A Comparative Study of Usage-Based Charging Schemes

Fotios C. Charmantzis\(^2\), Costas Courcoubetis\(^2\), Vasilios Siris\(^2\), George D. Stamoulis\(^2\)

1. Introduction and Objectives

Charging and accounting in modern high-speed networks are extremely vital for their successful operation and growth. Tariffs and pricing schemes are needed for the network to recover its costs in a fair way from the diverse population of users, and to effectively allocate network resources. Economic theory suggests that usage-based charging (or some scheme closely related to usage) will be employed by service providers in the case of perfect competition. This is the type of charging for network services we consider in this work.

Closely related to this approach is the notion of an effective bandwidth of a traffic stream; this summarizes the use of network resources by the stream, and might depend on a large number of parameters such as traffic statistics, QoS parameters (e.g. cell-loss probability, maximum delay), the multiplexing capability of the network, etc. When an effective bandwidth can be appropriately defined and measured, then the problem of usage-based charging greatly simplifies, and may reduce to standard optimization problems. However, analytical methods are usually hard to apply (if at all possible), particularly when real traffic sources are involved. Thus, simulation and on-line methods are invaluable tools for related investigations.

Ideally, accurate usage-based charging can be done on the basis of the effective bandwidth of each separate source, which can be estimated either analytically or through simulation and on-line estimation. Since such an estimation can be rather complicated, in this work, we investigate two simpler charging formulae which are based on bounds for the effective bandwidth. One of the formulae is based on a simple bound, while the other is based on a tighter one. We compare the charges derived through these formulae, as well as these charges to that based on the accurate calculation of the effective bandwidth of the source during the call. Three types of traffic sources are considered:

1. On/Off Markov Modulated Fluids,
2. 3-state Markov Modulated Fluids, and
3. MPEG traffic obtained through traces and models

We thus, assess appropriateness of the various formulae for charging each of the considered traffic sources. Our analysis is based on simulations carried out in a special purpose tool for Performance and Charging Evaluation developed at ICS-FORTH. Calculations of Cell Loss Probability, Effective Bandwidth and Charge, according to several methods can be carried out in

\(^{1}\) This work was partly supported by the ACTS (AC039) project CA$hMAN$ of the E.C. (see http://www.isoft.intranet.gr/cashman/)

\(^{2}\) Institute of Computer Science, Foundation for Research and Technology Hellas (ICS-FORTH) and Dept. of Computer Science, Univ. of Crete, Heraklion, Greece 711 10
the Tool's user-friendly and efficient environment. Graphical User Interfaces are supported by the TCL/TK (Tool Control Language / Tool Kit) programming system.

The remainder of this paper is organized as follows. In section 2, we briefly review the concept of effective bandwidth, we describe the traffic sources employed in our experiments, and we describe the charging formulae. In section 3, we present selected results from our experiments, together with related discussion. Finally in section 4, we provide some concluding remarks and we discuss some directions for further research.

2. Background Material

2.1 The notion of effective bandwidth

Suppose that \( J \) sources share a single unbuffered resource of capacity \( C \), and let \( X_j \) be the load produced by source \( j \). Assume that \( X_j, j = 1, \ldots, J \), are independent random variables, with possibly different distributions. Can the resource cope with the superposition of the \( J \) sources?

More precisely, can we impose a condition on the distributions of \( X_1, \ldots, X_J \) which ensures that

\[
P\left\{ \sum_{j=1}^{J} X_j > C \right\} \leq e^{-g}
\]

for a given value of \( g \)? The answer to this question is, by now, fairly well understood. There are constants \( s, C' \) (depending on \( g \) and \( C \)) such that if

\[
\sum_{j=1}^{J} \alpha(X_j) \leq C', \quad \text{where} \quad \alpha(X_j) = s^{-1} \log E[e^{sX_j}]
\]

(2)

The quantity \( \alpha(X_j) \) is referred to as the effective bandwidth of source \( j \), and summarizes the effect of each particular source to the rest of the sources multiplexed; this result is originally due to Hui [Hui88]. This approach is applicable rather broadly. In particular, it is by now known that for quite general models of sources and resources it is possible to associate an effective bandwidth with each source such that, provided the sum of the effective bandwidths of the sources using a resource is less than a certain level, then the resource can deliver a performance guarantee (for an overview and an extensive list of references see [CASHD96], chapter 3). In most of the cases, the relevant definition is of the form

\[
\alpha(X_j) = (st)^{-1} \log E[e^{sX_j[0,t]}]
\]

(3)

for particular choices of \( s \) and \( t \), where \( X_j[0,t] \) is the load produced by source \( j \) over an interval of length \( t \). In fact, there may be several constraints of the form in the leftmost part of (2), corresponding to different physical or logical resources within a network.

2.2 The Traffic Source Models

2.2.1 MMF on-off sources

According to this model, the source either generates traffic (in continuous time) with the peak rate \( h \) (while in the on state), or it is idle (while in the off state). The sojourn time in each state is exponentially distributed. The mean rate of the source is denoted by \( m \). The parameter \( t \) in (3) should be chosen small compared to the length of times for which the source is typically on and off. The exact choice of \( t \) is determined by the amount of buffer at the network ([CW95]), and a
reasonable choice could be the time it takes for a burst of the source to fill the buffer. For example, for a video source at a peak rate of 1.5Mps and a buffer of 200 cells, \( t \) is about 50ms.

An approximation of the effective bandwidth (3) of such a source is then

\[
\alpha(h,m) = \frac{1}{st} \log \left( 1 - \frac{m}{h} + \frac{m}{h} e^{\frac{m}{h}} \right)
\]  

(4)

Figure 2.1 depicts a particular sample of an discrete-time on/off fluid source.

Although the instantaneous bit-rate of real traffic sources may take one of multiple permissible values (rather than two), the on/off fluid is a popular model of traffic sources, since it captures the feature of activity-level fluctuations. Moreover, for given mean bit-rate and QoS requirements, on/off sources often constitute the worst case in terms of necessary resources.

### 2.2.2 MMF tri-state sources

This model generates traffic at three different rates, according to the source’s state: in the HIGH state the rate equals the peak rate \( h \), in the MED state the rate equals \( f_{M} \times h \) (where \( 0 < f_{M} < 1 \)) while in the source is idle in the OFF state. The sojourn time per state is again exponentially distributed. Figure 2.2 depicts a particular sample of a discrete-time tri-state fluid source.

### 2.2.3 MPEG Traffic

We investigated MPEG traces and developed models that follow the characteristics of a number of real MPEG video streams. Several different MPEG -streams (namely, cartoons, sports, computer animations, fractals, movies etc.) have been examined and analyzed using an appropriate MPEG video statistics gatherer; these movies were based on an MPEG-1 coding scheme with cyclic video frame pattern. The result of the analysis have been used to develop simple histogram based models for all the MPEG-streams. These models have been used to generate multiple independent traces, suitable for queueing and charging studies.
2.3 Charging Formulae

Ideally, under usage-based charging, the charge per call should be proportional to the source’s effective bandwidth and to the duration of the call. (The constant factor yielding the exact charge is to be determined by the pricing strategy of each provider.) Since estimation of the effective bandwidth may be inefficient, we resort to charging formulae based on bounds for the effective bandwidth.

2.3.1 The Simple Formula

This formula [Kel94b] is based on (4); the resulting tariff takes a strikingly simple form: a charge per unit time and a charge per unit volume of traffic carried. Before a connection’s admission, the network requires the user to announce a volume \( m \), and then charges for the call an amount \( f(m; M) \) per unit time, where \( M \) is the measured mean rate for the call. The charging mechanism encourage the user to access and to declare the expected mean rate \( m \), so the user has no incentive to “cheat” the network. So, under this mechanism, the best declaration for the user is the expected mean rate, because the expected cost per unit time under this declaration is equal to the effective bandwidth. In [Kel94b] it is shown that these requirements essentially characterize the tariff \( f(m; M) \) (i.e., the charge per unit time of the call) as

\[
f(m; M) = a(m) + b(m) M
\]

defined as the tangent to the effective bandwidth curve \( \alpha(M) \) at the point \( m = M \). The coefficients in expression (6) are given suitable expressions. That is, the total charge equals

\[
a(m)T + b(m)V
\]

2.3.2 The Tax-Band Formula

This charging formula depend on more complex traffic measurement and not just on total time and volume, as in the Simple Formula. This allows a tighter bound for the effective bandwidth, and is attained through the “tax-band” charging approach introduced in [CASHD96], chapter 5. This was called “tax-band” because of its similarity to the system of graduated (or banded) income tax adopted in most countries.

The main idea is to divide the duration of the connection, \( T \), into \( T/t \) intervals of length \( t \), where \( t \) is the parameter appearing in equation (3). Each such interval will be the charging interval for the traffic stream. For simplicity we only consider the case of two bands. Note that, since \( h \) is the maximum input rate, at any given interval the maximum number of cells is \( ht \). We denote by \( k \) the width of the first tax band. Charging intervals are classified as being of type I or II as the total volume of cells the source produces during the interval is either \( \leq kt \) or \( > kt \). Recall that \( T_1 \) and \( T_2 \) are the total duration of intervals of types I and II respectively; so \( T_1 + T_2 = T \). Similarly, are the total volumes of cells generated by the source during intervals of types I and II respectively; so \( V_1 + V_2 = V \). Assuming that \( \phi_i, m_i \) are the user’s estimates for the parameters \( \Phi_i = T_i / T, M_i = V_i / T, i = 1,2 \), the total charge for the call becomes

\[
a_i[...]T_i + a_2[...]T_2 + b_1[...]V_1 + b_2[...]V_2, \quad (7)
\]

where \([...]\) denotes a function of \( h, k \) and of \( \phi_1, \phi_2, m_1 \) and \( m_2 \) and the constants are given by the suitable formulae. If the user declares accurately the values of its traffic parameters \( \Phi_i, M_i, i = 1,2 \), then his charge will be the minimum possible (the charge will be
proportional to the effective bandwidth of the “worst possible” process with the same parameters). Such a charging scheme (even with two bands) may be more effective in accounting for burstiness than the simple scheme, which gives the same charge for two connections if they have the same values of $T$ and $V$. Thus, the tariff function underlying (7) is constructed as a tangent hyperplane to a bound for the effective bandwidth that is tighter compared to that leading to the simple formula. Therefore, the resulting charge (7) lies between the accurate usage charge (derived through estimation of the effective bandwidth) and that of the simple formula (6).

3. Results

In this section we investigate the charging formulae through numerical results. We compare the charging formulae under different traffic scenarios.

**Experiment A: Charge the MPEG-stream extracted from the considered models**

Suppose that the traffic offered to the link is extracted from three different MPEG video sources, and the duration of the calls is 5750 msec. The charging interval $t$ is 50 msec, and the choice $s = 0.00001 b^{-1}$ in expressions for effective bandwidths and charges is reasonable (see [CASHD96], chapter 3). The comparisons are based on optimal users’ declarations of parameters in both formulae. The band $k$ in the tax-band formula mechanism equals the actual mean rate of the source; this selection is justified in Experiment B. Note that charges are expressed in kbits, since conversion in absolute monetary units requires determination of the actual charge per kbit.

<table>
<thead>
<tr>
<th>Video</th>
<th>Peak Rate</th>
<th>Mean Rate</th>
<th>Effect. Bdw.</th>
<th>Tax-Band</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3148</td>
<td>1022</td>
<td>7085</td>
<td>7985 (+12.6%)</td>
<td>9951 (+40.4%)</td>
</tr>
<tr>
<td>2</td>
<td>1123</td>
<td>806</td>
<td>4750</td>
<td>4822 (+1.5%)</td>
<td>4975 (+4.7%)</td>
</tr>
<tr>
<td>3</td>
<td>1701</td>
<td>772</td>
<td>4661</td>
<td>4877 (+2.6%)</td>
<td>5468 (+15.1%)</td>
</tr>
</tbody>
</table>

Table 1: Comparison of charges for typical MPEG traffic

This table shows that the tax-band formula can be considerably more accurate that the simple formula. The maximum deviation of the tax-band charge from that derived through effective bandwidth is 12.6%, while the corresponding deviation for the simple formula is 40.4%.

**Experiment B: Optimal Selection of the Tax-Band (for MPEG)**

In order to select the band threshold $k$ optimally, we experiment with different percentages of the mean rate for the video 2. Again we assume that user’s parameter declarations are optimal. Our measurements (partly depicted in Figure 3.1) reveal that there is a rather flat optimal, with the minimum value appearing in the interval $(k \in [0.5*h, 0.6*h])$. Notice that the corresponding mean rate specified by the model equals $0.56*h$. Thus, selecting as $k$ the model’s mean rate is a reasonable choice; this could have been conjectured by looking more carefully at the how the tax-band approach approximates the effective bandwidth. Notice also that as $k$ approaches 0 or $h$, the charge increases rapidly to that derived through the simple formula; this is intuitively clear, since in both extreme cases, one of the tax-bands vanishes.
Experiment C: Charging the 3-State MMF Source

We charge streams generated by the 3-state MMF model with both considered formulae, assuming optimal parameter declarations, for a 1500 msec call, and a charging interval of 50 msec. (Note that a call duration of 10 times as much yielded essentially 10 times as much charge). We vary the rate at the MED state (which in Table 2 is expressed as a percentage of the source Peak Rate $h=64$ kbps), and we assume that the source can be at any state with equal probability. The charges derived are presented in Table 2. In each case, the band threshold $k$ in the tax-band formula is taken as the model’s mean rate (according to Experiment B).

<table>
<thead>
<tr>
<th>$f_{MED}$ * $h$</th>
<th>Effect. Bdw</th>
<th>Tax-Band</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0* $h$</td>
<td>31.7</td>
<td>31.7 (+0.06%)</td>
<td>32.1 (+1.2%)</td>
</tr>
<tr>
<td>0.1* $h$</td>
<td>35.0</td>
<td>35.0 (+0.05%)</td>
<td>35.4 (+1.1%)</td>
</tr>
<tr>
<td>0.2* $h$</td>
<td>38.2</td>
<td>38.3 (+0.05%)</td>
<td>38.6 (+1.0%)</td>
</tr>
<tr>
<td>0.3* $h$</td>
<td>41.5</td>
<td>41.5 (+0.04%)</td>
<td>41.9 (+0.9%)</td>
</tr>
<tr>
<td>0.4* $h$</td>
<td>44.8</td>
<td>44.8 (+0.04%)</td>
<td>45.1 (+0.8%)</td>
</tr>
<tr>
<td>0.5* $h$</td>
<td>48.0</td>
<td>48.0 (+0.03%)</td>
<td>48.4 (+0.7%)</td>
</tr>
<tr>
<td>0.6* $h$</td>
<td>51.3</td>
<td>51.3 (+0.03%)</td>
<td>51.6 (+0.6%)</td>
</tr>
<tr>
<td>0.7* $h$</td>
<td>54.5</td>
<td>54.6 (+0.03%)</td>
<td>54.9 (+0.5%)</td>
</tr>
<tr>
<td>0.8* $h$</td>
<td>57.8</td>
<td>57.8 (+0.02%)</td>
<td>58.1 (+0.4%)</td>
</tr>
<tr>
<td>0.9* $h$</td>
<td>61.1</td>
<td>61.1 (+0.02%)</td>
<td>61.3 (+0.3%)</td>
</tr>
<tr>
<td>1.0* $h$</td>
<td>64.3</td>
<td>64.3 (+0.02%)</td>
<td>64.5 (+0.2%)</td>
</tr>
</tbody>
</table>

Table 2: Optimal charges for 3-state MMF Sources with varying rate at the MED state

Clearly, both formulae are very accurate, with the tax-band charge being extremely close to that derived by direct estimation of the effective bandwidth.
Experiment D: Inaccurate in the Mean Rate Declaration for on-off Sources Charged with the Simple Formula

We charge on-off sources, each lasting for 15000 msec. The mean sojourn time for the on state is 200 msec, and 1000 msec for the off state. In the on phase the source produce cells with rate 64 kbps (peak). The model’s mean rate is 21.33 kbps, while the actual mean rate $M$ in the particular run was almost equal to this (namely, 21.38 kbps). When negotiating/selecting the tariff, the user declares a value of $m$ [employed in (6)] which may be different from both that specified by the model, and from the actual mean rate $M$. We have observed that when the declared value of $m$ is within ±15% of $M$, the overcharge (due to inaccurate parameter declaration) does not exceed 0.4% of the optimal charge under the simple formula [namely, that given by (6) with $M = m$].

4. Concluding Remarks and Open Issues

In this paper we have compared two usage-based charging formulae, for different types of traffic. In particular, a simple charging formula proved to very accurate (compared to the charge derived through direct estimation of the effective bandwidth), except for some cases of MPEG traffic, where the tax-band formula is considerably more accurate. We also investigated the selection of the band threshold $k$ for the two-band case of tax-band. Our measurements verify the reasonable guess that the mean rate of the source is a good selection for $k$. Both previous conclusions were derived under the assumption that the user declares the applicable parameters optimally (so as to incur the least charge). We also investigated the effect of inaccurate mean rate declaration for charging on-off sources under the simple formula. The next issues to be investigated in our work concern the effect of such inaccuracies for the tax-band formula as well as for charging tri-state MMF and MPEG sources, and the effect of the parameters $s$ and $t$ in the accuracy of the formulae. Selection of appropriate values of $s$ and $t$ for MPEG traffic is an issue deserving further analysis.

REFERENCES