Internet Economics, MIT Press, Cambridge MA & London, 1997.

[O'Reilly] Tim O'Reilly, *Publishing Models for Internet Commerce*, Communications of the ACM, Vol. 39, No. 6. (June 1996).

[RPI] J. MacKie-Mason et al., *Responsive Pricing in the Internet*, Internet Economics, MIT Press, Cambridge MA & London, 1997.

[Toffler84] Alvin Toffler, *Previews & premises : an interview*, Pan Books, London, 1984.

WWW References

Apache http://www.apache.org/docs/misc/API.html

Bangemann http://www2.echo.lu/eudocs/en/bangemann.html

DigiBox http://www.intertrust.com/architecture/overview.html

IPv6 http://ds1.internic.net/rfc/rfc1883.txt

NetCraft http://www.netcraft.com/survey/index.html

OMG http://www.omg.org/

Persapress http://ksgwww.harvard.edu/~itbspp/proj16/home.htm

PinaWeb http://www.cern.ch/WebOffice/Projects/Newspaper/index.html

RFC1272 http://ds1.internic.net/rfc/rfc1272.txt

TINAC http://www.tinac.com/

YAHOO http://www.yahoo.com/