Statistical Analysis of Feedback-Synchronization Signaling Delay for Multicast Flow Control

Xi Zhang and Kang G. Shin

Real-Time Computing Laboratory
Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109-2122, USA
Email: {*xizhang,kgshin*}*@eecs.umich.edu*

Abstract— **Feedback signaling plays a crucial role in flow control because the traffic source relies on the signaling information to make *correct* and *timely* flow-control decisions. However, it is difficult to design an efficient signaling algorithm since each signaling message can tolerate neither error nor latency. Multicast flow-control signaling imposes two additional challenges: *scalability* and *feedback synchronization*. Previous research on multicast signaling has mainly focused on the design and implementation of algorithms without analyzing their delay performances. To remedy this deficiency, we developed a binary-tree deterministic model [1] and an *independent*-marking statistical model [2] to study the delay performance of various multicast feedback-synchronization signaling algorithms.**

In this paper, we consider the general case in which the congestion markings at different links are *dependent*. Including congestion-marking dependencies in the analysis is usually much harder than that under the independence assumption. However, the analysis without assuming independent markings can more accurately capture statistical characteristics for many practical cases where the congestion markings are not independent. Specifically, we develop a Markov chain model defined by the link-marking state on each path in a multicast tree. The Markov chain can not only characterize link-marking dependencies, but also yield a tractable analytical model. We also develop a Markov-chain dependency-degree model which can be used to quantify/evaluate *all possible* Markov-chain dependency degrees without knowing *a priori* the dependency-degree information. Using the Markov-chain and dependency-degree models, we derive the general expressions for the probability distribution of each path being the multicast-tree bottleneck. Also derived are the closed-form expressions for the first and second moments of multicast signaling delays. The modeling accuracy and analytical findings have been confirmed by simulations. The proposed Markov chain is also shown to asymptotically reach an equilibrium, and its limiting state distributions converge to the link-marking marginal probabilities when the Markov chain is irreducible.

By applying these two models, we analyze and contrast the feedback-delay scalability of two representative multicast signaling protocols: *Soft-Synchronization Protocol* (SSP) [1], [3] and *Hop-By-Hop* (HBH) signaling algorithms [4], [5], [6]. The proposed modeling techniques are generic, and thus can be applied not only to the multicast signaling-delay analysis, but also to other Markov-chain-based analyses abstracted from other applications.

Index Terms— **Multicast flow control, feedback-synchronization signaling, Soft-Synchronization Protocol (SSP), scalability, REM, RED, Markov chain.**

I. Introduction

A flow-control algorithm consists of two basic components: rate control and flow-control signaling. These two components are conceptually separate from the flow-control theory's standpoint, but are often blended together in most flow-control algorithms. Rate control adjusts the source rate to the variation of bandwidth available in the network. Flow-control signaling conveys the congestion and rate-control information between the source and network/receivers. Consequently, this signaling is crucial to flow control because the source relies solely on the signaling information in making correct and timely flow-control decisions. However, designing an efficient flow-control signaling protocol is difficult because the signaling messages, unlike data or audio/video traffic, can tolerate neither error nor latency. A signaling message could be useless or even harmful if it is not accurate or its delay is unbounded. The delivery of signaling traffic must therefore be timely and reliable. For example, in ATM ABR service, flow-control signaling relies on RM (Resource Management) cells, which convey the rate-control and congestion information among the source rate-controller, network switches, and the receiver.

Signaling for multicast flow control introduces two additional problems: *scalability* and *feedback-synchronization*. These two problems are closely interwoven in the signaling protocol for multicast flow control. First, simultaneous feedback arrivals from all branches can cause *feedback implosion* [7] at the source as well as at branch nodes, especially when the multicast tree is large. Hence, it is important for each branch node to consolidate the congestion-information feedbacks from its downstream paths and then forward only the consolidated feedback to its upstream node. Second, we need a feedback-synchronization signaling algorithm for each branch node to consolidate feedbacks, because they may arrive at significantly different times.

The first-generation feedback consolidation algorithms [4], [5], [6] for multicast ABR flow control employ a simple Hop-By-Hop (HBH) mechanism to deal with the feedback-implosion problem. On receipt of one forward RM cell, each branch node sends only one consolidated feedback RM cell upward by a *single* hop, ensuring that at each node, the ratio of feedback RM cells to forward RM cells is no larger than 1. To reduce the RM-cell roundtrip time (RTT) and improve multicast signaling accuracy, the authors of [8], [9] proposed a different feedback-synchronization algorithm by accumulating feedbacks from *all* branches of each branch node. The authors of [10] proposed an algorithm to speed up the transient response by sending fast congestion feedback without waiting for all branches' feedback during the transient phase.

In [1], [3], we proposed a feedback-synchronization signaling algorithm, called the *Soft-Synchronization Protocol* (SSP), which derives a single consolidated RM cell at each branch node from feedback RM cells of different downstream branches that are not necessarily responses to the same forward RM cell in each synchronization cycle. The SSP is shown to not only scale well with the multicast-tree topology, but also readily detect and remove non-responsive branches.

All of the above-referenced work only focused on the design and implementation of feedback-synchronization signaling algorithms,

The work reported in this paper was supported in part by the U.S. Office of Naval Research under Grant N00014-99-1-0465.

without addressing their delay performances. To remedy this deficiency, we developed a binary-tree deterministic model [1] and an *independent*-marking statistical model [2] to study the delay performance of various multicast feedback-synchronization signaling algorithms.

The independent-marking statistical model [2] builds on the recently-proposed Random Early Marking (REM) [11], [12], [13], [14], [15] and widely-cited Random Early Detection (RED) [16] flow-control schemes.[1] The REM and RED schemes — originally proposed for unicasts — can also be extended to multicast environments. Moreover, unicast and multicast transmissions usually co-exist in a network. In RED or REM, each router marks a packet's ECN (Explicit Congestion Notification) bit with a probability that is exponential in REM, or proportional in RED, to the average queue length at the output link.

The independent-marking statistical model is suitable for signaling delay analysis for multicast flow control based on REM- or RED-like schemes, where link-markings at different links/routers are assumed to be independent. However, there are also cases where link-markings are not independent. In such a case, the independent-marking algorithm and analysis can only offer approximate results, and their performance and accuracy will be affected by the "degree of dependency" between link-markings. This paper addresses the general case of *dependent* link congestion markings. Including dependence in the analysis usually is much more difficult than that under the independent-marking assumption. However, the analysis without the independence assumption can capture more accurate delay statistical characteristics.

We develop a Markov chain model for the links at different levels in a multicast tree, and model the dependency degree between different link congestion markings by using a dependency-degree factor. Using the proposed Markov chain and dependency-degree models, we derive the probability distribution for a path to be the multicast-tree bottleneck. We also derive the first and second moments of a multicast-tree RTT delay.

The benefits of our modeling and evaluation technique are twofold. First, the technique enables a direct quantitative comparison of feedback-synchronization delays between different multicast signaling schemes. Second, the proposed modeling technique establishes a general framework for evaluating the signaling delay of various feedback-synchronization-based multicast flow-control algorithms. Although our evaluation focuses on ATM ABR multicast flow control, the modeling techniques can be applied to any feedback-synchronization-based multicast flow-control algorithm, and to other Markov-chain-model-based analyses as well.

The paper is organized as follows. Section II overviews SSP and Section III presents the proposed binary-tree model and the deterministic signaling-delay analysis. In Section IV, we develop the Markov-chain model and apply it to derive the multicast bottleneck probability distributions. Section V proposes the dependency-degree model to measure and calculate the Markov-chain dependency. In Section VI, we derive expressions for various statistical and dynamic characteristics of multicast signaling delays. Section VII explores the asymptotical behavior of the derived Markov chain and its dependency-degree models. Section VIII describes numerical evaluation results. The paper concludes with Section IX.

[1] The analytical techniques developed in this paper are also applicable to cases where a link's random congestion state is caused by flow-control schemes other than REM and RED.

```
00. On receipt of a feedback RM cell from i-th branch:
01.  if (conn_patt_vec(i) ≠ 1) { ! Only process connected branches
02.    resp_branch_vec(i) := 1; ! Mark connected and responsive branch
03.    MCI := MCI ∨ CI; !  CI is randomly marked at this switch
04.    MER := min{MER, ER}; ! ER information processing
05.    if (conn_patt_vec ⊕ resp_branch_vec = 1) { ! soft synchronization.
06.      send RM cell (dir := back, ER := MER,
07.                    CI := MCI); ! Send fully-consolidated RM-cell upstream.
08.      no_resp_timer := N_nrt;   ! Reset non-responsive timer.
09.      resp_branch_vec := 0);   ! Reset responsive branch vector.
10.      MCI := 0;   MER := ER;}} ! Reset RM-cell control variable.
```

Fig. 1. Pseudocode for switch feedback synchronization algorithm.

II. Description of SSP

We briefly introduce SSP,[2] the multicast feedback-synchronization signaling algorithm [1], [3]. At the center of SSP is a pair of connection-update vectors: (i) $conn\_patt\_vec$, the connection pattern vector, where $conn\_patt\_vec(i)$ = 0 (1) indicates the i-th output port of the switch is (not) a downstream branch of the multicast connection. Thus, $conn\_patt\_vec(i)$ = 0 (1) implies that a data copy should (not) be sent to the i-th downstream branch and a feedback RM cell is (not) expected from the i-th downstream branch;[3] (ii) $resp\_branch\_vec$, the responsive branch vector is initialized to $\underline{0}$ and reset to $\underline{0}$ whenever a consolidated RM cell is sent upward from the switch. $resp\_branch\_vec(i)$ is set to 1 if a feedback RM cell is received from the i-th downstream branch.

A simplified pseudocode (detailed in [1], [3]) of the switch RM-cell processing algorithm is given in Fig. 1. On receipt of a returned feedback RM-cell, the switch first marks its corresponding bit in the $resp\_branch\_vec$ and then conducts RM-cell consolidation operations. If the modulo-2 addition (the soft-synchronization operation), $conn\_patt\_vec \oplus resp\_branch\_vec$ equals $\underline{1}$, an all 1's vector, indicating all feedback RM cells are synchronized, then a fully-consolidated feedback RM cell is generated and sent upward. But, if the modulo-2 addition is not equal to $\underline{1}$, the switch needs to await other feedback RM cells for synchronization. Notice that since the synchronization algorithm allows feedback RM cells corresponding to different forward RM cells to be consolidated, the feedback RM cells are "softly-synchronized" at branch nodes.

III. Deterministic Path-Delay Analysis in a Multicast Tree

A. The Binary-Tree Model

To simplify the analysis of RM-cell RTT, we *quantize* the feedback delay by assuming each switch-hop to have an identical delay (including processing and propagation delays). This assumption can be readily relaxed because the difference in switch processing delay and link-propagation delay of different switch-hops can be translated into different numbers of switch-hops with the same delay. We use the *hop-delay*, τ_h, which is the sum of the switch-processing delay and link-propagation delay taken in each hop, as the *time unit* in our delay analysis. To study the worst case and enable performance comparison, we only consider two types of multicast trees: *balanced* and *unbalanced binary* trees (see Fig. 2 for the case of tree height $m = 4$). Since we are only concerned with

[2] We briefly overview SSP here for completeness, and it is detailed in [1], [3].

[3] Note that the negative logic is used for convenience of implementation.

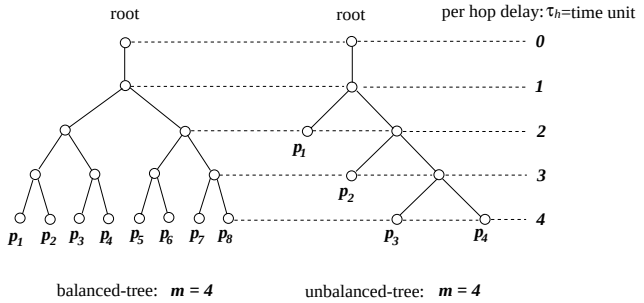


Fig. 2. Balanced and unbalanced binary multicast trees.

a path's RM-cell RTT in the multicast tree which is determined by its length, it suffices to consider binary trees only.

B. HBH and SSP Signaling Delay Properties for Each Path

Theorem 1: If an unbalanced[4] tree of height $m \geq 2$, as shown in Fig. 2, is flow-controlled by HBH with RM-cell update interval $\Delta \geq 1$ (τ_h), then the RM-cell RTT, denoted by $\tau_u(j, \Delta)$, of the j-th (counting from left to right) path, denoted by P_j, is given by:

$$\tau_u(j, \Delta) = 2 + j\ \Theta(\Delta), \tag{1}$$

where $\Theta(\Delta) \stackrel{\triangle}{=} \max\{2, \Delta\}$; $1 \leq \Delta \leq \tau_{max} = 2m$;[5] $1 \leq j \leq m-1$.

If the multicast tree is flow-controlled by SSP, then the following claims hold:

Claim 1. The number of P_j's feedback RM cells going through initial state is determined by:

$$k_j^* \stackrel{\triangle}{=} \max_{k \in \{0,1,\cdots\}} \{k \mid 2(m-j-1) - k\Delta \geq 0\}; \tag{2}$$

Claim 2. P_j's RM-cell RTT in steady state is determined by:

$$\tau_u(j, \Delta) = \tau_{max} - k_j^* \Delta. \tag{3}$$

where $1 \leq \Delta \leq \tau_{max} = 2m$ and $1 \leq j \leq m-1$.

Proof: The proof is provided in [1], [2]. ■

IV. The Markov-Chain Model for Dependent Congestion Markings

In random-marking schemes like REM/RED, and any other flow-control schemes, the marking/congestion state of a link is a function of its queue length. However, the queue lengths of different links carrying the same flows are generally not independent of each other. For instance, if a large (small) queue is built up at a congested upstream link in a multicast tree, the downstream links carrying the same flows are more likely to have large (small) queues.

For multicast flow control with *dependent* marking probabilities, we develop a Markov-chain model and a dependency-degree model for measuring and evaluating the degree of the Markov-chain dependency, in order to study the various statistical characteristics of multicast feedback-synchronization signaling delay. The proposed modeling techniques can not only be used to analyze the multicast signaling delay, but are also applicable to the general algorithm design/analysis for *both* multicast and unicast flow control.

[4]The formulas for RM-cell RTT of the balanced-tree cases are available in [2].

[5]Theorem 1 still holds for $\Delta > \tau_{max} = 2m$, but Δ is typically a fraction of the maximum RM-cell RTT $\tau_{max} = 2m$.

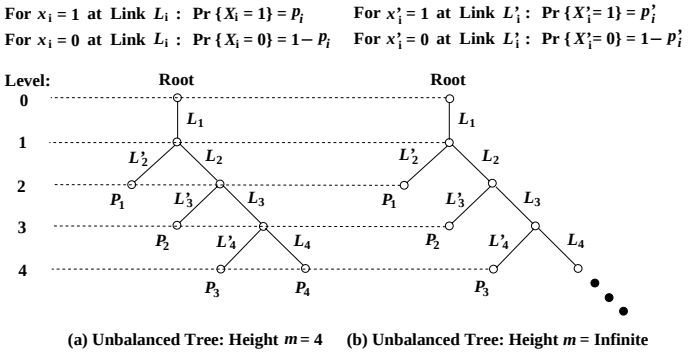


Fig. 3. Dependent random-marking unbalanced binary-tree model.

A. The Dependent Statistical Model

To analyze the multicast feedback-synchronization signaling delay with dependent marking probabilities, we introduce the following definition.

Definition 1: A ***dependent*** random-marking unbalanced binary tree of height m consists of a set, $\mathcal{L}$, of links which satisfy the following four conditions:

C1. All links in $\mathcal{L}$ are labeled as shown in Fig. 3(a) for $m < \infty$ (e.g., $m = 4$) and Fig. 3(b) for $m \to \infty$, respectively, such that

$$\mathcal{L} \stackrel{\triangle}{=} \begin{cases} \{L_1, L'_2, L_2, L'_3, L_3, \cdots, L'_m, L_m\}, & \text{if } m < \infty; \\ \{L_1, L'_2, L_2, L'_3, L_3, \cdots, L_\infty, L_\infty\}, & \text{if } m \to \infty; \end{cases} \tag{4}$$

which contains m paths, $P_1, P_2, \cdots, P_m$, as shown in Fig. 3. Denote each path by its component links as:

$$\begin{cases} P_k \stackrel{\triangle}{=} \{L_1, L_2, \cdots, L_k, L'_{k+1}\}, & \text{if } 1 \leq k \leq m-1; \\ P_m \stackrel{\triangle}{=} \{L_1, L_2, \cdots, L_m\}, & \text{if } k = m; \end{cases} \tag{5}$$

We define P_m as the *main-stream path* which takes only right-branch links at all branch nodes, and define each P_k, for $1 \leq k \leq m-1$, as a *branch-stream path* which consists of k right-branch links and one left-branch link at the last branch node (see Fig. 3). Links L_i and L'_i, $\forall i \geq 2$, are at the same level of the multicast tree.

C2. The marking state of link L_i (L'_i) ($i = 1, 2, \cdots$) is represented by a random variable X_i (X'_i) which takes value in $\{0, 1\}$ such that (see Fig. 3)

$$\Pr\{X_i = x_i\} = \begin{cases} p_i, & \text{for } x_i = 1; \\ 1 - p_i, & \text{for } x_i = 0; \end{cases} \tag{6}$$

$$\Pr\{X'_i = x'_i\} = \begin{cases} p'_i, & \text{for } x'_i = 1; \\ 1 - p'_i, & \text{for } x'_i = 0; \end{cases} \tag{7}$$

where p_i (p'_i) is the marking probability for L_i (L'_i) and is determined by

$$p_i = \begin{cases} 1 - \phi^{-\gamma \overline{q}_i}, & \text{if REM is used;} \\ \left[\dfrac{\overline{q}_i - min_{th}}{max_{th} - min_{th}}\right] p_{max}, & \text{if RED is used;} \end{cases} \tag{8}$$

$$p'_i = \begin{cases} 1 - \phi^{-\gamma \overline{q'}_i}, & \text{if REM is used;} \\ \left[\dfrac{\overline{q'}_i - min_{th}}{max_{th} - min_{th}}\right] p_{max}, & \text{if RED is used;} \end{cases} \tag{9}$$

where $0 < p_i, p'_i < 1$; $\overline{q}_i$ ($\overline{q'}_i$) is the average queue size at L_i (L'_i), $\gamma > 0$ is the step size, $\phi > 1$ for REM; p_{max} is the maximum marking probability, max_{th} (min_{th}) is the high (low) queue threshold for RED;

C3. The congestion marking states of all links at different multicast-tree levels are *dependent* and satisfy the following properties:

$$\begin{aligned}\Pr\{X_i = x_i \mid X_{i-1} = x_{i-1}, X'_{i-1} = x'_{i-1}, \\ X_{i-2} = x_{i-2}, X'_{i-2} = x'_{i-2}, \cdots, X_1 = x_1\} \\ = \Pr\{X_i = x_i \mid X_{i-1} = x_{i-1}\}; \qquad (10)\end{aligned}$$

$$\begin{aligned}\Pr\{X'_i = x'_i \mid X_{i-1} = x_{i-1}, X'_{i-1} = x'_{i-1}, \\ X_{i-2} = x_{i-2}, X'_{i-2} = x'_{i-2}, \cdots, X_1 = x_1\} \\ = \Pr\{X'_i = x'_i \mid X_{i-1} = x_{i-1}\} \qquad (11)\end{aligned}$$

C4. The congestion marking states within the same level are also *dependent* and satisfy the following properties:

$$\begin{aligned}\Pr\{X_i = x_i \mid X'_i = x'_i, X_{i-1} = x_{i-1}, X'_{i-1} = x'_{i-1}, \\ X_{i-2} = x_{i-2}, X'_{i-2} = x'_{i-2}, \cdots, X_1 = x_1\} \\ = \Pr\{X_i = x_i \mid X_{i-1} = x_{i-1}\}; \qquad (12)\end{aligned}$$

$$\begin{aligned}\Pr\{X'_i = x'_i \mid X_i = x_i, X_{i-1} = x_{i-1}, X'_{i-1} = x'_{i-1}, \\ X_{i-2} = x_{i-2}, X'_{i-2} = x'_{i-2}, \cdots, X_1 = x_1\} \\ = \Pr\{X'_i = x'_i \mid X_{i-1} = x_{i-1}\}. \qquad (13)\end{aligned}$$

where $2 \le i \le m$. ■

Remarks on Definition 1 (C3 and C4): We only consider the *upstream* and *same-level dependence* of link marking states as described by Eqs. (10), (11), (12), and (13), because the multicast-tree signaling delay analysis to be developed below need not consider the *downstream* dependence. The congestion information on the links above the immediate-next upstream link or on the link at the same level (see **C4**) is all concentrated into, and carried over by, the given congestion information on the immediate-next upstream link. Conditions **C3** and **C4** are reasonable because one link's congestion state depends most on its immediate upstream link's congestion state. The upstream's influence on a downstream link's congestion state propagates through its immediate upstream link which carries same flows, and thus, as long as the immediate upstream link's congestion state is given, the probability distribution at the downstream link is independent of the congestion state at links which are located above the immediate upstream link or at the same level as indicated by condition **C4** in Eqs. (12) and (13).

B. Probability Distribution of the Multicast-Tree Bottleneck Path

To ensure reliable data transmission, the multicast ABR service needs to adjust the source rate to the minimum available bandwidth share on the most congested path that is currently sensed (through feedback) by the source [3]. Clearly, based on the OR rule (see the multicast signaling algorithms given in Fig. 1, and the detailed versions in [1], [3]), among all existing bottleneck paths in a multicast tree, the *shortest* bottleneck path dominates the source's flow-control decisions and the RTT of flow-control feedback loop. To explicitly model this feature, we introduce the following definition.

Definition 2: Among all concurrent bottleneck paths sensed by the source in a multicast tree, the bottleneck path of the *minimum* length is called the *dominant bottleneck path* (also called *multicast-tree bottleneck path*), and its RM-cell RTT is called the *multicast-tree bottleneck RM-cell RTT* or simply *multicast-tree RTT*. ■

Based on Definitions 1 and 2, the following proposition lays a foundation for deriving the distribution of the dominant bottleneck path.

Proposition 1: The sequence of random marking states $\{X_1, X_2, \cdots, X_{m-1}, X_m\}$ (for the tree height $m < \infty$ and $m \to \infty$) in Definition 1 defines a 2-state discrete-indexed Markov chain over the links on the main-stream path $P_m = \{L_1, L_2, \cdots, L_m\}$, and the sequence of marking states $\{X_1, X_2, \cdots, X_k, X'_{k+1}\}$ in Definition 1 on each branch-stream path $P_k = \{L_1, L_2, \cdots, L_k, L'_{k+1}\}$, for $k = 1, 2, \cdots, m-1$, also define a 2-state discrete-indexed (finite-sequence) Markov chain.

Proof: The proof follows from Definition 1. ■

Remarks on Proposition 1: Unlike the traditional definition of Markov chain/process where the random-variable sequence index set is time, we define the Markov chain for every path (including the main- and branch-stream paths) which is indexed by the (discrete) link sequence number associated with that path.

Since the mathematical properties/treatments and random marking definitions for both the Markov chain defined over the main-stream path P_m and the Markov chain defined over the branch-stream paths P_k $(k = 1, 2, \cdots, m-1)$ are the same, except that the last link's marking state differs in labeling by a "$'$" symbol (see Proposition 1), we will henceforth use $\{X_i\}$ to represent the Markov chain defined over both the main- and branch-stream paths, and explicitly specify it wherever necessary.

Applying Proposition 1, the following theorem derives the probability distributions of the dominant-bottleneck path.

Theorem 2: If a *dependent*-marking multicast tree of height m as defined in Definition 1 is flow-controlled under SSP or HBH, then the following claims hold:

Claim 1. If $m \to \infty$, then there exists *one and only one* dominant bottleneck path, and the probability distribution, denoted by $\psi_d(P_k, \infty)$, that P_k becomes the dominant bottleneck path, is determined by

$$\psi_d(P_k, \infty) = \begin{cases} 1 - \Pr\{X_1 = 0\}\Pr\{X'_2 = 0 \mid X_1 = 0\}, & \text{if } k = 1; \\ \Pr\{X_1 = 0\}\Pr\{X'_k = 0 \mid X_{k-1} = 0\} \\ \quad \cdot \Big[\Pr\{X_k = 1 \mid X_{k-1} = 0\} \\ \quad\quad + \Pr\{X_k = 0 \mid X_{k-1} = 0\} \\ \quad\quad\quad \cdot \Pr\{X'_{k+1} = 1 \mid X_k = 0\}\Big] \\ \quad \cdot \prod_{i=1}^{k-2} \Big\{\Pr\{X_{i+1} = 0 \mid X_i = 0\} \\ \quad\quad \cdot \Pr\{X'_{i+1} = 0 \mid X_i = 0\}\Big\}, & \text{if } k \ge 2; \end{cases} \qquad (14)$$

And, $\psi_d(P_k, \infty)$ given in Eq. (14) satisfies:

$$\lim_{m \to \infty} \sum_{k=1}^{m} \psi_d(P_k, \infty) = 1; \qquad (15)$$

Claim 2. If $m < \infty$, then there exists *at most one* dominant bottleneck path, and the probability distribution, $\psi_d(P_k, m)$, that P_k becomes the dominant bottleneck path, is given by

$$\psi_d(P_k,m) = \begin{cases} 1-\Pr\{X_1=0\}\Pr\{X_2'=0 \mid X_1=0\}, \text{ if } k=1; \\ \Pr\{X_1=0\}\Pr\{X_k'=0 \mid X_{k-1}=0\} \\ \quad \cdot\Big[\Pr\{X_k=1 \mid X_{k-1}=0\} \\ \quad\quad +\Pr\{X_k=0 \mid X_{k-1}=0\} \\ \quad\quad\quad \cdot\Pr\{X_{k+1}'=1 \mid X_k=0\}\Big] \\ \quad \cdot\prod_{i=1}^{k-2}\Big\{\Pr\{X_{i+1}=0 \mid X_i=0\} \\ \quad\quad \cdot\Pr\{X_{i+1}'=0 \mid X_i=0\}\Big\}, \quad \text{if } k\geq 2; \\ \Pr\{X_1=0\}\Pr\{X_m=1 \mid X_{m-1}=0\} \\ \quad \cdot\Pr\{X_m'=0 \mid X_{m-1}=0\} \\ \quad \cdot\prod_{i=1}^{m-2}\Big\{\Pr\{X_{i+1}=0 \mid X_i=0\} \\ \quad\quad \cdot\Pr\{X_{i+1}'=0 \mid X_i=0\}\Big\}, \text{ if } k=m; \end{cases} \tag{16}$$

Proof: The proof is detailed in [17]. ■

Remarks on Theorem 2: We observe that by Eq. (14), $\lim_{k\to\infty}\psi_d(P_k,\infty)=0$. This is expected, since a longer bottleneck path is always dominated by a co-existing shorter bottleneck path, if any. Thus, when $k\to\infty$ as $m\to\infty$, P_∞ is always dominated by a shorter bottleneck path for $0<p_i,p_i'<1$, $i=1,2,\cdots,\infty$. That is, $\psi_d(P_\infty,\infty)=0$. In addition, notice that by Eq. (15) we have $\sum_{k=1}^{\infty}\psi_d(P_k,\infty)=1$, which also makes sense because as the unbalanced-tree's height $m\to\infty$ and $0<p_i,p_i'<1$, there always exists (with probability 1) one and only one dominant bottleneck path in a multicast tree. On the other hand, for the case of $m<\infty$, by Eqs. (16) and (15) we have $\sum_{k=1}^{m}\psi_d(P_k,m)\leq 1$, implying the possibility that there is no dominant bottleneck path in the multicast tree of height $m<\infty$. This is also expected because $0<p_i,p_i'<1$.

V. Markov-Chain Dependency-Degree Modeling

To use Eqs. (14) and (16), we need to derive explicit expressions for $\Pr\{X_i=x_i \mid X_{i-1}=x_{i-1}\}$ and $\Pr\{X_i'=x_i' \mid X_{i-1}=x_{i-1}\}$ used in the Eqs. (14) and (16). However, it is difficult to know/compute the accurate dependency between two random variables. To solve this problem, we propose to use a real-valued *dependency-degree factor* $\alpha\in[0,1]$ to quantify all possible degrees of dependency between the random variables in the Markov chain's one-step transition probabilities. Using this dependency-degree factor, one can evaluate any possible degree of dependency ranging from ***independent*** to ***perfectly dependent***, without knowing *a priori* the dependency degree of two random variables.

In general, two dependent random events can affect each other either *positively* or *negatively*. For instance, if occurrence of one event is likely to trigger another event, then they are said to be *positively dependent*. On the other hand, if occurrence of one event makes another event unlikely to occur, then they are said to be *negatively dependent*. As we discussed earlier, an upstream link's being congested (uncongested) state will make the downstream links carrying the same flows more likely (unlikely) to be congested. So, the positive dependence can more accurately characterize the dependence behavior of link markings. To quantitatively describe this feature, we introduce the following definition:

Definition 3: Two dependent link marking states X_i and X_{i+1} are said to be positively (negatively) dependent if $\Pr\{X_{i+1}=x \mid X_i=x\} > \Pr\{X_{i+1}=x \mid X_i=\overline{x}\}$ ($\Pr\{X_{i+1}=x \mid X_i=x\} < \Pr\{X_{i+1}=x \mid X_i=\overline{x}\}$), where $x\in\{0,1\}$. ■

Based on Definition 3, Theorem 3 models the dependency-degree between two random variables. Notice that Theorem 3 gives the results only for the case of $\Pr\{X_{i+1}=x_{i+1} \mid X_i=x_i\}$ and $\Pr\{X_i=1\}=p_i$, and it also holds for the case of $\Pr\{X_{i+1}'=x_{i+1}' \mid X_i=x_i\}$ and $\Pr\{X_i'=1\}=p_i'$ with the similar results that we omitted.

Theorem 3: Consider the Markov chain $\{X_i\}$ defined on link marking states on every path (for both main-stream and branch-stream) in the multicast tree specified by Definition 1. If Markov chain $\{X_i\}$ is *positively dependent*, and the link marking probability is equal to $\Pr\{X_i=1\}=p_i$, then the following claims hold:

Claim 1. The conditional distribution $\Pr\{X_{i+1}=x_{i+1} \mid X_i=x_i\}$, with $x_i, x_{i+1}\in\{0,1\}$, is upper- and lower-bounded by

$$1-p_{i+1} \leq \Pr\{X_{i+1}=0 \mid X_i=0\} \leq \begin{cases} 1, & \text{if } p_i\geq p_{i+1}; \\ \dfrac{1-p_{i+1}}{1-p_i}, & \text{if } p_i<p_{i+1}; \end{cases} \tag{17}$$

$$p_{i+1} \geq \Pr\{X_{i+1}=1 \mid X_i=0\} \geq \begin{cases} 0, & \text{if } p_i\geq p_{i+1}; \\ \dfrac{p_{i+1}-p_i}{1-p_i}, & \text{if } p_i<p_{i+1}; \end{cases} \tag{18}$$

$$1-p_{i+1} \geq \Pr\{X_{i+1}=0 \mid X_i=1\} \geq \begin{cases} \dfrac{p_i-p_{i+1}}{p_i}, & \text{if } p_i\geq p_{i+1}; \\ 0, & \text{if } p_i<p_{i+1}; \end{cases} \tag{19}$$

$$p_{i+1} \leq \Pr\{X_{i+1}=1 \mid X_i=1\} \leq \begin{cases} \dfrac{p_{i+1}}{p_i}, & \text{if } p_i\geq p_{i+1}; \\ 1, & \text{if } p_i<p_{i+1}; \end{cases} \tag{20}$$

Claim 2. $\exists\,\alpha_i(\alpha_i')\in[0,1]$ such that all possible dependency-degrees between X_i and X_{i+1} (X_{i+1}') can be measured by the real-valued ***dependency-degree factor*** α_i (α_i'), and [6]

$$\begin{cases} \alpha_i=0 \ \textit{iff}\ X_i \text{ and } X_{i+1} \text{ are independent;} \\ \alpha_i=1 \ \textit{iff}\ X_i \text{ and } X_{i+1} \text{ are } \textit{perfectly} \text{ dependent;} \end{cases} \tag{21}$$

and

$$\begin{cases} \alpha_i'=0 \ \textit{iff}\ X_i \text{ and } X_{i+1}' \text{ are independent;} \\ \alpha_i'=1 \ \textit{iff}\ X_i \text{ and } X_{i+1}' \text{ are } \textit{perfectly} \text{ dependent;} \end{cases} \tag{22}$$

Claim 3. The conditional distributions $\Pr\{X_{i+1}=x_{i+1} \mid X_i=x_i\}$, with $x_i, x_{i+1}\in\{0,1\}$, are determined by

$$\Pr\{X_{i+1}=0 \mid X_i=0\} =$$

[6] Examples of the *perfectly* dependent events discussed below include that two events are identical or one event is a sub-event of the other, see [17] for details.

$$\begin{cases} 1-(1-\alpha_i)\,p_{i+1}, & \text{if } p_i \geq p_{i+1}; \\ (1-\alpha_i)(1-p_{i+1}) + \alpha_i \left(\dfrac{1-p_{i+1}}{1-p_i}\right), & \text{if } p_i < p_{i+1}; \end{cases} \tag{23}$$

$$\Pr\{X_{i+1}=1 \mid X_i=0\} =$$

$$\begin{cases} (1-\alpha_i)\,p_{i+1}, & \text{if } p_i \geq p_{i+1}; \\ (1-\alpha_i)\,p_{i+1} + \alpha_i \left(\dfrac{p_{i+1}-p_i}{1-p_i}\right), & \text{if } p_i < p_{i+1}; \end{cases} \tag{24}$$

$$\Pr\{X_{i+1}=0 \mid X_i=1\} =$$

$$\begin{cases} (1-\alpha_i)(1-p_{i+1}) + \alpha_i \left(\dfrac{p_i-p_{i+1}}{p_i}\right), & \text{if } p_i \geq p_{i+1}; \\ (1-\alpha_i)(1-p_{i+1}), & \text{if } p_i < p_{i+1}; \end{cases} \tag{25}$$

$$\Pr\{X_{i+1}=1 \mid X_i=1\} =$$

$$\begin{cases} (1-\alpha_i)p_{i+1} + \alpha_i \left(\dfrac{p_{i+1}}{p_i}\right), & \text{if } p_i \geq p_{i+1}; \\ p_{i+1} + \alpha_i(1-p_{i+1}), & \text{if } p_i < p_{i+1}; \end{cases} \tag{26}$$

where $i = 1, 2, \cdots$, and α_i is defined in Claim 2.

Proof: The proof is available in [17]. ■

Remarks on Theorem 3: Claim 1 finds the upper and lower bounds of all 4 possible 2-state Markov chain one-step transition probabilities as functions of the marginal link-marking probabilities p_i and p_{i+1} given by networks. Claim 2 ensures the existence of a real-valued dependence-degree factor $\alpha_i \in [0,1]$. It also proves the completeness of the dependency-degree factor modeling by mapping all possible degrees of dependency onto the real-valued point set $[0,1]$. Claim 3 derives expressions for all 4 possible 2-state Markov chain one-step transition probabilities.

Applying Theorem 3 and Eqs. (23) and (24) to Theorem 2, we obtain the general-case (heterogeneous) expressions for calculating the multicast bottleneck path probability distributions, which are summarized in the following corollary.

Corollary 1: Let a dependent-marking multicast tree of height m as defined in Definition 1 be flow-controlled under SSP or HBH. If the one-step transition probability of the Markov chain $\{X_i\}$ defined over every path (including the main- and branch-stream paths) is specified by the dependency-factor vector $\vec{\alpha} \triangleq (\alpha_1, \alpha_1', \alpha_2, \alpha_2', \alpha_3, \alpha_3', \cdots)$ which is derived in Theorem 3, and further, denote the link marking probability vector by $\vec{p} \triangleq (p1, p_1', p_2, p_2', p_3, p_3', \cdots)$, then the following claims hold.

Claim 1. If $m \to \infty$, then there exists *one and only one* dominant bottleneck path, and the probability distribution, denoted by $\psi_d(P_k, \vec{\alpha}, \vec{p}, \infty)$, that P_k becomes the dominant bottleneck path, is determined by

$$\psi_d(P_k, \vec{\alpha}, \vec{p}, \infty) =$$

$$\begin{cases} 1-(1-p_1)\left[1-(1-\alpha_1')p_2'\right], & \text{if } k=1; \\ (1-p_1)\left[1-(1-\alpha_{k-1}')p_k'\right]\Big[(1-\alpha_{k-1})p_k \\ \quad +\left[1-(1-\alpha_{k-1})p_k\right](1-\alpha_k')p_{k+1}'\Big] \\ \quad \cdot \prod_{i=1}^{k-2}\Big\{\left[1-(1-\alpha_i)p_{i+1}\right] \\ \quad\quad \cdot\left[1-(1-\alpha_i')p_{i+1}'\right]\Big\}, & \text{if } k \geq 2; \end{cases} \tag{27}$$

and, $\psi_d(P_k, \infty)$ given in Eq. (27) satisfies:

$$\lim_{m\to\infty} \sum_{k=1}^{m} \psi_d(P_k, \vec{\alpha}, \vec{p}, \infty) = 1; \tag{28}$$

Claim 2. If $m < \infty$, then there exists *at most one* dominant bottleneck path, and the probability distribution, denoted by $\psi_d(P_k, \vec{\alpha}, \vec{p}, m)$, that P_k becomes the dominant bottleneck path, is determined by

$$\psi_d(P_k, \vec{\alpha}, \vec{p}, m) =$$

$$\begin{cases} 1-(1-p_1)\left[1-(1-\alpha_1')p_2'\right], & \text{if } k=1; \\ (1-p_1)\left[1-(1-\alpha_{k-1}')p_k'\right]\Big[(1-\alpha_{k-1})p_k \\ \quad +\left[1-(1-\alpha_{k-1})p_k\right](1-\alpha_k')p_{k+1}'\Big] \\ \quad \cdot \prod_{i=1}^{k-2}\Big\{\left[1-(1-\alpha_i)p_{i+1}\right] \\ \quad\quad \cdot\left[1-(1-\alpha_i')p_{i+1}'\right]\Big\}, & \text{if } k \geq 2; \\ (1-p_1)(1-\alpha_{m-1})p_m\left[1-(1-\alpha_{m-1}')p_m'\right] \\ \quad \cdot \prod_{i=1}^{m-2}\Big\{\left[1-(1-\alpha_i)p_{i+1}\right] \\ \quad\quad \cdot\left[1-(1-\alpha_i')p_{i+1}'\right]\Big\}, & \text{if } k = m; \end{cases} \tag{29}$$

Proof: The proof follows by plugging Eqs. (23) through (26) of Theorem 3 into Eqs. (14), (15) and (16) of Theorem 2. ■

Remarks on Corollary 1: We can use Eqs. (27) and (29), and tune up the dependence-degree factor $\vec{\alpha}$ to see how the system performs with different dependency degrees. More importantly, the completeness of the dependency-degree factor model derived in Theorem 3 guarantees that the actual unknown degree of dependency imposed by the practical problems can always be covered by tuning α in the interval [0,1]. Moreover, Eqs. (27) and (29) provide very general probability distribution expressions since one can arbitrarily select $\vec{\alpha}$ and $\vec{p}$ for different links to handle the heterogeneity. Eqs. (27) and (29) reduce to the probability distribution expressions of $\psi(P_k, m)$ derived for the multicast signaling delay analysis under *independent* random-marking [2] by letting $\vec{\alpha} = \vec{0}$ (independent), verifying the correctness of Eqs. (27) and (29).

VI. Statistical and Dynamic Properties of Multicast Signaling Delays under Dependent Markings

Using the probability distribution derived in Corollary 1 and Eqs. (1) and (3) of $\tau_u(j, \Delta)$ derived in Theorem 1, the following theorem derives the probability distributions, their dynamic properties, and the means and variances of multicast signaling delays under SSP and HBH for the homogeneous case and $m < \infty$.

Theorem 4: Let a dependent-marking multicast tree of height m as defined in Definition 1 be flow-controlled under SSP and HBH,

respectively, with the RM-cell interval Δ. If $m < \infty$, $0 < p_i = p_i' = p < 1$ and $0 \le \alpha_i = \alpha_i' = \alpha \le 1$, $\forall i$ (the homogeneous case),[7] then the following claims hold:

Claim 1. The probability distribution that P_k becomes the dominant bottleneck path, denoted by $\psi_d(P_k,\, \alpha, p, m)$, is determined by

$$\psi_d(P_k,\alpha,p,m) = \begin{cases} 1-(1-p)\left[1-(1-\alpha)p\right], & \text{if } k=1; \\ (1-\alpha)(1-p)p\left[2-(1-\alpha)p\right] & \\ \qquad \cdot\left[1-(1-\alpha)p\right]^{2k-3}, & \text{if } k \ge 2; \\ (1-\alpha)(1-p)p\left[1-(1-\alpha)p\right]^{2m-3}, & \text{if } k=m; \end{cases} \tag{30}$$

Claim 2. For each path P_k and a given α, $\psi_d(P_k,\alpha,p,m)$ attains the *unique* maximum at

$$p^* \triangleq \arg \max_{0<p<1} \psi_d(P_k,\alpha,p,m) = \begin{cases} 1, & \text{if } k=1; \\ \frac{m-(m-1)\alpha-\sqrt{(m-(m-1)\alpha)^2-(1-\alpha)(2m-1)}}{(1-\alpha)(2m-1)}, & \text{if } k=m; \end{cases} \tag{31}$$

and for $2 \le k \le (m-1)$, p^* is non-negative and no larger than 1 real-valued root of the following cubic equation:

$$2k(1-\alpha)^2p^3 + (1-\alpha)\left[(2k-1)\alpha - 6k\right]p^2 - 2\left[(2k-1)\alpha - 2k - 1\right]p - 2 = 0. \tag{32}$$

Claim 3. For each path P_k and a given p, $\psi_d(P_k,\alpha,p,m)$ attains the *unique* maximum at

$$\alpha^* \triangleq \arg \max_{0<\alpha<1} \psi_d(P_k,\alpha,p,m) = \begin{cases} \frac{p-1}{p} + \frac{1}{p}\sqrt{1-\frac{2}{2k-1}}, & \text{if } 2 \le k \le m-1 \text{ and } k \ge \left\lceil \frac{1}{2} + \frac{1}{p(2-p)} \right\rceil; \\ 1 - \frac{1}{2(m-1)p}, & \text{if } k=m \text{ and } k \ge \left\lceil 1 + \frac{1}{2p} \right\rceil; \end{cases} \tag{33}$$

Claim 4. If Markov chain dependency-factor $\alpha = \alpha_0 > 0$ for a given α_0, it shifts the probability distribution of multicast-tree bottleneck path from shorter paths to longer ones. If the tree height m satisfies:

$$m \ge \left\lfloor \frac{\log\sqrt{\frac{1}{1-\alpha_0}}}{\log\frac{1-(1-\alpha_0)p}{1-p}} + 2.5 \right\rfloor, \tag{34}$$

then there exists the *unique* "dependency-balanced path" $P_{\widetilde{k}}$ such that $2 \le \widetilde{k} \le m-1$ and

$$\begin{cases} \psi_d(P_k,\alpha,p,m)\,|_{\,\alpha=0} \ge \psi_d(P_k,\alpha,p,m)\,|_{\,\alpha=\alpha_0}, & \text{if } k \le \widetilde{k}; \\ \psi_d(P_k,\alpha,p,m)\,|_{\,\alpha=0} < \psi_d(P_k,\alpha,p,m)\,|_{\,\alpha=\alpha_0}, & \text{if } k > \widetilde{k}; \end{cases} \tag{35}$$

[7] The analytical results derived from the homogeneous case can be easily extended to the heterogeneous case where p_i and α_i are different $\forall i$.

where $\psi_d(P_k,\alpha,p,m)$ is given by Eq. (30), and the "dependency-balanced path number" $\widetilde{k}$ is determined by

$$\widetilde{k} = \left\lfloor \frac{\log\sqrt{\frac{2-p}{(1-\alpha_0)[2-(1-\alpha_0)p]}}}{\log\frac{1-(1-\alpha_0)p}{1-p}} + 1.5 \right\rfloor; \tag{36}$$

Claim 5. The means of multicast-tree bottleneck RM-cell RTT, denoted by $\overline{\tau}_{SSP}(\alpha,m)$ and $\overline{\tau}_{HBH}(\alpha,m)$ for the SSP and HBH schemes, respectively, are determined by:

$$\begin{aligned} \overline{\tau}_{SSP}(\alpha,m) = {} & \left[p+(1-\alpha)(p-p^2)\right]\left[2m - \left\lfloor\frac{2(m-2)}{\Delta}\right\rfloor\Delta\right] \\ & + 2m(1-p)\left[1-(1-\alpha)p\right]\Big\{1+\left[1-(1-\alpha)p\right]^{2(m-2)} \\ & \cdot\left[(1-\alpha)p-1\right]\Big\} - (1-\alpha)(1-p)p\left[2-(1-\alpha)p\right]\Delta \\ & \cdot\sum_{k=2}^{m-1}\left\{\left\lfloor\frac{2(m-k-1)}{\Delta}\right\rfloor\left[1-(1-\alpha)p\right]^{2k-3}\right\}, \end{aligned} \tag{37}$$

$$\begin{aligned} \overline{\tau}_{HBH}(\alpha,m) = {} & \frac{(1-p)\Theta(\Delta)}{(1-\alpha)p\left[2-(1-\alpha)p\right]}\Big\{2\left[1-(1-\alpha)p\right] \\ & - \left[1-(1-\alpha)p\right]^3 - m\left[1-(1-\alpha)p\right]^{2m-3} + (m-1) \\ & \cdot\left[1-(1-\alpha)p\right]^{2m-1}\Big\} + (1-p)\left[1-(1-\alpha)p\right]^{2m-3} \\ & \cdot\Big\{(1-\alpha)p\left[2+(m-1)\Theta(\Delta)\right]-2\Big\} + (2+\Theta(\Delta)) \\ & \cdot\left[p+(1-\alpha)(p-p^2)\right] + 2(1-p)\left[1-(1-\alpha)p\right]; \end{aligned} \tag{38}$$

where $\Theta(\Delta)$ is defined in Theorem 1.

Claim 6. The variances of multicast-tree bottleneck RM-cell RTT, denoted by $\sigma^2_{SSP}(\alpha,m)$ and $\sigma^2_{HBH}(\alpha,m)$ for the SSP and HBH schemes, respectively, are determined by:

$$\begin{aligned} \sigma^2_{SSP}(\alpha,m) = {} & 4m^2 + 4m^2(1-p)\left[1-(1-\alpha)p\right]^{2m-3} \\ & \cdot\left[(1-\alpha)p-1\right] - (1-\alpha)(1-p)p\left[2-(1-\alpha)p\right] \\ & \cdot\Bigg\{4m\Delta\sum_{k=2}^{m-1}\left\{\left\lfloor\frac{2(m-k-1)}{\Delta}\right\rfloor\left(1-(1-\alpha)p\right)^{2k-3)}\right\} \\ & -\Delta^2\sum_{k=2}^{m-1}\left\{\left\lfloor\frac{2(m-k-1)}{\Delta}\right\rfloor^2\left(1-(1-\alpha)p\right)^{2k-3)}\right\}\Bigg\} \\ & + p\Big[1+(1-\alpha)(1-p)\Big]\Bigg\{\Delta^2\left\lfloor\frac{2(m-2)}{\Delta}\right\rfloor^2 - 4m\Delta \\ & \cdot\left\lfloor\frac{2(m-2)}{\Delta}\right\rfloor\Bigg\} - \overline{\tau}^2_{SSP}(\alpha,m), \end{aligned} \tag{39}$$

$$\begin{aligned} \sigma^2_{HBH}(\alpha,m) = {} & \left[1+(1-\alpha)(1-p)\right]p(2+\Theta(\Delta))^2 + (1-\alpha) \\ & \cdot(1-p)p\left[1-(1-\alpha)p\right]^{2m-3}\left[2+(m-1)\Theta(\Delta)\right]^2 \\ & + 4(1-p)\left[1-(1-\alpha)p\right]\Big\{1-\left[1-(1-\alpha)p\right]^{2(m-2)}\Big\} \\ & + \frac{4\Theta(\Delta)(1-p)\left[1-(1-\alpha)p\right]}{(1-\alpha)p\left[2-(1-\alpha)p\right]}\Big\{2-\left[1-(1-\alpha)p\right]^2 \end{aligned}$$

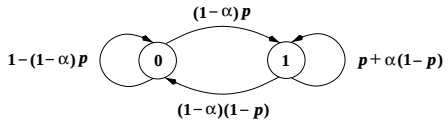


Fig. 4. Markov-chain model for dependent link-marking multicast flow control.

$$-m\left[1-(1-\alpha)p\right]^{2(m-2)}+\left[1-(1-\alpha)p\right]^{2(m-1)}$$
$$\cdot(m-1)\Big\}+\frac{(1-p)\,\Theta^2(\Delta)}{(1-\alpha)^2\left[2-(1-\alpha)p\right]^2\left[p^2-(1-\alpha)p^3\right]}$$
$$\cdot\Big\{1+\left[1-(1-\alpha)p\right]^2-\left[2-(1-\alpha)p\right]^3\left[(1-\alpha)p\right]^3$$
$$-m^2\left[1-(1-\alpha)p\right]^{2(m-1)}+(2m^2-2m-1)$$
$$\cdot\left[(1-\alpha)p-1\right]^{2m}+(2m-m^2-1)$$
$$\cdot\left[1-(1-\alpha)p\right]^{2(m+1)}\Big\}-\overline{\tau}^2_{HBH}(\alpha,m), \qquad (40)$$

where $\Theta(\Delta)$ is defined in Theorem 1, and $\overline{\tau}_{SSP}(\alpha,m)$ and $\overline{\tau}_{HBH}(\alpha,m)$ are given by Eqs. (37) and (38), respectively.

Proof: The proof is given in [17]. ■

Remarks on Theorem 4: Claim 1 derives formulas for multicast-tree bottleneck path distributions as a function of path length k, link-marking marginal probability p and dependency-degree factor α, and tree height m. Claim 2 examines the dynamic behavior of $\psi_d(P_k,\alpha,p,m)$ as p varies and observes that $\psi_d(P_k,\alpha,p,m)$ attains the unique maximum at p^* given by Eqs. (31) and (32), representing the link-marking probability that makes P_k the most likely multicast-tree bottleneck path. Claim 3 studies the behavior of $\psi_d(P_k,\alpha,p,m)$ from the viewpoint of α and indicates that $\psi_d(P_k,\alpha,p,m)$ can be either monotonic or non-monotonic, depending on the given k and p values. As long as k and p satisfy the condition specified in Eq. (33), $\psi_d(P_k,\alpha,p,m)$ achieves the maximum at α^* given by Eq. (33).

Claim 4 reveals the fact that the Markov-chain dependency ($\alpha > 0$) reduces the probabilities for shorter paths to be the bottleneck path while increasing the probabilities for longer paths to be the bottleneck path. This probability shift is also shown to be balanced at the unique path, $P_{\widetilde{k}}$ where $\psi_d(P_{\widetilde{k}},\alpha,p,m)\mid_{\alpha=0} = \psi_d(P_{\widetilde{k}},\alpha,p,m)\mid_{\alpha=\alpha_o}$, if the tree is high enough (see Eq. (34)). This claim also derives the condition for the existence of $P_{\widetilde{k}}$ and the equation to compute the dependency-balanced path number $\widetilde{k}$ as a function of Markov-chain dependency-factor α and the link-marking probability p. Claim 5 and Claim 6 show that the means and variances for SSP and HBH are the functions of Δ, p, α, and m. In addition, Eqs. (30), (31), (32) (37), (38), (39), and (40) all reduce to the analytical results derived for the multicast signaling delay analysis under *independent* random-marking [2] by letting $\alpha = 0$, confirming the correctness of the dependency-degree modeling and the associated equations.

VII. Asymptotical Analysis of Link-Marking Markov Chains

We now investigate the long-term behavior of the link-marking Markov chains based on the proposed dependency-degree factor modeling technique when m is large.

Theorem 5: Consider the Markov chain $\{X_i\}$ defined by the link-marking states on both main-stream and branch-stream paths in the multicast tree specified by Definition 1. If **(i)** the dependency degree of $\{X_i\}$ is specified by the dependency-degree factor vector $\vec{\alpha}=(\alpha_1,\alpha_1',\alpha_2,\alpha_2',\alpha_3,\alpha_3',\cdots)$ derived in Theorem 3; and **(ii)** the link-marking probability vector $\vec{p}=(p_1,p_1',p_2,p_2',p_3,p_3',\cdots)$ defined in Definition 1 and $\vec{\alpha}$ satisfy $0<p_i=p_i'=p<1$ and $0\le\alpha_i=\alpha_i'=\alpha\le 1$, $\forall i$, respectively, such that $\{X_i\}$ becomes a homogeneous Markov chain, then the following claims hold:

Claim 1. The n-step transition probability matrix, denoted by $P^{(n)}$, of the homogeneous Markov chain $\{X_i\}$ is determined by:

$$P^{(n)}\triangleq\left\{p_{jk}^{(n)}\right\}=\begin{bmatrix}1-(1-\alpha^n)p & (1-\alpha^n)p\\ (1-\alpha^n)(1-p) & \alpha^n(1-p)+p\end{bmatrix} \qquad (41)$$

where $j,k\in\{0,1\}$, $n\in\{1,2,\cdots\}$, and the case of $P^{(n)}$ with $n=1$ is shown in Fig. 4.

Claim 2. If $\alpha\in[0,1]$, then both link-marking states are *ergodic*, with

$$\limsup_{n\to\infty}p_{ii}^{(n)}=\lim_{n\to\infty}p_{ii}^{(n)}>0,\quad \lim_{n\to\infty}\sum_{r=1}^{n}p_{ii}^{(r)}=\infty, \qquad (42)$$

where $i\in\{0,1\}$, and

$$\lim_{n\to\infty}p_{ii}^{(n)}=\begin{cases}\Pr\{X_k=i\}=1-p, & \text{if } i=0,\ \alpha\in[0,1);\\ \Pr\{X_k=i\}=p, & \text{if } i=1,\ \alpha\in[0,1);\\ 1, & \text{if } i\in\{0,1\},\ \alpha=1;\end{cases} \qquad (43)$$

where $k\in\{1,2,\cdots\}$;

Claim 3. If $\alpha\in[0,1)$, then the Markov chain $\{X_i\}$ is *ergodic* and its limiting probabilities converge to the *unique* equilibrium state probabilities which are independent of both the initial state probabilities and dependency-degree α. The Markov chain's limiting probabilities, denoted by π_i, $i\in\{0,1\}$, converge to the marginal link-marking probabilities as follows:

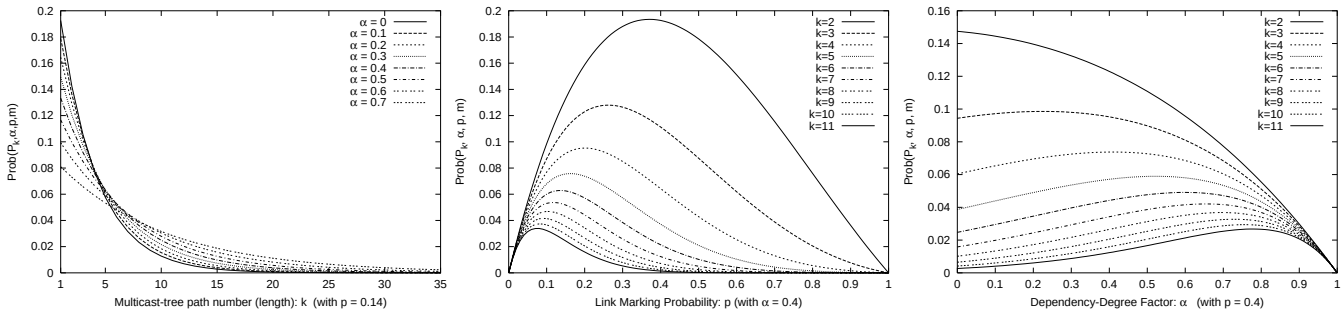
$$\vec{\pi}\ \triangleq\ \begin{bmatrix}\pi_0 & \pi_1\end{bmatrix}=\begin{bmatrix}(1-p) & p\end{bmatrix} \qquad (44)$$

i.e., $\pi_0=\Pr\{X_i=0\}=(1-p)$ and $\pi_1=\Pr\{X_i=1\}=p$;

Claim 4. If the Markov chain $\{X_i\}$ is "perfectly dependent", i.e., $\alpha=1$, then $\{X_i\}$ also converges to an equilibrium state, but the equilibrium state probabilities are not unique and are equal to the initial state probabilities. If the initial state probabilities are $\Pr\{X_i=0\}=1-p$ and $\Pr\{X_i=1\}=p$ (as in this paper addressed case), then $\pi_0=1-p$ and $\pi_1=p$ still hold.

Proof: The proof is presented in [17]. ■

Remarks on Theorem 5: Claim 1 fully specifies the long-term behavior of the Markov chain and determines the distribution of a bottleneck path in the homogeneous case. Claim 2 classifies the link-marking states as the dependency-factor α varies. It also shows that the Markov-chain state recurring probabilities converge asymptotically to the marginal link-marking probabilities (see Eq. (43)), if the Markov chain is not "perfectly" dependent ($\alpha\neq 1$).



(a) $\psi_d(P_k, \alpha, p, m)$ vs. k (b) $\psi_d(P_k, \alpha, p, m)$ vs. p (c) $\psi_d(P_k, \alpha, p, m)$ vs. α

Fig. 5. Impact of path length k, link-marking probability p, and dependency-degree α on multicast-tree bottleneck path probability distribution $\psi_d(P_k, \alpha, p, m)$.

Claim 3 ensures that our dependency-degree modeling converges asymptotically, and the long-term behavior of the resulting Markov chain is stable. Also, the ergodicity of the Markov chain enables us to evaluate its various statistics (ensemble average) through the sample averages in simulations or implementations. Moreover, this claim shows that the limiting probabilities converge to the marginal link-marking probabilities. This is also expected, because π_0 and π_1 represent the long-term proportion of Markov chain remaining at state 0 and 1, respectively, and is consistent with the definitions of $\Pr\{X_i = 0\}$ and $\Pr\{X_i = 1\}$, which verifies the validity of our dependency-degree modeling. Claim 4 says that when $\alpha = 1$, i.e., the link-marking states are perfectly dependent, the steady state distribution still exists, but is not unique, depending on the initial state probabilities. This is expected because when $\alpha = 1$, the Markov chain $\{X_i\}$ has two isolated classes (see Fig. 4). So it is not irreducible, and thus is no longer ergodic.

VIII. Numerical Evaluations

Based on the analytical results derived in the previous sections, the various multicast signaling delay properties are evaluated numerically as follows.

A. Multicast-Tree Bottleneck Path Distribution $\psi_d(P_k, \alpha, p, m)$

Fig. 5(a) plots $\psi_d(P_k, \alpha, p, m)$ vs. path length k while varying the dependency-degree factor α. $\psi_d(P_k, \alpha, p, m)$ is found to be a strictly monotonic decreasing function of k for both the independent ($\alpha = 0$) and dependent ($\alpha > 0$) cases. This is expected because the longer the bottleneck path, the more likely it will be dominated by shorter paths, as described in Definition 2.

Compared to the independent-marking case ($\alpha = 0$), the marking dependency is found to reduce the probability for shorter paths (with $k \leq 4$) to be the multicast-tree bottleneck path while increasing the probability for longer paths (with $k > 5$). This verifies the Claim 4 of Theorem 4, and the dependency-balanced path number: $\widetilde{k}$ is found to be around 4 and 5. Fig. 5(a) also shows that the larger α, the more this probability shifts from the shorter to longer paths. This is because the larger the link-marking dependence, the larger the probability that all links stay in the *same* congestion state.

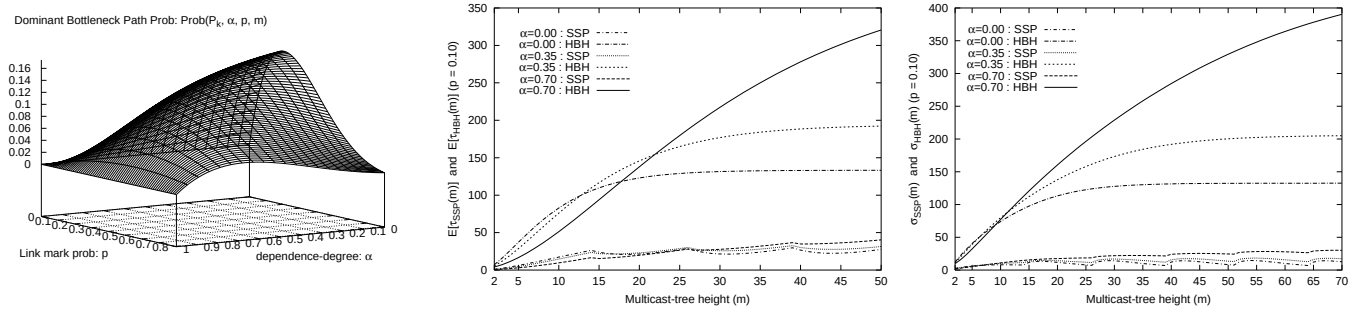
On the other hand, Theorems 2 and 4 state that for P_k to be the multicast bottleneck, all links on shorter paths $P_{k'}$ ($k' < k$) must be un-congested and P_k's last two links L_k *or* L'_{k+1} must be congested. Thus, $\psi_d(P_k, \alpha, p, m)$ is contributed by two events, $\{X_k = 1 \cup X'_{k+1} = 1\}$ and $\{\bigcap_{i=1}^{k-1}(X_i = 0, X'_{i+1} = 0)\}$, which must occur at the same time. But, link-marking dependence reduces the probability contribution from $\{X_k = 1 \cup X'_{k+1} = 1\}$ while increasing that from $\{\bigcap_{i=1}^{k-1}(X_i = 0, X'_{i+1} = 0)\}$. Then for $\alpha > 0$ the decaying rate of $\psi_d(P_k, \alpha, p, m)$ as k increases is slower than that for the case of $\alpha = 0$. Compared to the case of $\alpha = 0$, when k is small ($k \leq 4$), the decrease of probability contribution from $\{X_k = 1 \cup X'_{k+1} = 1\}$ due to $\alpha > 0$ cannot be compensated for by the increase in that from $\{\bigcap_{i=1}^{k-1}(X_i = 0, X'_{i+1} = 0)\}$. So, $\psi_d(P_k, \alpha, p, m) \mid_{\alpha>0} < \psi_d(P_k, \alpha, p, m) \mid_{\alpha=0}$ for small k ($k \leq 4$). When k is large ($k > 5$), the gain in probability contribution from $\{\bigcap_{i=1}^{k-1}(X_i = 0, X'_{i+1} = 0)\}$ is larger than the drop in that from $\{X_k = 1 \cup X'_{k+1} = 1\}$ due to $\alpha > 0$. Thus $\psi_d(P_k, \alpha, p, m) \mid_{\alpha>0} > \psi_d(P_k, \alpha, p, m) \mid_{\alpha=0}$ for large k ($k > 5$). When k becomes very large, both $\psi_d(P_k, \alpha, p, m) \mid_{\alpha>0}$ and $\psi_d(P_k, \alpha, p, m) \mid_{\alpha=0}$ converges to zero. So, $\psi_d(P_k, \alpha, p, m) \mid_{\alpha>0} = \psi_d(P_k, \alpha, p, m) \mid_{\alpha=0}$ as $k \to \infty$, which is confirmed by Fig. 5(a).

But, no matter how $\psi_d(P_k, \alpha, p, m)$ shifts as α changes, the normalization condition given by Eq. (28) is always satisfied, which is verified by the fact that the area under each plot for any given α always sums to 1 as shown in Fig. 5(a).

Fig. 5(b) shows that $\psi_d(P_k, \alpha, p, m)$ is inversely proportional to path length k, verifying the above observations. Fig. 5(b) also shows that there exists a unique maximum $\psi_d^*(P_k, \alpha, p^*, m)$ for any given k, verifying Claim 2 of Theorem 4. Fig. 5(c) indicates that for any given α, the larger the path length k, the smaller $\psi_d(P_k, \alpha, p, m)$. Fig. 5(c) also indicates that $\psi_d(P_k, \alpha, p, m)$ is not a monotonic function of α, but there can be a unique maximum $\psi_d^*(P_k, \alpha^*, p, m)$ as long as the given path length k and p satisfy the condition given in Eq. (33). As k gets larger, α^* increases. These also validate Claim 3 of Theorem 4. Fig. 6(a) shows a more complete dynamic-behavior picture of $\psi_d(P_k, \alpha, p, m)$ as a function of two independent variables (α, p). Fig. 6(a) shows that $\psi_d(P_k, \alpha, p, m)$ always has the maximum along the p-axis as α varies from 0 to 1. In contrast, $\psi_d(P_k, \alpha, p, m)$ can have the maximum along the α-axis only for a certain range of p values which satisfy the condition given in Eq. (33) in Theorem 4 for a given k.

B. Delay Statistics for HBH and SSP Schemes under the Dependent Markings

Fig. 6(b) plots the means, $\overline{\tau}_{SSP}(\alpha, m)$ and $\overline{\tau}_{HBH}(\alpha, m)$ calculated by Eqs. (37) and (38) against m for different α's. We observe that the link-marking dependency (α) has direct impact on the average multicast signaling delays. This impact gets more pronounced when the HBH scheme is used or the multicast tree's height increases, see Fig. 6(b). Consequently, the approximation



(a) $\psi_d(P_k, \alpha, p, m)$ vs. (α, p) (b) $\overline{\tau}_{SSP}(\alpha, m)$, $\overline{\tau}_{HBH}(\alpha, m)$ vs. m (c) $\sigma_{SSP}(\alpha, m)$, $\sigma_{HBH}(\alpha, m)$ vs. m

Fig. 6. Impact of dependency-degree factor α, link-marking probability p, and multicast-tree height m on multicast-tree bottleneck path probability $\psi_d(P_k, \alpha, p, m)$ and multicast signaling delay means and standard deviations. SSP is found to be more scalable with tree height than HBH in terms of signaling delay means and variations.

error in the multicast signaling delay analysis/calculation, which is caused by assuming "independent link-marking" while the actual congestion markings are dependent, is *not* negligible. This quantitatively justifies the necessity of the marking-dependency analysis for multicast signaling-delay performance evaluations, such as the Markov-chain and dependency-degree models developed in this paper. Fig. 6(b) also shows that for longer paths $(m > 20)$, the larger α, the larger the means while for shorter paths $(m < 12)$, the larger α the smaller the means, verifying that the multicast-tree bottleneck path probabilities shift from shorter to longer paths as α increases. Moreover, $\overline{\tau}_{HBH}(\alpha, m)$ is found to be much larger, and increase much faster with m, than $\overline{\tau}_{SSP}(\alpha, m)$ (see Fig. 6(b)).

Using Eqs. (39) and (40), Fig. 6(c) plots the standard deviations $\sigma_{SSP}(\alpha, m)$ and $\sigma_{HBH}(\alpha, m)$ against m while varying α. Like the average multicast signaling delays, similar trends in terms of effects of marking dependency and scalability are observed to hold for the multicast signaling delay variations (see Fig. 6(c)). Fig. 6(c) also shows that longer $(m \geq 10)$ paths' variances increase as α increases, while shorter $(m < 8)$ paths' variances decrease as α increases, also verifying that multicast-tree bottleneck probabilities shift from shorter to longer paths as α increases.

IX. Conclusion

We proposed statistical modeling approaches to the delay analysis of multicast feedback-synchronization signaling algorithms. Specifically, we developed a Markov-chain model to accurately characterize the multicast signaling delay when the congestion markings at different links are dependent. Using this model, we derived general expressions for the probability distributions of individual paths to be the multicast bottleneck. The derived Markov chain is also shown to have an equilibrium, and have its limiting state distributions converging to the link-marking marginal probabilities when the Markov chain is irreducible.

We also introduced a dependency-degree model to quantify and evaluate the dependency between different link congestion markings. Using this model, we derived equations for all one-step transition probabilities as functions of the marginal link-marking probabilities and the dependency-degree factors. The developed Markov-chain and dependency-degree models are generic and can be used not only for multicast signaling delay analysis, but for other Markov-chain-based analyses as well.

Using these two models, we derived the first and second moments of the multicast signaling delays for both HBH and SSP when link-markings are dependent. The numerical evaluations/analyses justified the necessity of the marking-dependency modeling/analysis, and also revealed that the marking-dependency tends to shift the bottleneck from shorter to longer paths, which is consistent with the definition of the positive link-marking dependency imposed by the nature of multicast signaling. The analytical results have also been confirmed by simulations [17].

Acknowledgment

The authors would like to thank Professor Steven H. Low of Caltech and Dr. David E. Lapsley of University of Melbourne for providing their preprints on REM.

References

[1] X. Zhang and K. G. Shin, "Performance analysis of feedback synchronization for multicast ABR flow control," in *Proc. of IEEE GLOBECOM*, Dec. 1999.

[2] X. Zhang and K. G. Shin, "Delay analysis of feedback-synchronization signaling for multicast flow control," *Technical Report, Real-Time Computing Labs., EECS Dept., The University of Michigan*, March 2000.

[3] X. Zhang, K. G. Shin, D. Saha, and D. Kandlur, "Scalable flow control for multicast ABR services," in *Proc. of IEEE INFOCOM*, pp. 837–846, 1999.

[4] L. Roberts, *Rate Based Algorithm for Point to Multipoint ABR Service*, ATM Forum contribution 94-0772, September 1994.

[5] K.-Y. Siu and H.-Y. Tzeng, "On max-min fair congestion control for multicast ABR services in ATM," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 3, pp. 545–556, April 1997.

[6] H. Saito, K. Kawashima, H. Kitazume, A. Koike, M. Ishizuka, and A. Abe, "Performance issues in public ABR service," *IEEE Communications magazine*, vol. 11, pp. 40–48, November 1996.

[7] J. Crowcroft and K. Paliwoda, "A multicast transport protocol," in *Proc. of ACM SIGCOMM*, pp. 247–256, August 1988.

[8] W. Ren, K.-Y. Siu, and H. Suzuki, "On the performance of congestion control algorithms for multicast ABR in ATM," *Proc. of IEEE ATM WORKSHOP'96*.

[9] Y.-Z. Cho and M.-Y. Lee, "An efficient rate-based algorithm for point-to-multipoint ABR service," in *Proc. of IEEE GLOBECOM*, November 1997.

[10] S. Fahmy, R. Jain, R. Goyal, B. Vandalor, and S. Kalyanaraman, "Feedbackback consolidation algorithms for ABR point-to-mulipoint connections in ATM networks," in *Proc. of IEEE INFOCOM*, April 1998.

[11] D. Lapsley and S. Low, "Random early marking: for Internet congestion control," in *Proc. of IEEE GLOBECOM*, pp. 1747–1752, December 1999.

[12] D. Lapsley and S. Low, "An optimization approach to ABR control," in *Proc. of IEEE International Conference on Communications*, June 1998.

[13] D. Lapsley and S. Low, "An IP implementation of optimization flow control," in *Proc. of IEEE GLOBECOM*, November 1998.

[14] S. Athuraliya, D. Lapsley, and S. Low, "An enhanced random early marking algorithm for Internet flow control," in *Proc. of IEEE INFOCOM*, pp. 1425–1434, March 2000.

[15] S. Athuraliya, S. Low, and D. Lapsley, "Random early marking," in *Proceedings of the First International Workshop on Quality of future Internet Services (QofIS'2000), Berlin, Germany*, September 2000.

[16] S. Floyd and V. Jacobson, "Random Early Detection gateways for congestion avoidance," *IEEE/ACM Trans. on Networking*, vol. 1, no. 4, pp. 397–413, August 1993.

[17] X. Zhang and K. G. Shin, "Markov-chain modeling for multicast signaling delay analysis," *Technical Report, Real-Time Computing Labs., EECS Dept., The University of Michigan*, November 2000.