The Estimation Error of Adaptive Deterministic Packet Marking

Lachlan Andrew*
Deptartment of Computer Science
Caltech
Pasadena CA 91125-8000 USA
l.andrew@ieee.org

Stephen Hanly**
Department of Electrical Engineering
University of Melbourne
Parkville VIC. 3010 Australia
s.hanly@ee.unimelb.edu.au

Abstract— **This paper is concerned with problem of signalling congestion link price information to a receiver using single bit marks. An efficient method was presented in [1] which exploits side information in the `IPid` field of the IP header to allow the maximum price on a flow's path to be estimated. In this paper we provide analysis to support the claim that the scheme can track a changing price. We consider a random walk model for the price, and provide a weak convergence result showing that the squared error (normalized by the drift) is asymptotically exponentially distributed, as the drift tends to zero.**

Index Terms— **congestion price, ECN, Explicit Congestion Notification, TCP, Transmission Control Protocol, flow control.**

I. Introduction

Many congestion control algorithms have been proposed for the Internet which require explicit feedback of congestion ("price") information from routers [2],[3] amongst many others. RFC 3168 [4] provides two "ECN" bits in the IP header for this purpose. Pricing information can be transmitted by randomly marking packets with these bits [2],[3],[5].

It has recently been proposed [6] that the process of setting these bits take into account "side information" contained in the IP header. This idea has been applied in a number of recent works, notably by Thommes and Coates [7], who provided an efficient, deterministic marking algorithm, using the 16-bit IP packet identifier ("IPid") to assist in conveying the base-two representation of the price. Based on that work, [8] and [9] proposed a similar scheme for estimating the maximum price, appropriate for max-min flow control.

As shown in [7]–[9], the deterministic marking schemes provide estimators that potentially have a much lower mean squared error (MSE) than the random marking schemes. However, these schemes suffer from the disadvantage that they must specify *a priori* how to trade resolution for agility. The fixed quantization of the range of possible prices means that the MSE is poor until a sufficient number of packets have been processed [1]. Further, the fixed quantization implies a square error floor for these schemes [1]. In contrast, the random marking schemes are adaptive, in that for a fixed price, the MSE consistently improves with the number of packets processed, allowing the estimator to track a changing price.

In recent work, [1] we provided an "adaptive" version of deterministic packet marking, ADPM, for estimating the maximum price seen by a flow along its path. This implicitly adapts its effective quantization resolution depending on the dynamics of the price. As for random marking of packets, numerical results in [1] show that the MSE consistently improves with the number of packets processed, for a fixed price. Numerical results show that static values can be estimated precisely, whilst rapidly changing values can be tracked quickly. These results also show that ADPM provides a MSE that is several orders of magnitude smaller than the estimators based on random marking of packets [2], [3], [5], or deterministic marking with static quantization [7]–[9].

The purpose of the present paper is to investigate in more detail the price tracking ability of ADPM. ADPM provides the receiver with information about the price every time that a packet is received. We use a discrete time model, where each discrete time unit represents the arrival of a packet at the receiver. If the arriving packet provides useful new information about the price, the receiver estimate is updated. Our interest is the MSE of this estimator, for a statistical model in which the price at the bottleneck router is executing a random walk. Routers have to estimate their price based on the random process of packet arrivals; it is for this reason that we use a random walk model for the price. The drift of the random walk models the average change in price between one received packet and the next: it is positive if congestion is building up, and negative when congestion is decreasing. We show that the error process is stationary, and we compute exact limiting distributions for the MSE, as the step size of the random walk tends to zero.

II. ADPM

The basic idea of ADPM is to transmit the *unary* representation of the *maximum* price seen by a packet as it traverses the network, appropriate for max-min flow control. Each packet that arrives at a router contains a threshold value, as provided by the `IPid` field. Each packet asks each router it encounters the same question: is your price greater than my threshold? The router answers "yes" or "no", providing a unary encoding of the price that is robust to packet loss, or to a reordering of the packet arrivals at the receiver.

In what follows, it is convenient to assume that prices have been mapped to lie in the unit interval $[0, 1]$; from now on, we will use the term "price" to refer to the mapped value.

Similarly, a mapping f is assumed, that maps `IPid` values to threshold values in $[0, 1]$. Following the terminology of [7], $i \equiv f(v)$ will be called the probe type of the packet. Implementation details behind the above assumptions are explained in Section V.

When a router with link price p forwards a packet of probe type i, it marks the packet if $p > i$, and leaves the mark unchanged otherwise. At the receiver, the mark of a packet of probe type i will be set if any router on the path had a price exceeding i. Decoding is simple. The receiver maintains a current estimate of the price, $\hat{p}$. If it sees a marked packet of probe type i with $i > \hat{p}$ or an unmarked packet of probe type i with $i < \hat{p}$, then it sets $\hat{p}$ to i.

In this algorithm, the interpretation of each mark is independent of the values of other marks. In contrast, with binary signalling [7], a price change from 3 (011) to 4 (100) could yield any price estimate from 000 to 111, depending on the order in which bits are signalled.

III. Performance Analysis

We will assume in this section that the probe types generated at the sender are independent and uniformly distributed on $[0, 1]$. We are interested in the estimation error after k packets have been received. The case of a fixed price is very simple to analyze, since for large k, the points picked out by the probe types are approximately a Poisson process of rate k on the interval $[0, 1]$. We can therefore easily show that the error is within a factor of 4 of the mean absolute quantization error of a k-level quantizer. More interesting is the analysis of a changing price.

A. Random walk model for price

A simple model is a discrete time random walk, in which the timeslots represent the times that packets arrive at the receiver. During each time-slot, the price at the bottleneck may change, and we use a random walk model. Let $p(n)$ be the price at time n. Then

$$p(n + 1) = p(n) + \delta J(n) \tag{1}$$

where $J(n)$ are an *i.i.d.* sequence of ± 1 random variables, each taking the value 1 with probability q, and -1 with probability $1 - q$. Thus, the drift of the random walk is $\mu\delta$, where $\mu = (2q - 1)$, and the variance of a jump is $\sigma^2\delta^2$, where $\sigma^2 = 4q(1 - q)$.

Let $\hat{p}(n)$ be the estimate of the price at the receiver immediately after the nth arriving packet has been processed. We assume that $\hat{p}(0)$ takes an arbitrary value in the interval $[0, 1]$.

To describe the error process, we begin with a definition that states precisely what we want to mean by the error, namely, $\epsilon(n) := p(n) - \hat{p}(n)$, although we will modify this definition slightly below. We note that when a packet arrives at the receiver, it may fail to cause an update in the receiver's estimate. In fact, there is an update of the estimate at time n in precisely two situations: either $\hat{p}(n-1) < i$ and the packet is marked, or $\hat{p}(n-1) > i$ and the packet is unmarked, where i is the probe type of the arriving packet at time n. In both

these cases, we get the assignment $\hat{p}(n) := i$. In all other cases, $\hat{p}(n) := \hat{p}(n - 1)$. Let $H(n)$, termed a "hit", be the event consisting of the two cases when the estimate changes occur, namely when the probe i falls into the interval between $\hat{p}(n - 1)$ and $p(n)$. The complement of $H(n)$ is denoted by $H^c(n)$. Since probes are uniformly distributed on $[0, 1]$, the probability of a hit at time n is $|\hat{p}(n - 1) - p(n)|$, and this hit will affect the estimate at time n.

We remark that we are using the very simple estimator that we described in Section II, and it is certainly not intended to be a maximum likelihood estimator. If the receiver knew the price drift it could do better than the scheme described here, but we do not wish to assume that the receiver has this information, and nor that it wishes to do significant computation on each received packet.

At time n, a packet arrives at the receiver, and may or may not cause an update in the price estimate at that instant. In any case, we denote by $\tilde{\epsilon}(n)$ the error in the estimate just *prior* to the processing of the packet received at time n, and by $\epsilon(n)$ the error immediately after the processing of that packet. These will be identical unless a hit occurs at time n, and if so, $\epsilon(n)$ will be the smaller of the two. Letting

$$[x]_{-1}^1 := \begin{cases} -1 & x < -1 \\ x & -1 \le x \le 1 \\ 1 & x > 1, \end{cases} \tag{2}$$

the price update relationship between $\epsilon(n)$ and $\tilde{\epsilon}(n)$ is expressed by

$$\tilde{\epsilon}(n) = [\epsilon(n - 1) + \delta J(n)]_{-1}^1. \tag{3a}$$

Also, the estimate update rule yields

$$\epsilon(n) = \begin{cases} \tilde{\epsilon}(n) & \text{if } H(n)^c \\ \tilde{\epsilon}(n) - U(n) & \text{if } H(n) \end{cases} \tag{3b}$$

where, conditional on $\tilde{\epsilon}(n)$,

$$U(n) \sim \mathcal{U}[0, \tilde{\epsilon}(n)] \tag{3c}$$
$$H(n) \sim \mathcal{B}(|\tilde{\epsilon}(n)|), \tag{3d}$$

where $\mathcal{U}$ denotes a uniform random variable, and $\mathcal{B}$ denotes a Bernoulli random variable.

Note that we are assuming that the estimate at time n, $\hat{p}(n)$, is calculated immediately *after* the probe arrives at time n. Using induction, we can see that $\epsilon(n)$ is always the error in this case. If instead $\hat{p}(n)$ is calculated immediately *before* the probe arrives at time n then it is $\tilde{\epsilon}(n)$ that represents the error. We will have use for both definitions of error process in this paper.

The above definition of a hit time involves a slight trick; strictly speaking, a hit should occur if a uniformly distributed probe lies in the interval between the estimated price (prior to processing the probe) and the true price. However, such a definition couples the hit events with the price process, which is itself nonstationary (when $\mu \ne 0$), and subject to the boundary conditions $p \le 1$ and $p \ge 0$. However, provided the boundary constraints are slack, the above definition is equivalent. The benefit of the above definition is that it avoids

boundary conditions, and allows the error processes to be stationary, as we will show. The definition is self contained, and from now on we can ignore the price process altogether.

To help visualize the error process, $\epsilon(n)$, consider the special case in which $q = 1$. The error process, ϵ increases at constant rate δ, until the random event of a "hit", and at this time, it makes a random-sized jump back towards zero. This process is illustrated in Figure 1.
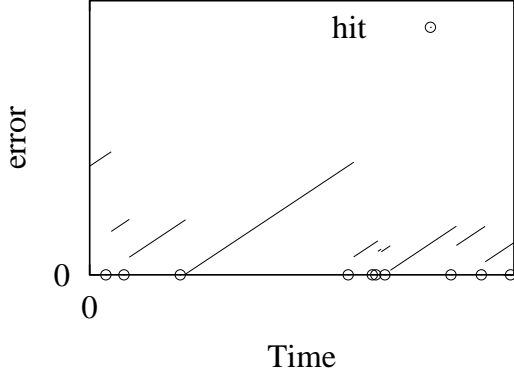

Fig. 1. Evolution of error process

Clearly, this particular $\epsilon(n)$ undergoes a zig-zag evolution, with steady increase at rate δ, followed by a jump in the slot when an update (hit) is detected. More generally, when $0 < q < 1$, $\epsilon(n)$ will undergo a more wiggly evolution due to random fluctuations in the price.

The case $q = 1$ is the special case in which the drift is $\mu\delta = \delta$. In general, if $1 > \mu > 0$, then the price tends to increase, modelling a scenario in which congestion is building up at the bottleneck. However, the increase is not in general deterministic, allowing for stochastic fluctuations in the price. If $-1 < \mu < 0$, then the price tends to decrease, modelling a scenario in which congestion is decreasing at the bottleneck. Oscillations in price are an inevitable consequence of two factors:

1) On a slow time-scale, the price can tend to increase, or tend to decrease, as a result of flows arriving and departing.
2) On a fast timescale, the price is affected by stochastic fluctuations in the packet arrival process at the bottleneck resource.

We capture both effects in our random walk model. The random steps of the walk model the stochastic effects, and the drift captures the slow timescale effect of flows arriving and departing. Note that if the flow control manages to reach equilibrium, then $\mu = 0$ at equilibrium.

The main result of this paper is as follows.

Theorem 1: The error process defined by (3) has an equilibrium distribution for any $\delta > 0$. Let F_δ be the equilibrium distribution of $\tilde{\epsilon}/\sqrt{\delta}$. Let $(\delta_n), n = 1, 2, \ldots$ be any positive monotonic sequence tending to zero as $n \to \infty$, with $\delta_1 < 1$. Let X_{δ_n} be a random variable with distribution F_{δ_n}. Then

for $\mu \neq 0$,

$$X_{\delta_n} \overset{d}{\leadsto} \mathcal{R}(\sqrt{|\mu|})\mu/|\mu| \qquad (4)$$

where $\overset{d}{\leadsto}$ denotes weak convergence, and $\mathcal{R}(\sqrt{|\mu|})$ is a Rayleigh random variable with parameter $\sqrt{|\mu|}$. If $\mu = 0$, then the sequence (F_{δ_n}) converges weakly to the distribution

$$F_X(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (5)$$

An immediate consequence of Theorem 1 is that the MSE of the estimate is asymptotically $2|\mu|\delta$ for small δ. Indeed, if there is no randomness in the walk ($|\mu| = 1$) then 2δ is an upper bound on the MSE for all δ. These results are expressed in the following corollary:

Corollary 2: 1) For all $-1 \leq \mu \leq 1$,

$$\mathbb{E}[\tilde{\epsilon}^2] = 2|\mu|\delta + o(\delta), \qquad (6)$$

2) If $|\mu| = 1$, then

$$\mathbb{E}[\tilde{\epsilon}^2] \in \left[\frac{2\delta}{1 + (2\delta)^{3/2}/(1-\delta)^2}, 2\delta \right]. \qquad (7)$$

B. Stationarity of the error processes

In this section, we show that the process $\tilde{\epsilon}(n)$ can be stationary, which establishes the first part of Theorem 1. The stationarity of the process $\epsilon(n)$ is an immediate consequence of this fact.

Note that we can rewrite (3b) as

$$\tilde{\epsilon}(n+1) = \begin{cases} [\tilde{\epsilon}(n) + \delta J(n+1)]^1_{-1} & \text{if } H(n)^c \\ [\tilde{\epsilon}(n) + \delta J(n+1) - U(n)]^1_{-1} & \text{if } H(n) \end{cases} \qquad (8)$$

where again, conditional on $\tilde{\epsilon}(n)$, $U(n) \sim U[0, \tilde{\epsilon}(n)]$.

Let the temporally homogeneous transition function for this Markov chain be $P(x, A) \equiv \mathbb{P}(\tilde{\epsilon}(2) \in A | \tilde{\epsilon}(1) = x)$ for $-1 < x < 1$ and A a Borel-measurable set contained in $[-1, 1]$. It follows from (3) and (8) that for $-1 + \delta < x < 1 - \delta$,

$$P(x, A) = (1 - |x|)(qI[x + \delta] \in A] + (1-q)I[x - \delta \in A])$$
$$+ |x| (q \, ||A \cap [\delta, x + \delta]|| + (1-q) \, ||A \cap [-\delta, x - \delta]||).$$

(For x close to the boundary, we get similar expressions, but need to invoke the clipping operator $[\cdot]^1_{-1}$.)

It is simple to calculate the drift function

$$\gamma_x \equiv \mathbb{E}[\tilde{\epsilon}(2) - x | \tilde{\epsilon}(1) = x].$$

For example, if $0 < x < 1 - \delta$, then

$$\gamma_x = x^3/2 - x^2 + (1-x)\mu\delta + x^2\mu\delta$$

Thus, in the limit as $\delta \downarrow 0$, and for x close to 1, we have γ_x close to $-1/2$. Similarly, it can be shown that in the limit as $\delta \downarrow 0$, and for x close to -1, we have γ_x close to $1/2$.

Stationarity follows from the compactness of the state-space, $[-1, 1]$, and the fact that the drift points inward from

the boundary. Technically, the conditions stated in Corollary 5.2 in [10] seem to require in addition that $P(x, A)$ is strongly continuous [10] for any Borel measurable set A, to conclude that $\tilde{\epsilon}$ is stationary, and this condition does not hold for our transition probability function. However, from the note added in proof in [10], it is in fact sufficient in our case to verify instead that the function $P(x, A)$ is weakly continuous for any Borel measurable set A, to conclude that $\tilde{\epsilon}$ is stationary. This weaker condition holds because our state-space is a Banach space. Weak continuity is the requirement that $\int g(y) P(x, dy)$ is a continuous bounded function of x, for any continuous, bounded function $g(y)$. This is the case for our transition function $P(x, A)$, and hence $\tilde{\epsilon}(n)$ can be stationary.

C. Asymptotic analysis of estimator mean square error

In this section, we derive the asymptotic form for the stationary distribution of the error process, ϵ.

Let $\tilde{F}_n$ and F_n denote the conditional distribution functions of $\tilde{\epsilon}(n)$ and $\epsilon(n)$, respectively. Clearly, for any x such that $-1 + \delta < x < 1 - \delta$, the mapping from F to $\tilde{F}$ is

$$\tilde{F}_n(x) = q F_n(x - \delta) + (1 - q) F_n(x + \delta). \quad (9)$$

A corresponding mapping from $\tilde{F}$ to F is obtained by conditioning (3b) on $\tilde{\epsilon}(n)$ and taking expectations. For $x \geq 0$,

$$
\begin{aligned}
F_{n+1}(x) &= \int_{-1}^{1} \mathbb{P}(\epsilon(n+1) \leq x | \tilde{\epsilon}(n)) \, d\tilde{F}(\tilde{\epsilon}(n)) \\
&= \int_{-1}^{x} 1 \, d\tilde{F}(\tilde{\epsilon}(n)) + \\
&\quad \int_{x}^{1} \tilde{\epsilon}(n) \mathbb{P}(\mathcal{U}(0, \tilde{\epsilon}(n)) \leq x | \tilde{\epsilon}(n)) \, d\tilde{F}(\tilde{\epsilon}(n)) \\
&= \tilde{F}_n(x) + \int_{x}^{1} x \, d\tilde{F}(\tilde{\epsilon}(n)).
\end{aligned}
$$

Similar manipulations for $x < 0$ yield

$$F_{n+1}(x) = \begin{cases} \tilde{F}_n(x) + x(1 - \tilde{F}_n(x)) & x \geq 0 \\ (1 + x)\tilde{F}_n(x) & x < 0 \end{cases} \quad (10)$$

The stationary distribution, F, must then satisfy the fixed point equation

$$F(x) =$$

$$\begin{cases} x + (1 - x)(q F(x - \delta) + (1 - q) F(x + \delta)) & x \geq 0 \\ (1 + x)(q F(x - \delta) + (1 - q) F(x + \delta)) & x < 0. \end{cases}$$

where $x \in [-1 + \delta, 1 - \delta]$. This is the stationary distribution of the error of our estimator, *i.e.* the error that is obtained after processing the incoming probe packet. Re-arranging, we obtain that for $x \in [0, 1 - \delta)$,

$$
\begin{aligned}
q(F(x) - F(x - \delta)) &+ (1 - q)(F(x) - F(x + \delta)) \\
&= x(1 - q F(x - \delta) - (1 - q) F(x + \delta)) \quad (11)
\end{aligned}
$$

and for $x \in (-1 + \delta, 0)$:

$$
\begin{aligned}
q(F(x) - F(x - \delta)) &+ (1 - q)(F(x) - F(x + \delta)) \\
&= x(q F(x - \delta) + (1 - q) F(x + \delta)). \quad (12)
\end{aligned}
$$

Let ϵ be a random variable with distribution function F, and define X_δ to be the random variable $\epsilon/\sqrt{\delta}$, with distribution function F_δ. Then we obtain equivalent fixed point equation for F_δ. Let

$$G_\delta(x) := q F_\delta(x - \sqrt{\delta}) + (1 - q) F_\delta(x + \sqrt{\delta})$$

Then for $x \in [0, 1/\sqrt{\delta} - \sqrt{\delta})$:

$$
\begin{aligned}
q(F_\delta(x) - F_\delta(x - \sqrt{\delta})) &+ (1 - q)(F_\delta(x) - F_\delta(x + \sqrt{\delta})) \\
&= x\sqrt{\delta}(1 - G_\delta(x)) \quad (13)
\end{aligned}
$$

and for $x \in (-1/\sqrt{\delta} + \sqrt{\delta}, 0)$:

$$
\begin{aligned}
q(F_\delta(x) - F_\delta(x - \sqrt{\delta})) &+ (1 - q)(F_\delta(x) - F_\delta(x + \sqrt{\delta})) \\
&= x\sqrt{\delta} G_\delta(x). \quad (14)
\end{aligned}
$$

Note that while F is the distribution function of a random variable taking values in an interval approximately $[-1, 1]$, F_δ is the distribution function of a random variable taking values in an interval approximately $[-1/\sqrt{\delta}, 1/\sqrt{\delta}]$. As δ tends to zero, the support becomes unbounded, and Theorem 1 confirms weak convergence to a distribution with unbounded support.

Before we prove the theorem, note that if such a convergence takes place to a limiting distribution $G(x)$, then taking formal limits as $\delta \downarrow 0$ in (13) and (14) suggest that G should satisfy the differential equations (15) given in the statement of Lemma 1 below. The following lemmas make this argument rigorous. Those proofs which are not here are in the appendix.

Lemma 1: Let $(\delta_n), n = 1, 2, \ldots$ be any positive monotonic sequence tending to zero as $n \to \infty$, with $\delta_1 < 1$. The sequence (F_{δ_n}) has a weakly convergent subsequence. Moreover, the limit of any weakly convergent subsequence satisfies the differential equations:

$$\mu \dot{G}(x) = \begin{cases} x(1 - G(x)) & x > 0 \\ x G(x) & x < 0 \end{cases} \quad (15)$$

If $\mu = 0$, then the sequence (F_{δ_n}) converges weakly to

$$G(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (16)$$

Lemma 2: If $\tilde{\epsilon}(k)$ is stationary, and the drift of the random walk (8) is $\mu\delta$ then in equilibrium,

$$E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon} > 0]}] - E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon} < 0]}] \approx 2\mu\delta,$$

in the sense that

$$|E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon} > 0]}] - E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon} < 0]}] - 2\mu\delta| \leq 2\delta \mathbb{P}(|\tilde{\epsilon}| > 1 - \delta) \mathbb{E}[|\tilde{\epsilon}|], \quad (17)$$

where $I_{[\cdot]}$ denotes the indicator function.

Proof: If $\mu = 0$, the result hold by symmetry about the origin. Consider now the case $\mu \neq 0$. Let $\mathbb{P}(H)$ be the equilibrium probability of a hit, averaged over the equilibrium statistics of $\tilde{\epsilon}$. Let $\mathbb{P}(H|x)$ denote the conditional probability of a hit, given $\tilde{\epsilon} = x$, which is given by $\mathbb{P}(H|x) = |x|$, since

the probe types are $U[0,1]$. Averaging over the statistics of $|\tilde{\epsilon}|$, we obtain

$$\mathbb{E}[|\tilde{\epsilon}|] = \mathbb{P}(H) \qquad (18)$$

Now consider two randomly chosen adjacent hit times, T_1 and T_2, and let $X = T_2 - T_1 > 0$ denote the time between these two hits. Clearly,

$$\mathbb{P}(H) = 1/\mathbb{E}[X] \qquad (19)$$

If $\tilde{\epsilon}$ is in equilibrium, then so is the embedded chain obtained by sampling at the hit times. Thus,

$$\mathbb{E}[\tilde{\epsilon}(T_1)] = \mathbb{E}[\tilde{\epsilon}(T_2)] \qquad (20)$$

and we denote the common value by $\mathbb{E}[\tilde{\epsilon}|H]$. However, consideration of the drift of the embedded chain (8) provides that

$$\mathbb{E}[\tilde{\epsilon}(T_2) - \tilde{\epsilon}(T_1)|\tilde{\epsilon}(T_1)] = \mu\delta\mathbb{E}[X|\tilde{\epsilon}(T_1)] - \frac{\tilde{\epsilon}(T_1)}{2} + \Delta \quad (21)$$

where

$$\Delta \in [-\delta I_{[\tilde{\epsilon}(T_1)<1-\delta]}, \delta I_{[\tilde{\epsilon}(T_1)>-(1-\delta)]}] \qquad (22)$$

accounts for the clipping operation $[\cdot]_{-1}^{1}$. Taking expectations in (21) and applying (20) yields

$$\mathbb{E}[\tilde{\epsilon}|H] = 2\mu\delta\mathbb{E}[X] + 2\mathbb{E}[\Delta]. \qquad (23)$$

Putting (18), (19) and (23) together, we obtain

$$\mathbb{E}[\tilde{\epsilon}|H]\mathbb{E}[|\tilde{\epsilon}|] = 2\mu\delta + 2\mathbb{E}[\Delta]\mathbb{E}[|\tilde{\epsilon}|] \qquad (24)$$

But by Bayes' Theorem,

$$\begin{aligned} E[\tilde{\epsilon}|H] &= \int_{-1}^{1} \varepsilon f_{\tilde{\epsilon}}(\varepsilon|H) \, d\varepsilon \\ &= \int_{-1}^{1} \varepsilon \frac{f_{\tilde{\epsilon}}(\varepsilon)}{P(H)} P(H|\varepsilon) \, d\varepsilon \\ &= \int_{0}^{1} \varepsilon^2 \frac{f_{\tilde{\epsilon}}(\varepsilon)}{P(H)} \, d\varepsilon - \int_{-1}^{0} \varepsilon^2 \frac{f_{\tilde{\epsilon}}(\epsilon)}{P(H)} \, d\varepsilon \\ &= \frac{(E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon}>0]}] - E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon}<0]}])}{E[|\tilde{\epsilon}|]}. \end{aligned} \qquad (25)$$

Combining (24) and (25) with

$$|\mathbb{E}[\Delta]| \le \delta\mathbb{P}(|\tilde{\epsilon}| > 1 - \delta) \qquad (26)$$

gives the stated result. ∎

Corollary 3:

$$\limsup_{\delta\downarrow 0} \mathbb{E}\left[\left(\frac{\tilde{\epsilon}}{\sqrt{\delta}}\right)^2\right] + \mathbb{P}(|\tilde{\epsilon}| > 1 - \delta)\mathbb{E}[|\tilde{\epsilon}|] \ge 2|\mu| \quad (27)$$

Proof: Follows immediately from Lemma 2. ∎

We are now ready to prove Theorem 1.

Proof: It follows from Lemma 1 that (F_{δ_n}) has a weakly convergent subsequence, converging to a limiting distribution function G, which satisfies the differential equations (15). It remains to be shown that G is uniquely defined by these equations, and that there is no discontinuity at $x = 0$.

Consider the region $x \ge 0$. Setting $H(x) := 1 - G(x)$, the equivalent ODE is

$$\dot{H}(x) = -\frac{x}{\mu}H(x)$$

which has solutions $H(x) = A_1 \exp(-x^2/(2\mu))$. Hence, $G(x) = 1 - A_1 \exp(-x^2/(2\mu))$, where A_1 is a constant satisfying $0 \le A_1 \le 1$.

For the region $x < 0$, (15) can be rewritten

$$\dot{G}(x) = -\frac{x}{-\mu}G(x)$$

which has solutions $G(x) = A_2 \exp(-x^2/(-2\mu))$, for $0 \le A_2 \le 1$. Thus,

$$G(x) = \begin{cases} 1 - A_1 \exp(-x^2/(2\mu)) & x \ge 0 \\ A_2 \exp(-x^2/(-2\mu)) & x < 0 \end{cases} \qquad (28)$$

It remains to find the constants A_1, and A_2, using the fact that $G(x)$ is a distribution. If $\mu > 0$ then $\exp(-x^2/(-2\mu))$ is unbounded as $x \downarrow -\infty$, and so $A_2 = 0$ in that case. Conversely, if $\mu < 0$ then $A_1 = 0$ in that case by the same reasoning.

Consider first the case that $\mu > 0$, whence $A_2 = 0$, and G is the distribution of a random variable $Y = BR$, where B is a Bernoulli random variable with mean A_1, and R is an independent Rayleigh random variable, with parameter $\sqrt{\mu}$. Note that

$$\mathbb{E}[Y^2] = A_1 2\mu. \qquad (29)$$

Weak convergence of F_δ implies that for any $x \ne 0$ (where G is continuous), $F_\delta(x) \to G(x)$ along the subsequence $\mathcal{S}$. Continuity of G at x also implies (see Lemma 3 in the appendix) that

$$qF_\delta(x - \delta^{-1/2}) + (1-q)F_\delta(x + \delta^{1/2}) \to G(x).$$

Defining

$$\tilde{F}_\delta(x) := \mathbb{P}\left(\frac{\tilde{\epsilon}}{\sqrt{\delta}} \le x\right)$$

we obtain that $\tilde{F}_\delta$ converges weakly to G along the subsequence also. Together with (29) we obtain that

$$\mathbb{E}\left[\left(\frac{\tilde{\epsilon}}{\sqrt{\delta}}\right)^2\right] \to A_1 2\mu \qquad (30)$$

and that

$$\mathbb{P}(|\tilde{\epsilon}| > 1 - \delta)\mathbb{E}[|\tilde{\epsilon}|] \to 0. \qquad (31)$$

But Corollary 3 then implies that $A_1 = 1$, and hence G is Rayleigh with parameter $\sqrt{\mu}$.

A very similar argument applies in the case $\mu < 0$ to show that $A_2 = 1$ in that case. In either case, let us label the unique solution F_X. If $\mu > 0$, then

$$F_X(x) = \begin{cases} 1 - \exp(-x^2/|2\mu|) & x > 0 \\ 0 & x < 0 \end{cases} \qquad (32)$$

which is a Rayleigh distribution with parameter $\sqrt{\mu}$. If $\mu < 0$, then

$$F_X(x) = \begin{cases} 1 & x > 0 \\ \exp(-x^2/|2\mu|) & x < 0. \end{cases} \qquad (33)$$

Note that in both these cases, there is a density function, valid for all x.

The case $\mu = 0$ is just a restatement of the corresponding result in Lemma 1. (Identify G in (16) with F_X in (5).)

Since the limiting distribution is in all cases unique, it follows that all convergent subsequences must converge to F_X, and hence $(F_{\delta_n}), n = 1, 2, \ldots$ converges weakly to F_X. Note that if $\mu < 0$, then $(-X_{\delta_n}), n = 1, 2, \ldots$ converges in distribution to a Rayleigh distribution with parameter $\sqrt{-\mu}$. ∎

We can now prove Corollary 2

Proof: Part (i) of Corollary 2 follows immediately from the definition of weak convergence. For part (ii), note that the equilibrium values are independent of the initial value $\tilde{\epsilon}(0)$. Consider now the case $\mu = 1$. In this case, $\tilde{\epsilon}(0) > 0$ implies $\tilde{\epsilon}(n) > 0$ for all n. Thus the support of the equilibrium distribution for $\tilde{\epsilon}$ is the positive reals and the clipping in (8) is always down, giving $\Delta \leq 0$.

Note also that $\mathbb{P}(|\tilde{\epsilon}| > 1 - \delta) \leq \mathbb{E}[\tilde{\epsilon}^2]/(1 - \delta)^2$. Together, $\Delta \leq 0$, (24), (25) and (26) imply

$$\mathbb{E}[\tilde{\epsilon}^2] - 2\delta \in \left[-2\delta \frac{\mathbb{E}[|\tilde{\epsilon}|]\mathbb{E}(\tilde{\epsilon}^2)}{(1 - \delta)^2}, 0\right].$$

Hence $\mathbb{E}[|\tilde{\epsilon}|] \leq \sqrt{2\delta}$ and

$$\mathbb{E}[\tilde{\epsilon}^2]\left(1 + \frac{2\delta\sqrt{2\delta}}{(1 - \delta)^2}\right) \geq 2\delta$$

giving the result. The $\mu = -1$ case follows similarly. ∎

Clearly, Theorem 1 provides precise asymptotics for the square error when $\mu \neq 0$. Let X be a random variable with the distribution F_X, which represents the estimator error on the scale $O(\sqrt{\delta})$ as δ tends to zero. It follows that on a scale of $O(\delta)$, the squared error of the estimator converges in distribution to an exponential distribution with mean 2μ, provided $\mu \neq 0$. When $\mu = 0$, Corollary 2 implies that the squared error is $o(\delta)$, but the corollary does not characterize the precise order in this case. Part (ii) of Corollary 2 provides a special case in which a pre-asymptotic result is available, namely when the walk is not random, and $|\mu| = 1$. In this intuitively worst case, the upper bound

$$\mathbb{E}[\tilde{\epsilon}^2] \leq 2\delta \tag{34}$$

holds for all δ.

Finally, we can strengthen the statement of Lemma 2 as follows. From (31) and Lemma 2, it follows that

$$E\left[\frac{\tilde{\epsilon}^2}{\delta}I_{[\tilde{\epsilon}>0]}\right] - E\left[\frac{\tilde{\epsilon}^2}{\delta}I_{[\tilde{\epsilon}<0]}\right] \to 2\mu \tag{35}$$

But by Corollary 2 we have that

$$E\left[\frac{\tilde{\epsilon}^2}{\delta}I_{[\tilde{\epsilon}>0]}\right] + E\left[\frac{\tilde{\epsilon}^2}{\delta}I_{[\tilde{\epsilon}<0]}\right] \to 2|\mu| \tag{36}$$

Hence, if $\mu > 0$,

$$E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon}>0]}] = 2\mu\delta + o(\delta), \quad E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon}<0]}] = o(\delta)$$

and, if $\mu < 0$,

$$E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon}<0]}] = -2\mu\delta + o(\delta), \quad E[\tilde{\epsilon}^2 I_{[\tilde{\epsilon}>0]}] = o(\delta).$$

IV. BIT REVERSED COUNTING

The analysis we have provided in the present paper applies directly to ADPM when `IPid` values provide random thresholds. It is certainly the case that some operating systems use uniformly distributed pseudo-random values for the `IPid` (*e.g.* NetBSD) making our analysis directly applicable. However, many operating systems assign sequential `IPid` values to packets. This allows the sequence of probes types to form a sequence of "bit-reversed counting" (BRC) values, that is, $100\ldots$, $010\ldots$, $110\ldots$, $\ldots$, preceded by a "binary point". That is, using the sequence $R(1)$, $R(2)$, $R(3)$,$\ldots$, where $R(\sum_{i=0}^{\infty} a_i 2^i) = \sum_{i=0}^{\infty} a_i 2^{-i}$ for any sequence of bits $\{a_i\}$.

If the probe type sequence is bit-reversed counting, $(R(1), R(2), R(3),\ldots)$ then after $k = 2^j - 1$ probe packets, probes 2^{-j}, 2×2^{-j}, $\ldots$, $1 - 2^{-j}$ will have been received. This divides the possible values for p into uniformly spaced "uncertainty intervals" of width 2^{-j}. The ADPM receiver knows which interval the price is in, but not where within that interval. Thus the estimation error after k packets is bounded above by

$$|\hat{p} - p| < 2^{-\lfloor \log_2 k \rfloor} \in [1/k, 2/k). \tag{37}$$

Thus, BRC bounds the error without increasing its mean. Numerical results in [1] show that for a fixed price, it provides better performance than that predicted by the model of uniformly distributed probes.

It is likely that BRC provides a better sequence of probes for tracking a changing price, although we conjecture that the asymptotics for the random walk model, as δ tends to zero, will be the same as for random probes. However, we have done no numerical or analytical work to study this case: a topic for future work.

V. IMPLEMENTATION DETAILS

IP has two ECN bits (RFC 3168), which a source sets to 00 to indicate that it does not understand ECN, or to either 01 or 10 to indicate an unmarked packet, while routers mark packets with 11 to indicate congestion, which the source must treat like a packet loss. Sending 11 frequently [5]–[7],[9] will cripple standards-compliant flows. To prevent this, ADPM uses 01 to indicate no mark, and 10 to indicate a mark due to pricing. As such, ADPM is "ECN-friendly".

The function f that maps `IPid` values into probe types can be implemented either at the source or in the routers. If it is to be implemented at the routers, we must verify that it will be compatible with different `IPid` sequences produced by different operating systems. For operating systems which assign sequential `IPid` values to packets, the f corresponding to bit reverse counting can be implemented in the routers in a simple and scalable way by simply reversing the order of the `IPid` bits and comparing this with the price. For operating systems which use uniformly distributed pseudo-random values (*e.g.* NetBSD), the routers can reverse the bits

before the comparison with the price, without affecting the uniform distribution for the resulting probe type.

Using pseudo-random `IPid` values provides worse performance than that obtained by bit reversed counting; however, our analysis of this case (above) shows that performance is still very good. If necessary, `IPid` values can be replaced at the source without changing their primary function, providing nearby packets have unique identifiers.

Note that the mapping of prices and `IPid` to $[0, 1]$ is purely a mathematical convenience, and does not limit the actual range of prices. Either mapping may be nonlinear. In particular, the mapping of prices can be selected to get better quantization and faster estimation of the true price at the receiver.

VI. CONCLUSION

We have provided an analysis of the error of an estimator based on the recently proposed scheme ADPM [1]. The task is to estimate the maximum congestion price seen along a path in the Internet, as required by a congestion control algorithm for a flow using that path. The constraint is that the routers can only mark packets with a single bit. In this paper, we model the price at the bottleneck router as a random walk with drift. We show that as the step size, δ, of the random walk tends to zero, and provided the drift is nonzero, the distribution of the squared error converges weakly to an Exponential distribution with mean $2|\mu|\delta$, where $\mu\delta$ is the drift of the random walk. Thus, the MSE is of the same order as the step size of the random walk. If the drift is zero, we show that the MSE is $o(\delta)$, but do not characterize its precise order in this case.

Since ADPM is a signalling technique, not a flow control technique, its usefulness will depend on the flow control control technique it is applied to. Currently, ADPM is being integrated with the MaxNet flow control algorithm [11] using the WAN-in-Lab infrastructure [12].

APPENDIX I
PROOFS OF LEMMAS

The proof of Lemma 1 will make use of the following two lemmas.

Lemma 3: Let (G_n) be a sequence of distribution functions converging weakly to G. If G is continuous on an open interval I and (x_n) is a sequence in I converging to a point x in I, then $G_n(x_n) \to G(x)$.

Proof: Let $\varepsilon > 0$ be sufficiently small that both $x - \varepsilon$ and $x + \varepsilon$ lie in I. For n sufficiently large, we have $x - \varepsilon < x_n < x + \varepsilon$, and thus $G_n(x - \varepsilon) \le G_n(x_n) \le G_n(x + \varepsilon)$. By continuity of G on I, and the Portmanteau theorem, it follows that $G_n(x - \varepsilon) \to G(x - \varepsilon)$, and $G_n(x + \varepsilon) \to G(x + \varepsilon)$. But by continuity of G at x, the difference between $G(x - \varepsilon)$ and $G(x + \varepsilon)$ can be made arbitrarily small. It follows that $G_n(x_n) \to G(x)$. ∎

Lemma 4: For all $x > 1$ and $0 < \delta < \min(1/x^2, 1/4)$,
1) $F(-x\sqrt{\delta}) < 2/x$ and
2) $F(x\sqrt{\delta}) > 1 - 2/x$.

Proof: The proof divides a sub-interval of $[-x\sqrt{\delta}, 0]$ (respectively $[0, x\sqrt{\delta}]$) into strips of width

$$K_\delta := \lfloor 1/\sqrt{\delta} \rfloor. \tag{38}$$

A bound is found concerning the change in F over each of these strips. By summing these bounds, it is shown that if the lemma were false then F would exceed 1 at some point, and could not be a distribution function.

To prove part 1), assume for the sake of argument that $F(-x\sqrt{\delta}) \ge 2/x$. For any $1 \le k \le K_\delta$ it follows from (12) that

$$\begin{aligned}
&\Big(F(-x\sqrt{\delta} + k\delta) - F(-x\sqrt{\delta} + (k-1)\delta) \Big) q + \\
&\Big(F(-x\sqrt{\delta} + k\delta) - F(-x\sqrt{\delta} + (k+1)\delta) \Big) (1-q) \\
&= \Big(-x\sqrt{\delta} + k\delta \Big) \Big(q(F(-x\sqrt{\delta} + (k-1)\delta)) + \\
&\qquad (1-q)(F(-x\sqrt{\delta} + (k+1)\delta)) \Big). \tag{39}
\end{aligned}$$

By the monotonicity of F and the fact that $k \ge 1$, the contradiction hypothesis implies

$$F(-x\sqrt{\delta} + (k+1)\delta) \ge F(-x\sqrt{\delta} + (k-1)\delta) \ge 2/x,$$

whence the second factor of the right hand side of (39) is at least $2/x$. Since $-x\sqrt{\delta} + k\delta \le 0$, (39) implies

$$\begin{aligned}
&\Big(F(-x\sqrt{\delta} + k\delta) - F(-x\sqrt{\delta} + (k-1)\delta) \Big) q + \\
&\Big(F(-x\sqrt{\delta} + k\delta) - F(-x\sqrt{\delta} + (k+1)\delta) \Big) (1-q) \\
&\le -2\sqrt{\delta} + 2k\delta/x. \tag{40}
\end{aligned}$$

Summing (40) over $k = 1, 2, \ldots, K_\delta$, gives

$$\begin{aligned}
&(F(-x\sqrt{\delta} + K_\delta\delta) - F(-x\sqrt{\delta}))q + \\
&(F(-x\sqrt{\delta} + \delta) - F(-x\sqrt{\delta} + (K_\delta + 1)\delta))(1-q) \\
&\le -2K_\delta\sqrt{\delta} + K_\delta(K_\delta + 1)\delta/x \tag{41} \\
&< -1 + 2/x, \tag{42}
\end{aligned}$$

as $2\lfloor 1/\sqrt{\delta} \rfloor\sqrt{\delta} > 1$ and $\lfloor 1/\sqrt{\delta} \rfloor(\lfloor 1/\sqrt{\delta} \rfloor + 1)\delta < 2$ for $\delta < 1/4$. But by assumption,

$$F(-x\sqrt{\delta} + K_\delta\delta) \ge F(-x\sqrt{\delta} + \delta) \ge 2/x. \tag{43}$$

Together with the monotonicity of F, (42) and (43) imply that $F(-x\sqrt{\delta} + (K_\delta + 1)\delta) > 1$, a contradiction. Thus $F(-x\sqrt{\delta}) \le 2/x$.

The proof of part 2) is analogous, applying (11) around the point

$$F(x\sqrt{\delta} - k\delta)$$

and reversing the inequalities in (40)-(42). ∎

We can now prove Lemma 1.

Proof: In the following, we will use the same symbol to denote both the distribution function and the probability measure induced by it. Thus $G(x)$ is equivalent to $G((-\infty, x])$.

To show that the sequence of distribution functions F_{δ_n} is tight [13], let $\gamma > 0$, and consider the compact set $K_\gamma = [-2/\gamma, 2/\gamma]$. Lemma 4 implies that for all n,

$\mathbb{P}(X_{\delta_n} \in K_\gamma^c) < 2\gamma$. This implies that the sequence $(F_{\delta_n})_n$ is tight, and hence has a weakly convergent subsequence, $\mathcal{S}$, converging to a limiting distribution function, G, by Prokhorov's theorem.

We now rule out the possibility that $G(x)$ jumps at any $x > 0$. Denote the jump of G at x by $\triangle G(x)$. Let $\varepsilon > 0$ be sufficiently small that $x - \varepsilon > 0$. Consider δ small enough that $K := \lfloor > \varepsilon/\sqrt{\delta} \rfloor > 1$, and note that from (13), for any integer k from $-(K-1)$ to $K-1$, we have that:

$$q(F_\delta(x + k\sqrt{\delta}) - F_\delta(x + (k-1)\sqrt{\delta})) +$$
$$(1-q)(F_\delta(x + k\sqrt{\delta}) - F_\delta(x + (k+1)\sqrt{\delta}))$$
$$= (x + k\sqrt{\delta})\sqrt{\delta}(1 - G_\delta(x + k\sqrt{\delta})) \qquad (44)$$

Adding up these $2K + 1$ equations we obtain the following upper bounds:

$$
\begin{aligned}
L_\delta :=&\ (F_\delta(x + (K-1)\sqrt{\delta}) - F_\delta(x - K\sqrt{\delta}))q + \\
&\ (F_\delta(x - (K-1)\sqrt{\delta}) - F_\delta(x + K\sqrt{\delta}))(1-q) \quad (45) \\
<&\ (x + (K-1)\sqrt{\delta})\sqrt{\delta} \sum_{k=-(K-1)}^{K-1} (1 - G_\delta(x + k\sqrt{\delta})) \\
\leq&\ (x + \varepsilon)\sqrt{\delta} \left(1 + \sum_{k=-(K-2)}^{K-1} 1 - F_\delta(x + (k-1)\sqrt{\delta}) \right) \\
\leq&\ (x + \varepsilon)\sqrt{\delta}(1 + (2K-2)(1 - F_\delta(x - (K-1)\sqrt{\delta}))) \\
\leq&\ (x + \varepsilon)(2\varepsilon - (2\varepsilon - 4\sqrt{\delta})F_\delta(x - (K-1)\sqrt{\delta})) \\
\leq&\ (x + \varepsilon)(2\varepsilon - (2\varepsilon - 4\sqrt{\delta})F_\delta((-\infty, x - \varepsilon))) \quad (46)
\end{aligned}
$$

where the second inequality uses $G_\delta(x) \geq F_\delta(x - \sqrt{\delta})$, and the fourth uses $\varepsilon - \sqrt{\delta} \leq \sqrt{\delta}K \leq \varepsilon$. Using the Portmanteau theorem [13] applied to open sets, we can take the limsup of the right hand side of this bound, as $\delta \downarrow 0$, to obtain the asymptotic upper bound of

$$(x + \varepsilon)2\varepsilon(1 - G((-\infty, x - \varepsilon))). \qquad (47)$$

This term will upper bound the $\liminf$ of L_δ as δ tends to zero.

Note that

$$(x - K\sqrt{\delta}, x + (K-1)\sqrt{\delta}) \subseteq (x - \varepsilon + \sqrt{\delta}, x + \varepsilon - 2\sqrt{\delta})$$

and

$$(x - (K-1)\sqrt{\delta}, x + K\sqrt{\delta}) \supseteq (x - \varepsilon + \sqrt{\delta}, x + \varepsilon).$$

Thus for any $\xi > 0$,

$$
\begin{aligned}
\liminf_{\delta \downarrow 0} L_\delta \geq&\ \liminf_{\delta \downarrow 0} \big[qF_\delta((x - \varepsilon + \sqrt{\delta}, x + \varepsilon - 2\sqrt{\delta})) \\
&\ -(1-q)F_\delta([x - \varepsilon + \sqrt{\delta}, x + \varepsilon])\big] \\
\geq&\ \liminf_{\delta \downarrow 0} \big[qF_\delta((x - \varepsilon + \xi, x + \varepsilon - \xi)) \\
&\ -(1-q)F_\delta([x - \varepsilon, x + \varepsilon])\big] \\
\geq&\ qG((x - \varepsilon + \xi, x + \varepsilon - \xi)) \\
&\ -(1-q)G([x - \varepsilon, x + \varepsilon]) \quad (48)
\end{aligned}
$$

where the third inequality follows from the Portmanteau theorem applied to open and closed sets. By taking ξ to zero,

and applying both this asymptotic lower bound (48), and the asymptotic upper bound (47), we obtain the inequality:

$$qG((x - \varepsilon, x + \varepsilon)) - (1-q)G([x - \varepsilon, x + \varepsilon])$$
$$\leq (x + \varepsilon)(2\varepsilon)(1 - G(-\infty, x - \varepsilon)) \qquad (49)$$

By taking ε to zero, we see that $2\mu\Delta G(x) \leq 0$ which, when $\mu \neq 0$, implies that $\Delta G(x) = 0$. The same argument can be applied when $x < 0$. We conclude that G is continuous at x for all nonzero x when $\mu \neq 0$.

Using the fact that G is continuous for positive x, and the assumption that $x - \varepsilon > 0$, we can replace (49) with:

$$\mu G((x - \varepsilon, x + \varepsilon)) \leq 2\varepsilon(x + \varepsilon)(1 - G(x - \varepsilon)) \qquad (50)$$

We now wish to obtain a similar inequality, but a lower bound, rather than upper bound, to the left hand side of (50). Returning to (44), adding up the $2K - 1$ equations, but lower bounding the result, we obtain that for any $\xi > 0$, and for δ sufficiently small,

$$
\begin{aligned}
L_\delta >&\ (x - (K-1)\sqrt{\delta})\sqrt{\delta} \sum_{k=-(K-1)}^{K-1} (1 - G_\delta(x + k\sqrt{\delta})) \\
\geq&\ (x - (K-1)\sqrt{\delta})\sqrt{\delta} \\
&\ \left((2K-1) - \frac{\xi}{\sqrt{\delta}} - \sum_{k=-(K-2)}^{K-1} F_\delta(x + k\sqrt{\delta}) \right) \\
\geq&\ (x - (K-1)\sqrt{\delta})\sqrt{\delta} \\
&\ \left((2K-1) - \frac{\xi}{\sqrt{\delta}} - (2K-1)F_\delta(x + (K-1)\sqrt{\delta}) \right) \\
\geq&\ (x - \varepsilon)(2\varepsilon - \sqrt{\delta} - \xi - (2\varepsilon + \sqrt{\delta})F_\delta(x + \varepsilon)) \ (51)
\end{aligned}
$$

Using the continuity of G, and the Portmanteau theorem, we obtain that an asymptotic lower bound (as $\delta \downarrow 0$) to L_δ is given by

$$(x - \varepsilon)(2\varepsilon - \xi - 2\varepsilon G(x + \varepsilon)). \qquad (52)$$

But

$$
\begin{aligned}
\limsup_{\delta \downarrow 0} L_\delta & \qquad\qquad\qquad\qquad\qquad (53) \\
\leq&\ \limsup_{\delta \downarrow 0} \big[qF_\delta([x - \varepsilon - \sqrt{\delta}, x + \varepsilon])) \\
&\ -(1-q)F_\delta((x - \varepsilon + \sqrt{\delta}, x + \varepsilon - \sqrt{\delta}))\big] \\
\leq&\ qG([x - \varepsilon - \xi, x + \varepsilon]) - \\
&\ (1-q)G((x - \varepsilon + \xi, x + \varepsilon - \xi)) \quad (54) \\
=&\ qG((x - \varepsilon - \xi, x + \varepsilon)) - \\
&\ (1-q)G((x - \varepsilon + \xi, x + \varepsilon - \xi)) \quad (55)
\end{aligned}
$$

where the second last equality follows from the Portmanteau theorem, and the last inequality follows from the continuity of G. By applying both this asymptotic upper bound (55), and the asymptotic lower bound (52), and taking ξ to zero, we obtain the inequality, valid for $x - \varepsilon > 0$:

$$\mu G((x - \varepsilon, x + \varepsilon)) \geq 2\varepsilon(x - \varepsilon)(1 - G(x + \varepsilon)) \qquad (56)$$

Identical reasoning for $x + \varepsilon < 0$ provides the following two bounds, analogous to (50) and (56) respectively:

$$\mu G((x - \varepsilon, x + \varepsilon)) \leq 2\varepsilon(x + \varepsilon)G(x - \varepsilon) \qquad (57)$$

$$\mu G((x - \varepsilon, x + \varepsilon)) \geq 2\varepsilon(x - \varepsilon)G(x + \varepsilon) \qquad (58)$$

When $\mu \neq 0$, taking limits as ε tends to zero in (50), (56), (55) and (58) we obtain that G is differentiable at any $x \neq 0$, and satisfies the differential equations (15).

When $\mu = 0$, the proof of continuity for $x \neq 0$ no longer applies. However, (50), (54) still apply. For any $\varepsilon > 0$, there exists a $\hat{\varepsilon} \in (0, \varepsilon]$ such that

$$G((x - \hat{\varepsilon}, x + \hat{\varepsilon})) = G([x - \hat{\varepsilon}, x + \hat{\varepsilon}]),$$

since G has at most countably many jumps. Thus, by the limit of (54) as $\xi \to 0$, and by (52)

$$\begin{aligned} 0 &= \mu G((x - \hat{\varepsilon}, x + \hat{\varepsilon})) \\ &\geq 2\mu\hat{\varepsilon}(x - \hat{\varepsilon})(1 - G(x + \hat{\varepsilon})) \end{aligned}$$

which implies that $G(x + \hat{\varepsilon}) = 1$. Since $\varepsilon > 0$ can be arbitrarily small, this proves that $G(x) = 1$ for all $x > 0$. The same reasoning for $x + \epsilon < 0$, using (49) implies that $G(x) = 0$ for all $x < 0$. Thus, G is uniquely characterized by (16), since a distribution function is right continuous. ∎

REFERENCES

[1] L. L. H. Andrew, S. V. Hanly, S. Chan, and T. Cui, "Adaptive deterministic packet marking," *IEEE Commun. Lett.*, to appear. [Online]. Available: http://netlab.caltech.edu/~lachlan/abstract/ADPM_CL.pdf

[2] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[3] S. H. Low and D. E. Lapsley, "Optimization flow control I: Basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, no. 6, pp. 861–875, Dec. 1999.

[4] K. Ramakrishnan, S. Floyd, and D. Black, "The addition of explicit congestion notification (ECN) to IP," IETF, RFC 3168, Sept. 2001.

[5] S. Athuraliya, V. H. Li, S. H. Low, and Q. Yin, "REM: Active queue management," *IEEE Network*, vol. 15, no. 3, pp. 48–53, May/June 2001.

[6] M. Adler, J.-Y. Cai, J. K. Shapiro, and D. Towsley, "Estimation of congestion price using probabilistic packet marking," in *Proc. IEEE INFOCOM*, 2003, pp. 2068–2078.

[7] R. W. Thommes and M. J. Coates, "Deterministic packet marking for time-varying congestion price estimation," *IEEE/ACM Trans. Networking*, vol. 14, no. 3, pp. 592–602, June 2006.

[8] ——, "Deterministic packet marking for maximum link price estimation," in *Proc. Canadian Workshop on Information Theory*, Montreal, Canada, June 2005. [Online]. Available: http://www.tsp.ece.mcgill.ca/Networks/projects/pdf/thommes_CWIT05.pdf

[9] H.-K. Ryu and S. Chong, "Deterministic packet marking for max-min flow control," *IEEE Commun. Lett.*, vol. 9, no. 9, pp. 856–858, Sept. 2005.

[10] R. L. Tweedie, "Sufficient conditions for ergodicity and recurrence of Markov chains on a general state-space," *Stoch. Process. Appl.*, vol. 3, pp. 385–403, 1975.

[11] L. L. H. Andrew, K. Jacobsson, S. H. Low, M. Suchara, R. Witt, and B. P. Wydrowski, "Maxnet: Theory and implementation," Tech. Rep., 2006. [Online]. Available: http://netlab.caltech.edu/maxnet/MaxNet_Implementation_TechReport.pdf

[12] "WAN in Lab," July 2005. [Online]. Available: http://wil.cs.caltech.edu

[13] P. Billingsley, *Convergence of Probability Measures.* Wiley, 1999.