SAND97-1927C
CONF-980212--

The Effects of User Mobility on Usage Parameter Control (UPC) in Wireless ATM Systems

Peter Sholander, Luis Martinez and Lawrence Tolendino
Sandia National Lab
PO Box 5800, Mail Stop 0449
Albuquerque, NM 87185-0449
USA
{peshola, lgmarti, lftolen}@sandia.gov

Bruce A. Mah
Sandia National Lab
PO Box 969, Mail Stop 9011
Livermore, CA 94551-0969
USA
bmah@ca.sandia.gov

ABSTRACT

Wireless Asynchronous Transfer Mode (WATM) networks pose new traffic management problems. One example is the effect of user mobility on Usage Parameter Control (UPC). If the UPC algorithm resets after each handoff between wireless-cells, then users can cheat on their traffic contract. This paper derives explicit relationships between a user's traffic parameters (Peak Cell Rate, Sustained Cell Rate and Maximum Burst Size), their transit time per wireless-cell, their maximum sustained cheating-rate and the Generic Cell Rate Algorithm's (GCRA's) Limit (L) parameter. It also shows that the GCRA can still effectively police Constant Bit Rate (CBR) traffic, but not some types of realistic Variable Bit Rate (VBR) traffic.

1.0 INTRODUCTION

This section briefly reviews traffic management in Asynchronous Transfer Mode (ATM) networks. Since ATM networks are connection-oriented [1], a connection-setup phase occurs before the flow of user-data begins. During connection-setup, the user may signal various Quality of Service (QoS) parameters and traffic characteristics to the network via the User-Network Interface (UNI) protocol. For end-to-end transmission, the sender segments the transmitted user-data into ATM cells. Each of those 53 byte ATM cells has a five byte cell-header, and can carry up to 48 bytes of user-data. Hence, the QoS

This work was supported by the U.S. Department of Energy under Contract DE-AC04-94AL85000.

DISCLAIMER

parameters are cell-based ones such as Cell Transfer Delay (CTD), Cell Delay Variation (CDV) and Cell Loss Ratio (CLR).

Different applications may need different QoS levels from an ATM network. For example, voice applications need bounded end-to-end delay since conversation becomes difficult once the connection's end-to-end CTD exceeds a few hundred milliseconds. Other non-interactive applications, such as broadcast video distribution, can tolerate much larger end-to-end CTD. However, they still require bounded delay-variation (or CDV) if the receiving terminal must provide a constant bit-rate input to the user's display device. Finally, voice is relatively tolerant of cell-loss, since the receiving terminal can use error-masking techniques. Indeed, an ATM CLR of 1e-3 may be acceptable for many voice applications. However, TCP/IP-based file transfers may require much lower cell-loss rates. While TCP does guarantee error-free end-to-end transmission, it also interprets cell-loss (and hence TCP segment loss) as network congestion. This causes the TCP protocol to temporarily reduce its transmission rate. As such, a large CLR within an ATM network can greatly reduce the end-to-end TCP performance, or "goodput".

If the user requests a given QoS, or traffic contract, from an ATM network then the user must also supply the traffic characteristics for that connection to the network. The network then does Call Admission Control (CAC) based on the network's CAC algorithm, the requested QoS, those traffic characteristics and the contracted QoS for other existing connections. If the network can provide the requested QoS, without violating the contracted QoS for the existing connections, then it usually accepts the new connection. Otherwise, it typically rejects that connection. While the ATM Forum's (ATMF's) Private Network-Network Interface (PNNI) [2] does specify a Generic Call Admission Control (GCAC) algorithm, each network's CAC algorithm is typically network-specific. In any event, specific CAC algorithms are outside this paper's scope.

If an ATM network does accept a new connection then it may need to "police" that connection's traffic contract. Policing involves monitoring the connection to determine if it abides by its traffic contract and also possibly taking action if it does not. Public networks often do policing for two reasons.

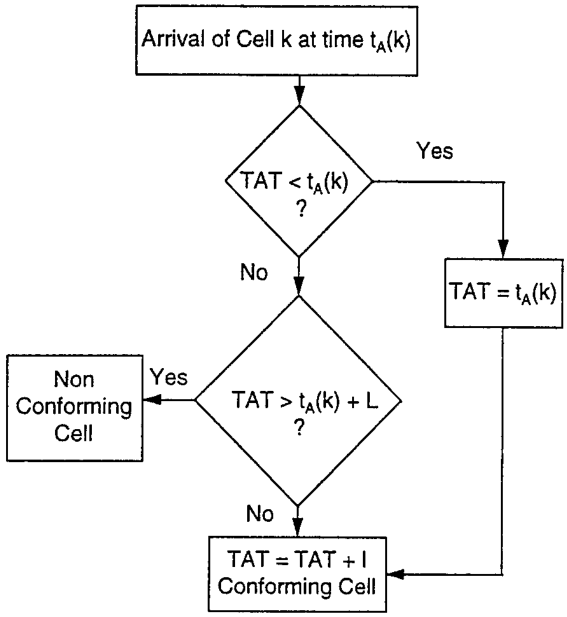
The first is billing. Users may only receive the services that they paid for. The second reason is network protection or "fairness". A malicious user should not be able to impact other users' contracted QoS by flooding the public network with excess traffic. (In contrast, private networks may not need to police traffic contracts, since administrative procedures usually limit network abuse.) Policing, or Usage Parameter Control (UPC), can take two forms. If the user exceeds the traffic characteristics provided in the connection setup request (e.g., sends data too fast) then the network may drop the excess (or "non-conforming") traffic at the UNI interface. Alternatively, each ATM cell's five-byte header contains a Cell Loss Priority (CLP) bit [1]. The UPC function may just set the CLP bit in the headers of the excess cells. Those "marked" cells will then be preferentially dropped during network congestion. This allows users to exceed their traffic contract, so long as it does not inconvenience conforming users (whose cell's CLP were not set by their UNI's UPC function). Finally, the UPC algorithm and cell-dropping policies are network specific. The only requirement is that the UPC must not mark conforming cells as non-conforming ones.

The remainder of this paper is organized as follows. Section 1.1 describes the Usage Parameter Control (UPC), or policing, function in terms of the Generic Cell Rate Algorithm (GCRA). Section 1.2 then relates the traffic contracts for Constant Bit Rate (CBR) and Variable Bit Rate (VBR) services to that GCRA. After these introductory discussions, Section 2 discusses the effects of mobility on UPC convergence for both CBR and VBR service. The interesting result is that UPC convergence is probably not a problem for typical CBR service in proposed Wireless ATM (WATM) systems. However, it may be a problem for some VBR traffic contracts. Section 3 then concludes this paper with some recommendations.

1.1 Generic Cell Rate Algorithm (GCRA)

The exact implementation for the Usage Parameter Control (UPC) function is network-specific. Any UPC algorithm may be used, so long as it does not mark conforming cells as non-conforming.

However, the ATM Forum's Traffic Management specification [3] does specify a generic cell-based UPC algorithm -- namely the Generic Cell Rate Algorithm (GCRA). That GCRA can be expressed as either a virtual scheduling algorithm or a continuous-state leaky bucket algorithm. The virtual scheduling form [3], shown in Figure 1, simplifies the analysis given in Section 2.



TAT = Theoretical Arrival Time I = Increment

$t_A(k)$ = Actual arrival time for cell k L = Limit

(Note: At the time of the first cell's arrival, TAT = $T_A(1)$)

Figure 1: Virtual Scheduling Algorithm for the Generic Cell Rate Algorithm [3]

The GCRA has two parameters: the Increment (I) and the Limit (L). Both parameters have units of seconds. One simple explanation for the GCRA(I,L) involves a CBR traffic stream with an ideal cell-interarrival time of I. If the previous cells have all been conformant (i.e., their interarrival times were all $\geq$ I) then the next cell will be conformant if its interarrival time is at least (I-L). The GCRA also bounds the time interval for which the CBR connection's average rate can exceed 1/I before its cells start being marked as non-conformant. The next section relates the ATMF's traffic parameters for CBR and VBR services to the GCRA's I and L variables.

1.2 Traffic Contracts for CBR and VBR Traffic

The ATMF's Traffic Management specification [3] specifies four cell-based traffic parameters -- namely the Peak Cell Rate (PCR), Sustainable Cell Rate (SCR), Maximum Burst Size (MBS) and Minimum Cell Rate (MCR). Only the first three are relevant to this paper. The ATMF's Available Bit Rate (ABR) service uses the MCR parameter [3].

The PCR is maximum rate at which the user will emit cells. Its inverse, the minimum cell-interarrival time (1/PCR), may be easier to measure in practice. The Sustained Cell Rate (SCR) is an upper bound on the possible conforming "average rate" for an ATM connection [3], where the average rate is simply the number of cells transmitted divided by the connection's "duration". (To be precise, the connection's "duration" is the time from the first cell's emission until the time when the state of the GCRA, for that SCR, returns to zero after the emission of the connection's last cell [3].) For ideal Constant Bit Rate (CBR) traffic, the PCR equals the SCR. For Variable Bite Rate (VBR) traffic, the SCR is typically less than the PCR.

The Maximum Burst Size (MBS) is the maximum number of back-to-back cells that the connection will send at its PCR. However, the connection's contracted SCR also dictates a minimum inter-burst spacing. So, the GCRA for VBR service actually uses the Burst Tolerance (BT) parameter described below for its Limit parameter. Equation (1) will give a relationship between the MBS and that BT parameter.

A CBR traffic contract includes the user's PCR and also a Cell Delay Variation Tolerance (CDVT). Various ATM layer functions, such as multiplexing between different user connections, can introduce CDV, into a connection's cell-stream, between the source and its UNI interface [3]. The CDVT accounts for those effects by allowing the minimum cell-interarrival time to be 1/PCR - CDVT, as long as the average cell-rate is still less than the PCR. Hence, a CBR connection could use a GCRA(I,L) with I = 1/PCR and L = CDVT as its UPC conformance test.

A traffic contract for VBR service includes three parameters -- namely PCR, SCR and MBS. In this case, the UPC must check both the PCR and SCR for conformance. The PCR conformance can still use a GCRA(1/PCR, CDVT). However, the SCR conformance is more complex. The SCR measurement still suffers from the CDV induced by the ATM layer. However, the user can also burst traffic at their PCR. This causes a short-term increase in the measured SCR. So, let τ_s denote the "Burst Tolerance (BT)". Also, define T and T_s to be 1/PCR and 1/SCR, respectively. Then, the SCR conformance test [3] can use a GCRA(T_s, τ_s + CDVT), where the Maximum Burst Size and the Burst Tolerance are related by Equation (1).

$$MBS = \left\lfloor 1 + \frac{\tau_s}{T_s - T} \right\rfloor \text{ cells} \qquad (1)$$

However, equation (1) only specifies the BT value to within the half-closed interval [(MBS-1)(T_s - T), MBS (T_s - T)). So, the ATMF convention [3] is to use the minimum value of the BT. An example of a traffic pattern that conforms to GCRA(T_s, τ_s) may help. Consider an "on-off" VBR source that transmits B cells at its PCR with intervening inter-burst spacings of T_1 = (B*(T_s - T) + T). That VBR source has a PCR of 1/T, an SCR of 1/ T_s and an MBS of B [3]. Hence, it can use the GCRA(T,0) and the GCRA(T_s, τ_s) for PCR and SCR conformance, respectively.

2.0 ANALYSIS

This section discusses the effects of user mobility on UPC conformance testing in proposed WATM systems. If the GCRA's limit parameter L is non-zero then mobile users can "cheat" on their traffic contract. As such, this section's main results are simple relationships between the GCRA's limit parameter, L, the user's transit time per wireless-cell, t_w, and the user's maximum cheating factor, Δ_m.

(The cheating factor, Δ, is defined in the next subsection.) An interesting secondary result is that this effect is probably not significant for CBR services. However it may be problematic for VBR services. Section 3 will propose solutions for this new network impairment.

2.1 GCRA Convergence Time for Mobile CBR Users

Assume CBR service with PCR = I and CDVT = L. However, also assume that the user is "cheating" on their traffic contract by a constant cheating-factor, Δ, over their contracted rate, 1/I. Thus they are actually sending their ATM cells at a rate of $(1+\Delta)/I$ cells/sec instead of sending at their contracted PCR of 1/I cells/sec. (Hence, Δ is defined as a dimensionless quantity. However $100*\Delta$ does equal a percentage of the contracted PCR.) In that case:

$$TAT_k = kI = k^{th} \text{ Theoretical Arrival Time (in seconds)} \quad (2)$$

$$t_A(k) = kI/(1+\Delta) = k^{th} \text{ Actual Arrival Time (in seconds)} \quad (3)$$

Based on the GCRA(I,L) algorithm given in Section 1.1 and Figure 1, the first non-conforming cell occurs when the j^{th} actual arrival time is less than the j^{th} Theoretical Arrival Time minus the Limit parameter L (or, in other words, when $TAT_j > t_A(j) + L$). Hence the maximum number of conformant cells, Nc, that can be sent at the non-conformant rate $(1+\Delta)/I$ is the greatest integer less than j. Or:

$$N_c = \left\lfloor \left(\frac{1+\Delta}{\Delta}\right)\left(\frac{L}{I}\right) \right\rfloor \text{ cells} \quad (4)$$

Hence, in a normal "fixed" ATM network, stationary users can not cheat indefinitely. Eventually, the UPC function will mark their cells as non-conformant. (As previously stated, the treatment of non-conformant cells is network-specific. Those cells might be dropped immediately at the UNI interface. Alternatively, those cells might be dropped only during network congestion -- so as to preserve the contracted QoS for conformant connections.)

Mobile users modify this picture however, since their network point-of-attachment may change. If the GCRA algorithm resets after each such change (or handoff), then there may be some combinations of wireless-cell sizes, user mobility rates and CBR traffic contracts that can not be policed by a GCRA(I,L) algorithm. In particular, let a mobile user have a CBR traffic contract with (I,L) = (PCR, CDVT) and a transit time for each wireless cell of t_w seconds. (Note: some WATM system proposals terminate the UNI at each basestation. Other proposals terminate the UNI at a mobile-enhanced ATM switch, where that switch then controls several wireless basestations. In second case, the basestations are cheaper but an additional signaling protocol is required between the mobile-enhanced ATM switch and its basestations. This paper's examples assume the first case. Hence, t_w is the transit time per wireless cell. However, this paper's equations still apply to the second case if t_w denotes the transit time across the cluster of basestations associated with each mobile-enhanced ATM switch.) In that case, the number N_A of ATM-cells transmitted by that user in each wireless-cell is approximately t_w/I. (The number of ATM-cells transmitted in each wireless-cell must, of course, be an integer.) For simplicity, assume that N_A is indeed an integer and that equality holds in Equation (4) without taking the integer-part of the right-hand side. These assumptions then provide a simple relationship between the GCRA's Limit parameter (L), the user's transit time, t_w, and the user's maximum cheating factor, Δ_m.

$$N_c = \left(\frac{1+\Delta_m}{\Delta_m}\right)\left(\frac{L}{I}\right) = \frac{t_w}{I} = N_A \quad \text{cells} \tag{5}$$

or:

$$\Delta_m = \frac{L}{t_w - L} \quad , \text{ for } t_w > L \tag{6}$$

So, mobile users can indeed cheat on their CBR traffic contracts by transmitting at the non-conformant rate $(1 + \Delta_m)$*PCR cells/sec, or less. (Interestingly, the maximum cheating factor is independent of the CBR connection's PCR. It just depends on the GCRA's Limit parameter L.) However, reasonable values for WATM wireless-cell sizes, user mobility rates and CDVT produce very small values for Δ_m. For example, consider a micro-cellular or LAN system with 30m wireless cells [4], a CDVT of 3 ms [4] and a maximum user speed of 3 m/s [5]. Then, for a reasonably worst-case mobility pattern (e.g., constant-velocity, linear-motion), t_w is 10 seconds and Δ_m is equal to only 3e-4. Hence, it should take a large number of malicious (or, more likely, oblivious) users to have a noticeable impact (say even 1%) on the overall system capacity. The situation for VBR traffic contracts is less clear. There may be useful VBR traffic contracts, such as VBR MPEG-2 video, that have much larger values for Δ_m.

2.2 GCRA Convergence Time for Mobile VBR Users

The UPC function for VBR traffic-contracts must police both the PCR and SCR. The results given by Equation (6) apply to PCR cheating for VBR service also. Hence, PCR cheating should not be a problem for either VBR or CBR services.

SCR conformance testing is more complex, though. Equation (6) still applies, with (τ_s + CDVT) substituted for L. However, there is an auxiliary relationship between the GCRA's Burst Tolerance parameter (τ_s) and the VBR traffic parameters. Section 1.2 gave the following relationship between τ_s, the PCR (which is equal to 1/T), the SCR (which is equal to $1/T_s$) and the Maximum Burst Size (MBS) [3].

$$\tau_s = (\text{MBS} - 1)(T_s - T) \quad \text{seconds} \qquad (7)$$

Assume a best-case scenario of CDVT = 0. Then substituting Equation (7) into Equation (6) and defining a peaking factor (P = PCR/SCR), yields:

$$\Delta_m = \frac{(\text{MBS} - 1)(\text{P} - 1)}{t_w(\text{PCR}) - (\text{MBS} - 1)(\text{P} - 1)} \qquad (8)$$

or:

$$\text{MBS} = 1 + \frac{\Delta_m \, t_w(\text{PCR})}{(\text{P} - 1)(1 + \Delta_m)} \quad \text{cells} \qquad (9)$$

So, the relationship between the monitored parameters (PCR, SCR, MBS) and the maximum cheating factor, Δ_m, is no longer rate-independent! In addition, one user can now achieve a non-trivial cheating factor. For example, let the PCR be 2500 cells/second (which is 960 Kb/s of user data), the SCR be 500 cells/second and the MBS be 1250 cells, or about 500,000 bits of user data.. Let t_w be equal to 10 seconds, as in Section 2.1. In that case, by Equation (8), Δ_m is 0.25. (Hence, the mobile user can substainably exceed their SCR contract by 25 %.) Furthermore, some recent VBR MPEG-2 traffic models have had similar (PCR,SCR) combinations and MBS's of several hundred thousand cells [6]. In that case, UPC may be really problematic as Δ_m is greater than 25. The next section discusses some solutions for this new network impairment.

3.0 CONCLUSIONS and SOLUTIONS

This paper has identified a new network impairment for proposed Wireless ATM systems -- namely the effects of user mobility on Usage Parameter Control (UPC) -- that is not present in existing ATM networks and cellular networks. Section 2 derived approximations for the relationship between the user's traffic parameters (Peak Cell Rate, Sustained Cell Rate and Maximum Burst Size), the user's transit time per wireless cell and the user's maximum possible "cheating factor", Δ_m. Those approximations show that GCRA-based UPC functions are still adequate for mobile CBR users. However, there are realistic VBR traffic contracts that can not be adequately policed by a GCRA-based UPC function if the GCRA algorithm resets after each handoff between wireless-cells. This is problematic since a non-conforming VBR user could then send excess traffic that degrades the contracted QoS for conformant users. This violates one of the core principles of ATM traffic management. Since wireless spectrum is a scarce resource, VBR video may become more popular than CBR video in WATM systems. Hence, this network impairment deserves further study.

There are at least three possible solutions for this new network impairment. The naive answer is to *not* reset the GCRA algorithm after each handoff. However, this requires the handoff process to signal the previous wireless-cell's GCRA state (current TAT, or Last Compliance Time for the GCRA's Leaky Bucket formulation [3]) to the new wireless-cell's UNI. Hence, it probably requires new standards-based signaling messages, or at least new Information Elements (IEs) within existing ATMF, or ITU-T, signaling messages. Obtaining these new signaling messages and/or IEs can be a time-consuming, multi-year process. A more fundamental technical problem with this approach is that it requires a global time standard in the network. Networks with a SONET/SDH physical layer can indeed provide synchronized clocks at each ATM switch in the network. However, not all networks provide a global-time service. The mobile-enhanced ATM network might use global-time distribution protocols, such as the Internet's Network Time Protocol (NTP). However, those software-based protocols probably have both resolution and accuracy issues.

The other two practical solutions are not standards-based. First, the network could disallow, or re-negotiate downwards, any traffic contracts, for mobile users, that cannot be policed to within some fraction, Δ_{max}, of the contracted PCR or SCR. This paper gave approximations for Δ_{max} for both CBR and VBR traffic contracts, based on the GCRA algorithm and a linear user-mobility pattern. If network designers choose this solution then this paper's analysis should, of course, be re-done for their network's UPC function and user-mobility patterns. One widely used user-mobility pattern is the constant-velocity, random-direction model given in [7]. The other non-standards-based solution is a variant on this. The network acknowledges that some mobile users can indeed cheat by a factor, Δ_m. The network's Call Admission Control (CAC) algorithm could then derate a mobile VBR user's requested traffic contract by their factor Δ_m. Hence, the network might do CAC for VBR connections based on SCR*(1+ Δ_m), rather than the actual requested SCR. The choice between solutions two and three is network-dependent. Solution two may increase the connection setup time, while solution three probably wastes more wireless bandwidth.

REFERENCES:

[1] C. Partridge, *Gigabit Networking*, Addison Wesley, New York, 1994.

[2] *Private Network-Network Interface Specification, Version 1.0, PNNI 1.0*, ATM Forum af-pnni-0055.00, March 1996.

[3] *Traffic Management Specification Version 4.0*, ATM Forum af-tm-0056.000, April 1996.

[4] T. Phipps, et al, "Requirements for Wireless ATM at 5 GHz", ATM Forum Contribution 97-0006, San Diego, February 10-14, 1997.

[5] C. Lind, "PHY/MAC Requirements for WATM", ATM Forum Contribution 96-1777, Vancouver, December 2-6, 1996.

[6] P. Key, R. Macfadyen and M. Nilsson, "VBR Video, Policing and Dimensioning" ATM Forum Contribution 97-0308, Chicago, April 28 - May 2, 1997.

[7] R. Thomas, H. Gilbert, G. Mazziotto, " Influence of the Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network", Proc. 3rd Nordic Seminar, Paper 9.4, Copenhagen, Denmark, September 1988.

M97007974

Report Number (14) SAND--97-1927C
CONF-980212--

Publ. Date (11) 199707
Sponsor Code (18) DOE/MA, XF
UC Category (19) UC-900, DOE/ER

DOE