Rarest-First with Probabilistic-Mode-Suppression (RFwPMS)

Nouman Khan, *Student Member, IEEE,* Mehrdad Moharrami, and Vijay G. Subramanian, *Senior Member, IEEE*

***Abstract*—Recent studies suggested that the BitTorrent's rarest-first (RF) protocol, owing to its work-conserving nature, can become unstable in the presence of non-persistent users. Consequently, for any provably stable protocol, many peers, at some point, have to be forced to hold off their file-download activity. In this work, we propose a tunable piece-selection policy that minimizes this (undesirable) requisite by combining the (work-conserving but not stabilizing) RF protocol with only an appropriate share of the (stabilizing but not work-conserving) mode-suppression (MS) protocol. We refer to this policy as "Rarest-First with Probabilistic Mode-Suppression" or simply RFwPMS. We study RFwPMS using a stochastic abstraction of the BitTorrent network that is general enough to capture a multi-swarm setting of non-persistent users—each swarm having its own altruistic preferences that may or may not overlap with those of other swarms. Using Lyapunov drift analysis, we show that for all kinds of inter-swarm behaviors and all arrival-rate configurations, RFwPMS is stable. Then, using the Kingman's moment bound technique, we further show that the steady-state expected sojourn time of RFwPMS is independent of the arrival-rate in the single-swarm case (under a mild additional assumption). Finally, our simulation-based performance evaluation confirms our theoretical findings, and shows that the steady-state expected sojourn time[1] is linear in the file-size (compared to our loose estimate of a polynomial with degree 6). Overall, an improved performance is observed in comparison to previously proposed stabilizing schemes like MS.**

***Index Terms*—P2P File-Sharing, BitTorrent, Rarest-First, Mode-Suppression, Foster-Lyapunov Theorem**

I. INTRODUCTION

CONSIDER the task of distributing a large file to peers in a peer-to-peer (P2P) network. The file is initially available with a distinguished peer (usually termed as the seed) and each peer can initiate a transfer connection with any other peer [1]. One method to perform this distribution task is to chop the file into a large number of small and roughly equally-sized pieces/chunks, and to allow peers to share the pieces with each other. Chopping the file allows peers to distribute parts of it before possessing it completely—this is the key idea behind the popular "BitTorrent protocol" [2]. Such an upload-while-download scheme reduces the average file-download time and more importantly, enables the network to scale its throughput with the number of peers. As a result, the BitTorrent protocol has gained large popularity over the years. Even today, despite the growth of streaming services like Netflix, Hulu, and Youtube, BitTorrent sharing remains a significant source of internet traffic [3]. In the research literature also, the protocol has gained extensive interest. For instance, on the theoretical side, various mathematical models have been recently studied [1], [4]–[16]—each model providing a high-level abstraction of the detailed workings of the actual protocol.

Once a file has been chopped into pieces, peers in the network make contacts and exchange pieces with each other. The set of rules that describe how peers exchange pieces with each other, and which pieces are chosen during exchange, are referred to as the *piece-exchange mechanism* and the *piece-selection policy*, respectively—both being critical to the network's performance. The original BitTorrent protocol uses the *Rarest-First (RF)* piece-selection policy in combination with *Tit-for-Tat* and *Optimistic-Unchoking* piece-exchange mechanisms.[2] Tit-for-tat, as the name indicates, are interactions where both peers share the file-contents with each other based on mutual benefit; on the other hand, opportunistic-unchoking interactions are those in which a peer altruistically offers pieces to other peers. Both these interactions are important for BitTorrent's performance; opportunistic-unchoking helps the incoming empty-peers to get some pieces of the file whereas the tit-for-tat interactions help ward off *free-riders* (users who download the file but do not contribute any of it back). In the RF piece-selection policy, peers prefer downloading pieces that are rarest in the network.

A common occurrence in BitTorrent-like networks is that a peer usually spends relatively more time downloading the last few pieces of the file. This phenomenon, referred to as the *delay-in-endgame-mode* [2], is because the last few pieces are often the rare ones in the network. Inspired by this, Hajek and Zhu [5] studied a stochastic model and showed that an unstructured[3] BitTorrent-like network that employs a work-conserving[4] piece-selection policy (e.g., Random-Novel (RN), Rarest-First (RF)) becomes unstable if peers are non-persistent[5] and their arrival-rate exceeds the fixed seed's upload rate.

The cause behind instability is a phenomenon called the *Missing Piece Syndrome* or *Last Piece Syndrome (LPS)* [5], wherein the network can get trapped in the *one-club scenario*. In this scenario, there are a large number of peers who possess all the pieces of the file except one (such peers are called *one-club peers*), a very small number of peers who possess that one piece (such peers are called *infected peers*), and a very small number of peers that are in neither of these two groups (they

A short version of this work appeared in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 1153-1162.

[1]The time a peer remains in the system collecting all the pieces of the file.

[2]See [2] for an overview on how a BitTorrent network operates.

[3]Each peer contacts another peer uniformly at random.

[4]In a work-conserving piece-selection policy, a piece transfer always happens if the uploading peer has a piece that the downloading peer needs.

[5]Each peer departs immediately upon completing their own file-download.

are called *young peers*). Since a vast majority of the peers are one-club peers, owing to random peer contacts and work-conserving nature of the piece-selection policy, all the infected peers leave the network quickly, and almost all young peers (including the new-comers) join the one-club peers. Inevitably, the fixed seed is tasked with uploading that one rare piece to the entire one-club whose growth rate is larger than its upload capacity. The network thus remains trapped in this configuration and the size of the one-club grows to infinity, causing instability. This result was later substantiated with experimental checks performed by Mendes, de Souza e Silva, Menasché, Leão, and Towsley [17], where it was observed that when seeds have a small effective service capacity, or when seeds are intermittent, the throughput saturates as the population size grows. Importantly, the authors also demonstrated that LPS can unfold in a closed BitTorrent network wherien each departure causes an arrival of an empty peer.

Another typical phenomenon in BitTorrent-like networks is the *low availability of chunks in the start-up phase* when only one user (the seed) has the complete contents of the file and the rest are empty.[6] Reference [17] calls this scenario the *first block problem*.[7] Understandably, the duration of this phase, for any choice of piece-selection algorithm, depends on the upload capacity of the seed. However, if the incoming peers choose to leave the network immediately upon download completion, then the choice of piece-selection policy is important—it should hold off download activity at times (to endogenously enforce some form of peer-persistence).

Finally, multiple file downloads occurring simultaneously is commonly observed in practice. This is the multi-swarm setting [7], [8] for which there isn't as yet a piece-selection policy that guarantees stability for all arrival-rates, particularly when the swarms interact with each other.

In the above discussion, we have highlighted multiple design criteria for BitTorrent-like file-sharing networks. These are:

- stability guarantees for all arrival-rates in both single-swarm and multi-swarm settings;
- improved steady-state file-delivery times; and
- improved transient responses (e.g. low flush-out times in flash-crowds).

Our goal in this paper is to propose a flexible piece-selection policy that can attain the aforementioned design criteria. To this end, in Section I-A, we will first go through the existing work and then in Section I-B, list the contributions of this manuscript.

A. Related Work

A series of recent works have appeared of which the relevant papers are [1], [4]–[12].

- Zhu and Hajek [6], in a follow-up to [5], showed that if, after completing their file-download, each peer remains in the system long enough to upload one additional piece, then the network is stable under any positive seed uploading capacity and any peer arrival-rate. This demands persistence of peers, which may not hold, especially with wireless users who are sensitive to their energy and bandwidth usages. Additionally, current implementations of BitTorrent allow peers to depart immediately after their own download is completed. Thus, recent works have also studied stability when peers are strictly non-persistent.
- Massoulié and Vojnović [4] considered a BitTorrent-like stochastic system where before entering the system, users obtain one piece (referred to as *coupon* in [4]) from a central bootstrap server. *Under this assumption*, they showed that the performance of the network does not depend critically on either persistence of users or on load balancing piece-selection policies like RF.
- Norros, Reittu, and Eirola [1] proposed the *Enforced Friedman algorithm* in which a peer makes three contacts simultaneously (with replacement), and if there are 'minority pieces' (pieces possessed by exactly one of the three peers), then the peer downloads one of them uniformly at random. If there are no minority pieces, then the peer waits for the next triple contact. The stability of the protocol was shown for a two-chunk system whereas for the case of a multi-chunk system, it was left as a conjecture, with numerical simulations providing evidence of stability. In a follow-up work, Oguz, Anantharam, and Norros [9] proved the stability of the Enforcement Friedman protocol for multi-chunk systems, and also proposed a provably stable improvement under the name of *Common Chunk protocol*. In this protocol, only new peers who arrive with no pieces follow the rules in [1], and peers who lack only one piece, can download it only if every piece they have is also present with at least two of the three contacted peers.
- Bilgen and Wagner [10] proposed the *Group Suppression Protocol*. Here, peers who share the same piece profile are defined as a *group*, and the group with the largest population is defined as the *largest club*. In this protocol, a peer belonging to the largest club uploads only to those peers who hold more pieces than it does, and refuses the upload to all other peers. Stability of the protocol was proven for a two-chunk file-sharing system, and stability for multi-chunk systems was left as a conjecture.
- Reddyvari, Parag, and Shakkottai [11] took a chunk-level viewpoint and proposed the *Mode-Suppression* (MS) protocol. Here, the transfer of pieces in the mode (present with the most number of peers) is prohibited, except when all pieces are in the mode, and a random-novel piece not in the mode (if any) is sent. In a follow-up [12], Reddyvari, Parag and Shakkottai prove the stability and scalability of *Threshold Mode-Suppression* (TMS) wherein pieces in the mode are prohibited from transfer when the largest-mismatch (difference between chunk-counts of the most-abundant and the rarest chunks, see 3) crosses a certain fixed threshold. This line of work showed stability and scalability for all arrival-rates in the single-swarm setting.
- Previous works have also considered bundling different swarms together in the same network and then allowing content-sharing across them. Such sharing can be useful in multi-priority cache networks—wherein all users store

[6] Flash-crowd is a start-up phase when a large number of users enter the network and only a few existing users possess the complete file.

[7] Indeed, [17] argues that the last piece syndrome (LPS) is overestimated whereas the first block problem is much more critical in practice.

the highest priority cache content. This was proposed by Zhou, Ioannidis, and Massoulié [7] with the claim that such “universal swarms” can increase the stability region of a BitTorrent network. Then, Zhu, Ioannidis, Hegde, and Massoulié [8] formally characterized the stability region of such networks under work-conserving piece-selection policies. The stability region is indeed larger than in the single-swarm setting, but, yet again, it doesn’t include all arrival-rate configurations.

B. Contributions

The contributions of this paper are listed below:

- Developing a tractable model (that extends the stochastic model from [5] and [6]) in Section II-C for the analysis of file-sharing P2P networks that employ tit-for-tat and optimistic-unchoke mechanisms (such as the BitTorrent network);
- Demonstrating that instability occurs when a hard tit-for-tat rule is used without optimistic-unchoking. Here, instability occurs due to a *first piece syndrome (FPS)* wherein newly arriving peers have to rely solely on the seed to get their first piece (Proposition 1);
- Showing that soft tit-for-tat or optimistic-unchoking mechanisms cannot ward off LPS whenever a work-conserving piece-selection policy is used (Proposition 2);
- In a general multi-swarm setting, proposing the (swarm-based) RFwPMS piece-selection policy and proving its stability using a novel Lyapunov function (Theorem 1). The Lyapunov function that we use is designed carefully by viewing the multi-swarm model as a system of coupled simple harmonic oscillators—see Sections II-D2 and III-A for detailed intuition behind RFwPMS and our choice of Lyapunov function.
- Demonstrating the scalability of swarm-based RFwPMS in the single-swarm case—i.e., the expected steady-state sojourn time is upper-bounded by a constant which is independent of the arrival-rate of the incoming peers.[8] (Theorem 4). Contrary to the standard approach in using the Kingman’s moment bound technique, (where the Lyapunov function for stability determines performance bounds), our bounds are obtained by using the drift of two other non-negative functions.
- Showing improved performance of swarm-based RFwPMS in steady-state and flash-crowds via an extensive numerical investigation; see Section VI.

A preliminary version of this work was presented in [14] with a partial theoretical analysis and a special case of the general stochastic model studied here. The current work adds to the contributions of [14] – items a) through f) above – and provides complete detailed intuition and methods for the stability analysis of (swarm-based) RFwPMS.

C. Organization

Since the single-swarm model is a special case of the multi-swarm model, henceforth we only consider the multi-swarm model, except when we discuss our scalability results. The remainder of the paper is organized as follows. In Section II, we introduce the multi-swarm model (building upon the model in [8]). Section III presents the main stability theorem along with the preliminary setup needed for the detailed proof (the detailed proof is provided in Appendix E). Section V presents our scalability result (and its proof) for the single-swarm setting. Section IV discusses the working of RFwPMS as well as a few types of inter-swarm behaviors that can be relevant in specific P2P environments and follow naturally as a result of the general nature of our assumption on inter-swarm behaviors. Section VI presents few important snapshots of numerical results on stability, scalability and performance of RFwPMS. Finally, in Section VII, we give concluding remarks and a few future-work directions.

D. Notation

To aid in the reading experience, we refer the reader to the list of symbols given in Appendix G.

II. System Model

In this section, we introduce the stochastic model and describe our proposed piece-selection policy.

A. Key Model Assumptions

A **master-file**, denoted by $\mathcal{F}$, is chopped into at least two equally-sized pieces, i.e., $\mathcal{F} = [K]$[9] where $K \geqslant 2$. There is a distinguished peer, the **seed**, which holds the master-file $\mathcal{F}$ and stays in the network indefinitely. The existence of the seed for indefinite period of time ensures that every piece is available in the network at all times, thus allowing us to study the long-term behavior of the network. We define **file-**W as any non-empty subset of the master-file with at least two pieces, thus, $W \neq \emptyset$, $W \subseteq \mathcal{F}$, and $|W| \geqslant 2$. The number of pieces in file-W is denoted by K_W, i.e., $K_W = |W|$. With this definition of file, **swarm**-W is defined as the set of peers who are *primarily* interested in downloading (pieces of) file-W. We note that peers entering the network can be interested in any file, i.e., the files need not be disjoint subsets of $\mathcal{F}$. Besides their primary interest in file-W, swarm-W peers may also have a secondary preference for some other pieces of the master-file. Thus, the set of pieces that a swarm-W peer can download during its stay in the network is given by $\mathcal{F}_W$ where $W \subseteq \mathcal{F}_W \subseteq \mathcal{F}$. It is assumed that peers enter the network according to independent Poisson processes, i.e., a swarm-W peer enters the network according to a Poisson process of rate $\lambda_W > 0$, independent of other swarms. We denote the set of all swarms entering the network by $\mathcal{W}$, and let $\boldsymbol{\lambda} := (\lambda_W : W \in \mathcal{W})$ denote the vector of their arrival-rates. The total arrival-rate is denoted by $|\boldsymbol{\lambda}|$, i.e., $|\boldsymbol{\lambda}| := \sum_{W \in \mathcal{W}} \lambda_W$.

Now, we list three important assumptions of the model.

a) *Empty Cache upon Arrival*: Each peer maintains a **cache** to store the pieces it downloads. The cache is empty upon arrival and has a capacity of $|\mathcal{F}_W|$ pieces. The part of the

[8]Showing scalability in the general multi-swarm case is for future work.

[9]For $c, d \in \mathbb{Z} := \{\ldots, -1, 0, 1, \ldots\}$ and $c < d$, we use the notation $[c, d] := \{c+1, \ldots, d\}$ and $[d]$ for $[0, d]$ when $d \in \mathbb{N} := \{1, 2, \ldots\}$.

cache that is devoted for the pieces of secondary interest is called the **excess-cache** (where pieces from the set $\mathcal{F}_W \backslash W$ are stored). In the context of [4], the empty cache assumption aligns with the case when the central bootstrap server is bottlenecked, for example, in the case of a high peer arrival-rate or during a flash-crowd.[10]

b) *Ally Swarms*: While peers interested in the same file (i.e., belonging to the same swarm) exchange pieces with each other, they may or may not prefer to collaborate with peers who are interested in other files (i.e. peers belonging to a different swarm). Thus, each swarm has an associated set of receiving-ally-swarms. Formally, the **receiving-ally-set** of swarm-W, denoted by $\mathcal{W}_W^{(ally\text{-}\uparrow)}$, is a non-empty subset of $\mathcal{W}$ that consists of swarm-W as well as any other swarms to which its peers upload pieces ($W \in \mathcal{W}_W^{(ally\text{-}\uparrow)} \subseteq \mathcal{W}$). Equivalently, the **donor-ally-set** of swarm-W is given by

$$\mathcal{W}_W^{(ally\text{-}\downarrow)} := \left\{ V \in \mathcal{W} : W \in \mathcal{W}_V^{(ally\text{-}\uparrow)} \right\}.$$

c) *Strictly Non-persistent Peers*: Once a peer finishes downloading their pieces of primary interest, they leave the network immediately.

B. State Description

The notation used in this paper is a combination of related notations in [8] and [11]. We classify peers into types according to the swarm they belong to, and the set of pieces in their cache. So, a peer in swarm-W holding $S \subseteq \mathcal{F}_W$ is said to be of type (W, S). We denote the number of (W, S)-type peers at time $t \geqslant 0$ by $x_W^{(S)}(t) \in \mathbb{Z}_{\geqslant 0} := \{0, 1, \dots\}$. Then, the **state** of the network at time t is given by the vector,

$$\mathbf{x}(t) := \left(x_W^{(S)} : W \in \mathcal{W}, S \subseteq \mathcal{F}_W, \text{ and } W \backslash S \neq \varnothing \right). \quad (1)$$

Note that as a result of aforementioned assumptions, the cache-profile S of a swarm-W peer always satisfies the conditions, $S \subseteq \mathcal{F}_W$ and $W \backslash S \neq \varnothing$, with the latter condition capturing strictly non-persistent peers. For the sake of brevity, from hereon, we will omit writing these two conditions. Also, since the fixed seed is always present in the network, we do not include it in $\mathbf{x}(t)$.

The **population-size** of swarm-W (number of swarm-W peers) at time t is given by

$$|\mathbf{x}|_W(t) := \sum_S x_W^{(S)}(t). \quad (2)$$

Similarly, the total number of peers at time t is given by

$$|\mathbf{x}|(t) := \sum_{W \in \mathcal{W}} |\mathbf{x}|_W(t). \quad (3)$$

From hereon, for brevity, we will write $\mathbf{x}(t)$ as $\mathbf{x}$ (since the dependence on time t will be clear from the context).

[10] A situation in which the network suddenly encounters a very large number of empty peers; this is commonly seen with torrents of popular files.

C. Peer-Contact Method

Consistent with the stochastic models of [1], [4]–[11], we assume that the network employs random peer-contacts. Specifically, it is assumed that each peer has a fixed number of contact-links, denoted by $L \in \mathbb{N}$. In normal peers, the first $\left(L - \mathbb{1}\left[Y^{(opt)} = 1\right]\right)$ of these links are reserved for tit-for-tat based piece exchanges whereas the L^{th} link is used for optimistic-unchoking if and only if $Y^{(opt)} = 1$; otherwise it is also used for tit-for-tat based piece exchanges. Here, $Y^{(opt)}$ is a binary parameter that is set to 1 if optimistic-unchoking is desired in the network. In contrast to normal peers, all the L links of the fixed seed are used for optimistic-unchokes.

For normal peers, we assume that each tit-for-tat link is activated according to an independent Poisson process of rate $\mu^{(\mathrm{tft})} > 0$, and the optimistic-unchoke link is activated according to another independent Poisson process of rate $\mu^{(\mathrm{opt})} > 0$. Upon activation of the link, the peer contacts some other (normal) peer from the network uniformly at random.[11] For the fixed seed, each of its (optimistic-unchoke) links is activated according to a Poisson process of rate $U > 0$. We assume that the transfer of piece/s occurs instantaneously with the contact (in reality, this will take more time than initiating a contact).

To define the tit-for-tat and optimistic-unchoking mechanisms, let us imagine a contact in which a peer, say peer-(1), which is of type (W_1, S_1), has contacted another peer, say peer-(2), which is of type (W_2, S_2). The interaction between the two peers depends on the type of contact that peer-(1) has made with peer-(2). Thus, there are two cases.

a) *Tit-for-Tat Contact*: In this case, both peers first check the swarm-identity of each other after which they reveal their cache-profiles (based on how they view each other's swarm). Specifically, for each $k = 1, 2$, the following events happen sequentially.

   i) Peer-(k) shares its swarm-identity W_k with peer-$(-k)$.[12]

   ii) If $W_{-k} \notin \mathcal{W}_{W_k}^{(ally\text{-}\uparrow)}$, peer-$(k)$ reveals an empty cache-profile, $\widehat{S}_k = \varnothing$[13], otherwise it shows its true cache-profile, $\widehat{S}_k = S_k$. We call $\widehat{S}_k(\cdot)$ the **revealed cache-profile** of peer-(k).

   iii) Peer-(k) checks if the contact is potentially useful, i.e., if $(\widehat{S}_{-k} \cap W_k) \backslash S_k$ is non-empty.[14] If $(\widehat{S}_{-k} \cap W_k) \backslash S_k$ is non-empty, peer-(k) commits to transfer some piece to peer-$(-k)$ from $\widehat{S}_k$. If $(\widehat{S}_{-k} \cap W_k) \backslash S_k$ is empty, the network forces peer-(k) to conduct a Bernoulli(p) trial, only upon the success of which, it must commit to transfer some piece to peer-$(-k)$ from $\widehat{S}_k$. Once peer-(k) has committed to transfer a piece to peer-$(-k)$

[11] The motivation behind introducing a separate rate for the optimistic-unchoke link comes from how BitTorrent operates. In practice, by default, the number of links L is 5, and which peer is optimistically-unchoked is rotated roughly every third tit-for-tat period (see [2]). By introducing $\mu^{(\mathrm{tft})}$ and $\mu^{(\mathrm{opt})}$, this can be captured in our model by setting $\mu^{(\mathrm{opt})} = \frac{\mu^{(\mathrm{tft})}}{3}$.

[12] Here, $-k$ denotes the element in $\{1, 2\} \backslash \{k\}$.

[13] This ensures that no piece is transferred to a non-ally peer.

[14] Checking $(\widehat{S}_{-k} \cap W_k) \backslash S_k$ instead of $(\widehat{S}_{-k} \cap \mathcal{F}_{W_k}) \backslash S_k$ matches with our assumption that extra pieces (in $\mathcal{F}_{W_k} \backslash W_k$) are given secondary preference.

from $\widehat{S}_k$, we say that "peer-(k) has **push-contacted** peer-$(-k)$ using revealed cache-profile $\widehat{S}_k$" or equivalently "peer-$(-k)$ has pull-contacted peer-(k) having revealed cache-profile $\widehat{S}_k$". Here, importantly, we note that the probability of this push-contact is at least p.

b) *Optimistic Unchoke*: In this case, peer-(1) checks the swarm-identity W_2 and then push-contacts peer-(2) with revealed cache-profile S_1 (if $W_2 \in \mathcal{W}_{W_1}^{(ally\text{-}\uparrow)}$) or $\varnothing$ (if $W_2 \notin \mathcal{W}_{W_1}^{(ally\text{-}\uparrow)}$).

From a) and b), it is clear that no piece is transferred from peer-(k) to peer-$(-k)$ if $W_{-k} \notin \mathcal{W}_{W_k}^{(ally\text{-}\uparrow)}$ $(k = 1, 2)$. Hence, we need not consider push-contacts from a peer to its non-ally peer. Once a push-contact has been made, which exact piece is chosen for transfer, and whether the transfer is successful or not, is determined by the network's piece-selection policy. This is described next.

D. Piece-Selection Policy (Swarm-based RFwPMS)

Suppose that at time $t \geqslant 0$, a (V,T)-peer has push-contacted its receiving-ally peer, say a (W,S)-peer, using revealed cache-profile $\widehat{T}$. To describe swarm-based RFwPMS, some definitions (*a'la* [11], [12]) are needed.

Definition 1: The **frequency** of piece i in swarm-W is denoted by $\pi_W^{(i)}(\mathbf{x})$ and is defined as follows:

$$\pi_W^{(i)}(\mathbf{x}) := \begin{cases} \frac{1}{|\mathbf{x}|_W} \sum\limits_{S:i\in S} x_W^{(S)}(t) & \text{if } |\mathbf{x}|_W > 0, \\ 0 & \text{if } |\mathbf{x}|_W = 0. \end{cases} \tag{4}$$

The **chunk-count** of piece i in swarm-W is then given by

$$c_W^{(i)}(\mathbf{x}) := \pi_W^{(i)}(\mathbf{x})|\mathbf{x}|_W. \tag{5}$$

The maximum and minimum chunk frequencies in swarm-W are denoted by $\overline{\pi}_W(\mathbf{x})$ and $\underline{\pi}_W(\mathbf{x})$ respectively, i.e.,

$$\overline{\pi}_W(\mathbf{x}) := \max_{i\in W} \pi_W^{(i)}(\mathbf{x}) \quad \text{and} \quad \underline{\pi}_W(\mathbf{x}) := \min_{i\in W} \pi_W^{(i)}(\mathbf{x}). \tag{6}$$

Similar definitions hold for maximum and minimum chunk-counts of swarm-W, denoted respectively by $\overline{c}_W(\mathbf{x})$ and $\underline{c}_W(\mathbf{x})$, i.e.,

$$\overline{c}_W(\mathbf{x}) := \overline{\pi}_W(\mathbf{x})|\mathbf{x}|_W \quad \text{and} \quad \underline{c}_W(\mathbf{x}) := \underline{\pi}_W(\mathbf{x})|\mathbf{x}|_W. \tag{7}$$

Importantly, both are computed over file-W and not $\mathcal{F}_W$.

Definition 2: The **total chunk-count** in swarm-W is given by

$$P_W^{(\mathrm{tot})}(\mathbf{x}) := \sum_{i\in W} c_W^{(i)}(\mathbf{x}). \tag{8}$$

Definition 3: The **mismatch** of piece i in swarm-W is given by

$$m_W^{(i)}(\mathbf{x}) := \overline{c}_W(\mathbf{x}) - c_W^{(i)}(\mathbf{x}). \tag{9}$$

Importantly, we shall be interested in the **largest-mismatch**

$$\overline{m}_W(\mathbf{x}) := (\overline{c}_W - \underline{c}_W)(\mathbf{x}), \tag{10}$$

and the **total-mismatch**

$$M_W(\mathbf{x}) := \sum_{i\in W} m_W^{(i)}(\mathbf{x}). \tag{11}$$

Here, we note that

$$\overline{m}_W(\mathbf{x}) \leqslant M_W(\mathbf{x}) \leqslant (K_W - 1)\overline{m}_W(\mathbf{x}). \tag{12}$$

Definition 4: For swarm-W, the **complementary chunk-count** of piece i, denoted by $d_W^{(i)}(\mathbf{x})$, is its number of copies in other donor-ally swarms of swarm-W. That is,

$$d_W^{(i)}(\mathbf{x}) := \sum_{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)}\setminus\{W\}} c_V^{(i)}(\mathbf{x}). \tag{13}$$

Definition 5: The set of **rare pieces** in swarm-W, denoted by $R_W(\mathbf{x})$, is defined as follows:

$$R_W(\mathbf{x}) := \begin{cases} \{i \in W : c_W^{(i)}(\mathbf{x}) < \bar{c}_W(\mathbf{x})\} & \text{if } \bar{c}_W(\mathbf{x}) \neq \underline{c}_W(\mathbf{x}), \\ W & \text{if } \bar{c}_W(\mathbf{x}) = \underline{c}_W(\mathbf{x}). \end{cases} \tag{14}$$

Definition 6: The set of **non-rare pieces** in swarm-W is given by $R_W^c(\mathbf{x}) := W \setminus R_W(\mathbf{x})$.

Definition 7: The set of **extra pieces** for swarm-W is given by $\mathcal{F}_W \setminus W$.

1) Rules of Swarm-Based RFwPMS: Unless otherwise noted, we will refer to a rare piece by r, a non-rare piece by n, and importantly many times, the rarest piece (one with the lowest chunk-count) by $\underline{r}$. We now list the rules of swarm-based RFwPMS for a possible piece transfer when the (V,T)-peer has push-contacted (its receiving-ally) (W,S)-peer using revealed cache-profile $\widehat{T}$.

a) *Download of a Rare Piece*: If a novel rare piece is available for transfer, i.e., $\left(\widehat{T} \cap R_W(\mathbf{x})\right) \setminus S \neq \varnothing$, then the (V,T)-peer uploads the rarest piece it can offer, i.e., a piece chosen uniformly at random from the set,

$$\underline{R}_{(\widehat{T},W,S)}(\mathbf{x}) := \underset{j\in\left(\widehat{T}\cap R_W(\mathbf{x})\right)\setminus S}{\arg\min} c_W^{(j)}(\mathbf{x}). \tag{15}$$

b) *Download of a Non-rare Piece*: In the case that no novel rare piece is available for transfer but a novel non-rare piece is, i.e., $\left(\widehat{T} \cap R_W(\mathbf{x})\right) \setminus S = \varnothing$ and $(\widehat{T} \cap W) \setminus S \neq \varnothing$, then the (V,T)-peer chooses some non-rare piece $n \in (\widehat{T} \cap W) \setminus S$ uniformly at random, and uploads it only if the result of a Bernoulli$\left(\zeta_W^{(n)}(\mathbf{x})\right)$ trial is a success. We refer to $\zeta_W^{(n)}(\mathbf{x})$ as the *non-rares' sharing factor*, and define it as follows.

$$\zeta_W^{(n)}(\mathbf{x}) := \begin{cases} \exp\left(-\frac{\left(\overline{m}_W(\mathbf{x}) + \left(d_W^{(n)}(\mathbf{x})\right)^{\alpha_W}\right)}{\beta_W K_W}\right) & \text{if } \beta_W > 0, \\ 0 & \text{if } \beta_W = 0. \end{cases} \tag{16}$$

Here, $\beta_W \geqslant 0, \alpha_W \in (0, 1]$ are tuning parameters. (As an aside, we note that MS piece-selection policy is covered as we allow $\beta_W = 0$.)

Intuition Behind The Non-Rares' Sharing Factor: Let us develop some intuition about our choice of $\zeta_W^{(n)}(\mathbf{x})$. As indicated earlier in Section I, the MS protocol [11] strictly forbids the replication of non-rare pieces. This strict rule does a good job of maintaining a uniform chunk-distribution throughout the network's evolution. However, there is an accompanying undesirable effect—no pieces of file-W are transferred in all those push-contacts where only non-rare pieces are novel. As shown in Section VI, this can incur a high penalty on the file-delivery time during a flash-crowd.[15] Besides flash-crowds, even under normal operating conditions, completely suppressing non-rare pieces is unnecessary and, as indicated in [11], [12], a trade-off exists between their suppression and sharing. Too much suppression will force peers to linger in the system longer (to replicate the rare pieces), thus increasing the expected sojourn time, whereas too much sharing (trying to be work-conserving) will lead to causing instability. Swarm-based RFwPMS allows for tuning this trade-off via $\zeta_W^{(n)}(\mathbf{x})$. Even though different swarms are coupled together in the same system, one can optimistically expect that if each swarm tries to keep close its own maximum and minimum chunk-counts, the stochastic system should hopefully converge to some form of equilibrium (after possibly some temporary transient behavior due to the effects of other swarms). Keeping this in mind, the first term of $\zeta_W^{(n)}(\mathbf{x})$,

$$\exp\left(-\frac{\overline{m}_W(\mathbf{x})}{\beta_W K_W}\right),$$

is intended to increase the (probabilistic) suppression of non-rare pieces as the largest-mismatch in swarm-W gets larger.

Ideally, we would have liked $\zeta_W^{(n)}(\mathbf{x})$ to consist of the first term only. But showing stability of our multi-swarm model in that case is hard. The key technical difficulty is the form of the Lyapunov function that we use (see (30), (31)), and our analysis of the unit-transition drift (where we are essentially trying to decouple the inter-swarm effects). Owing to these reasons, if we consider states with large $\overline{m}_W(\mathbf{x})$ and a large complementary chunk-count $d_W^{(n)}(\mathbf{x})$, then the suppression through $\exp\left(-\frac{\overline{m}_W(\mathbf{x})}{\beta_W K_W}\right)$ turns out to be insufficient to satisfy the unit-transition drift conditions.[16] We circumvent this technical issue by introducing the second term,

$$\exp\left(-\frac{\left(d_W^{(n)}(\mathbf{x})\right)^{\alpha_W}}{\beta_W K_W}\right),$$

where, by choosing $\alpha_W \in (0,1]$ sufficiently small, $\zeta_W^{(n)}(\mathbf{x})$ effectively becomes a function of the ratio of swarm-W's largest-mismatch to its file-size K_W. The higher this ratio, the lesser the likelihood that the (non-rare) piece n gets replicated. The choice of the ratio instead of just the largest-mismatch matches the intuition that a file with larger number of pieces should allow relatively more sharing of non-rare pieces.

c) *Download of an Extra Piece*: If, from rules (b) and (c), no piece of file-W could be uploaded to the (W,S)-peer, then the (V,T)-peer uploads a novel extra piece (if it exists), chosen uniformly at random—i.e., a piece from the set $(\widehat{T}\cap \mathcal{F}_W)\backslash(S\cup W)$.

2) Intuition Behind Swarm-Based RFwPMS: One may get an intuition of how swarm-based RFwPMS operates by viewing the multi-swarm network as a collection of coupled simple harmonic oscillators—where the coupling arises due to the inter-swarm interactions—, and a given system's displacement from its equilibrium region, say that of system-W, arises due to swarm-W peers greedily downloading file-W (thus, increasing the largest-mismatch $\overline{m}_W(\mathbf{x})$)—see Fig. 1). Work-conserving policies which always cater to this greediness of users, are doomed to keep increasing this mismatch in high arrival-rate regimes, thus causing instability. On the other hand, swarm-based RFwPMS, via deterministic download of rarest-pieces and appropriate suppression of non-rare pieces, provides sufficient restoring force to keep the system near equilibrium.

Algorithm 1 Transferable-set of swarm-based RFwPMS.

Input: $\mathbf{x}, W, S, \widehat{T}$.
Output: Transferable-set, $A^{(\mathrm{trf})}(\mathbf{x}, \widehat{T}, W, S)$.

Set $A \leftarrow \varnothing$.
Set $H_1 \leftarrow \left(\widehat{T}\cap R_W(\mathbf{x})\right)\backslash S$.
Set $H_2 \leftarrow \left(\widehat{T}\cap W\right)\backslash S$.
Set $H_3 \leftarrow \left(\widehat{T}\cap \mathcal{F}_W\right)\backslash (S\cup W)$.
if $H_1 \neq \varnothing$ **then**
  Choose r randomly from $\underline{R}_{(\widehat{T},W,S)}(\mathbf{x})$.
  Set $A \rightarrow A\cup\{r\}$.
else if $H_2 \neq \varnothing$ **then**
  Choose n randomly from $\left(\widehat{T}\cap W\right)\backslash S$.
  Conduct Bernoulli trial with success probability $\zeta_W^{(n)}(\mathbf{x})$.
  if Success **then**
    Set $A \leftarrow A\cup\{n\}$.
  else
    Set $A \leftarrow A\cup$ ChooseExtraPiece(H_3).
else
  Set $A \leftarrow A\cup$ ChooseExtraPiece(H_3).
return A.

function ChooseExtraPiece(H_3)
  if $H_3 \neq \varnothing$ **then**
    Choose e randomly from H_3.
    return $\{e\}$.
return $\varnothing$.

With the above three rules, the **transferable-set** $A^{(\mathrm{trf})}(\mathbf{x}, \widehat{T}, W, S)$ for swarm-based RFwPMS can be computed according to Algorithm 1. By transferable set, we mean the set of pieces chosen for transfer by the piece-selection policy – in our case, the transferable set will either be empty or a singleton set.

In order to extend our rules to the seed, we assume that

[15] A situation in which the network suddenly encounters a very large number of empty peers; this is commonly seen with torrents of popular files.

[16] The term $\exp\left(-\beta_W^{-1}K_W^{-1}\overline{m}_W(\mathbf{x})\right)$ cannot counter a polynomial term in $d_W^{(n)}(\mathbf{x})$.

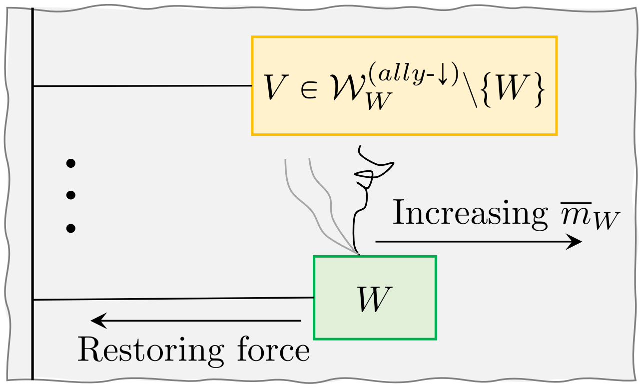


Fig. 1. Intuition behind swarm-based RFwPMS and the role of α_W and β_W in the definition of non-rares sharing factor $\zeta_W^{(n)}(\mathbf{x})$. The restoring force (in response to $\overline{m}_W$) arises through RF (for download of rare pieces) and probabilistic MS (for download of non-rare pieces). The parameter α_W helps tackle the coupling with other swarms, and β_W functions as a proportionality constant.

the seed is a $(\dagger, \mathcal{F})$-type peer[17] with the receiving-ally-set $\mathcal{W}_\dagger^{(ally\text{-}\uparrow)} = \mathcal{W}$. Thus, the set of rare pieces when the seed push-contacts a normal peer is given by

$$\underline{R}_{(\mathcal{F},W,S)}(\mathbf{x}) = \arg\min_{j \in R_W(\mathbf{x}) \backslash S} c_W^{(j)}(\mathbf{x}).$$

Remark: Swarm-based RFwPMS is not a work-conserving scheme—when there is no (novel) rare piece at offer, a (novel) non-rare piece $n \in (T \cap W) \backslash S$ is transferred only *probabilistically*.

Remark: With rules a), b) and c), we note that each swarm maintains its own chunk-table and also prioritizes its primary pieces over other pieces of the master-file. For these two reasons, we call our policy, "*swarm-based*" RFwPMS.

E. Process Description / Suitable Bounds on Transition Rates

For the sake of notational simplicity, from hereon, we will write function $f(\cdot)$ as f when the dependence on the argument is clear. Given a state $\mathbf{x}$, the next state of the network depends solely on $\mathbf{x}$ because the piece-selection policy solely depends on $\mathbf{x}$ and the new arrivals (of empty peers) are determined by independent Poisson processes. Hence, the evolution of the network described by the process $\{\mathbf{x}(t) : t \geqslant 0\}$ is a continuous-time, time-homogeneous, and irreducible[18] Markov chain with state space,

$$\mathcal{S} := \mathbb{Z}_{\geqslant 0}^{\sum_{W \in \mathcal{W}} |\{S : S \subseteq \mathcal{F}_W \text{ and } W \backslash S \neq \varnothing\}|}. \tag{17}$$

A typical element of $\mathcal{S}$ represents the number of peers of each type (see (1)).

Given a state $\mathbf{x}$, there are different events that can lead to state-transitions, namely, the arrival of a new empty peer, the download of a single-piece or a (two-sided) piece-exchange (download of two pieces simultaneously as a result of some tit-for-tat contact). Here, we do not list the transition rates of (two-sided) piece-exchanges: later in Section III, we will see that with our choice of Lyapunov function, such transitions can be viewed as two separate single-piece-download events.

a) *Arrival of an Empty-peer*: A swarm-W peer with an empty cache enters the network according to a Poisson process of rate λ_W. This results in a unit increase in the number of swarm-W peers with no pieces. Let us denote this transition by $\{(\varnothing, +)_W\}$ and its corresponding rate by $q_W^{(\varnothing,+)}$. Then,

$$q_W^{(\varnothing,+)} = \lambda_W. \tag{18}$$

b) *Download of Piece of Primary Interest*: Consider the event that a (W, S)-peer missing piece $i \in W$ downloads piece i. The necessary condition for this event is that (W, S)-peer gets push-contacted by a (V, T)-peer such that $V \in \mathcal{W}_W^{(ally\text{-}\downarrow)}$ and $i \in T$. There are three distinct ways in which that can happen.

i) The (W, S)-peer gets push-contacted by the (V, T) peer as a result of a tit-for-tat contact initiated by the (V, T)-peer. By the superposition and thinning properties of Poisson processes, the rate at which a donor-ally peer of swarm-W with piece i makes a tit-for-tat contact with a (W, S)-peer is

$$\left((L - \mathbb{1}\left[Y^{(opt)} = 1\right]) \mu^{(\mathrm{tft})} \sum_{\substack{V \in \mathcal{W}_W^{(ally\text{-}\downarrow)}, \\ T : i \in T}} x_V^{(T)} \right) \times \left(\frac{x_W^{(S)}}{|\mathbf{x}|} \frac{|\mathbf{x}|}{|\mathbf{x}| - 1} \right).$$

Here, we recall that $Y^{(opt)}$ is the binary parameter reserved for optimistic-unchoking; the term in the first parenthesis is the aggregate rate with which a donor-ally peer of swarm-W having piece i activates one of its tit-for-tat links; and the term in the second parenthesis is the probability that a (W, S)-peer is chosen after the activation.

ii) The (W, S)-peer gets push-contacted by the (V, T)-peer as a result of an optimistic-unchoke (initiated by the (V, T)-peer). The rate at which a (W, S)-peer gets contacted in an optimistic-unchoke by a donor-ally peer having piece i is given by

$$UL \cdot \frac{x_W^{(S)}}{|\mathbf{x}|} + \left(\mathbb{1}\left[Y^{(opt)} = 1\right] \mu^{(\mathrm{opt})} \sum_{\substack{V \in \mathcal{W}_W^{(ally\text{-}\downarrow)}, \\ T : i \in T}} x_V^{(T)} \right) \times \frac{x_W^{(S)}}{|\mathbf{x}|} \frac{|\mathbf{x}|}{|\mathbf{x}| - 1}.$$

Here, the first term is the aggregate rate with which the seed contacts a (W, S)-peer; and the second term is the aggregate rate with which a normal donor-ally peer of swarm-W with piece i contacts a (W, S)-peer.

iii) The (W, S)-peer gets push-contacted by the (V, T)-peer as a result of a tit-for-tat contact initiated by the (W, S)-peer. The rate at which a (W, S)-peer makes a

[17] For example, $\dagger$ can be $\mathcal{F} \cup \{K + 1\}$.

[18] Since the seed is present indefinitely in the network, has all the pieces of the master-file, and has a positive contact rate, there is a positive probability of reaching a zero population state.

tit-for-tat contact with a donor-ally peer that has piece i is given by

$$\Big(\big(L - \mathbb{1}\left[Y^{(opt)} = 1\right]\big)\mu^{(\mathrm{tft})}x_W^{(S)}\Big)$$
$$\times \left(\frac{\sum_{\substack{V \in \mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T: i \in T}} x_V^{(T)}}{|\mathbf{x}|} \frac{|\mathbf{x}|}{|\mathbf{x}|-1} \right).$$

Here, the term in the first parenthesis is the aggregate rate with which a tit-for-tat link of a (W,S)-peer gets activated; and the term in the second parenthesis is the probability that a normal donor-ally peer of swarm-W with piece i is chosen after the activation.

Let us denote by $\Gamma_W^{(i)} = \Gamma_W^{(i)}(\mathbf{x})$ the aggregate rate at which a donor-ally peer of swarm-W with piece i **push-contacts** in the network. The exact form of $\Gamma_W^{(i)}$ is complicated, however, from i), ii), iii) and our descriptions of tit-for-tat and optimistic-unchoke, we can bound it both from above and below. Let

$$\underline{\Gamma}_W^{(i)}(\mathbf{x}) := LU + \kappa_p \xi \left(c_W^{(i)} + d_W^{(i)}\right) \geqslant LU + \kappa_p c_W^{(i)}, \tag{19}$$

$$\text{and } \overline{\Gamma}_W^{(i)}(\mathbf{x}) := LU + \kappa_1 \xi \left(c_W^{(i)} + d_W^{(i)}\right) \leqslant LU + 2\kappa_1 \left(c_W^{(i)} + d_W^{(i)}\right), \tag{20}$$

where, we have used the shortand,

$$\xi := \xi(\mathbf{x}) = \frac{|\mathbf{x}|}{|\mathbf{x}|-1}, \qquad (|\mathbf{x}| \geqslant 2)$$

$$\text{and } \kappa_a := 2\left(L - \mathbb{1}\left[Y^{(opt)} = 1\right]\right)\mu^{(\mathrm{tft})}a + \mathbb{1}\left[Y^{(opt)} = 1\right]\mu^{(\mathrm{opt})}. \quad (a \in [0,1]). \tag{21}$$

Intuitively, for any $a \in [0,1]$, κ_a denotes the aggregate rate with which a normal peer would push-contact another normal peer if the network were to enforce a push-contact probability of a in all tit-for-tat contacts (regardless of the cache-contents of the interacting peers). It is clear that

$$\underline{\Gamma}_W^{(i)} \leqslant \Gamma_W^{(i)} \leqslant \overline{\Gamma}_W^{(i)}. \tag{22}$$

and

$$\frac{\overline{\Gamma}_W^{(i)}}{\underline{\Gamma}_W^{(i)}} = \frac{LU + \kappa_1 \xi \left(c_W^{(i)} + d_W^{(i)}\right)}{LU + \kappa_p \xi \left(c_W^{(i)} + d_W^{(i)}\right)} \leqslant \Upsilon := \begin{cases} 1 + \frac{2(L-1)\mu^{(\mathrm{tft})} + \mu^{(\mathrm{opt})}}{2(L-1)\mu^{(\mathrm{tft})}p + \mu^{(\mathrm{opt})}} & \text{if } Y^{(opt)} = 1, \\ 1 + \frac{1}{p} & \text{if } Y^{(opt)} = 0. \end{cases} \tag{23}$$

Here, $\Upsilon = \Upsilon(L, Y^{(opt)}, \mu^{(\mathrm{tft})}, \mu^{(\mathrm{opt})}, p)$ is a constant that depends on the model parameters.

After the (W,S)-peer has downloaded piece i, depending on S, it will either remain in the network or leave it immediately. Therefore, we have two cases.

i) If $W \backslash S \supsetneq \{i\}$, the peer stays in the system. Let us denote this transition by $\{(S, i+)_W\}$ and the corresponding rate by $q_W^{(S,i+)}$. Based on the description of (swarm-based) RFwPMS, $q_W^{(S,i+)}$ depends on whether $i \in R_W$ or $i \in R_W^c$. If $i = r$ is some rare piece in R_W, then we have

$$\frac{x_W^{(S)}}{|\mathbf{x}|}\underline{\Gamma}_W^{(r)}\Upsilon \geqslant \frac{x_W^{(S)}}{|\mathbf{x}|}\overline{\Gamma}_W^{(r)} \geqslant q_W^{(S,r+)} \geqslant$$
$$\geqslant \frac{x_W^{(S)}}{|\mathbf{x}|}\left[\frac{LU\mathbb{1}\left[r \in \underline{R}_{(\mathcal{F},W,S)}\right]}{|\underline{R}_{(\mathcal{F},W,S)}|} + \kappa_p \xi \sum_{\substack{V \in \mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T: r \in T}} \frac{\mathbb{1}\left[r \in \underline{R}_{(T,W,S)}\right]}{|\underline{R}_{(T,W,S)}|} x_V^{(T)}\right], \tag{24}$$

where the two indicator terms ensure that piece r is transferred *only if it is the rarest of all the available (novel) rare pieces*. Importantly, when the chunk-distribution in swarm-W is uniform i.e., $\overline{c}_W = \underline{c}_W$, by definition, $R_W = W$ and the above expression assumes a more tractable form. Therefore, for each swarm-W, we partition the state space into two regions, namely $\mathcal{S}_W^{(1)}$ and $\mathcal{S}_W^{(2)}$, where $\mathcal{S}_W^{(1)} := \{\mathbf{x} \in \mathcal{S} : R_W(\mathbf{x}) \subsetneq W\}$, and $\mathcal{S}_W^{(2)} := \{\mathbf{x} \in \mathcal{S} : R_W(\mathbf{x}) = W\}$.[19] If $i = n$ is some non-rare piece in R_W^c, then

$$\underline{q}_W^{(S,n+)} \leqslant q_W^{(S,n+)} \leqslant \frac{x_W^{(S)}}{|\mathbf{x}|}\overline{\Gamma}_W^{(n)}\zeta_W^{(n)} \leqslant \frac{x_W^{(S)}}{|\mathbf{x}|}\underline{\Gamma}_W^{(n)}\Upsilon\zeta_W^{(n)}, \tag{25}$$

where,

$$\underline{q}_W^{(S,n+)} := \frac{x_W^{(S)}}{|\mathbf{x}|}\left[\frac{LU\mathbb{1}\left[R_W \backslash S = \varnothing\right]}{|W \backslash S|} + \kappa_p \xi \sum_{\substack{V \in \mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T: n \in T}} \frac{x_V^{(T)}\mathbb{1}\left[(T \cap R_W) \backslash S = \varnothing\right]}{|(T \cap W) \backslash S|}\right]\zeta_W^{(n)}. \tag{26}$$

ii) If $W \backslash S = \{i\}$, the only piece of file-W that is missing with the (W,S)-peer is piece i. Consequently, for both $\mathbf{x} \in \mathcal{S}_W^{(1)}$ and $\mathbf{x} \in \mathcal{S}_W^{(2)}$, piece i is the rarest piece transferable to the (W,S)-peer—which leaves the network immediately upon downloading it. Let us denote this transition by $\{(S, i-)_W\}$ and its corresponding rate by $q_W^{(S,i-)}$. Then,

$$\frac{|\mathbf{x}|_W}{|\mathbf{x}|}\frac{x_W^{(S)}}{|\mathbf{x}|_W}\underline{\Gamma}_W^{(i)}a_W^{(i)} \leqslant q_W^{(S,i-)} \leqslant \frac{|\mathbf{x}|_W}{|\mathbf{x}|}\frac{x_W^{(S)}}{|\mathbf{x}|_W}\underline{\Gamma}_W^{(i)}\Upsilon a_W^{(i)}, \tag{27}$$

where, we have used the shorthand,

$$a_W^{(i)} = a_W^{(i)}(\mathbf{x}) := \zeta_W^{(i)}\mathbb{1}\left[i \in R_W^c\right] + \mathbb{1}\left[i \in R_W\right]. \tag{28}$$

[19] When $\mathbf{x} \in \mathcal{S}_W^{(2)}$, both the indicator terms in (24) evaluate to 1.

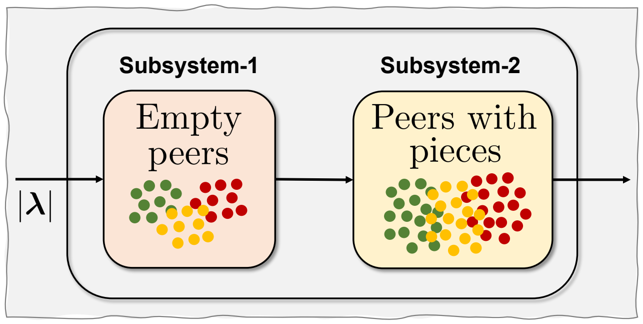


Fig. 2. Subsystems 1 and 2 in tandem.

c) *Download of Extra Piece*: Recall that extra pieces are preferred only when no pieces from the file of primary interest are transferable. We shall see that the drift analysis of our Lyapunov function does not depend on the download of extra pieces. Hence, we skip listing the associated rates.

F. Last Piece Syndrome vs. Tit-for-Tat

While it is clear that a tit-for-tat mechanism helps combat the *last piece syndrome* (LPS), the issue of quantifying its efficacy via a suitable model was left as an open problem in [5]. Our stochastic model described above is partly motivated by this question. Here, we show that introducing tit-for-tat does not prevent instability if a work-conserving piece-selection policy (like RF and RN) is used.

Proposition 1: Consider the multi-swarm model as described in Section II, with a hard tit-for-tat mechanism ($p = 0$), and without optimistic-unchoking ($Y^{(opt)} = 0$). If $|\boldsymbol{\lambda}| > LU$, then the network is unstable under any piece-selection policy.

Proof: We divide the network into two subsystems, where subsystem-1 consists of empty peers and subsystem-2 consists of the rest of the peers. The two subsystems are connected in tandem as shown in Fig. 2. The arrival-rate of incoming peers in subsystem 1 is $|\boldsymbol{\lambda}|$ and the departure rate is upper-bounded by LU (at each tick of the fixed seed's contact-link, at most 1 empty peer can depart from subsystem-1 and there are L such links). It is well-known that subsystem-1 is unstable if $|\boldsymbol{\lambda}| > LU$, rendering the network unstable. Since this instability is manifested by the build-up of empty peers, we call this phenomenon, *first piece syndrome (FPS).* ∎

Fig.3 shows FPS manifesting in a single-swarm network when it uses a hard tit-for-tat mechanism without optimistic-unchoking.

Proposition 2: Consider a single-swarm model (swarm-W) of the type presented in Section II, (a) with a soft tit-for-tat mechanism ($p > 0$) or, (b) with a hard tit-for-tat mechanism ($p = 0$) and optimistic-unchoking ($Y^{(opt)} = 1$). If a work-conserving piece-selection policy is used, then the network is unstable whenever $\lambda_W > LU$.

Proof: The method of the proof remains the same as in [5, Proposition 2.1 (i)]. In soft tit-for-tat mechanism, a young peer can get a piece from a one-club peer with probability at least $p > 0$ whereas in a hard tit-for-tat mechanism with optimistic-unchoking, it can get a piece from a one-club peer via the optimistic-unchoke. In either case, by taking the initial size of the one-club peers suitably large, it can be shown that the LPS event has a positive probability, establishing the transience of the system. ∎

Table I summarizes the results of Propositions 1 and 2 together which motivate the design of a piece-selection policy that is provably stable in file-sharing networks whose piece-exchange mechanism is either "soft tit-for-tat" or "hard tit-for-tat with optimistic-unchoking".

III. Stability of Swarm-Based RFwPMS in the Multi-Swarm

In this section, we present our main result on the stability of swarm-based RFwPMS. The proof is established using the Foster-Lyapunov theorem [18], [19]; see Proposition 6 in Appendix F.

Theorem 1: For the multi-swarm model with non-persistent peers as described in Section II, swarm-based RFwPMS is stable over the parameter region,

$$\mathcal{P}^{(stab)} := \{U > 0, \boldsymbol{\lambda} > \mathbf{0}, L \in \mathbb{N}, \kappa_p > 0, K_W \in \mathbb{N}\backslash\{1\} \,\forall\, W \in \mathcal{W}\}. \tag{29}$$

Remark: Here, $\kappa_p > 0$ includes

- "soft tit-for-tat": $p\mu^{(\text{tft})} > 0$ AND ($L = 1, Y^{(opt)} = 0$ OR $L > 1$);
- "hard tit-for-tat with optimistic-unchoking": $p = 0$ AND $Y^{(opt)}\mu^{(\text{opt})} > 0$; and
- stability result of our earlier work [14]: $Y^{(opt)} = 1, L = 1, \mu^{(\text{opt})} > 0$.

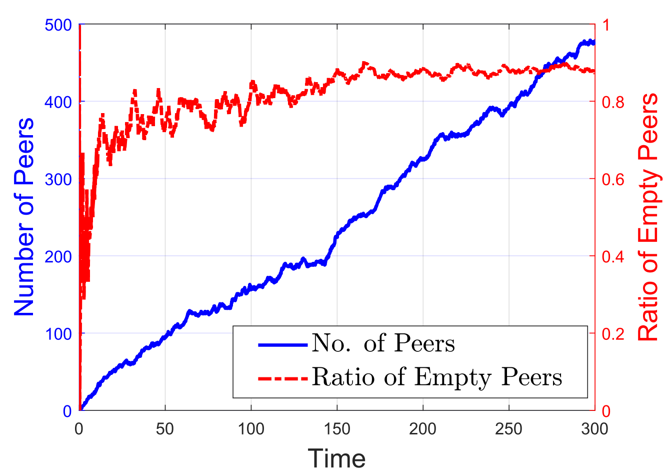


Fig. 3. First Piece Syndrome (FPS) in a single-swarm Network with hard tit-for-tat ($p = 0$) and no optimistic-unchoking ($Y^{(opt)} = 0$). Assignment of other model parameters was $W = [10]$, $\lambda_W = 4$, $U = \mu^{(\text{tft})} = 1$, $\mu^{(\text{opt})} = \frac{1}{3}$, $L = 3$.

TABLE I
Cause of Instability for High Arrival-Rate $|\boldsymbol{\lambda}|$ Under Work-Conserving Piece-Selection Policies

Tit-for-tat	Optimistic-unchoking	
	Yes	No
Yes	LPS	LPS if $p > 0$. FPS if $p = 0$.[a]
No	LPS	—

[a]Holds for all piece-selection policies.

The proof of Theorem 1 uses a novel Lyapunov function that is based on the refinement of our earlier intuition in [14]. The key ideas of the proof are as follows:

- In the case of large total-mismatch, extract negative drift from the download of rare pieces, and use the non-rares sharing factor to upper-bound the positive component of the drift from the download of non-rare pieces; (this is proved in Lemma 7 in Appendix B and Lemma 9 in Appendix D); and
- When the total-mismatch is not too large, extract negative drift from the download of all primary pieces (this is incorporated into the proof of Lemma 8 in Appendix C).

The Lyapunov function we use is given by

$$V(\mathbf{x}) := \sum_{W\in\mathcal{W}} V_W(\mathbf{x}), \tag{30}$$

where

$$\begin{aligned} V_W(\mathbf{x}) &:= V_{W,1}(\mathbf{x}) + V_{W,2}(\mathbf{x}) + V_{W,3}(\mathbf{x}),\\ V_{W,1}(\mathbf{x}) &:= \left((M_W - \eta\bar{c}_W)^+\right)^2,\\ V_{W,2}(\mathbf{x}) &:= C_W^{(1)}\left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right),\\ V_{W,3}(\mathbf{x}) &:= C_W^{(2)}\left(C_W^{(3)} - P_W^{(\mathrm{tot})}\right)^+.\text{[20]} \end{aligned} \tag{31}$$

Here, $\eta \in (0,1)$, and $C_W^{(1)}, C_W^{(2)}, C_W^{(3)} \in \mathbb{R}_{>0}$ are suitable constants. Instead of presenting the relations satisfied by $C_W^{(1)}$, $C_W^{(2)}$, and $C_W^{(3)}$, in the onset, we will go through the analysis and present them naturally as and when they are needed.

A. Intuition Behind Choice of Lyapunov Function

The exact details of the Lyapunov function given by (30) and (31) are complicated. However, the following lines will prove to be of substantial value in following the proof of Theorem 1.

- Given each swarm maintains its own chunk-counts, V being the superposition of V_W's, reflects the intuition that the network shall be stable if each swarm is stable.
- To develop some intuition behind the choice of V_W, let us consider a multi-swarm network in which each swarm's receiving-ally-set contains that swarm only (later in Section IV, we call this the setting of selfish swarms). In this case, the non-rares sharing factor $\zeta_W^{(n)}$ (assuming $\beta_W > 0$) takes the form $\exp\left(-\frac{\overline{m}_W}{\beta_W K_W}\right)$.
  - Given that swarm-based RFwPMS would allow the download of non-rare pieces with high probability whenever the largest-mismatch $\overline{m}_W \geqslant \frac{M_W}{K_W-1}$ is relatively small, it is reasonable, as a starting point, to guess a term of the form $(M_W)^2$. Unfortunately, as it turns out, with the bounds that we have derived, working with $(M_W)^2$ fails in satisfying the unit-transition drift stability conditions. The primary reason behind this is the fact *the rarity of the rarest piece and the abundance of the non-rare pieces are not related to each other*, that is, knowing that $\bar{c}_W$ is large does not necessarily imply that the largest-mismatch $\overline{m}_W$ is large as well. We circumvent this issue by using $V_{W,1}$ (see (31)) as a "proxy" for $(M_W)^2$. Specifically, by using $\left((M_W - \eta\bar{c}_W)^+\right)^2$ instead of $(M_W)^2$, we avoid any drift penalization (i.e., positive drift) whenever M_W is less than $\eta\bar{c}_W$. Therefore, in essence, whenever $M_W < \eta\bar{c}_W$, as concerns $V_{W,1}$, our Lyapunov function discards the distinction between rare and non-rare pieces. On the other hand, if $M_W \geqslant \eta\bar{c}_W$, this implies that $\overline{m}_W \geqslant \frac{\eta}{(K_W-1)}\bar{c}_W$, and now, the abundance of non-rare pieces, i.e., large $\bar{c}_W$, also implies a high largest-mismatch $\overline{m}_W$. This will trigger our non-rares sharing factor $\zeta_W^{(n)}$ which will ensure that the positive component of the drift from the non-rare pieces decays to zero with increasing $\bar{c}_W$.
  - In contrast to $V_{W,1}$, the second term, $V_{W,2}$, promotes the download of every piece of file-W. The intuition behind including $V_{W,2}$ is to ensure a negative drift is generated from every piece download. This helps us generate sufficient negative drift in the case when $V_{W,1}$ is inactive, i.e., when $M_W < \eta\bar{c}_W$. Furthermore, an important technical purpose that it serves is to ensure that V has finite level-sets.
  - Finally, the last term $V_{W,3}$ handles all the states in which swarm-W is seriously piece-deprived—states in which very few pieces of file-W are present with swarm-W).

Next, we present Lemmas 2 and 3 together which establish that V has finite level-sets.

Lemma 2 (Bounds on Chunk-Counts): For all $\mathbf{x}$, the total chunk-count in swarm-W is upper-bounded by $(K_W-1)|\mathbf{x}|_W$. Consequently, the fraction of swarm-W peers who have the rarest piece, i.e., $\underline{\pi}_W$, is upper-bounded by $(K_W-1)/K_W$ and the fraction of swarm-W peers who are missing the rarest piece is lower-bounded by $1/K_W$.

Proof: For all $\mathbf{x}$, we can lower bound the total chunk-count in swarm-W as follows,

$$\begin{aligned} K_W\underline{c}_W \leqslant \sum_{i\in W} c_W^{(i)} = P_W^{(\mathrm{tot})} = \sum_S x_W^{(S)}|S\cap W| \leqslant \dots\\ \leqslant (K_W-1)\sum_S x_W^{(S)} = (K_W-1)|\mathbf{x}|_W. \end{aligned}$$

Hence, $\underline{\pi}_W \leqslant \frac{K_W-1}{K_W}$ and $1-\underline{\pi}_W \geqslant \frac{1}{K_W}$. The second inequality is true because $W\backslash S \neq \varnothing$. ■

Lemma 3 (Finite Level-Sets of V): $V(\mathbf{x}) \to \infty$ as $|\mathbf{x}| \to \infty$.

Proof: We have $P_W^{(\mathrm{tot})} \leqslant (K_W-1)|\mathbf{x}|_W$ for every $\mathbf{x}$. So, $V_W(\mathbf{x}) \geqslant V_{W,2}(\mathbf{x}) \geqslant C_W^{(1)}|\mathbf{x}|_W \to \infty$ as $|\mathbf{x}|_W \to \infty$. Since $|\mathbf{x}| \to \infty$ only if $|\mathbf{x}|_W \to \infty$ for some $W \in \mathcal{W}$, it follows that $V(\mathbf{x}) \to \infty$ as $|\mathbf{x}| \to \infty$. Consequently, for all $C \in \mathbb{R}_{>0}$, the set $\{\mathbf{x} : V(\mathbf{x}) \leqslant C\}$ is finite. ■

B. Upper Bounds on Potential Changes

Having established the finite level-sets of V, we evaluate the potential change for each possible transition. Note that, with our choice of V, any transition that occurs in swarm-W, affects the term V_W only.

[20]For all $d \in \mathbb{R}$, we use the notation $(d)^+ := \max\{d, 0\}$.

1) Arrival of an Empty Peer: The arrival of a $(W,\varnothing)$-peer results in a unit increase in $|\mathbf{x}|_W$ but does not affect any chunk-count. Therefore, $M_W, \bar{c}_W, P_W^{(\mathrm{tot})}$ stay the same. The potential change as a result of this transition, denoted by $\Delta V_W^{(\varnothing,+)}$, is component-wise given by

$$\Delta V_{W,1}^{(\varnothing,+)} = 0.$$
$$\Delta V_{W,2}^{(\varnothing,+)} = K_W C_W^{(1)}.$$
$$\Delta V_{W,3}^{(\varnothing,+)} = 0.$$

Overall,

$$\Delta V_W^{(\varnothing,+)} = K_W C_W^{(1)}. \tag{32}$$

2) Download of a Single Piece $i \in W$ with Peer-Departure: Assume that a (W,S)-peer with $W \backslash S = \{i\}$ downloads piece i and leaves the system. The departure causes a unit-decrease in $|\mathbf{x}|_W$; a unit-decrease in chunk-count of every piece $j \in W\backslash\{i\}$. The chunk-count of piece i, i.e., $c_W^{(i)}$, stays the same. Therefore, $P_W^{(\mathrm{tot})}$ decreases by $K_W - 1$. The resulting potential change, denoted by $\Delta V_W^{(S,i-)}$, can be component-wise upper-bounded as

$$\Delta V_{W,2}^{(S,i-)} \leqslant -C_W^{(1)}.$$
$$\begin{aligned}
&\Delta V_{W,3}^{(S,i-)} \\
&= C_W^{(2)}\left[\left(C_W^{(3)} - P_W^{(\mathrm{tot})} + (K_W - 1)\right)^+ \right. \\
&\qquad\qquad\qquad \left. - \left(C_W^{(3)} - P_W^{(\mathrm{tot})}\right)^+\right] \\
&\leqslant C_W^{(2)}(K_W-1)\mathbb{1}\left[C_W^{(3)} + K_W - 1 \geqslant P_W^{(\mathrm{tot})}\right] \\
&\leqslant C_W^{(2)} K_W \mathbb{1}\left[C_W^{(3)} + 2K_W \geqslant P_W^{(\mathrm{tot})}\right] =: I_W^{(2)}(\mathbf{x}),
\end{aligned}$$

where $\Delta V_{W,1}^{(S,i-)}$ depends on whether i is a rare or non-rare piece. When $i = r$ is some rare piece, $\Delta V_{W,1}^{(S,r-)}$ further depends on whether swarm-W's chunk-distribution is uniform or non-uniform, i.e., whether $\mathbf{x} \in \mathcal{S}_W^{(1)}$ or $\mathbf{x} \in \mathcal{S}_W^{(2)}$. If $\mathbf{x} \in \mathcal{S}_W^{(1)}$, then $c_W^{(i)} \neq \bar{c}_W$, therefore the departure reduces both the total mismatch and the highest chunk-count by 1. This gives

$$\begin{aligned}
&\Delta V_{W,1}^{(S,r-)} \\
&= \left((M_W - \eta\bar{c}_W - (1-\eta))^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2 \\
&\leqslant \begin{cases} (1-\eta)^2 - 2(1-\eta)\left(M_W - \eta\bar{c}_W\right) & \text{if } M_W \geqslant \\ & (1-\eta) + \eta\bar{c}_W \\ 0 & \text{otherwise.} \end{cases} \\
&\leqslant \begin{cases} 2(1-\eta)^2 - 2(1-\eta)\left(M_W - \eta\bar{c}_W\right) & \text{if } M_W \geqslant \\ & 2(1-\eta) + \eta\bar{c}_W \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

If $\mathbf{x} \in \mathcal{S}_W^{(2)}$, the download of piece r makes it the only non-rare piece in the next state with its chunk-count equal to $\bar{c}_W$. The chunk-count of every other piece decreases to $\bar{c}_W - 1$. The total mismatch M_W changes from 0 to $K_W - 1$. Consequently,

$$\begin{aligned}
\Delta V_{W,1}^{(S,r-)} &= \left((K_W - 1 - \eta\bar{c}_W)^+\right)^2 - \left((-\eta\bar{c}_W)^+\right)^2 \\
&\leqslant K_W^2 \leqslant 4K_W^2.
\end{aligned}$$

Combining the two cases for $i = r$, we have

$$\begin{aligned}
&\Delta V_{W,1}^{(S,r-)} \\
&\leqslant \begin{cases} 2(1-\eta)^2 - 2(1-\eta)\left(M_W - \eta\bar{c}_W\right) \leqslant 0 & \text{if } M_W \geqslant \\ & 2(1-\eta) + \eta\bar{c}_W, \\ 4K_W^2 & \text{otherwise.} \end{cases} \\
&=: \underline{\psi}_W(\mathbf{x}).
\end{aligned}$$

When $i = n$ is some non-rare piece[21], then the highest chunk-count $\bar{c}_W = c_W^{(i)}$ stays the same and the total mismatch M_W increases by $K_W - 1$. Overall, this causes an increment of $K_W - 1 \leqslant K_W$ in $M_W - \eta\bar{c}_W$. Therefore,

$$\begin{aligned}
&\Delta V_{W,1}^{(S,n-)} \\
&= \left((M_W - \eta\bar{c}_W + K_W)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2 \\
&\leqslant \begin{cases} K_W^2 + 2K_W\left(M_W - \eta\bar{c}_W\right) & \text{if } M_W - 2(1-\eta) \geqslant \eta\bar{c}_W, \\ (K_W+2)^2 & \text{otherwise.} \end{cases} \\
&\leqslant \begin{cases} K_W^2 + 2K_W\left(M_W - \eta\bar{c}_W\right) & \text{if } M_W - 2(1-\eta) \geqslant \eta\bar{c}_W, \\ 4K_W^2 & \text{otherwise} \end{cases} \\
&=: \overline{\psi}_W(\mathbf{x}).
\end{aligned}$$

To utilize the negative drift from the download of swarm-W's rare pieces, we partition the state space into regions $\mathcal{R}_W^{(1)} := \{\mathbf{x} : M_W - 2(1-\eta) \geqslant \eta\bar{c}_W\}$ and $\mathcal{R}_W^{(2)} := \mathcal{S}\backslash\mathcal{R}_W^{(1)}$. (Note that $\mathcal{R}_W^{(1)} \subseteq \mathcal{S}_W^{(1)}$ and $\mathcal{R}_W^{(2)} \supseteq \mathcal{S}_W^{(2)}$). We can now write

$$\begin{aligned}
&\Delta V_W^{(S,i-)} \\
&\leqslant \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] - C_W^{(1)} + I_W^{(2)} \\
&= \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 + \dots \\
&\quad + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right]\left(\overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]\right) - \dots \\
&\quad - C_W^{(1)} + I_W^{(2)} =: \Psi_W^{(i-)}(\mathbf{x}).
\end{aligned} \tag{33}$$

3) Download of a Single Piece $i \in W$ without Peer Departure: Here, a (W,S)-peer with $W \backslash S \supsetneq \{i\}$ downloads piece i and remains in the system. The total number of swarm-W peers remains the same and the chunk-count of piece i (and thus $P_W^{(\mathrm{tot})}$) gets incremented by 1. The resulting potential change, denoted by $\Delta V_W^{(S,i+)}$ is component-wise upper-bounded as

$$\Delta V_{W,2}^{(S,i+)} = -C_W^{(1)},$$
$$\begin{aligned}
\Delta V_{W,3}^{(S,i+)} &= C_W^{(2)}\left[\left(C_W^{(3)} - P_W^{(\mathrm{tot})} - 1\right)^+ - \left(C_W^{(3)} - P_W^{(\mathrm{tot})}\right)^+\right] \\
&\leqslant -C_W^{(2)}\mathbb{1}\left[C_W^{(3)} - 1 \geqslant P_W^{(\mathrm{tot})}\right] \\
&\leqslant -C_W^{(2)}\mathbb{1}\left[C_W^{(3)} - 2 \geqslant P_W^{(\mathrm{tot})}\right] =: -I_W^{(1)}(\mathbf{x}).
\end{aligned}$$

Like $\Delta V_{W,1}^{(S,i-)}$, one can show that $\Delta V_{W,1}^{(S,i+)}$ is upper-bounded by $\overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]$. Therefore,

$$\Delta V_W^{(S,i+)}$$

[21] Non-rare pieces in swarm-W exist if and only if $\mathbf{x} \in \mathcal{S}_W^{(1)}$.

$$
\begin{aligned}
&\leqslant \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] - C_W^{(1)} - I_W^{(1)} \\
&= \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 + \ldots \\
&\quad + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right]\left(\overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]\right) \\
&\quad - C_W^{(1)} - I_W^{(1)} =: \Psi_W^{(i+)}(\mathbf{x}). \qquad (34)
\end{aligned}
$$

Looking at (34) and (33), we note that by choosing $C_W^{(1)} > 4K_W^2$, say $C_W^{(1)} \geqslant 8K_W^2$, we ensure that the term $\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)}$ is non-positive. Intuitively, over the region $\mathcal{R}_W^{(2)}$, as concerns the potential-change, our upper-bounds treat the non-rare and rare pieces in the same way.

4) Download of a Single Extra Piece: It can be observed that for any swarm-W, no potential change is induced in V_W by the download of its extra pieces.

5) Two-Sided Piece Exchanges: Now, we consider the state-transitions that involve two mutually-ally peers, each transferring a piece to the other, as a result of some tit-for-tat contact initiated by one of them. In 4), we noted that no potential change is induced by the download of extra pieces. Therefore, as concerns the potential changes, all those transitions that involve the transfer of two extra pieces can be ignored. Similarly, those transitions that involve the transfer of an extra-piece coupled with some rare/non-rare piece can be viewed as (single) download of the accompanying rare/non-rare piece, which we have covered in 1)-3). We therefore focus only on those transitions that involve the transfer of two non-extra pieces. For this, consider two mutually-ally peers, say a (W,S)-peer missing piece $i \in W$ (and holding piece $j \in V$) and a (V,T)-peer missing piece j (and holding piece i), who download the respective missing pieces as a result of a tit-for-tat contact initiated by one of them. We denote this transition by $\{(S,i\pm)_W,(T,j\pm)_V\}$ where "$-$" indicates departure of the peer and "$+$" indicates otherwise. From 1)-3), we note that the potential change induced by this piece exchange will depend on

- The type of piece i ($i = r \in R_W$ or $i = n \in R_W^c$),
- The type of piece j ($j = r' \in R_V$ or $j = n' \in R_V^c$),
- Whether (W,S)-peer stays in or leaves the system, ($W \backslash S \supsetneq \{i\}$ or $W \backslash S = \{i\}$), and
- Whether (V,T)-peer stays in or leaves the system ($V \backslash T \supsetneq \{j\}$ or $V \backslash T = \{j\}$).

Table II summarizes the upper-bounds on the potential change associated with each of those cases. Importantly, in each case, the upper-bound has been decomposed as the sum of two upper-bounds for the corresponding single piece download events. The bounds are obvious for the case $W \neq V$ and with some work, are fairly easy to establish when $W = V$. For completeness, they are derived in Appendix A.

C. Upper Bounds on Unit-Transition Drifts

Having established upper-bound estimates of the potential changes, we can now proceed to evaluate the unit-transition drift $QV(\mathbf{x})$. We have established that for any swarm W, the upper-bound on potential-change induced by the download of piece $i \in W$ depends only on its type ($i \in R_W$ or $i \in R_W^c$), whether or not the download is accompanied with the departure of the peer ($i+$ or $i-$), and the current chunk-distribution in swarm-W ($\Psi_W^{(r+)}$ and $\Psi_W^{(r-)}$ depend on whether $\mathbf{x} \in \mathcal{R}_W^{(1)}$ or $\mathbf{x} \in \mathcal{R}_W^{(2)}$). Thus, for any state $\mathbf{x}$, we can write

$$
\begin{aligned}
QV(\mathbf{x}) = &\sum_{W \in \mathcal{W}} QV_W^{(\varnothing,+)} + \sum_{W \in \mathcal{W}} \underbrace{QV_W^{(R_W,+)} + QV_W^{(R_W^c,+)}}_{:=QV_W^{(W,+)}} \\
&+ \sum_{W \in \mathcal{W}} \underbrace{QV_W^{(R_W,-)} + QV_W^{(R_W^c,-)}}_{:=QV_W^{(W,-)}}, \qquad (35)
\end{aligned}
$$

where

$$
\begin{aligned}
QV_W^{(\varnothing,+)}(\mathbf{x}) &:= q_W^{(\varnothing,+)} \Delta V_W^{(\varnothing,+)}, \\
QV_W^{(R_W,+)}(\mathbf{x}) &:= \sum_{\substack{r \in R_W, \\ W \backslash S \supsetneq \{r\}}} q_W^{(S,r+)} \Delta V_W^{(S,r+)}, \\
QV_W^{(R_W,-)}(\mathbf{x}) &:= \sum_{\substack{r \in R_W, \\ W \backslash S = \{r\}}} q_W^{(S,r-)} \Delta V_W^{(S,r-)}, \qquad (36)\\
QV_W^{(R_W^c,+)}(\mathbf{x}) &:= \sum_{\substack{n \in R_W^c, \\ W \backslash S \supsetneq \{n\}}} q_W^{(S,n+)} \Delta V_W^{(S,n+)}, \\
QV_W^{(R_W^c,-)}(\mathbf{x}) &:= \sum_{\substack{n \in R_W^c, \\ W \backslash S = \{n\}}} q_W^{(S,n-)} \Delta V_W^{(S,n-)}.
\end{aligned}
$$

1) Empty Peer Arrivals: Let $C^{(1)} := \max_{W \in \mathcal{W}} K_W C_W^{(1)}$. Using (18) and (32),

$$
\sum_{W \in \mathcal{W}} QV_W^{(\varnothing,+)} \leqslant |\boldsymbol{\lambda}| C^{(1)}. \qquad (37)
$$

2) Downloads of Primary Pieces without Peer Departure: The potential-change from the download of rare pieces without accompanying peer-departure via the rarest-first (RF) mechanism is non-positive. ($\Psi_W^{(r+)}$ is always non-positive). In Lemma 7 in Appendix B, we upper-bound the corresponding drift by what we would have obtained if we were to replace RF by random-novel (RN), i.e.,

$$
\begin{aligned}
QV_W^{(R_W,+)} \leqslant &\sum_{\substack{r \in R_W, \\ W \backslash S \supsetneq \{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|} \left[\frac{LU}{|R_W \backslash S|} \right. \\
&\left. + \kappa_p \xi \sum_{\substack{V \in \mathcal{W}_W^{(ally\text{-}\downarrow)}, \\ T: r \in T}} \frac{x_V^{(T)}}{|(T \cap R_W) \backslash S|} \right] \Psi_W^{(r+)}. \qquad (38)
\end{aligned}
$$

Remark: Since (38) is also an upper-bound for the download of rare pieces using the random-novel (RN) mechanism, the

TABLE II
UPPER-BOUNDS ON POTENTIAL CHANGES ASSOCIATED WITH SIMULTANEOUS DOWNLOAD OF PIECES i AND j BY MUTUALLY ALLY-PEERS (W,S) AND (V,T) RESPECTIVELY

Category	Transition	Potential Change
1	$(i,+)_W,(j,+)_V$	$\Psi_W^{(i+)} + \Psi_V^{(j+)}$
2	$(i,+)_W,(j,-)_V$	$\Psi_W^{(i+)} + \Psi_V^{(j-)}$
	$(i,-)_W,(j,+)_V$	$\Psi_W^{(i-)} + \Psi_V^{(j+)}$
3	$(i,-)_W,(j,-)_V$	$\Psi_W^{(i-)} + \Psi_V^{(j-)}$

stability proof of RFwPMS also works for "RNwPMS" – a piece-selection policy same as RFwPMS except that rare pieces are downloaded randomly.

Using (25), (26), (34), and the fact that $\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)} \leqslant 0$, we get

$$QV_W^{(R_W^c,+)} \leqslant \sum_{\substack{n \in R_W^c, \\ W \setminus S \supsetneq \{n\}}} q_W^{(S,n+)} \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)}\right] + \sum_{\substack{n \in R_W^c, \\ W \setminus S = \{n\}}} \frac{x_W^{(S)}}{|\mathbf{x}|} \underline{\Gamma}_W^{(n)} \Upsilon \overline{\psi}_W \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \zeta_W^{(n)}. \quad (39)$$

Lemma 8 in Appendix C combines (38) and (39) to give

$$QV_W^{(W,+)} \leqslant \frac{|\mathbf{x}|_W}{|\mathbf{x}|} \sum_{i \in W} \left(\frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W} \left(1 - \pi_W^{(i)} - \gamma_W^{(i)}\right) \times \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]\right) + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)}\right] \right), \quad (40)$$

where, we have introduced $\gamma_W^{(i)}$ as a short-hand for $\frac{\sum_{S: W\setminus S = \{i\}} x_W^{(S)}}{|\mathbf{x}|_W}$, i.e., the fraction of swarm-W peers missing only piece i of file-W.

3) Downloads of Primary Pieces with Peer Departure: Using (27), (33), and the fact that $\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] \leqslant 0$, we have

$$QV_W^{(W,-)} \leqslant \frac{|\mathbf{x}|_W}{|\mathbf{x}|} \sum_{i \in W} \underline{\Gamma}_W^{(i)} a_W^{(i)} \gamma_W^{(i)} \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \left(\Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]\right) + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} + \Upsilon I_W^{(2)}\right]. \quad (41)$$

4) Combining Everything: Let us define

$$\Psi_W^{(i)}(\mathbf{x}) := \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]\right) - C_W^{(1)}. \quad (42)$$

Then, using (37), (40), and (41), in (35), we get

$$QV(\mathbf{x}) \leqslant \sum_{W \in \mathcal{W}} \frac{|\mathbf{x}|_W}{|\mathbf{x}|} \widetilde{QV}_W, \quad (43)$$

where

$$\widetilde{QV}_W(\mathbf{x}) := |\boldsymbol{\lambda}| C^{(1)} + \sum_{i \in W} \frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W} \left[\left(1 - \pi_W^{(i)}\right) \Psi_W^{(i)} - \left(1 - \pi_W^{(i)} - \gamma_W^{(i)}\right) I_W^{(1)} + \gamma_W^{(i)} K_W \Upsilon I_W^{(2)}\right]. \quad (44)$$

With this initial setup in place, the rest of the proof is provided as a succession of lemmas in Appendix E.

IV. Discussion

A. Sharing vs Suppression Trade-off for Non-Rare Pieces

Note that with $\alpha_W \in (0, 1]$ sufficiently small, if we take the limit $\beta_W \to \infty$, swarm-based RFwPMS converges to a swarm-based version of RF (which is unstable in high arrival-rate regime), whereas if we take the limit $\beta_W \to 0$, swarm-based RFwPMS converges to a swarm-based version of MS [11]. Thus, our expectation is that by choosing α_W sufficiently small, and β_W appropriately, a good trade-off between the sharing and suppression of non-rare pieces can be found. Via numerical simulations, we found that the expected sojourn time appears to be minimized with small $\alpha_W \approx 0$ and $\beta_W \approx 1.5$.

Remark: Swarm-based RFwPMS like MS is a centralized scheme as it requires the knowledge of piece distribution in the network; we assume that in practice, peers can keep such estimates either via gossiping or via a centralized tracker.

B. Rarest-First vs Random-Selection for Rare Pieces

Another notable observation about mode-suppression (from [11]) is that it performs random-selection on the set of available rare pieces. We assert that knowing the relative chunk-counts of the rare pieces should allow for more advantage than just random-selection (RN), and that a load-balancing scheme like RF will reduce the duration of the transient phase. This also gets manifested in Section VI-D2 where we compare the flash-crowd response of RNwPMS (replacing the selection of the rarest available rare piece by random-selection) with RFwPMS.

C. Inter-Swarm Collaboration

Our multi-swarm model in section II is general in terms of inter-swarm behavior of peers and their secondary download preferences. Here, we discuss three different behaviors that are covered:

a) *Altruistic Swarms*: In most wired P2P environments (e.g., the Internet), peers are generally insensitive to the consumption of their download and upload bandwidths, and they may download extra pieces in order to help other swarms. Such behavior can be captured in our model by setting $\mathcal{W}_W^{(ally\text{-}\uparrow)} = \mathcal{W}$ and $\mathcal{F}_W = \mathcal{F}$ for every $W \in \mathcal{W}$. From [8], a network in which all swarms are altruistic is called a *universal swarm network*.

b) *Opportunistic Swarms*: A different type of altruism is when peers do not download any extra pieces but share their pieces with those of other swarms who need them. We obtain this by setting $\mathcal{W}_W^{(ally\text{-}\uparrow)} = \mathcal{W}$ and $\mathcal{F}_W = W$ for every $W \in \mathcal{W}$.

c) *Selfish Swarms*: In wireless P2P networks, peers may be sensitive to the consumption of their download and upload bandwidths. This holds by setting $\mathcal{W}_W^{(ally\text{-}\uparrow)} = \{W\}$ and $\mathcal{F}_W = W$. Thus, the peers do not download any extra pieces nor do they upload any piece to other swarms.

D. Piece-Selection Policies for Excess-Cache

In our original version of RFwPMS, random-selection was assumed for the download of extra pieces, but the stability result in Theorem 1 extends to any piece-selection policy one may use for the excess-cache. Without loss of generality, consider a *piece-selection policy for the excess-cache*, whose transferable-set is denoted by $E\left(\mathbf{x}, \widehat{T}, W, S\right)$, i.e., E satisfies

$$E\left(\mathbf{x}, \widehat{T}, W, S\right) \subseteq \left(\widehat{T} \cap \mathcal{F}_W\right) \setminus (S \cup W).$$

Recall that the Lyapunov function given by (30) and (31) is not affected by a piece download from the set E. Consequently, all the bounds on $QV(\mathbf{x})$ including Lemma 15 still hold. The final step of combining Lemma 15 and Proposition 6 has to be modified, however, because under some policies, the Markov process may no longer be irreducible with the current definitions of the state space (17) and state (1). For instance, take a single-swarm network in which $E \equiv \varnothing$, and the incoming swarm, say swarm-W, satisfies $\mathcal{F}_W = \mathcal{F} \supsetneq W$. Then, the set of all states in which some peer holds a piece from the set $\mathcal{F} \backslash W$ is not reachable from all other states. In such cases, the set of all states that are reachable from the zero population state is a closed irreducible set of states and should be defined as the state space of the network. Stability then follows from combining Lemma 15 and Proposition 6. We summarize this discussion in the following proposition.

Proposition 3: For the multi-swarm network model with non-persistent peers as described in Section II, swarm-based RFwPMS with any piece-selection policy for the excess-cache, is stable over the parameter region, $\mathcal{P}^{(stab)}$.

E. Autonomous Swarms

One more case to consider is when all the swarms in the network operate in isolation from each other. Specifically, a peer belonging to swarm-W contacts and exchanges pieces with peers in the same swarm. The fixed seed, on the other hand, divides its uploading capacity across swarms, providing a static non-zero fraction of its total capacity to each swarm; optimal partition of the seed capacity is for future work. Such swarms are called *autonomous swarms* in [8]. The stability of swarm-based RFwPMS holds for such swarms as well.

Corollary 3.1: Consider a multi-swarm network where each swarm $W \in \mathcal{W}$ behaves autonomously and the seed has allocated a static non-zero fraction of its total capacity for each swarm-W (say $U_W > 0$). Then, the network is stable under swarm-based RFwPMS over the parameter region, $\mathcal{P}^{(stab)}$.

Proof: Each swarm-W can be considered as an isolated single-swarm network with fixed seed of capacity $U_W > 0$. Stability follows by applying Theorem 1 to each swarm. ∎

F. Optimistic-Unchoking and Tit-for-Tat

In our description of the Peer-Contact Policy in Section II-C, peer-(1) makes a tit-for-tat contact with peer-(2), it is assumed that the probability of peer-(k) committing to transfer a piece to peer-$(-k)$, in the event that it does not benefit from peer-$(-k)$ is some fixed value, $p \in [0, 1]$. However, from the stability analysis in Appendix E, one may note that this assumption can be relaxed in two ways. *i*) If the network enforces optimistic-unchoking on the incoming peers, i.e., $Y^{(opt)} = 1$, then the system is inherently stable regardless of what p value is chosen by each peer in any of its tit-for-tat contacts – indeed it may depend on the history of peer-(k)'s interactions with peer-$(-k)$. (Υ in (23) remains bounded). *ii*) In the case that $Y^{(opt)} = 0$, the stability of the network still holds with variable p values as long as peers are forced to choose p values larger than or equal to some positive threshold value $p_0 \in (0, 1]$. (Υ in (23) remains bounded). In particular, this means that our stability result holds for a history based tit-for-tat mechanism where each peer keeps track of its interactions with other peers. From *i*) and *ii*), we get the following proposition.

Proposition 4: Consider the multi-swarm network model with non-persistent peers (as described in Section II) with the change that now each peer, in any tit-for-tat contact, can choose a variable p value in $[p_0, 1]$. Then, swarm-based RFwPMS is stable over the parameter region,

$$\left\{U > 0, \boldsymbol{\lambda} > \mathbf{0}, L \in \mathbb{N}, \kappa_{p_0} > 0, K_W \in \mathbb{N}\backslash\{1\}\, \forall\, W \in \mathcal{W}\right\}.$$

G. Comparison with Work of [11], [12]

As mentioned in Section I-A, [12] introduced a threshold based version of mode-suppression, where the non-rare pieces are (completely) suppressed only when the largest-mismatch crosses a fixed constant threshold. Compared with this, swarm-based RFwPMS inhibits the replication of non-rare pieces by smoothly decreasing the non-rares sharing factor $\zeta_W^{(n)}$ in proportion with the largest-mismatch value. In Section VI-D2, we note that a smoother-suppression like this should be preferred compared to a threshold-based suppression.

Another notable observation is that [11], [12] perform random-selection on the set of available rare pieces. We assert that knowing the chunk distribution should allow for more advantage than just random-selection, and that a load-balancing scheme like rarest-first should do a better job in reducing the duration of transient phases. This gets manifested in Section VI-D2 and is also in line with the practical conclusions made in [20] (albeit without a theoretical backing).

H. Effect of Policy Parameters $\{\alpha_W\}$:

An interesting question is whether or not the stability result for swarm-based RFwPMS can be extended to the case when $\alpha_W = 0$ for at least one of the swarms. From Lemma 9, we see that the positive component of $QV_W^{(R_W^c)}$ can be upper-bounded by a fixed constant, namely

$$\begin{aligned}
D_W &= D_W^{(1)} + D_W^{(2)} \mathbb{1}\left[\mathcal{W}_W^{(ally\text{-}\downarrow)} \backslash \{W\} \neq \varnothing\right], \\
D_W^{(1)} &= 12 e^{-2} \eta^{-2} \Upsilon(LU + \kappa_p) \beta_W^2 K_W^7, \\
D_W^{(2)} &= 3 \frac{e^{-\left(1+\alpha_W^{-1}\right)}}{\alpha_W^{\alpha_W^{-1}}} \eta^{-1} \Upsilon(LU + \kappa_p) \beta_W^{1+\alpha_W^{-1}} K_W^{5+\alpha_W^{-1}}.
\end{aligned} \tag{45}$$

If a swarm-W does not have any donor-ally swarm besides itself, then D_W is independent of α_W. This immediately gives the below proposition.

Proposition 5: For the multi-swarm model with non-persistent peers as described in Section II, swarm-based RFwPMS is stable over the parameter region, $\mathcal{P}^{(stab)}$, where α_W can be chosen to be zero for any swarm which does not have any donor-ally swarms (besides itself).

When swarm-W has other donor-ally swarms, the constant $D_W \to \infty$ as $\alpha_W \to 0$. Based on this, one may speculate that the stability of swarm-based RFwPMS is coupled with the choice of $\{\alpha_W\}$'s in the sense that the state of the system can possibly spiral out into ever increasing loads if $\alpha_W = 0$ (for swarms W that have other donor-ally swarms also). However, this does not seem to be the case in the numerical simulations we have performed. Fig.4 shows one such numerical snapshot for illustration. This corroborates the assertion that the parameter α_W is a technical artifact for upper-bounding the positive drift from the download of non-rare pieces for ever increasingly large states which have large $d_W^{(n)}$. For the aforementioned reasons, we believe that swarm-based RFwPMS is stable in the setting when α_W's can assume zero values. We state this as a conjecture below.

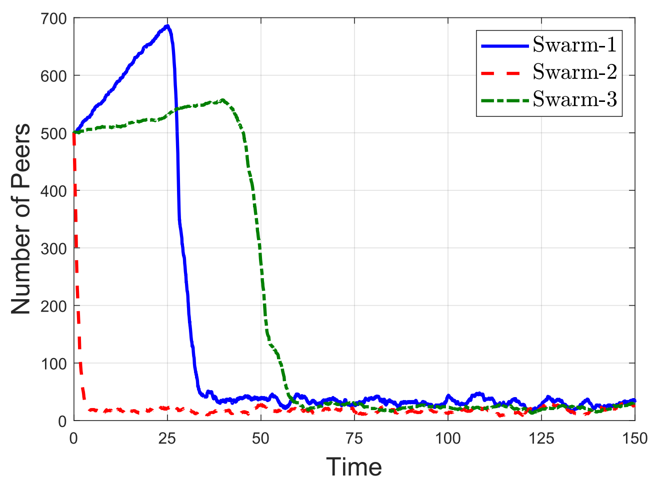


Fig. 4. Stability-check of a three-swarm network when $\alpha_W = 0$. The file-configurations were $W_1 = [10], W_2 = [6, 15], W_3 = [16, 30]$ and all swarms interacted altruistically. Initial state of each swarm was one-club—all peers in the swarm missing one specific piece of the file. Assignment of other model parameters was $\boldsymbol{\lambda} = (8, 4, 2), U = \frac{1}{3}, \mu^{(\text{tft})} = 1, \mu^{(\text{opt})} = \frac{1}{3}, L = 3, p = 0.5, Y^{(opt)} = 1, \beta_W = 1.5$.

Conjecture 1: For the multi-swarm model with non-persistent peers as described in Section II, swarm-based RFwPMS is stable over the parameter region, $\mathcal{P}^{(stab)}$, where α_W can be chosen to be zero for any $W \in \mathcal{W}$.

I. Scalability of swarm-based RFwPMS

Scalability is also necessary for a P2P network, i.e., the system-wide throughput should scale with the number of peers. For the multi-swarm model presented in Section II, this means that the expected steady-state sojourn time in the network should be upper-bounded by a constant independent of $\boldsymbol{\lambda}$. Unfortunately, analytically showing the scalability of stochastic models similar to the one we considered in this work is a hard endeavor and we leave this as an open problem. However, using an approach similar to that used in [12], we are able to show (under a mild assumption) the scalability of RFwPMS in the single-swarm case. This result is formally stated below.

Theorem 4: For the single-swarm case of the multi-swarm model described in Section II (i.e., $\mathcal{W} = \{W\}, \mathcal{F} = W, |\mathbf{x}|_W = |\mathbf{x}|$), provided that the empty-peers are always push-contacted in a tit-for-tat contact, swarm-based RFwPMS is scalable over its stability region. That is, the steady-state expected sojourn time, denoted by $\overline{T}_W$, is upper-bounded by a constant independent of the arrival-rate of empty-peers. In particular,

$$\overline{T}_W = \begin{cases} O\left(K_W^3\right) & \text{if } 2K_W^2\lambda_W \leqslant LU, \\ O\left(K_W^6\right) & \text{if } 2K_W^2\lambda_W > LU, \end{cases}$$

when $\beta_W = O\left(K_W^{-3}\right)$.

The proof is established using Kingman's moment bound [18] which uses the well-known technique of designing a suitable non-negative function and setting the steady-state expectation of its unit-transition drift equal to zero. Provided that the Markov chain is positive-recurrent, one usually hopes to obtain some convenient bound on the queue-length—although determining a good function (for a tight bound) is usually a challenging task. For reference, please see Proposition 6 in Appendix F.

In the next section, we provide the proof of Theorem 4.

V. Scalability of swarm-based RFwPMS in the Single-Swarm

Positive-recurrence of $\{\mathbf{x}(t) : t \geqslant 0\}$ is guaranteed by Theorem 1. Thus, to establish scalability, the key novelty is to first partition the stability region into two distinct load-regimes based on the size of total arrival-rate λ_W in comparison to the seed's aggregate upload capacity LU and then design a suitable non-negative function for each. The regimes are

i) $\mathcal{L}_1 := \mathcal{P}^{(stab)} \cap \left\{\lambda_W < 0.5K_W^{-2}LU\right\}$; and
ii) $\mathcal{L}_2 := \mathcal{P}^{(stab)} \cap \left\{\lambda_W \geqslant 0.5K_W^{-2}LU\right\}$.

Let us denote the stationary distribution of $\{\mathbf{x}(t) : t \geqslant 0\}$ by $\{\rho(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, so that the expected number of peers in steady-state is given by $\mathbb{E}_\rho[|\mathbf{x}|] = \mathbb{E}_\rho[|\mathbf{x}|_W]$, and by Little's Law, $\overline{T}_W = \lambda_W^{-1}\mathbb{E}_\rho[|\mathbf{x}|_W]$.

A. Low Load Regime $\mathcal{L}_1$

For regime $\mathcal{L}_1$, we have the following Lemma.

Lemma 5: For the single-swarm case of the multi-swarm model described in Section II, with swarm-based RFwPMS as the piece-selection policy, if the network parameters belong to $\mathcal{L}_1$, then the steady-state expected sojourn time $\overline{T}_W$ is independent of λ_W. In particular, $\overline{T}_W = O(K_W^3)$.

Proof: The intuition behind this upper-bound is as follows: In $\mathcal{L}_1$, the total arrival-rate λ_W is sufficiently small compared to the seed's aggregate upload-rate LU; therefore, we expect the seed to quickly flush the system without any contribution from the normal peers. Keeping this in mind, the $(\mu^{(\text{tft})}, \mu^{(\text{opt})})$-independent bound, $\frac{K_W^2(K_W+1)}{LU}$, can be established by using Kingman's moment bound on a suitable non-negative function. Let

$$V_1(\mathbf{x}) = \left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right)^2 + K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}. \quad (46)$$

It can be observed that V_1 heavily penalizes the states which have large number of missing pieces.[22] The arrival of an empty-peer in swarm W causes a unit-increase in $|\mathbf{x}|_W$. So, the potential change associated with the arrival of an empty-peer is given by

$$\begin{aligned}\Delta V_1^{(\varnothing,+)} &\leqslant \Phi^{(\varnothing,+)} \\ &:= K_W^2 + 2K_W\left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right) + K_W.\end{aligned} \quad (47)$$

The download of a piece $i \in W$ by any (W,S) peer causes a unit-decrease in the difference $K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}$. Therefore, the associated potential change is given by

$$\Delta V_1^{(S,i\pm)} \leqslant \underline{\Phi} := -2\left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right). \quad (48)$$

Upper-bounding $QV_1(\mathbf{x})$ by considering peer-arrivals, and downloads of the rarest-pieces (from $\underline{R}_{(\mathcal{F},W,S)}$) by the fixed seed only,

$$\begin{aligned}
&QV_1(\mathbf{x}) \\
&\leqslant \lambda_W \Phi^{(\varnothing,+)} + \sum_{\underline{r}\in\underline{R}_{(\mathcal{F},W,S)}} \left(1-\underline{\pi}_W\right) \frac{LU}{|\underline{R}_{(\mathcal{F},W,S)}|}\underline{\Phi} \\
&\overset{(a)}{=} \lambda_W \Phi^{(\varnothing,+)} + LU K_W^{-1}\underline{\Phi} \\
&\overset{(b)}{=} \lambda_W\left(K_W^2 + 2K_W\left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right) + K_W\right) \\
&\qquad - 2LUK_W^{-1}\left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right) \\
&= \lambda_W(K_W^2 + K_W) \\
&\qquad - 2\left(LUK_W^{-1} - \lambda_W K_W\right)\left(K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}\right) \\
&\overset{(c)}{\leqslant} \lambda_W(K_W^2 + K_W) - K_W^{-1}LU|\mathbf{x}|_W.
\end{aligned} \quad (49)$$

Here, (a) uses $1 - \underline{\pi}_W \geqslant K_W^{-1}$ (Lemma 2); (b) uses (47) and (48); (c) follows from $\lambda_W < 0.5K_W^{-2}LU$ and $|\mathbf{x}|_W \leqslant K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}$.

Now, applying Kingman's moment-bound to (49) with $f(\mathbf{x}) = K_W^{-1}LU|\mathbf{x}|_W$ and $g(\mathbf{x}) = \lambda_W(K_W + K_W^2)$ gives

$$\begin{aligned}\mathbb{E}_\rho[|\mathbf{x}|_W] &\leqslant \frac{\lambda_W K_W^2(K_W+1)}{LU}, \text{ and} \\ \overline{T}_W &\leqslant \frac{K_W^2(K_W+1)}{LU} = O\left(K_W^3\right).\end{aligned}$$

■

B. High Load Regime $\mathcal{L}_2$

For regime $\mathcal{L}_2$, we have the following Lemma.

Lemma 6: For the single-swarm case of the multi-swarm model described in Section II, with swarm-based RFwPMS as the piece-selection policy, if the network parameters belong to $\mathcal{L}_2$ and the empty-peers are always push-contacted in a tit-for-tat contact, then the steady-state expected sojourn time $\overline{T}_W$ is independent of λ_W. In particular, $\overline{T}_W = O(K_W^6)$ for $\beta_W = O(K_W^{-3})$.

Proof: Let

$$\begin{aligned}V_2(\mathbf{x}) &= V_{W,1}(\mathbf{x}) + V_{W,2}(\mathbf{x}) \\ &= \left((M_W - \eta\bar{c}_W)^+\right)^2 + C_W^{(1)}\left(|\mathbf{x}|_W - \bar{c}_W\right).\end{aligned} \quad (50)$$

The above function uses the terms $V_{W,1}$ and $V_{W,2}$ of V_W given in (31). Therefore, by ignoring the terms associated with $V_{W,3}(\mathbf{x})$ and noting that $\mathcal{W} = \{W\}$, from (43), we get

$$QV_2(\mathbf{x}) \leqslant \lambda_W K_W C_W^{(1)} + \sum_{i\in W}\frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\Psi_W^{(i)}\right].$$

Case 1 - $\mathbf{x} \in \mathcal{R}_W^{(2)}$: Here,

$$\begin{aligned}
&QV_2(\mathbf{x}) \\
&\overset{(a)}{\leqslant} \lambda_W K_W C_W^{(1)} + \sum_{i\in W}\frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(4K_W^2 - C_W^{(1)}\right)\right] \\
&\overset{(b)}{\leqslant} \lambda_W K_W C_W^{(1)} + \frac{\underline{\Gamma}_W^{(\underline{r})}}{K_W}\left[(1-\underline{\pi}_W)\left(-0.5C_W^{(1)}\right)\right] \\
&\overset{(c)}{\leqslant} \lambda_W K_W C_W^{(1)} + K_W^{-2}\underline{\Gamma}_W^{(\underline{r})}\left[-0.5C_W^{(1)}\right] \\
&\overset{(d)}{\leqslant} \lambda_W K_W C_W^{(1)} - 0.5C_W^{(1)}K_W^{-2}(LU + \kappa_p\underline{c}_W) \\
&\overset{(e)}{\leqslant} \lambda_W K_W C_W^{(1)} - 0.5C_W^{(1)}K_W^{-2}(LU + \kappa_p(1-\eta)(\bar{c}_W - 2)) \\
&\leqslant \lambda_W K_W C_W^{(1)} - (1-\eta)K_W^{-2}\bar{c}_W\left[0.5C_W^{(1)}\kappa_p\right] \\
&\qquad + K_W^{-2}\left(2(1-\eta)\kappa_p - LU\right)^+\left[0.5C_W^{(1)}\right].
\end{aligned} \quad (51)$$

Here, (a) uses the definition of $\Psi_W^{(i)}$ (see (42)); (b) uses $C_W^{(1)} \geqslant 8K_W^2$ and then upper-bounds the summation by considering only the rarest-piece (denoted by $\underline{r}$); (c) uses $1 - \underline{\pi}_W \geqslant K_W^{-1}$ (Lemma 2); (d) uses $\underline{\Gamma}_W^{(\underline{r})} \geqslant LU + \kappa_p\underline{c}_W$ (see (19)); (e) uses $\underline{c}_W > (1-\eta)(\bar{c}_W - 2)$ because $\mathbf{x} \in \mathcal{R}_W^{(2)}$.

Case 2 - $\mathbf{x} \in \mathcal{R}_W^{(1)}$: Here,

$$\begin{aligned}
&QV_2(\mathbf{x}) \\
&\overset{(a)}{\leqslant} \lambda_W K_W C_W^{(1)} + \sum_{i\in W}\frac{\underline{\Gamma}_W^{(r)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\qquad \left.\times\left(K_W\Upsilon\overline{\psi}_W\mathbb{1}\left[i\in R_W^c\right] + \underline{\psi}_W\mathbb{1}\left[i\in R_W\right] - C_W^{(1)}\right)\right] \\
&\overset{(b)}{\leqslant} \lambda_W K_W C_W^{(1)} + \sum_{n\in R_W^c}\frac{\underline{\Gamma}_W^{(n)}\zeta_W^{(n)}}{K_W}\left[(1-\overline{\pi}_W)K_W\Upsilon\overline{\psi}_W\right] \\
&\qquad + \sum_{r\in R_W}\frac{\underline{\Gamma}_W^{(r)}}{K_W}\left[\left(1-\pi_W^{(r)}\right)\left(\underline{\psi}_W - C_W^{(1)}\right)\right] \\
&\overset{(c)}{\leqslant} \lambda_W K_W C_W^{(1)} + D_W^{(1)} + \sum_{r\in R_W}\frac{\underline{\Gamma}_W^{(r)}}{K_W}\left[\left(1-\pi_W^{(r)}\right)\underline{\psi}_W - C_W^{(1)}\right] \\
&\overset{(d)}{\leqslant} \lambda_W K_W C_W^{(1)} + D_W^{(1)} \\
&\qquad - 2(1-\eta)K_W^{-2}\left[2(1-\eta)\left(LU - 2(1-\eta)\kappa_p\right)\right.
\end{aligned}$$

[22] $K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})}$ is the number of pieces of file-W that are missing in swarm-W.

$$+\left((K_W-\eta)\bar{c}_W-\sum_{i\neq \underline{r}}c_W^{(i)}\right)(LU\wedge 2(1-\eta)\kappa_p)\Bigg]$$

$$\overset{\text{(e)}}{\leqslant}\lambda_W K_W C_W^{(1)}+D_W^{(1)}$$
$$-2(1-\eta)K_W^{-2}\left[2(1-\eta)\left(LU-2(1-\eta)\kappa_p\right)\right.$$
$$\left.+(1-\eta)\bar{c}_W(LU\wedge 2(1-\eta)\kappa_p)\right]$$
$$=\lambda_W K_W C_W^{(1)}+D_W^{(1)}$$
$$-(1-\eta)K_W^{-2}\bar{c}_W\left[2(1-\eta)\left(LU\wedge 2(1-\eta)\kappa_p\right)\right]$$
$$+K_W^{-2}\left(2(1-\eta)\kappa_p-LU\right)^+\left[4(1-\eta)^2\right]. \qquad (52)$$

Here, (a) uses the definition of $\Psi_W^{(i)}$ (see (42)); (b) uses $C_W^{(1)}\geqslant 0$; (c) follows from Lemma 9; (d) follows from Lemma 13; (e) uses $c_W^{(i)}\leqslant \bar{c}_W$.

Combining (51) and (52), we get

$$QV_2(\mathbf{x})\leqslant \lambda_W K_W C_W^{(1)}+D_W^{(1)}$$
$$+K_W^{-2}\left(2(1-\eta)\kappa_p-LU\right)^+\left(0.5C_W^{(1)}+4(1-\eta)^2\right)$$
$$-K_W^{-2}(1-\eta)\bar{c}_W\min\left\{0.5C_W^{(1)}\kappa_p,2(1-\eta)LU,\right.$$
$$\left.4(1-\eta)^2\kappa_p\right\}$$
$$\overset{\text{(a)}}{\leqslant}8\lambda_W K_W^3+D_W^{(1)}+5\left(2(1-\eta)\kappa_p-LU\right)^+$$
$$-2(1-\eta)^2K_W^{-2}\bar{c}_W\min\{LU,2(1-\eta)\kappa_p\}. \qquad (53)$$

Here, (a) follows from choosing $C_W^{(1)}=8K_W^2$ and using $4(1-\eta)^2\leqslant K_W^2$.

Applying Kingman's moment bound to (53) with $f(\mathbf{x})=2(1-\eta)^2K_W^{-2}\min\{LU,2(1-\eta)\kappa_p\}\bar{c}_W$ and $g(\mathbf{x})=8\lambda_W K_W^3+D_W^{(1)}+5\left(2(1-\eta)\kappa_p-LU\right)^+$, we get

$$\mathbb{E}_\rho[|\mathbf{x}|_W-x_W^{(\varnothing)}]\overset{\text{(a)}}{\leqslant}\mathbb{E}_\rho[P_W^{(\mathrm{tot})}]\leqslant K_W\mathbb{E}_\rho[\bar{c}_W]$$
$$\leqslant\frac{\left(8\lambda_W K_W^3+5\left(2(1-\eta)\kappa_p-LU\right)^++D_W^{(1)}\right)K_W^3}{2(1-\eta)^2\min\{LU,2(1-\eta)\kappa_p\}}$$
$$=\frac{K_W^3\left(8\lambda_W K_W^3+5\left(2(1-\eta)\kappa_p-LU\right)^+\right)+D_W^{(1)}K_W^3}{2(1-\eta)^2\min\{LU,2(1-\eta)\kappa_p\}}.$$

Here, (a) uses the fact that each non-empty peer must possess at least one piece of the file.

Now, we divide the single-swarm system into two subsystems; one with empty-peers and the other with peers who hold some piece of file-W. The two subsystems are connected in tandem as shown in Fig. 5. Only the empty peers arrive in subsystem-1, so the arrival rate in subsystem-1 is λ_W. Since both systems are stable in steady-state, the departure rate of subsystem-1 (which is also the arrival rate of subsystem 2) and the departure rate of subsystem-2 are λ_W. Let $\overline{T}_{W,1}$ and $\overline{T}_{W,2}$ denote the steady-state expected sojourn-times in subsystems 1 and 2 respectively, so that by Little's Law, $\mathbb{E}_\rho\left[|\mathbf{x}|_W\right]=\lambda_W\left(\overline{T}_{W,1}+\overline{T}_{W,2}\right)$. Under swarm-based RFwPMS, consider two situations – one in which a (W,S)-peer contacts a (W,T)-peer where $T\neq\varnothing$ and the other in which it contacts a $(W,\varnothing)$-peer. Then under the assumption, that empty-peers are always push-contacted, (W,S)-peer will push-contact $(W,\varnothing)$-peer with probability 1. This combined with the fact that $S\backslash T\subseteq S\backslash\varnothing$ ensures that the probability of an empty peer downloading a piece in the next state transition is always lower-bounded by that of a peer from subsystem-2. Thus, $\overline{T}_{W,1}\leqslant\overline{T}_{W,2}$, which gives

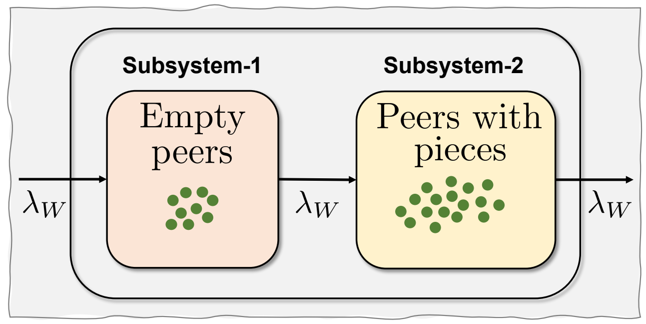


Fig. 5. Subsystems 1 and 2 in tandem (Single-swarm version of the multi-swarm model described in Section II).

$$\mathbb{E}_\rho[|\mathbf{x}|_W]\leqslant 2\lambda_W\overline{T}_{W,2}=2\mathbb{E}[|\mathbf{x}|_W-x_W^{(\varnothing)}],\text{ and}$$
$$\overline{T}_W\leqslant\frac{K_W^3\left(8K_W^3+5\lambda_W^{-1}\left(2(1-\eta)\kappa_p-LU\right)^+\right)}{(1-\eta)^2\min\{LU,2(1-\eta)\kappa_p\}}$$
$$+\frac{\lambda_W^{-1}D_W^{(1)}K_W^3}{(1-\eta)^2\min\{LU,2(1-\eta)\kappa_p\}}.$$

Using $\lambda_W\geqslant 0.5K_W^{-2}LU$, the definition of $D_W^{(1)}$ from (45), and the choice of $\beta_W=O(K_W^{-3})$, we get

$$\overline{T}_W=O\left(K_W^6\right).$$

■

VI. Simulation Based Performance Evaluation

Next, we investigate the stability, scalability and sojourn times of (swarm-based) RFwPMS via numerical simulations. Since our stability result holds for any number of swarms, we will evaluate performance in both single-swarm and multi-swarm settings. Unless otherwise noted, in all cases, we set α_W to 10^{-9} and β_W to 1.5. Furthermore, the tabulated steady-state sojourn times are based on samples which were taken after the simulated Markov chain hit stationarity (simulation end-times were chosen long enough to collect sufficiently large number of samples from the stationary distribution).

A. Stability Check

We illustrate the stability of (swarm-based) RFwPMS by simulating four instances of a two-swarm network with the setting of "hard tit-for-tat ($p=0$) and opportunistic-unchoking ($Y^{(opt)}=1$)" – the four instances correspond to altruistic, opportunistic, selfish, and autonomous inter-swarm behaviors. In each instance, the network consists of a master-file $\mathcal{F}$ of 25 pieces, i.e., $\mathcal{F}=[25]$. The two swarms entering the network are denoted by W_1 and W_2, each having an arrival-rate of 20. Peers from swarm-W_1 wish to download file $W_1=[15]$ whereas those from swarm-W_2 are interested in file $W_2=[10,25]$. We initiate the network in a state where both swarms are in the one-club scenario: both have 500 peers with all in swarm-W_1 missing piece 1, and all in swarm-W_2 missing piece 11. For the autonomous case, the seed's upload capacity is divided evenly among the two swarms, i.e., $(U_{W_1}, U_{W_2}) = (0.5, 0.5)$.

Fig. 6 shows the evolution of the swarms' population-sizes for the four inter-swarm behaviors. It is observed that each swarm is able to escape the one-club in finite time, after which a stable regime persists with minimal fluctuations. In the case of altruistic and opportunistic swarms, the population of swarm-W_2 suddenly drops in the beginning. This is because piece 11, which is missing in swarm-W_2, is widely available in swarm-W_1; due to altruism of swarm-W_1 peers, the one-club peers in swarm-W_2 quickly grab piece 11 and leave the network. In the opportunistic case, the population of swarm-W_2 increases almost linearly after the big initial drop. This is because once most of the one-club peers in swarm-W_2 have left, the network is dominated by swarm-W_1 peers. Therefore, most of the contacts of the new-comers in swarm-W_2 are with swarm-W_1 peers. Since the files are not identical, these new-comers cannot accumulate all pieces from swarm-W_1 peers and are forced to linger in the system – till swarm-W_1's population-size reduces enough and they get more opportunities for useful contacts.

B. Scalability Check

Here, we simulate twelve instances of a two-swarm network with the setting of "soft tit-for-tat (choice of $p = 0.5$) and no optimistic-unchoking ($Y^{(opt)} = 0$)". In each instance, the network's master-file comprises of 18 pieces ($\mathcal{F} = [18]$), and the two swarms entering the network are interested in files $W_1 = [10]$ and $W_2 = [8, 18]$. The twelve instances correspond to four inter-swarm behaviors (altruistic, opportunistic, selfish, and autonomous) and three arrival-rate configurations ($\boldsymbol{\lambda} = (4, 2)$, $4\boldsymbol{\lambda}$ and $16\boldsymbol{\lambda}$). For the autonomous case, the seed's upload capacity is again split equally between the two swarms.

Table III lists the (steady-state) expected sojourn times for the twelve model instances. We note that the expected sojourn time practically stays constant in all the four inter-swarm behaviors, and improves as swarms interact more altruistically. Another observation is that the autonomous setting has lower expected sojourn times than the altruistic setting. This is not unexpected though; in autonomous setting, every peer makes contacts within their own swarm (no inter-swarm interactions), thus the likelihood that the next contact is useless is generally lower.

C. Steady-State Expected Sojourn Times in Multiple Swarms

How does the (steady-state) expected sojourn time scale with the file-size is another important performance measure. For this, in Section V, we obtained an upper-bound estimate of $O\left(K_W^6\right)$, for choice of $\beta_W = O\left(K_W^{-3}\right)$. In Fig. 7, we verify this for a two-swarm network in a soft tit-for-tat setting (choice of $p = 0.5$) with optimistic-unchoke ($Y^{(opt)} = 1$)—once again, under the four different inter-swarm behaviors.

Fixing the arrival-rate vector to $\boldsymbol{\lambda} = (6, 2)$, increasing configurations of the two file-sizes were simulated such that for every choice of $W_1 = [K_{W_1}]$, we set W_2 to $[\frac{1}{2}K_{W_1}, \frac{3}{2}K_{W_1}]$. (Again the seed's upload capacity was split evenly in the autonomous case). The sojourn times for each swarm are observed to scale linearly with the file-size. Additionally, we note that for each swarm, the altruistic and autonomous cases are *Pareto* better than opportunistic and selfish cases.

D. RFwPMS vs. Mode Suppression

Via simulations MS [11] has been shown to outperform other previously proposed piece-selection policies. Here, we do a base-line comparison of RFwPMS and MS in a single-swarm network.

1) Steady-State Expected Sojourn Times: Table IV compares the (steady-state) expected sojourn times of RFwPMS, MS, and TMS[23] for different values of file-size K_W with the arrival-rate fixed at $\lambda_W = 4$. It is noted that using RFwPMS (with $\beta_W = 1.7$), the expected sojourn time is indeed reduced compared to MS. This reduction is observed to be less significant when K_W is large (roughly 100 or more)—the increased chunk diversity in steady-state reduces the number of contacts in which only the non-rare pieces are offered. The expected sojourn times of RFwPMS and TMS, on the other hand, are comparable for all values of K_W.

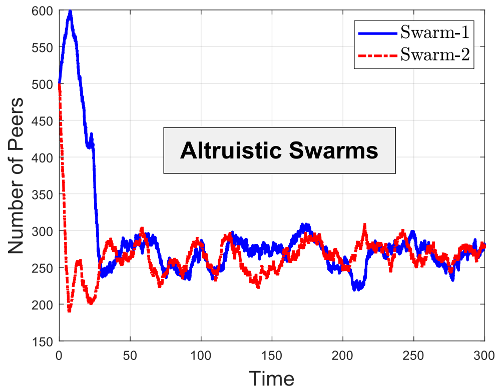
2) Flash-Crowd Responses: Real torrents seldom experience steady-state behavior as the arrival-rates vary over time with intermittent bursts of empty peers. Fig. 8 compares the flash-crowd response of MS, TMS, RFwPMS, and RNwPMS (replacing RF by RN in our policy) for $K_W = 100$ ("large") and an initial population of 500 empty peers ($|\mathbf{x}|_W(0) = x_W^{(\varnothing)}(0) \gg K_W$); RFwPMS flushes out the system in the least amount of time (about half the time as MS) whereas RNwPMS takes the most amount of time (see Fig. 8a). The reason behind such responses can be understood from Fig. 8b. Initially, there are no pieces in the system. Therefore, the time till all K_W pieces are introduced into the system depends solely on the seed's uploading capacity which is set to 1. Thus, for every policy, if K_W is sufficiently large, it will take on average K_W

[23] Threshold Mode-Suppression with suppression-threshold set to $2K_W$.

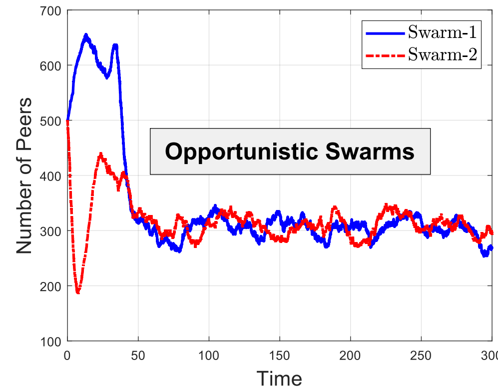
TABLE III
STEADY-STATE EXPECTED SOJOURN TIMES
FOR DIFFERENT ARRIVAL RATE VECTORS.
$(U = \mu^{(\mathrm{tft})} = 1, \mu^{(\mathrm{opt})} = \frac{1}{3}, L = 3, p = 0.5, Y^{(opt)} = 0)$.

Swarms' Behavior	**Arrival Rates**[a]	$\mathbb{E}_\rho$**[Sojourn Time]**	
		$W_1 = [10]$	$W_2 = [8, 18]$
Altruistic	$\boldsymbol{\lambda}$	2.927	4.400
	$4\boldsymbol{\lambda}$	3.088	3.990
	$16\boldsymbol{\lambda}$	3.134	3.971
Opportunistic	$\boldsymbol{\lambda}$	3.704	5.042
	$4\boldsymbol{\lambda}$	3.832	5.341
	$16\boldsymbol{\lambda}$	3.956	5.570
Selfish	$\boldsymbol{\lambda}$	4.378	6.394
	$4\boldsymbol{\lambda}$	4.590	6.482
	$16\boldsymbol{\lambda}$	4.667	6.604
Autonomous	$\boldsymbol{\lambda}$	2.791	3.769
	$4\boldsymbol{\lambda}$	2.712	2.667
	$16\boldsymbol{\lambda}$	2.788	2.740

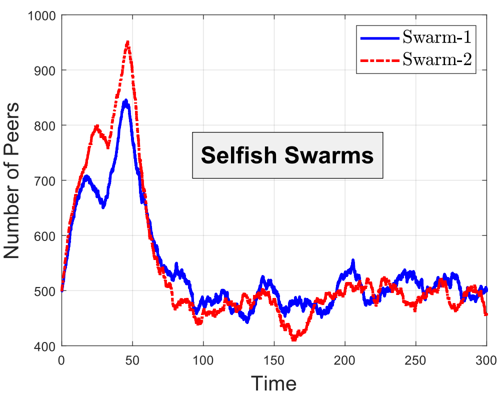
[a] $\boldsymbol{\lambda} = (4, 2)$.
Simulation end-time: 1000 units.
The confidence intervals are not shown due to their negligible magnitude (even with a choice of 99.9 percent confidence level).



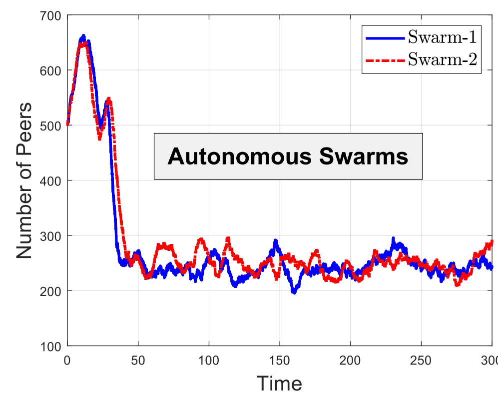
(a) Altruistic swarms.



(b) Opportunistic swarms.

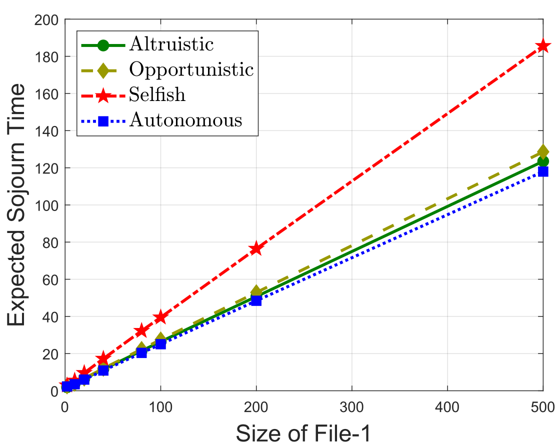


(c) Selfish swarms.

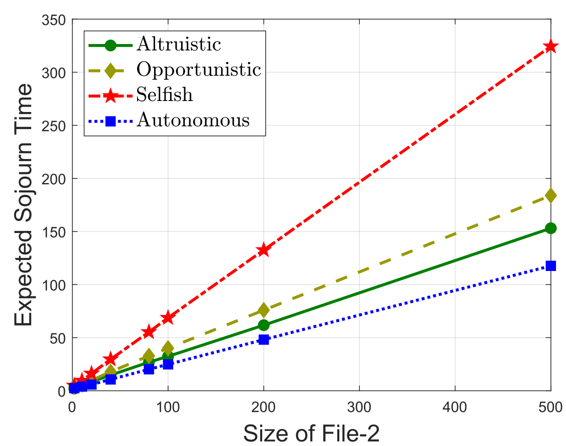


(d) Autonomous swarms

Fig. 6. Number of peers in a two-swarm network ($W_1 = [15], W_2 = [10, 25]$) for different inter-swarm behaviors. Other model parameters were set to $U = \mu^{(\text{tft})} = 1, \mu^{(\text{opt})} = \frac{1}{3}, L = 3, p = 0, Y^{(opt)} = 1$. In the case of autonomous-swarms, the seed's upload capacity was divided equally among the swarms.



(a) Expected Sojourn Times of Swarm-1.



(b) Expected Sojourn Times of Swarm-2.

Fig. 7. Steady-state expected sojourn times vs file-sizes. Fig. 7a shows the expected sojourn times of swarm-1 and Fig. 7b shows those of swarm-2. Assignment of remaining model parameters was $\boldsymbol{\lambda} = (6, 2)$, $U = \mu^{(\text{tft})} = 1$, $\mu^{(\text{opt})} = \frac{1}{3}$, $L = 3$, $p = 0.5$, $Y^{(opt)} = 1$. The error-bars (for confidence intervals) are not shown due to their negligible magnitude (even with a choice of 99.9 percent confidence level).

time units till every piece has been introduced into the system (see Fig. 8b). MS forbids the download of non-rare pieces. Thus, there is exactly one copy of each piece when the seed has introduced the last piece into the system. After that, the number of copies of the non-rare piece and the rarest piece go hand-in-hand with at most a difference of 1. TMS, on the

TABLE IV
STEADY-STATE EXPECTED SOJOURN TIMES FOR RFWPMS AND MS. ($\lambda_W = 4, U = \mu^{(\mathrm{opt})} = L = Y^{(opt)} = 1, \beta_W = 1.7$).

K_W	$\mathbb{E}_\rho$[Sojourn Time]			Perc. Improv. over MS
	MS	*TMS*	*RFwPMS*	
2	6.246	5.022	5.178	17.103
10	18.250	12.546	12.525	31.367
20	31.741	23.020	23.058	27.356
40	55.648	43.775	43.750	21.382
80	100.300	84.374	84.421	15.831
100	121.804	104.849	104.610	14.116
200	226.998	205.300	205.176	9.613
500	533.737	506.480	506.351	5.131

Simulation end-time: 5000 units.
The confidence intervals are not shown due to their negligible magnitude (even with a choice of 99.9 percent confidence level).

other hand, allows the download of non-rare pieces till the largest-mismatch equals the threshold (twice the size of the file). After that, no peer is allowed to download those pieces. Therefore, by the time, the last piece is introduced into the system, there are many peers who have yet to download a large number of pieces to depart.

Comparing with MS and TMS, both RFwPMS and RNwPMS allow the (probabilistic) download of non-rare pieces, and since K_W is sufficiently large, by the time the last piece is introduced into the system, the largest-mismatch $\overline{m}_W$ has increased considerably coupled with proportional decrease in $\zeta_W^{(n)}$. At this point, the largest-mismatch is very large, but owing to the RF piece-selection, RFwPMS quickly stabilizes the this mismatch by giving preference to *the rarest of all available rare pieces* in all peer encounters. On the other hand, RNwPMS does not consider the frequency differences of the rare pieces. Thus, if K_W is sufficiently large, the system is likely to get trapped in a one-club type state where almost all peers have all the pieces except the rarest one. Once this happens, only the seed can flush such peers from the system, which leads to a larger file-delivery time than that of MS.

It is worth noting that RFwPMS can also get trapped in such one-club like states. This is specially true when file-size is much larger than the size of the flash-crowd ($K_W \gg |\mathbf{x}|_W(0) = x_W^{(\emptyset)}(0)$). As an example, Fig. 9 shows such a flash-crowd situation—RFwPMS gets locked into the one-club like state (Fig. 9a) whereas MS avoids this (Fig. 9b). (Note that the flush-out time of MS is still larger than RFwPMS). The locking-in of RFwPMS in this one-club like state is because the non-rares' sharing factor $\zeta_W^{(n)}$ defined in (16) suppresses non-rare pieces via the ratio $\frac{\overline{m}_W}{K_W}$. Since $\overline{m}_W$ is upper-bounded by the network's population-size (which in this case is much smaller than K_W), the suppression is not enough to slow down the download of non-rare pieces.[24] This observation while simple is worth noting as it is reflective of what should happen in practice. Usually, files shared over such networks would have chunk-sizes that are much larger than the flash-crowd. In such transient situations, to obtain a good trade-off between suppression and sharing, it would make sense to have a non-rares' sharing factor that takes into account the flash-crowd's size (besides network's parameters and the file-size). One way we could choose a $\zeta_W^{(n)}$ that can work for both types of flash-crowds would be to replace it by

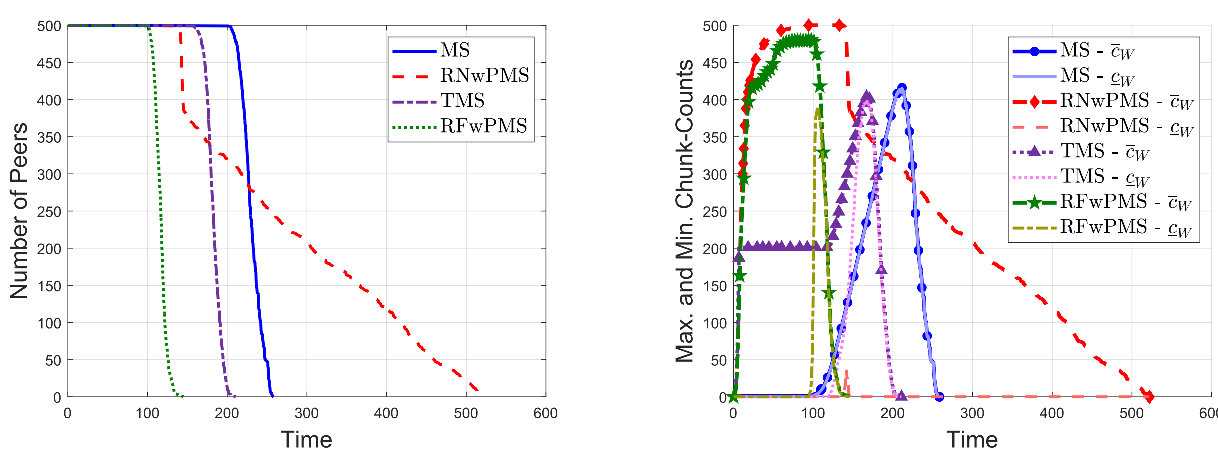
$$\exp\left(-\frac{\kappa_1}{\beta_W L U}\frac{\overline{m}_W + d_W^{(i)}}{\min\{K_W, |\mathbf{x}|\}}\right). \tag{54}$$

Fig. 10 shows the flash-crowd response of RFwPMS for the aforementioned choice of $\zeta_W^{(n)}$ with different values of β_W namely 0.2, 0.4, and 0.5. When β_W is set to 0.4, it seems to replicate the response of RFwPMS in Fig. 8 and attains the least flush-out time.

VII. CONCLUSION

In this work, we proposed and studied a piece-selection policy, (swarm-based) RFwPMS, for a BitTorrent-like P2P

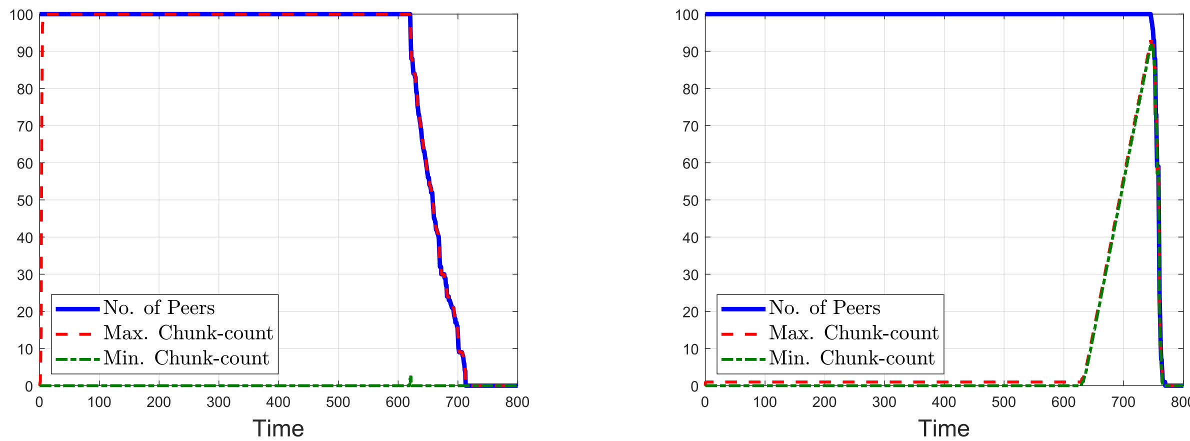
[24]This locking-in of RFwPMS would still occur if we suppress each piece i through the ratio $\frac{c_W^{(i)} - \underline{c}_W}{K_W}$.



(a) Number of Peers vs. Time.

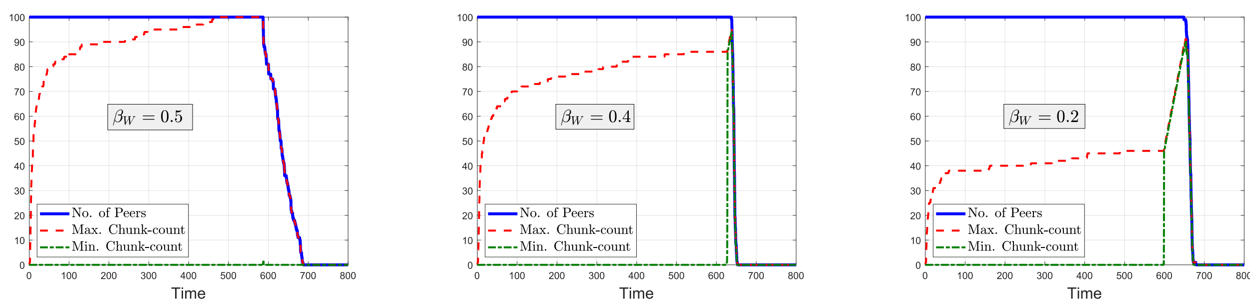
(b) Chunk-counts vs Time.

Fig. 8. Flash-crowd responses of MS, RNwPMS, TMS, and RFwPMS when flash-crowd size is much larger than file-size. Initial population and file-size were set to 500 and 100 respectively. Assignment of other model parameters was $\lambda_W = 0$, $U = \mu^{(\mathrm{opt})} = L = Y^{(opt)} = 1$).



(a) RFwPMS. (b) MS.

Fig. 9. Flash-crowd responses of RFwPMS and MS when file-size is much larger than the size of the flash-crowd. Initial population and file sizes were set to 100 and 600 respectively. Assignment of other model-parameters was $\lambda_W = 0$, $U = \frac{1}{3}$, $\mu^{(\text{tft})} = \mu^{(\text{opt})} = 1$, $L = 3$, $p = 0.5$, $Y^{(opt)} = 1$.



(a) $\beta_W = 0.5$. (b) $\beta_W = 0.4$. (c) $\beta_W = 0.2$.

Fig. 10. Flash-crowd responses of RFwPMS when file-size is much larger than the size of the flash-crowd. The non-rares' sharing factor given by (54) was used with three different choices of β_W. Other model parameters were the same as given in the caption of Fig. 9.

network with multiple swarms of non-persistent users. RFwPMS combines rarest-first (for rare pieces) with an adjustable sharing-versus-suppression choice for non-rare pieces. Using Lyapunov drift analysis, we proved the stability of RFwPMS in a multi-swarm setting, independent of how peers use their excess cache, and whether or not swarms collaborate with each other. By using the Kingman's moment bound technique, we established the scalability of RFwPMS in the single-swarm case. Our numerical simulations demonstrated evidence that RFwPMS can reduce the steady-state expected sojourn times as well as file-delivery times during a flash-crowd.

Lastly, since RFwPMS uses RF with a minor modification (use of the non-rares' sharing factor), it is amenable to be incorporated into BitTorrent-like protocols.

In regards to future-work, proving (or disproving) Conjecture 1 and determining the multi-swarm scaling behavior could be topics for future work.

Appendix A
Two-sided Piece Exchanges in the Same Swarm

Here, we consider the event that (W, S) and (W, T) peers download pieces i and j respectively.

A. Change in $V_{W,1}$

Case 1 - Both i and j are swarm-W's non-rare pieces: If both peers remain in the system, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ increase by 1; $c_W^{(i')}$ stays the same for all $i' \in W\backslash\{i,j\}$. If both peers depart, $\bar{c}_W$, $c_W^{(i)}$, $c_W^{(j)}$ decrease by 1; $c_W^{(i')}$ decreases by 2 for all $i' \in W\backslash\{i,j\}$. If one peer departs and the other stays, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ stay the same; $c_W^{(i')}$ decreases by 1 for all $i' \in W\backslash\{i,j\}$. Overall, in all cases, the total mismatch M_W increases by $K_W - 2$ and the highest chunk-count reduces by at most 1. Overall, this causes an increment of at most $K_W - 2 + \eta \leqslant K_W$ in $M_W - \eta\bar{c}_W$. Therefore,

$$\begin{aligned}
&\Delta V_{W,1}^{((S,i\pm),(T,j\pm))} \\
&\leqslant \left((M_W - \eta\bar{c}_W + K_W)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2 \\
&\leqslant \overline{\psi}_W \leqslant 2\overline{\psi}_W.
\end{aligned}$$

Case 2A - i and j are swarm-W's rare pieces and $\mathbf{x} \in \mathcal{S}_W^{(1)}$: If both peers remain in the system, $\bar{c}_W$ stays the same; $c_W^{(i)}$, $c_W^{(j)}$ increase by 1; $c_W^{(i')}$ stays the same for all $i' \in W\backslash\{i,j\}$. If both peers depart the system, $\bar{c}_W$ decreases by 2; $c_W^{(i)}$ and $c_W^{(j)}$ each decrease by 1; $c_W^{(i')}$ also decreases by 2 for all

$i' \in W\backslash\{i,j\}$. If one peer departs and the other stays, $\bar{c}_W$ and $c_W^{(i')}$ ($i' \in W\backslash\{i,j\}$) decrease by 1; $c_W^{(i)}$ and $c_W^{(j)}$ remain the same. Overall, in all cases, the total mismatch M_W reduces by 2 and $\bar{c}_W$ decreases by 2 at most. Therefore,

$$\begin{aligned}
&\Delta V_{W,1}^{((S,i\pm),(T,j\pm))}\\
&\leqslant \left((M_W - \eta\bar{c}_W - 2(1-\eta))^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&= \begin{cases} 2\left[2(1-\eta)^2 - 2(1-\eta)\left(M_W - \eta\bar{c}_W\right)\right] & \text{if } M_W \geqslant 2(1-\eta) + \eta\bar{c}_W\\ 0 & \text{otherwise.}\end{cases}\\
&\leqslant 2\underline{\psi}_W.
\end{aligned}$$

Case 2B - i and j are swarm-W's rare pieces and $\mathbf{x} \in \mathcal{S}_W^{(2)}$: If both peers remain in the system, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ increase by 1; $c_W^{(i')}$ remains the same for all $i' \in W\backslash\{i,j\}$. If both peers depart the system, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ decrease by 1; $c_W^{(i')}$ decreases by 2 for all $i' \in W\backslash\{i,j\}$. If one peer departs and the other stays, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ stay the same; $c_W^{(i')}$ decreases by 1 for all $i' \in W\backslash\{i,j\}$. Overall, in all cases, the total mismatch M_W increases from 0 to $K_W - 2$ and $\bar{c}_W$ decreases by at most 1. Therefore,

$$\begin{aligned}
&\Delta V_{W,1}^{((S,i\pm),(T,j\pm))}\\
&\leqslant \left((K_W - 2 + \eta - \eta\bar{c}_W)^+\right)^2 - \left((-\eta\bar{c}_W)^+\right)^2\\
&\leqslant (K_W+2)^2 \leqslant 2(K_W+2)^2 \leqslant 2\underline{\psi}_W.
\end{aligned}$$

Case 3 - i is a rare piece and j a non-rare piece: If both peers stay in the system, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ increase by 1; $c_W^{(i')}$ stays the same for all $i' \in W\backslash\{i,j\}$. If both peers depart, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ decrease by 1; $c_W^{(i')}$ decreases by 2 for all $i' \in W\backslash\{i,j\}$. If one peer departs and the other stays, $\bar{c}_W, c_W^{(i)}, c_W^{(j)}$ stay the same; $c_W^{(i')}$ decreases by 1 for all $i' \in W\backslash\{i,j\}$. Overall, in all cases, the total mismatch M_W increases by $K_W - 2$ and $\bar{c}_W$ decreases by at most 1. Therefore,

$$\begin{aligned}
&\Delta V_{W,1}^{((S,i\pm),(T,j\pm))}\\
&= \left((M_W - \eta\bar{c}_W + K_W - 2 + \eta)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&\overset{(a)}{\leqslant} \left((M_W - \eta\bar{c}_W + K_W - 1)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&\overset{(b)}{\leqslant} \left((M_W - \eta\bar{c}_W + K_W)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&+ \left((M_W - \eta\bar{c}_W - 1)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&\overset{(c)}{\leqslant} \left((M_W - \eta\bar{c}_W + K_W)^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&+ \left((M_W - \eta\bar{c}_W - (1-\eta))^+\right)^2 - \left((M_W - \eta\bar{c}_W)^+\right)^2\\
&\leqslant \bar{\psi}_W + \underline{\psi}_W.
\end{aligned}$$

Here, step (a) uses $\eta < 1$; (c) uses $\eta > 0$; and step (b) uses the inequality

$$\begin{aligned}
&\left((x+a)^+\right)^2 - \left((x)^+\right)^2 \leqslant \left((x+1+a)^+\right)^2 - \left((x)^+\right)^2\\
&+ \left((x-1)^+\right)^2 - \left((x)^+\right)^2 \;\; \forall x \in \mathbb{R}, a \geqslant 0.
\end{aligned}$$

The case when i is a non-rare piece and j is a rare piece is similar.

The above inequalities ensure that we can write

$$\Delta V_{W,1}^{((S,i\pm),(T,j\pm))} \leqslant \sum_{i'=i,j} \bar{\psi}_W \mathbb{1}\left[i' \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i' \in R_W\right]. \tag{A.55}$$

B. Change in $V_{W,2}$

For $V_{W,2}$, note that $K_W|\mathbf{x}|_W - P_W^{(\mathrm{tot})} = \sum_{i\in W} |\mathbf{x}|_W - c_W^{(i)}$. If both peers stay in the system, $|\mathbf{x}|_W$ stays the same; $c_W^{(i)}, c_W^{(j)}$ increase by 1; $c_W^{(i')}$ stays the same for all $i' \in W\backslash\{i,j\}$. If both peers depart, $|\mathbf{x}|_W$ decreases by 2; $c_W^{(i)}, c_W^{(j)}$ decrease by 1; $c_W^{(i')}$ decrease by 2 for all $i' \in W\backslash\{i,j\}$. If one peer leaves and the other departs; $|\mathbf{x}|_W$ decreases by 1; $c_W^{(i)}, c_W^{(j)}$ stay the same; $c_W^{(i')}$ decreases by 1 for all $i' \in W\backslash\{i,j\}$. In all cases, the deficit $|\mathbf{x}|_W - c_W^{(i')}$ stays the same for all $i' \in W\backslash\{i,j\}$ and decreases by 1 for $i' \in \{i,j\}$. Thus, we can write

$$\Delta V_{W,2}^{\{(S,i\pm),(T,j\pm)\}} = -2C_W^{(2)}. \tag{A.56}$$

C. Change in $V_{W,3}$

The change in $V_{W,3}$ depends on whether peers depart or not, and does not depend on the type of i and j. There are three cases. If both peers stay in the system, $P_W^{(\mathrm{tot})}$ increases by 2. Therefore,

$$\Delta V_{W,3}^{\{(S,i+),(T,j+)\}} \leqslant -2C_W^{(2)} \mathbb{1}\left[C_W^{(3)} - 2 \geqslant P_W^{(\mathrm{tot})}\right] = -2I_W^{(1)}. \tag{A.57}$$

If only one of the two peers depart, then $P_W^{(\mathrm{tot})}$ decreases by $K_W - 2$. Therefore,

$$\begin{aligned}
&\Delta V_{W,3}^{\{(S,i+/-),(T,j-/+)\}}\\
&\leqslant C_W^{(2)}(K_W - 2)\mathbb{1}\left[C_W^{(3)} + K_W - 2 \geqslant P_W^{(\mathrm{tot})}\right]\\
&\leqslant C_W^{(2)} K_W \mathbb{1}\left[C_W^{(3)} + 2K_W \geqslant P_W^{(\mathrm{tot})}\right] - C_W^{(2)} \mathbb{1}\left[C_W^{(3)} - 2 \geqslant P_W^{(\mathrm{tot})}\right]\\
&= I_W^{(2)} - I_W^{(1)}.
\end{aligned} \tag{A.58}$$

If both peers depart the system, $P_W^{(\mathrm{tot})}$ decreases by $2(K_W - 1)$. Therefore,

$$\begin{aligned}
&\Delta V_{W,3}^{\{(S,i-),(T,j-)\}}\\
&\leqslant C_W^{(2)}(2(K_W-1))\mathbb{1}\left[C_W^{(3)} + 2(K_W-1) \geqslant P_W^{(\mathrm{tot})}\right]\\
&\leqslant 2C_W^{(2)} K_W \mathbb{1}\left[C_W^{(3)} + 2K_W \geqslant P_W^{(\mathrm{tot})}\right] = 2I_W^{(2)}.
\end{aligned} \tag{A.59}$$

Equations (A.55) to (A.59) establish the bounds in Table II.

Appendix B
Rarest-First vs Random-Novel over Rare Pieces

Lemma 7: For all $\mathbf{x}$,

$$QV_W^{(R_W,+)} \leqslant \sum_{\substack{r\in R_W,\\ S:W\backslash S \supsetneq \{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|} \left[\frac{LU}{|R_W\backslash S|}\right.$$

$$+\kappa_p\xi \sum_{\substack{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T:r\in T}} \frac{x_V^{(T)}}{|(T\cap R_W)\backslash S|}\Bigg]\Psi_W^{(r+)}.$$

Proof: Using (24), (34) and the fact that $\Psi_W^{(r+)}\leqslant 0$ for every $r\in R_W$,

$$\begin{aligned}
QV_W^{(R_W,+)} &\leqslant \sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\Bigg[\frac{LU\mathbb{1}\{r\in \underline{R}_{(\mathcal{F},W,S)}\}}{|\underline{R}_{(\mathcal{F},W,S)}|} \\
&+\kappa_p\xi \sum_{\substack{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T:r\in T}} \frac{x_V^{(T)}\mathbb{1}\{r\in \underline{R}_{(T,W,S)}\}}{|\underline{R}_{(T,W,S)}|}\Bigg]\Psi_W^{(r+)} \\
&\overset{(a)}{=} \underbrace{\sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\Bigg[\frac{LU\mathbb{1}\{r\in \underline{R}_{(\mathcal{F},W,S)}\}}{|\underline{R}_{(\mathcal{F},W,S)}|}\Psi_W^{(+)}}_{\text{First Term}} \\
&\underbrace{+\frac{x_W^{(S)}}{|\mathbf{x}|}\kappa_p\xi \sum_{\substack{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T:r\in T}} \frac{x_V^{(T)}\mathbb{1}\{r\in \underline{R}_{(T,W,S)}\}}{|\underline{R}_{(T,W,S)}|}\Bigg]\Psi_W^{(+)}}_{\text{Second Term}}.
\end{aligned} \tag{B.60}$$

Here, (a) uses the fact that $\Psi_W^{(r+)}$ is the same for all $r\in R_W$, which we have denoted by $\Psi_W^{(+)}$. Then, we can upper bound the first term as follows.

$$\begin{aligned}
\frac{\text{First Term}}{LU} &= \sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{\mathbb{1}\{r\in \underline{R}_{(\mathcal{F},W,S)}\}}{|\underline{R}_{(\mathcal{F},W,S)}|}\Psi_W^{(+)} \\
&\overset{(a)}{=} \sum_{\substack{S:|S\cap W|<K_W-1\\ r:r\in R_W\backslash S}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{\mathbb{1}\{r\in \underline{R}_{(\mathcal{F},W,S)}\}}{|\underline{R}_{(\mathcal{F},W,S)}|}\Psi_W^{(+)} \\
&= \sum_{\substack{S:|S\cap W|<K_W-1\\ r:r\in \underline{R}_{(\mathcal{F},W,S)}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{1}{|\underline{R}_{(\mathcal{F},W,S)}|}\Psi_W^{(+)} \\
&= \sum_{S:|S\cap W|<K_W-1} \frac{x_W^{(S)}}{|\mathbf{x}|}\Psi_W^{(+)} \\
&= \sum_{S:|S\cap W|<K_W-1} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{|R_W\backslash S|}{|R_W\backslash S|}\Psi_W^{(+)} \\
&= \sum_{\substack{S:|S\cap W|<K_W-1\\ r:r\in R_W\backslash S}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{1}{|R_W\backslash S|}\Psi_W^{(+)} \\
&\overset{(b)}{=} \sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{1}{|R_W\backslash S|}\Psi_W^{(+)}.
\end{aligned} \tag{B.61}$$

In steps (a) and (b), we change the order of summation. Similarly, we can upper bound the second term as follows.

$$\begin{aligned}
\frac{\text{Second Term}}{\kappa_p\xi} &= \sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\},\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:r\in T}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{x_V^{(T)}\mathbb{1}\{r\in \underline{R}_{(T,W,S)}\}}{|\underline{R}_{(T,W,S)}|}\Psi_W^{(+)} \\
&\overset{(a)}{=} \sum_{\substack{S:|S\cap W|<K_W-1,\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:(T\cap R_W)\backslash S\neq\varnothing,\\ r:r\in(T\cap R_W)\backslash S}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{x_V^{(T)}\mathbb{1}\{r\in \underline{R}_{(T,W,S)}\}}{|\underline{R}_{(T,W,S)}|}\Psi_W^{(+)} \\
&= \sum_{\substack{S:|S\cap W|<K_W-1,\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:(T\cap R_W)\backslash S\neq\varnothing,\\ r:r\in \underline{R}_{(T,W,S)}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{x_V^{(T)}}{|\underline{R}_{(T,W,S)}|}\Psi_W^{(+)} \\
&= \sum_{\substack{S:|S\cap W|<K_W-1,\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:(T\cap R_W)\backslash S\neq\varnothing}} \frac{x_W^{(S)}}{|\mathbf{x}|}x_V^{(T)}\Psi_W^{(+)} \\
&= \sum_{\substack{S:|S\cap W|<K_W-1,\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:(T\cap R_W)\backslash S\neq\varnothing}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{|(T\cap R_W)\backslash S|x_V^{(T)}}{|(T\cap R_W)\backslash S|}\Psi_W^{(+)} \\
&= \sum_{\substack{S:|S\cap W|<K_W-1,\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:(T\cap R_W)\backslash S\neq\varnothing,\\ r:r\in(T\cap R_W)\backslash S}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{x_V^{(T)}}{|(T\cap R_W)\backslash S|}\Psi_W^{(+)} \\
&\overset{(b)}{=} \sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\},\\ V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},T:r\in T}} \frac{x_W^{(S)}}{|\mathbf{x}|}\frac{x_V^{(T)}}{|(T\cap R_W)\backslash S|}\Psi_W^{(+)}.
\end{aligned} \tag{B.62}$$

In steps (a) and (b), we change the order of summation. Using (B.61) and (B.62) in (B.60), we get

$$\begin{aligned}
QV_W^{(R_W,+)} &\leqslant \sum_{\substack{r\in R_W,\\ S:W\backslash S\supsetneq\{r\}}} \frac{x_W^{(S)}}{|\mathbf{x}|}\Bigg[\frac{LU}{|R_W\backslash S|} \\
&+\kappa_p\xi \sum_{\substack{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T:r\in T}} \frac{x_V^{(T)}}{|(T\cap R_W)\backslash S|}\Bigg]\Psi_W^{(r+)}.
\end{aligned}$$

■

Appendix C
Combining $QV_W^{(R_W,+)}$ and $QV_W^{(R_W^c,+)}$

Lemma 8: For all $\mathbf{x}$,

$$\begin{aligned}
QV_W^{(W,+)} &\leqslant \frac{|\mathbf{x}|_W}{|\mathbf{x}|}\sum_{i\in W}\frac{\underline{\Gamma}_W^{(i)}a_W^{(i)}}{K_W}\left(1-\pi_W^{(i)}-\gamma_W^{(i)}\right) \\
&\times\left[\mathbb{1}\left[\mathbf{x}\in\mathcal{R}_W^{(1)}\right]\left(K_W\Upsilon\overline{\psi}_W\mathbb{1}\left[i\in R_W^c\right]+\underline{\psi}_W\mathbb{1}\left[i\in R_W\right]\right)\right. \\
&\left.+\mathbb{1}\left[\mathbf{x}\in\mathcal{R}_W^{(2)}\right]4K_W^2-C_W^{(1)}-I_W^{(1)}\right].
\end{aligned}$$

Proof: Let

$$g(i,\mathbf{x})=\mathbb{1}\left[\mathbf{x}\in\mathcal{R}_W^{(2)}\right]4K_W^2-C_W^{(1)}-I_W^{(1)}$$

$$+\,\mathbb{1}\left[\mathbf{x}\in\mathcal{R}_W^{(1)}, i\in R_W\right]\underline{\psi}_W \leqslant 0.$$

Adding (38) with the first term in (39), and using the definitions of $\underline{q}_W^{(S,n+)}$ (see (26)) and $\Psi_W^{(i+)}$ (see (34)), we get

$$\begin{aligned}
&\sum_{\substack{i\in W,\\ S:W\backslash S\supsetneq\{i\}}}\frac{x_W^{(S)}}{|\mathbf{x}|}\Bigg[LU\Bigg(\frac{\zeta_W^{(i)}\mathbb{1}\left[i\in R_W^c, R_W\backslash S=\varnothing\right]}{|W\backslash S|}\\
&\quad+\frac{\mathbb{1}\left[i\in R_W\right]}{|R_W\backslash S|}\Bigg)+\kappa_p\xi\sum_{\substack{V:W\in\mathcal{W}_V^{(ally\text{-}\uparrow)}\\ T:i\in T}}x_V^{(T)}\\
&\quad\times\Bigg(\frac{\zeta_W^{(i)}\mathbb{1}\left[i\in R_W^c,(T\cap R_W)\backslash S=\varnothing\right]}{|(T\cap W)\backslash S|}\\
&\quad+\frac{\mathbb{1}\left[i\in R_W\right]}{|(T\cap R_W)\backslash S|}\Bigg)\Bigg]g(i,\mathbf{x}).
\end{aligned}\tag{C.63}$$

Here,

$$\begin{aligned}
&\frac{\text{First Term}}{LU}\\
&=\left[\sum_{\substack{i\in R_W^c,\\ W\backslash S\supsetneq\{i\},\\ R_W\backslash S=\varnothing}}\frac{\zeta_W^{(i)}}{|W\backslash S|}+\sum_{\substack{i\in R_W\\ W\backslash S\supsetneq\{i\}}}\frac{1}{|R_W\backslash S|}\right]\frac{x_W^{(S)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(a)}{=}\left[\sum_{\substack{|S\cap W|<K_W-1,\\ R_W\backslash S=\varnothing, i\in W\backslash S}}\frac{\zeta_W^{(i)}}{|W\backslash S|}\right.\\
&\qquad\left.+\sum_{\substack{|S\cap W|<K_W-1,\\ R_W\backslash S\neq\varnothing, i\in R_W\backslash S}}\frac{1}{|R_W\backslash S|}\right]\frac{x_W^{(S)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(b)}{\leqslant}\left[\sum_{\substack{|S\cap W|<K_W-1,\\ R_W\backslash S=\varnothing, i\in W\backslash S}}\frac{\zeta_W^{(i)}}{|W\backslash S|}\right.\\
&\qquad+\sum_{\substack{|S\cap W|<K_W-1,\\ R_W\backslash S\neq\varnothing, i\in R_W^c\backslash S}}\frac{\zeta_W^{(i)}}{|W\backslash S|}\\
&\qquad\left.+\sum_{\substack{|S\cap W|<K_W-1,\\ R_W\backslash S\neq\varnothing, i\in R_W\backslash S}}\frac{1}{|W\backslash S|}\right]\frac{x_W^{(S)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&=\left[\sum_{\substack{|S\cap W|<K_W-1,\\ i\in R_W^c\backslash S}}\frac{\zeta_W^{(i)}}{|W\backslash S|}\right.\\
&\qquad\left.+\sum_{\substack{|S\cap W|<K_W-1,\\ R_W\backslash S\neq\varnothing, i\in R_W\backslash S}}\frac{1}{|W\backslash S|}\right]\frac{x_W^{(S)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(c)}{=}\left[\sum_{\substack{i\in R_W^c,\\ W\backslash S\supsetneq\{i\}}}\frac{\zeta_W^{(i)}}{|W\backslash S|}+\sum_{\substack{i\in R_W,\\ W\backslash S\supsetneq\{i\}}}\frac{1}{|W\backslash S|}\right]\frac{x_W^{(S)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(d)}{\leqslant}\sum_{\substack{i\in W,\\ W\backslash S\supsetneq\{i\}}}\frac{x_W^{(S)}}{|\mathbf{x}|}\frac{a_W^{(i)}}{K_W}g(i,\mathbf{x}).
\end{aligned}\tag{C.64}$$

Steps (a) and (c) change the order of summation; step (b) uses $\zeta_W^{(i)}\leqslant 1$ and $0\geqslant g(n,\mathbf{x})\geqslant g(r,\mathbf{x})$ for all $n\in R_W^c, r\in R_W$; (d) uses $|W\backslash S|\leqslant K_W$. Similarly,

$$\begin{aligned}
&\frac{\text{Second Term}}{\kappa_p\xi}\\
&=\sum_{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)}}\left[\sum_{\substack{i\in R_W^c, W\backslash S\supsetneq\{i\},\\ i\in T,(T\cap R_W)\backslash S=\varnothing}}\frac{\zeta_W^{(i)}}{|(T\cap W)\backslash S|}\right.\\
&\qquad\left.+\sum_{\substack{i\in R_W, W\backslash S\supsetneq\{i\},\\ i\in T}}\frac{1}{|(T\cap R_W)\backslash S|}\right]\frac{x_W^{(S)}x_V^{(T)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(a)}{=}\sum_{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)}}\left[\sum_{\substack{S:|S\cap W|<K_W-1,\\ (T\cap R_W)\backslash S=\varnothing, i\in(T\cap W)\backslash S}}\frac{\zeta_W^{(i)}}{|(T\cap W)\backslash S|}\right.\\
&\qquad\left.+\sum_{\substack{|S\cap W|<K_W-1,\\ (T\cap R_W)\backslash S\neq\varnothing,\\ i\in(T\cap R_W)\backslash S}}\frac{1}{|(T\cap R_W)\backslash S|}\right]\frac{x_W^{(S)}x_V^{(T)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(b)}{\leqslant}\sum_{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)}}\left[\sum_{\substack{S:|S\cap W|<K_W-1,\\ (T\cap R_W)\backslash S=\varnothing, i\in(T\cap W)\backslash S}}\frac{\zeta_W^{(i)}}{|(T\cap W)\backslash S|}\right.\\
&\qquad+\sum_{\substack{|S\cap W|<K_W-1,\\ (T\cap R_W)\backslash S\neq\varnothing, i\in(T\cap R_W^c)\backslash S}}\frac{\zeta_W^{(i)}}{|(T\cap W)\backslash S|}\\
&\qquad\left.+\sum_{\substack{|S\cap W|<K_W-1,\\ (T\cap R_W)\backslash S\neq\varnothing, i\in(T\cap R_W)\backslash S}}\frac{1}{|(T\cap W)\backslash S|}\right]\frac{x_W^{(S)}x_V^{(T)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(c)}{=}\sum_{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)}}\left[\sum_{\substack{i\in R_W^c, W\backslash S\supsetneq\{i\}\\ T:i\in T}}\frac{\zeta_W^{(i)}}{|(T\cap W)\backslash S|}\right.\\
&\qquad\left.+\sum_{\substack{i\in R_W, W\backslash S\supsetneq\{i\},\\ T:i\in T}}\frac{1}{|(T\cap W)\backslash S|}\right]\frac{x_W^{(S)}x_V^{(T)}}{|\mathbf{x}|}g(i,\mathbf{x})\\
&\overset{(d)}{\leqslant}\sum_{\substack{i\in W,\\ W\backslash S\supsetneq\{i\}}}\frac{x_W^{(S)}}{|\mathbf{x}|}\frac{a_W^{(i)}}{K_W}\sum_{\substack{V\in\mathcal{W}_W^{(ally\text{-}\downarrow)},\\ T:i\in T}}x_V^{(T)}g(i,\mathbf{x}).
\end{aligned}\tag{C.65}$$

Steps (a) and (c) change the order of summation; step (b) uses $\zeta_W^{(i)}\leqslant 1$ and $0\geqslant g(n,\mathbf{x})\geqslant g(r,\mathbf{x})$ for all $n\in R_W^c, r\in R_W$;

and step (d) uses $|(T \cap W)\backslash S| \leqslant K_W$. Adding (C.64) and (C.65) gives the below upper-bound on (C.63),

$$\sum_{\substack{i \in W, \\ W\backslash S \supsetneq \{i\}}} \frac{x_W^{(S)}}{|\mathbf{x}|} \frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W} \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)} \right.$$
$$\left. + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}, i \in R_W\right] \underline{\psi}_W \right]. \tag{C.66}$$

Adding (C.66) with the second term in (39), we get

$$\begin{aligned}
QV_W^{(W,+)} \leqslant & \sum_{\substack{i \in W, \\ W\backslash S \supsetneq \{i\}}} \frac{x_W^{(S)}}{|\mathbf{x}|} \frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W} \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \right. \\
& \times \left(\mathbb{1}\left[i \in R_W^c\right] K_W \Upsilon \overline{\psi}_W + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] \right) \\
& \left. + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)} \right] \\
= & \frac{|\mathbf{x}|_W}{|\mathbf{x}|} \sum_{\substack{i \in W, \\ W\backslash S \supsetneq \{i\}}} \frac{x_W^{(S)}}{|\mathbf{x}|_W} \frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W} \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \right. \\
& \times \left(\mathbb{1}\left[i \in R_W^c\right] K_W \Upsilon \overline{\psi}_W + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] \right) \\
& \left. + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)} \right] \\
= & \frac{|\mathbf{x}|_W}{|\mathbf{x}|} \sum_{i \in W} \frac{\underline{\Gamma}_W^{(i)} a_W^{(i)}}{K_W} \left(1 - \pi_W^{(i)} - \gamma_W^{(i)}\right) \left[\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] \right. \\
& \times \left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] \right) \\
& \left. + \mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right] 4K_W^2 - C_W^{(1)} - I_W^{(1)} \right],
\end{aligned}$$

where we recall that $\gamma_W^{(i)}$ is the fraction of swarm-W peers who have all the pieces of file-W except i. ■

Appendix D
Exponential Decay of Positive Drift from Non-Rare Pieces in $\mathcal{R}_W^{(1)}$

Lemma 9: For all $\mathbf{x} \in \mathcal{R}_W^{(1)}$,

$$\sum_{n \in R_W^c} \frac{\underline{\Gamma}_W^{(n)} \zeta_W^{(n)}}{K_W} \left[(1 - \overline{\pi}_W) K_W \Upsilon \overline{\psi}_W \right] \leqslant D_W.$$

where

$$\begin{aligned}
D_W &= D_W\left(\eta, \Upsilon, L, U, \kappa_p, K_W, \mathcal{W}_W^{(ally\text{-}\downarrow)}\right) \\
&:= D_W^{(1)} + D_W^{(2)} \mathbb{1}\left[\mathcal{W}_W^{(ally\text{-}\downarrow)} \backslash \{W\} \neq \varnothing\right], \\
D_W^{(1)} &= 12e^{-2}\eta^{-2}\Upsilon\left(LU + \kappa_p\right) \beta_W^2 K_W^7, \\
D_W^{(2)} &= 3\frac{e^{-\left(1+\alpha_W^{-1}\right)}}{\alpha_W^{\alpha_W^{-1}}} \eta^{-1} \Upsilon\left(LU + \kappa_p\right) \beta_W^{1+\alpha_W^{-1}} K_W^{5+\alpha_W^{-1}}.
\end{aligned}$$

Proof:

$$\begin{aligned}
& \sum_{n \in R_W^c} \frac{\underline{\Gamma}_W^{(n)} \zeta_W^{(n)}}{K_W} \left[(1 - \overline{\pi}_W) K_W \Upsilon \overline{\psi}_W \right] \\
& \overset{(a)}{\leqslant} \sum_{n \in R_W^c} \left(LU + 2\kappa_p \left(\overline{c}_W + d_W^{(n)} \right) \right) \Upsilon \left(3K_W^2 \overline{c}_W \right) \zeta_W^{(n)} \\
& \overset{(b)}{\leqslant} 3K_W^2 \Upsilon \left(LU + 2\kappa_p \right) \sum_{n \in R_W^c} \left(\overline{c}_W + d_W^{(n)} \right) \overline{c}_W \zeta_W^{(n)}.
\end{aligned} \tag{D.67}$$

Here, (a) uses $\overline{\psi}_W \leqslant 3K_W^2 \overline{c}_W$ in $\mathcal{R}_W^{(1)}$; (b) uses $1 \leqslant \overline{c}_W + d_W^{(n)}$. Now, in $\mathcal{R}_W^{(1)}$, $M_W \geqslant \eta \overline{c}_W$. This implies that

$$\overline{m}_W = \overline{c}_W - \underline{c}_W \geqslant \frac{\eta \overline{c}_W}{K_W}.$$

Therefore,

$$\zeta_W^{(n)} \leqslant \exp\left(-\frac{\eta \overline{c}_W K_W^{-1} + \left(d_W^{(n)}\right)^{\alpha_W}}{\beta_W K_W} \right).$$

If $\beta_W = 0$, then $\zeta_W^{(n)} = 0$ (non-rare pieces are completely suppressed) and the Lemma follows trivially. If $\beta_W > 0$, then

$$\begin{aligned}
\sum_{n \in R_W^c} \overline{c}_W^2 \zeta_W^{(n)} &\leqslant K_W \max_{y \geqslant 0} y^2 \exp\left(-\frac{\eta}{\beta_W K_W^2} y \right) \\
&= K_W \left(2\frac{\beta_W K_W^2}{\eta} \right)^2 e^{-2} \\
&= 4e^{-2}\eta^{-2}\beta_W^2 K_W^5,
\end{aligned} \tag{D.68}$$

and

$$\begin{aligned}
& \sum_{n \in R_W^c} \overline{c}_W d_W^{(n)} \zeta_W^{(n)} \\
& = \sum_{n \in R_W^c} \overline{c}_W d_W^{(n)} \zeta_W^{(n)} \mathbb{1}\left[\mathcal{W}_W^{(ally\text{-}\uparrow)} \backslash \{W\} \neq \varnothing\right] \\
& \leqslant K_W \max_{y \geqslant 0} y \exp\left(-\frac{\eta}{\beta_W K_W^2} y \right) \\
& \quad \times \max_{z \geqslant 0} z \exp\left(-\frac{z^{\alpha_W}}{\beta_W K_W} \right) \mathbb{1}\left[\mathcal{W}_W^{(ally\text{-}\uparrow)} \backslash \{W\} \neq \varnothing\right] \\
& = K_W \left(\frac{\beta_W K_W^2}{\eta} \right) e^{-1} \\
& \quad \times \left(\frac{\beta_W K_W}{\alpha_W} \right)^{\frac{1}{\alpha_W}} e^{-\frac{1}{\alpha_W}} \mathbb{1}\left[\mathcal{W}_W^{(ally\text{-}\uparrow)} \backslash \{W\} \neq \varnothing\right] \\
& = \frac{e^{-\left(1+\alpha_W^{-1}\right)}}{\alpha_W^{\alpha_W^{-1}}} \eta^{-1} \beta_W^{1+\alpha_W^{-1}} K_W^{3+\alpha_W^{-1}} \mathbb{1}\left[\mathcal{W}_W^{(ally\text{-}\uparrow)} \backslash \{W\} \neq \varnothing\right].
\end{aligned} \tag{D.69}$$

Using (D.68) and (D.69) in (D.67), the Lemma follows. ■

Appendix E
Proof of Theorem 1

For each $W \in \mathcal{W}$, let $\delta_W \in (0, 1)$ be sufficiently small number to be chosen appropriately, and recall that the two blocks $\mathcal{R}_W^{(1)} = \{\mathbf{x} \in \mathcal{S} : M_W - 2(1-\eta) \geqslant \eta \overline{c}_W\}$ and $\mathcal{R}_W^{(2)} = \mathcal{S} \backslash \mathcal{R}_W^{(1)}$ form a partition of the state space $\mathcal{S}$. For each $k \in [2]$, we further divide the block $\mathcal{R}_W^{(k)}$ into three regions, namely

$$\mathcal{R}_W^{(k1)} = \mathcal{R}_W^{(k)} \cap \left\{ (K_W - \eta)\overline{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} \geqslant \delta_W |\mathbf{x}|_W \right\},$$

$$\mathcal{R}_W^{(k2)} = \mathcal{R}_W^{(k)} \cap \left\{ \delta_W |\mathbf{x}|_W > (K_W - \eta)\overline{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} \right.$$

$$\geqslant \frac{(C_W^{(3)}-2)(1-\eta)}{K_W}\Bigg\},$$

$$\mathcal{R}_W^{(k3)} = \mathcal{R}_W^{(k)} \cap \Bigg\{(K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} < \delta_W|\mathbf{x}|_W \text{ and }$$
$$(K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} < \frac{(C_W^{(3)}-2)(1-\eta)}{K_W}\Bigg\}.$$

Below we prove Theorem 1 through a series of lemmas, together which establish that the unit-transition drift conditions are satisfied by the candidate Lyapunov function given by (30) and (31).

Lemma 10: For all $W \in \mathcal{W}$, $k \in [2]$, and $\mathbf{x} \in \mathcal{R}_W^{(k1)}$, the highest frequency $\overline{\pi}_W$ is lower-bounded by $\frac{\delta_W}{K_W-\eta}$. Therefore, the total chunk-count $P_W^{(\text{tot})}$ is lower-bounded by $\frac{\delta_W}{K_W-\eta}|\mathbf{x}|_W$.

Proof: We can lower-bound $P_W^{(\text{tot})}$ by $\overline{c}_W$. Given $(K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} \geqslant \delta_W|\mathbf{x}|_W$ in $\mathcal{R}_W^{(k1)}$, the Lemma is obvious. ∎

Lemma 11: For all $W \in \mathcal{W}$, $k \in [2]$, and $\mathbf{x} \in \mathcal{R}_W^{(k2)} \cup \mathcal{R}_W^{(k3)}$, the fraction of peers missing piece $i \in W$, i.e., $1-\pi_W^{(i)}$ is lower-bounded by 0.5 provided $\delta_W \leqslant 0.5(1-\eta)$.

Proof: In $\mathcal{R}_W^{(k2)} \cup \mathcal{R}_W^{(k3)}$, we have $(K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} < \delta_W|\mathbf{x}|_W$. Since $c_W^{(i)}$ is at most $\overline{c}_W$, this implies that $(1-\eta)\overline{c}_W < \delta_W|\mathbf{x}|_W \implies \overline{\pi}_W < \frac{\delta_W}{1-\eta}$. Now, for any piece $i \in W$, we have $1-\pi_W^{(i)} \geqslant 1-\overline{\pi}_W > 1-\frac{\delta_W}{1-\eta}$. By choosing $\delta_W \leqslant 0.5(1-\eta)$, we ensure that $1-\pi_W^{(i)} \geqslant 0.5$. ∎

Lemma 12: For all $W \in \mathcal{W}$, $k \in [2]$, and $\mathbf{x} \in \mathcal{R}_W^{(k2)} \cup \mathcal{R}_W^{(k3)}$, the fraction of peers who are missing only piece $i \in W$, i.e., $\gamma_W^{(i)} = \sum_{S:W\setminus S=\{i\}} \frac{x_W^{(S)}}{|\mathbf{x}|_W}$ is upper bounded by $\frac{\delta_W}{1-\eta}$.

Proof: A swarm-W peer missing only piece i of file-W has all the other pieces of the file, therefore,

$$\begin{aligned}\sum_{S:W\setminus S=\{i\}} x_W^{(S)} &= \gamma_W^{(i)}|\mathbf{x}|_W \\ &\leqslant \pi_W^{(\hat{i})}|\mathbf{x}|_W \text{ for all } \hat{i}\in W \text{ such that } \hat{i}\neq i \\ &\leqslant \overline{\pi}_W|\mathbf{x}|_W.\end{aligned}$$

Given $\overline{\pi}_W < \frac{\delta_W}{1-\eta}$ in $\mathcal{R}_W^{(k2)} \cup \mathcal{R}_W^{(k3)}$, the inequality follows. ∎

Lemma 13: For all $W \in \mathcal{W}$, $\theta_1 > 0$, and $\theta_2 \geqslant 2(1-\eta)^2$, let g be defined as

$$g(\theta_1,\theta_2) = \sum_{r\in R_W} \frac{\Gamma_W^{(r)}}{\theta_1}\left[\left(1-\pi_W^{(r)}\right)\left(\underline{\psi}_W - \theta_2\right)\right]$$

Then, for all $\mathbf{x} \in \mathcal{R}_W^{(1)}$, g is upper-bounded by

$$-\frac{2(1-\eta)}{\theta_1 K_W}\Bigg[2(1-\eta)\left(LU-2(1-\eta)\kappa_p\right) + \left((K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)}\right)\left(LU \wedge 2(1-\eta)\kappa_p\right)\Bigg].$$

Consequently, for any θ and $\epsilon > 0$, there exists $N_W^{(1)} = N_W^{(1)}(\theta,\theta_1,\theta_2,\epsilon) \in \mathbb{R}_{>0}$ such that for all $\mathbf{x} \in \mathcal{R}_W^{(1)}$,

$$\theta + g(\theta_1,\theta_2) \leqslant -\epsilon$$

if $(K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} \geqslant N_W^{(1)}$.

Proof: Since $\mathbf{x} \in \mathcal{R}_W^{(1)}$, $\underline{\psi}_W \leqslant 0$. Given $\theta_2 > 0$, each term in the summation is negative. Let us denote the rarest-piece by $\underline{r}$, i.e., $c_W^{(\underline{r})} = \underline{c}_W$. By Lemma 2, $1-\pi_W^{(\underline{r})}$ is bounded from below by $\frac{1}{K_W}$. Upper-bounding the summation in g by considering only the rarest-piece $\underline{r}$, we get

$$\begin{aligned}&g(\theta_1,\theta_2) \\ &\leqslant \frac{\Gamma_W^{(\underline{r})}}{\theta_1 K_W}\left[2(1-\eta)^2 - 2(1-\eta)\left(M_W-\eta\overline{c}_W\right) - \theta_2\right] \\ &\overset{(a)}{=} -\frac{2(1-\eta)\Gamma_W^{(\underline{r})}}{\theta_1 K_W}\left[\left((K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} - \underline{c}_W\right)\right] \\ &\overset{(b)}{\leqslant} -\frac{2(1-\eta)}{\theta_1 K_W}\left(LU+\kappa_p\underline{c}_W\right) \\ &\qquad \times\left((K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} - \underline{c}_W\right) \\ &:= \tilde{g}(\underline{c}_W;\theta_1),\end{aligned}$$

where, (a) uses $\theta_2 \geqslant 2(1-\eta)^2$ and the definition of M_W (see (11)); and (b) uses $\Gamma_W^{(\underline{r})} \geqslant LU + \kappa_p\underline{c}_W$ (see (19)).

Since $\mathbf{x} \in \mathcal{R}_W^{(1)}$, we have $M_W - 2(1-\eta) \geqslant \eta\overline{c}_W$, which implies that $\underline{c}_W \leqslant (K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} - 2(1-\eta)$. It can be verified that $\tilde{g}(\underline{c}_W;\theta_1)$ is a quadratic and strictly convex function of $\underline{c}_W$. Therefore, it has a unique minimum and attains its maximum at the boundary points, 0 and $(K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} - 2(1-\eta)$. This gives

$$\begin{aligned}&g(\theta_1,\theta_2) \leqslant \tilde{g}(0) \wedge \tilde{g}\left((K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)} - 2(1-\eta)\right) \\ &= -\frac{2(1-\eta)}{\theta_1 K_W}\Bigg[2(1-\eta)\left(LU-2(1-\eta)\kappa_p\right) \\ &\quad + \left((K_W-\eta)\overline{c}_W - \sum_{i\neq \underline{r}} c_W^{(i)}\right)\left(LU \wedge 2(1-\eta)\kappa_p\right)\Bigg].\end{aligned}$$

∎

Lemma 14: For any $\epsilon > 0$ and any $W \in \mathcal{W}$, there exists a countable set $\mathcal{A}_W$ and a finite constant $B_W \geqslant 0$ such that

$$\begin{aligned}\widetilde{QV}_W &\leqslant -\epsilon\mathbb{1}\left[\mathbf{x}\in\mathcal{A}_W^c\right] + B_W\mathbb{1}\left[\mathbf{x}\in\mathcal{A}_W\right], \text{ and} \\ N_W &= \max_{\mathbf{x}\in\mathcal{A}_W} |\mathbf{x}|_W < \infty.\end{aligned}$$

Proof: We will show that $\widetilde{QV}_W$ is upper bounded by $-\epsilon$ over the region $\mathcal{R}_W^{(k)}$ for each $k \in [2]$, except possibly, for a finite and bounded population of swarm-W.

Case 1 - $\mathbf{x} \in \mathcal{R}_W^{(1)}$: From (44) and (42), we have

$$\begin{aligned}\widetilde{QV}_W = |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\Big[&\left(1-\pi_W^{(i)}\right)\Psi_W^{(i)} \\ &- \left(1-\pi_W^{(i)}-\gamma_W^{(i)}\right)I_W^{(1)} + \gamma_W^{(i)}K_W\Upsilon I_W^{(2)}\Big],\end{aligned}$$

where

$$\Psi_W^{(i)}(\mathbf{x}) = K_W\Upsilon\overline{\psi}_W\mathbb{1}\left[i\in R_W^c\right] + \underline{\psi}_W\mathbb{1}\left[i\in R_W\right] - C_W^{(1)}.$$

Region $\mathcal{R}_W^{(11)}$: Here, $P_W^{(\text{tot})} \geqslant \bar{c}_W \geqslant \frac{\delta_W}{K_W-\eta}|\mathbf{x}|_W$ (see Lemma 10). Therefore, for all large $|\mathbf{x}|_W$, $I_W^{(1)}, I_W^{(2)}$ are both zero. This gives

$$\begin{aligned}
\widetilde{QV}_W &\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(r)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\left.\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] - C_W^{(1)}\right)\right] \\
&\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{n \in R_W^c} \frac{\Gamma_W^{(n)} \zeta_W^{(n)}}{K_W}\left[\left(1-\overline{\pi}_W\right) K_W \Upsilon \overline{\psi}_W\right] \\
&+ \sum_{r \in R_W} \frac{\Gamma_W^{(r)}}{K_W}\left[\left(1-\pi_W^{(r)}\right)\left(\underline{\psi}_W - C_W^{(1)}\right)\right] \\
&\leqslant |\boldsymbol{\lambda}|C^{(1)} + D_W + \sum_{r \in R_W} \frac{\Gamma_W^{(r)}}{K_W}\left[\left(1-\pi_W^{(r)}\right)\underline{\psi}_W\right],
\end{aligned}$$

where $D_W = D_W(\eta, \Upsilon, L, U, \kappa_p, \beta_W, K_W)$ is given by Lemma 9. From Lemma 13, it follows that $\widetilde{QV}_W \leqslant -\epsilon$ if $(K_W-\eta)\bar{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} \geqslant N_W^{(1)}(\theta = |\boldsymbol{\lambda}|C^{(1)} + D_W, \theta_1 = K_W, \theta_2 = C_W^{(1)}, \epsilon)$. Since $(K_W-\eta)\bar{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} \geqslant \delta_W|\mathbf{x}|_W$, this is true for all large $|\mathbf{x}|_W$.

Region $\mathcal{R}_W^{(12)}$: Here, we use the bounds, $-I_W^{(1)} \leqslant 0$, $I_W^{(2)} \leqslant K_W C_W^{(2)}$, and $\gamma_W^{(i)} < \frac{\delta_W}{1-\eta}$ for every $i \in W$ (Lemma 12). Then, for all large $|\mathbf{x}|_W$,

$$\begin{aligned}
\widetilde{QV}_W &\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\times\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] - C_W^{(1)}\right) \\
&\left.+\frac{\delta_W}{1-\eta} K_W^2 \Upsilon C_W^{(2)}\right] \\
&\overset{\text{(a)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right] - C_W^{(1)}\right. \\
&\left.\left.+\frac{2\delta_W}{1-\eta} K_W^2 \Upsilon C_W^{(2)}\right)\right] \\
&\overset{\text{(b)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right] + \underline{\psi}_W \mathbb{1}\left[i \in R_W\right]\right. \\
&\left.\left.-0.5 C_W^{(1)}\right)\right] \\
&\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{n \in R_W^c} \frac{\Gamma_W^{(n)} \zeta_W^{(n)}}{K_W}\left[\left(1-\overline{\pi}_W\right) K_W \Upsilon \overline{\psi}_W\right] \\
&+ \sum_{r \in R_W^c} \frac{\Gamma_W^{(r)}}{K_W}\left[\left(1-\pi_W^{(r)}\right)\left(\underline{\psi}_W - 0.5 C_W^{(1)}\right)\right] \\
&\overset{\text{(c)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + D_W + \sum_{r \in R_W} \frac{\Gamma_W^{(r)}}{K_W}\left[\left(1-\pi_W^{(r)}\right)\underline{\psi}_W\right].
\end{aligned}$$

Here, (a) uses $1-\pi_W^{(i)} \geqslant 0.5$ for every $i \in W$ (Lemma 11); (b) follows from choosing δ_W small enough so that $2\frac{\delta_W}{1-\eta} K_W^2 \Upsilon C_W^{(2)} \leqslant 0.5 C_W^{(1)}$; (c) follows from Lemma 9. From Lemma 13, it follows that $\widetilde{QV}_W \leqslant -\epsilon$ if $(K_W-\eta)\bar{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} \geqslant N_W^{(1)}(\theta = |\boldsymbol{\lambda}|C^{(1)} + D_W, \theta_1 = K_W, \theta_2 = 0.5 C_W^{(1)}, \epsilon)$. Since $(K_W-\eta)\bar{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} \geqslant \frac{(C_W^{(3)}-2)(1-\eta)}{K_W}$, choosing $C_W^{(3)}$ large enough so that $\frac{(C_W^{(3)}-2)(1-\eta)}{K_W} \geqslant N_W^{(1)}(\theta = |\boldsymbol{\lambda}|C^{(1)} + D_W, \theta_1 = K_W, \theta_2 = 0.5 C_W^{(1)}, \epsilon)$ ensures that $\widetilde{QV}_W \leqslant -\epsilon$ for large enough $|\mathbf{x}|_W$.

Region $\mathcal{R}_W^{(13)}$: Here, $(K_W-\eta)\bar{c}_W - \sum_{i \neq \underline{r}} c_W^{(i)} < \frac{(C_W^{(3)}-2)(1-\eta)}{K_W}$. This implies that $\bar{c}_W < \frac{C_W^{(3)}-2}{K_W}$. Therefore, the total chunk-count $P_W^{(\text{tot})}$ is upper-bounded by $K_W \bar{c}_W \leqslant C_W^{(3)} - 2$. Consequently, $I_W^{(1)}$ and $I_W^{(2)}$ are both non-zero. We use the bounds $\underline{\psi}_W \mathbb{1}[i \in R_W] - C_W^{(1)} \leqslant 0$ and $\gamma_W^{(i)} < \frac{\delta_W}{1-\eta}$ for every $i \in W$ (Lemma 12). Then, for all large $|\mathbf{x}|_W$,

$$\begin{aligned}
\widetilde{QV}_W &\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\times\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right]\right) - \left(1-\pi_W^{(i)} - \frac{\delta_W}{1-\eta}\right) C_W^{(2)} \\
&\left.+\frac{\delta_W}{1-\eta} K_W^2 \Upsilon C_W^{(2)}\right] \\
&\overset{\text{(a)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right]\right)\right. \\
&\left.-0.5 C_W^{(2)} + \frac{\delta_W}{1-\eta} C_W^{(2)}\left(1 + K_W^2 \Upsilon\right)\right] \\
&\overset{\text{(b)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\right. \\
&\left.\times\left(K_W \Upsilon \overline{\psi}_W \mathbb{1}\left[i \in R_W^c\right]\right) - 0.25 C_W^{(2)}\right] \\
&\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{n \in R_W^c} \frac{\Gamma_W^{(n)} \zeta_W^{(n)}}{K_W}\left[\left(1-\overline{\pi}_W\right)\left(K_W \Upsilon \overline{\psi}_W\right)\right] \\
&+ \sum_{r \in R_W} \frac{\Gamma_W^{(r)}}{K_W}\left[-0.25 C_W^{(2)}\right] \\
&\overset{\text{(c)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + D_W + \sum_{r \in R_W} \frac{\Gamma_W^{(r)}}{K_W}\left[-0.25 C_W^{(2)}\right] \\
&\overset{\text{(d)}}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + D_W - 0.25 K_W^{-1} L U C_W^{(2)} \\
&\overset{\text{(e)}}{\leqslant} -\epsilon.
\end{aligned}$$

Here, (a) uses $1-\pi_W^{(i)} \geqslant 0.5$ for $i \in W$ (Lemma 11); (b) follows from choosing δ_W small enough so that $\frac{\delta_W}{1-\eta} C_W^{(2)}(1 + K_W^2 \Upsilon) \leqslant 0.25 C_W^{(2)}$; (c) follows from Lemma 9; (d) upper-bounds the summation by considering only the rarest piece and using $\Gamma_W^{(i)} \geqslant LU$ (see (19)); (e) follows by choosing $C_W^{(2)}$ large enough so that $0.25 K_W^{-1} L U C_W^{(2)} \geqslant |\boldsymbol{\lambda}|C^{(1)} + D_W + \epsilon$.

Case 2 - $\mathbf{x} \in \mathcal{R}_W^{(2)}$: From (44) and (42), we can write

$$\widetilde{QV}_W \leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i \in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(4K_W^2 - C_W^{(1)}\right)\right.$$

$$-\left(1-\pi_W^{(i)}-\gamma_W^{(i)}\right) I_W^{(1)} + \gamma_W^{(i)} K_W \Upsilon I_W^{(2)}\Big].$$

Region $\mathcal{R}_W^{(21)}$: Like in $\mathcal{R}_W^{(11)}$, $I_W^{(1)}, I_W^{(2)}$ are both zero for large $|\mathbf{x}|_W$. Since $\mathbf{x} \in \mathcal{R}_W^{(2)}$, $M_W - 2(1-\eta) < \eta \bar{c}_W$. This implies that $\overline{m}_W < \eta \bar{c}_W + 2(1-\eta) \implies \underline{c}_W > (1-\eta)\bar{c}_W - 2 \geqslant \frac{\delta_W(1-\eta)}{K_W-\eta}|\mathbf{x}|_W - 2$. Then, for all large $|\mathbf{x}|_W$,

$$\begin{aligned}
\widetilde{QV}_W &\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(4K_W^2 - C_W^{(1)}\right)\right] \\
&\overset{(a)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \frac{\underline{\Gamma}_W^{(r)}}{K_W}\left[\left(1-\underline{\pi}_W\right)\left(-0.5C_W^{(1)}\right)\right] \\
&\overset{(b)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + K_W^{-2}\underline{\Gamma}_W^{(r)}\left[-0.5C_W^{(1)}\right] \\
&\overset{(c)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} - 0.5C_W^{(1)}K_W^{-2}\kappa_p \underline{c}_W \\
&\overset{(d)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} - 0.5C_W^{(1)}K_W^{-2}\kappa_p\left(\frac{\delta_W(1-\eta)}{K_W-\eta}|\mathbf{x}|_W - 2\right) \\
&\overset{(e)}{\leqslant} -\epsilon.
\end{aligned}$$

Here, (a) uses $4K_W^2 - C_W^{(1)} \leqslant -0.5C_W^{(1)}$ and then upper-bounds the summation by considering only the rarest-piece, that is denoted by $\underline{r}$; (b) uses $1-\underline{\pi}_W \geqslant K_W^{-1}$ (Lemma 2); (c) uses $\underline{\Gamma}_W^{(r)} \geqslant \kappa_p \underline{c}_W$ (see (19)); (d) uses $\underline{c}_W \geqslant \frac{\delta_W(1-\eta)}{K_W-\eta}|\mathbf{x}|_W - 2$; (e) uses the fact that $|\mathbf{x}|_W$ is large enough.

Region $\mathcal{R}_W^{(22)}$: Like $\mathcal{R}_W^{(12)}$ we use the bounds, $-I_W^{(1)} \leqslant 0$, $I_W^{(2)} \leqslant K_W C_W^{(2)}$, and $\gamma_W^{(i)} < \frac{\delta_W}{1-\eta}$ for every $i \in W$ (Lemma 12). Then,

$$\begin{aligned}
\widetilde{QV}_W &\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(4K_W^2 - C_W^{(1)}\right)\right. \\
&\qquad \left. + \frac{\delta_W}{1-\eta}K_W^2 \Upsilon C_W^{(2)}\right] \\
&\overset{(a)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(-0.5C_W^{(1)}\right)\right. \\
&\qquad \left. + \frac{\delta_W}{1-\eta}K_W^2 \Upsilon C_W^{(2)}\right] \\
&\overset{(b)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(-0.5C_W^{(1)}\right.\right. \\
&\qquad \left.\left. + \frac{2\delta_W}{1-\eta}K_W^2 \Upsilon C_W^{(2)}\right)\right] \\
&\overset{(c)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[\left(1-\pi_W^{(i)}\right)\left(-0.25C_W^{(1)}\right)\right] \\
&\overset{(d)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[-0.125C_W^{(1)}\right] \\
&\overset{(e)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} - 0.125C_W^{(1)}K_W^{-1}\underline{\Gamma}_W^{(r)} \\
&\overset{(f)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} - 0.125C_W^{(1)}K_W^{-1}\kappa_p \underline{c}_W \\
&\overset{(g)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} - 0.125C_W^{(1)}K_W^{-1}\kappa_p\left(\frac{(C_W^{(3)}-2)(1-\eta)^2}{K_W(K_W-\eta)} - 2\right) \\
&\overset{(h)}{\leqslant} -\epsilon.
\end{aligned}$$

Here, (a) uses $4K_W^2 - C_W^{(1)} \leqslant -0.5C_W^{(1)}$; (b) uses $1-\pi_W^{(i)} \geqslant 0.5$ for every $i \in W$ (Lemma 11); (c) follows from choosing δ_W small enough so that $\frac{2\delta_W}{1-\eta}K_W^2 \Upsilon C_W^{(2)} \leqslant 0.25C_W^{(1)}$; (d) uses $1-\pi_W^{(i)} \geqslant 0.5$ for every $i \in W$ (Lemma 11); (e) upper-bounds the summation by considering only the rarest piece (denoted by $\underline{r}$); (f) uses $\underline{\Gamma}_W^{(r)} \geqslant \kappa_p \underline{c}_W$ (see (19)); (g) uses $\underline{c}_W \geqslant (1-\eta)\bar{c}_W - 2$ and $\bar{c}_W \geqslant \frac{(C_W^{(3)}-2)(1-\eta)}{K_W(K_W-\eta)}$; (h) follows by choosing $C_W^{(3)}$ large enough so that $0.125C_W^{(1)}K_W^{-1}\kappa_p\left(\frac{(C_W^{(3)}-2)(1-\eta)^2}{K_W(K_W-\eta)} - 2\right) \geqslant |\boldsymbol{\lambda}|C^{(1)} + \epsilon$.

Region $\mathcal{R}_W^{(23)}$: Like $\mathcal{R}_W^{(13)}$, $I_W^{(1)}$ and $I_W^{(2)}$ are both non-zero. We use the bounds, $I_W^{(2)} \leqslant K_W C_W^{(2)}$, $4K_W^2 - C_W^{(1)} \leqslant 0$, $\gamma_W^{(i)} < \frac{\delta_W}{1-\eta}$ for every $i \in W$ (Lemma 12). Then,

$$\begin{aligned}
\widetilde{QV}_W &\leqslant |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[-\left(1-\pi_W^{(i)} - \frac{\delta_W}{1-\eta}\right)C_W^{(2)}\right. \\
&\qquad \left. + \frac{\delta_W}{1-\eta}K_W^2 \Upsilon C_W^{(2)}\right] \\
&\overset{(a)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[-0.5C_W^{(2)}\right. \\
&\qquad \left. + \frac{\delta_W}{1-\eta}C_W^{(2)}\left(1+K_W^2\Upsilon\right)\right] \\
&\overset{(b)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} + \sum_{i\in W} \frac{\Gamma_W^{(i)} a_W^{(i)}}{K_W}\left[-0.25C_W^{(2)}\right] \\
&\overset{(c)}{\leqslant} |\boldsymbol{\lambda}|C^{(1)} - 0.25K_W^{-1}LUC_W^{(2)} \\
&\overset{(d)}{\leqslant} -\epsilon.
\end{aligned}$$

Here, (a) uses $1-\pi_W^{(i)} \geqslant 0.5$ for every $i \in W$ (Lemma 11); (b) follows from choosing δ_W small enough so that $\frac{\delta_W}{1-\eta}C_W^{(2)}(1+K_W^2\Upsilon) \leqslant 0.25C_W^{(2)}$; (c) upper-bounds the summation by considering only the rarest piece and uses $\underline{\Gamma}_W^{(i)} \geqslant LU$ (see (19)); (d) follows by choosing $C_W^{(2)}$ large enough so that $0.25K_W^{-1}LUC_W^{(2)} \geqslant |\boldsymbol{\lambda}|C^{(1)} + \epsilon$. ∎

Lemma 15: For any $\epsilon' > 0$, there exists a finite set $\mathcal{A}$ and a finite constant B such that

$$QV(\mathbf{x}) \leqslant -\epsilon' \mathbb{1}\left[\mathbf{x} \in \mathcal{A}^c\right] + B\mathbb{1}\left[\mathbf{x} \in \mathcal{A}\right].$$

Proof: Fix $\epsilon' > 0$. For any $\epsilon > 0$ and any $W \in \mathcal{W}$, it follows from Lemma 14 that $\widetilde{QV}_W \leqslant -\epsilon$, except possibly over $\mathcal{A}_W$ where its population and corresponding term $\widetilde{QV}_W$ are bounded from above by N_W and $B_W \geqslant 0$ respectively. We can write the state space as $\mathcal{S} = \bigcup_{\mathcal{H}:\mathcal{H}\subseteq\mathcal{W}} \mathcal{S}_{\mathcal{H}}$, where

$$\begin{aligned}
\mathcal{S}_{\mathcal{H}} := \{&|\mathbf{x}|_{W_1} \leqslant N_{W_1} \text{ for all } W_1 \in \mathcal{H}, \text{ and} \\
&|\mathbf{x}|_{W_2} > N_{W_2} \text{ for all } W_2 \in \mathcal{W}\backslash\mathcal{H}\}.
\end{aligned}$$

Case 1 - $\mathcal{H} = \varnothing$: Let $\mathbf{x} \in \mathcal{S}_{\varnothing}$. Since $|\mathbf{x}|_W > N_W$ for all $W \in \mathcal{W}$, from Lemma 14, it follows that $\widetilde{QV}_W \leqslant -\epsilon$ for all $W \in \mathcal{W}$. From (43), this gives $QV \leqslant -\epsilon$. Choosing $\epsilon \geqslant \epsilon'$ ensures $QV \leqslant -\epsilon'$.

Case 2 - $\mathcal{H} = \mathcal{W}$: The set $\mathcal{S}_{\mathcal{W}}$ is finite. Therefore, for any $\mathbf{x} \in \mathcal{S}_{\mathcal{W}}$, we can write $QV \leqslant B_{\mathcal{W}} := \max_{\mathbf{x}\in\mathcal{S}_{\mathcal{W}}} (QV)^+ < \infty$.

Case 3 - $\varnothing \neq \mathcal{H} \subsetneq \mathcal{W}$: Let $\mathbf{x} \in \mathcal{S}_{\mathcal{H}}$. We can upper-bound QV as follows.

$$QV = \sum_{W_1 \in \mathcal{H}} \frac{|\mathbf{x}|_{W_1}}{|\mathbf{x}|} \widetilde{QV}_{W_1} + \sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} \frac{|\mathbf{x}|_{W_2}}{|\mathbf{x}|} \widetilde{QV}_{W_2}$$
$$\leqslant \frac{\sum_{W_1 \in \mathcal{H}} N_{W_1} B_{W_1}}{\sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2}} + \frac{\sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2}}{\sum_{W_1 \in \mathcal{H}} N_{W_1} + \sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2}} (-\epsilon).$$

Note that $\sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2} \to \infty$ over the set $\mathcal{S}_{\mathcal{H}}$, which implies

$$\frac{\sum_{W_1 \in \mathcal{H}} N_{W_1} B_{W_1}}{\sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2}} \to 0 \text{ and}$$
$$\frac{\sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2}}{\sum_{W_1 \in \mathcal{H}} N_{W_1} + \sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2}} \to 1.$$

Therefore, for any $\phi > 0$, there exists $N_{\mathcal{H}} = N_{\mathcal{H}}(\phi) \in \mathbb{R}_{>0}$ such that $QV \leqslant \phi + (1-\phi)(-\epsilon)$ for any $\mathbf{x} \in \mathcal{S}_{\mathcal{H}}$ with $\sum_{W_2 \in \mathcal{W}\setminus\mathcal{H}} |\mathbf{x}|_{W_2} \geqslant N_{\mathcal{H}}$. Choosing $\epsilon = 2\epsilon'$ and $0 < \phi \leqslant \frac{\epsilon/2}{1+\epsilon}$ ensures $QV \leqslant -\epsilon'$.

For all $\mathcal{H}$ such that $\varnothing \neq \mathcal{H} \subsetneq \mathcal{W}$, let us define

$$\mathcal{A}_{\mathcal{H}} := \mathcal{S}_{\mathcal{H}} \cap \left\{ \sum_{U \in \mathcal{H}} |\mathbf{x}|_U < N_{\mathcal{H}} \right\} \qquad \text{(finite)},$$
$$\text{and then } \mathcal{A} := \left(\bigcup_{\varnothing \neq \mathcal{H} \subsetneq \mathcal{W}} \mathcal{A}_{\mathcal{H}} \right) \cup \mathcal{S}_{\mathcal{W}} \qquad \text{(finite)},$$
$$B := \max_{\mathbf{x} \in \mathcal{A}} \left(QV(\mathbf{x}) \right)^+ < \infty.$$

Then, for any state $\mathbf{x}$, we can write

$$QV(\mathbf{x}) \leqslant -\epsilon' \mathbb{1}\left[\mathbf{x} \in \mathcal{A}^c\right] + B \mathbb{1}\left[\mathbf{x} \in \mathcal{A}\right].$$

establishing the result. ■

Lemma 15 is the final stage. Combining Lemma 15 and Proposition 6, we conclude that Theorem 1 holds. As a final step, we now illustrate how the constants $\{C_W^{(1)}, C_W^{(2)}, C_W^{(3)}, \delta_W\}_{W \in \mathcal{W}}$ can be set consistently.

Setting $\{C_W^{(1)}, C_W^{(2)}, C_W^{(3)}, \delta_W\}_{W \in \mathcal{W}}$: Set some $\epsilon' > 0$, $\epsilon = 2\epsilon'$, and $\eta \in (0,1)$.

- For all $W \in \mathcal{W}$, individually set $C_W^{(1)} = 8K_W^2$.
- Set $C^{(1)} = \max_{W \in \mathcal{W}} K_W C_W^{(1)}$.
- For all $W \in \mathcal{W}$, individually set $C_W^{(2)}$ such that
$$0.25 K_W^{-1} L U C_W^{(2)} \geqslant |\boldsymbol{\lambda}| C^{(1)} + D_W + \epsilon.$$
- For all $W \in \mathcal{W}$, individually set δ_W such that
$$\delta_W \leqslant 0.5(1-\eta),$$
$$\frac{2\delta_W}{1-\eta} K_W^2 \Upsilon C_W^{(2)} \leqslant 0.25 C_W^{(1)}, \text{ and}$$
$$\frac{\delta_W}{1-\eta} C_W^{(2)} (1 + K_W^2 \Upsilon) \leqslant 0.25 C_W^{(2)}.$$
- For all $W \in \mathcal{W}$, set $C_W^{(3)}$ so that
$$C_W^{(3)} - 2 \geqslant K_W \frac{N_W^{(1)}\left(|\boldsymbol{\lambda}| C^{(1)} + D_W, K_W, 0.5 C_W^{(1)}, \epsilon\right)}{1-\eta} \text{ and}$$
$$0.125 C_W^{(1)} K_W^{-1} \kappa_p \left(\frac{(C_W^{(3)} - 2)(1-\eta)^2}{K_W (K_W - \eta)} - 2 \right) \geqslant |\boldsymbol{\lambda}| C^{(1)} + \epsilon,$$

where $N_W^{(1)}(\cdot)$ is as specified in Lemma 13.

Appendix F
Foster-Lyapunov Theorem

Proposition 6: Let $\{\mathbf{x}(t) : t \geqslant 0\}$ be a continuous-time, time-homogeneous and irreducible Markov chain with state space $\mathcal{S}$ and generator matrix Q. Suppose there exist a non-negative function $V : \mathcal{S} \to \mathbb{R}_{>0}$, an $\epsilon' > 0$, a finite set $\mathcal{A}$, and a finite constant B such that $\{\mathbf{x} : V(\mathbf{x}) \leqslant C\}$ is finite for all $C \in \mathbb{R}_{>0}$, and the unit-transition drift $QV(\mathbf{x})$ is upper-bounded as

$$QV(\mathbf{x}) \leqslant -\epsilon' \mathbb{1}\left[\mathbf{x} \in \mathcal{A}^c\right](\mathbf{x}) + B \mathbb{1}\left[\mathbf{x} \in \mathcal{A}\right](\mathbf{x}).$$

Then, the process $\{\mathbf{x}(t) : t \geqslant 0\}$ is positive recurrent. The unit-transition drift $QV(\mathbf{x})$ is given by

$$QV(\mathbf{x}) := \sum_{\mathbf{y} \in \mathcal{S}, \mathbf{y} \neq \mathbf{x}} q(\mathbf{x}, \mathbf{y}) \left(V(\mathbf{y}) - V(\mathbf{x}) \right). \tag{F.70}$$

Now, suppose V', f, and g are non-negative functions on $\mathcal{S}$, and suppose $QV'(\mathbf{x}) \leqslant -f(\mathbf{x}) + g(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{S}$. In addition, suppose $\{\mathbf{x}(t) : t \geqslant 0\}$ is positive-recurrent, so that the means, $\overline{f} = \pi f$ and $\overline{g} = \pi g$ are well-defined. Then $\overline{f} \leqslant \overline{g}$. (In particular, if g is bounded, then $\overline{g}$ is finite, and therefore $\overline{f}$ is finite).

Appendix G
List of Important Symbols

- RF: Rarest-First.
- MS: Mode-Suppression.
- PMS: Probabilistic Mode-Suppression.
- RFwPMS: Rarest-First with Probabilistic-Mode-Suppression.
- P2P: Peer-to-peer.
- RN: Random-Novel.
- LPS: Last Piece Syndrome.
- TMS: Threshold Mode-Suppression.
- FPS: First Piece Syndrome.
- $\mathcal{F}$: Master-file.
- K: Number of pieces in $\mathcal{F}$, i.e., $|\mathcal{F}|$. It is assumed that $K \in \mathbb{N}\setminus\{1\}$.
- W: Denotes a file or the corresponding swarm.
- K_W: $|W|$. It is assumed that $K_W \in \mathbb{N}\setminus\{1\}$.
- $\mathcal{F}_W$: Set of all pieces downloadable by swarm-W.
- λ_W: Arrival-rate of an empty swarm-W peer.
- $\mathcal{W}$: Set of all swarms entering the network.
- $\boldsymbol{\lambda}$: $(\lambda_W : W \in \mathcal{W})$.
- $|\boldsymbol{\lambda}|$: $\sum_{W \in \mathcal{W}} \lambda_W$.
- $\mathcal{W}_W^{(ally\text{-}\uparrow)}$: Receiving-ally-set of swarm-W.
- $\mathcal{W}_W^{(ally\text{-}\downarrow)}$: Donor-ally-set of swarm-W.
- $x_W^{(S)}$: Number of (W, S)-type peers.
- $\mathbf{x}$: State of the network. See (1).
- $|\mathbf{x}|_W$: Number of peers in swarm-W. See (2).
- $|\mathbf{x}|$: Number of peers in the network. See (3).

- L: Number of contact links with each peer (seed included).
- $Y^{(opt)}$: Binary parameter for optimistic-unchoke.
- $\mu^{(\mathrm{tft})}$: Ticking rate of a tit-for-tat link.
- $\mu^{(\mathrm{opt})}$: Ticking rate of a normal peer's optimistic-unchoke link.
- U: Ticking rate of each link of the fixed seed.
- p: In a given tit-for-tat contact, p is a lower-bound on the probability of a peer push-contacting the other peer.
- i: Typically used to denote a file's piece.
- $\pi_W^{(i)}$: Frequency of piece i in swarm-W. See (4).
- $c_W^{(i)}$: Chunk-count of piece i in swarm-W. See (5).
- $\overline{\pi}_W$ and $\underline{\pi}_W$: Maximum and minimum chunk-frequencies in swarm-W (computed over pieces of file-W). See (6).
- $\overline{c}_W$ and $\underline{c}_W$: Maximum and minimum chunk-counts in swarm-W (computed over pieces of file-W). See (7).
- $P_W^{(\mathrm{tot})}$: Total chunk-count in swarm-W. See (8).
- $m_W^{(i)}$: Mismatch of piece i in swarm-W. See (9).
- $\overline{m}_W$: Largest-mismatch in swarm-W. See (10).
- M_W: Total-mismatch in swarm-W. See (11) and (12).
- $d_W^{(i)}$: Swarm-W's complementary chunk-count of piece i. See (13).
- R_W: Set of rare pieces in swarm-W. See (14).
- R_W^c: $W\backslash R_W$, set of non-rare pieces in swarm-W.
- r: Typically used to denote rare piece in a swarm.
- n: Typically used to denote a non-rare piece in a swarm.
- $\underline{r}$: Typically used to denote a piece with the smallest chunk-count in a given set of rare pieces.
- $\underline{R}_{(\widehat{T},W,S)}$: Set of rarest novel pieces transferable from revealed cache-profile $\widehat{T}$ to (W,S)-peer. See (15).
- $\zeta_W^{(n)}$: Non-rares sharing factor for swarm-W and non-piece $n \in R_W^c$. See (16).
- α_W and β_W. Tuning parameters in the definition of $\zeta_W^{(n)}$.
- $A^{(\mathrm{trf})}(\mathbf{x}, \widehat{T}, W, S)$: Transferable-set. See Algorithm 1.
- $\mathcal{S}$: State space of $\{\mathbf{x}(t) : t \geqslant 0\}$. See (17).
- $q_W^{(\varnothing,+)}$: λ_W. See (18).
- $\Gamma_W^{(i)}$, $\underline{\Gamma}_W^{(i)}$, and $\overline{\Gamma}_W^{(i)}$: Exact, lower, and upper estimates for the aggregate rate at which some donating ally-peer of swarm-W push-contacts in the network. See (19), (20), and (22).
- ξ, κ_a: See (21).
- Υ: An upper-bound on the ratio $\overline{\Gamma}_W^{(i)}/\underline{\Gamma}_W^{(i)}$ for all $W \in \mathcal{W}$ and $i \in W$. See (23).
- $q_W^{(S,i+)}$: Aggregate rate at which a (W,S)-peer downloads piece i without departing the system. See (24)-(26).
- $\underline{q}_W^{(S,n+)}$: Lower estimate for aggregate rate of (W,S)-peer downloading non-rare piece $n \in R_W^c$. See (26).
- $\mathcal{S}_W^{(1)}$: $\{\mathbf{x} : R_W(\mathbf{x}) \subsetneq W\}$.
- $\mathcal{S}_W^{(2)}$: $\{\mathbf{x} : R_W(\mathbf{x}) = W\}$.
- $q_W^{(S,i-)}$: Aggregate rate at which a (W,S)-peer downloads piece i and departs the system. See (27).
- $a_W^{(i)}$: See (28).
- $\mathcal{P}^{(stab)}$: See (29).
- V: $\sum_{W \in \mathcal{W}} V_W(\mathbf{x})$. See (30).
- V_W: See (31).
- $\eta, C_W^{(1)}, C_W^{(2)}, C_W^{(3)}$: Suitable constants in the definition of V_W. See (31).
- $\mathcal{R}_W^{(1)}$: $\{\mathbf{x} : M_W - 2(1-\eta) \geqslant \eta \overline{c}_W\}$.
- $\mathcal{R}_W^{(2)}$: $\mathcal{S}\backslash\mathcal{R}_W^{(1)}$.
- $I_W^{(2)}$: $C_W^{(2)} K_W \mathbb{1}\left[C_W^{(3)} + 2K_W \geqslant P_W^{(\mathrm{tot})}\right]$.
- $\underline{\psi}_W$:
$$\left(2(1-\eta)^2 - 2(1-\eta)\left(M_W - \eta\overline{c}_W\right)\right)\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] + 4K_W^2\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right].$$
- $\overline{\psi}_W$:
$$\left(K_W^2 + 2K_W\left(M_W - \eta\overline{c}_W\right)\right)\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(1)}\right] + 4K_W^2\mathbb{1}\left[\mathbf{x} \in \mathcal{R}_W^{(2)}\right].$$
- $\Psi_W^{(i-)}$: Upper bound on the potential change due to download of piece i in swarm-W accompanied by departure of a peer. See (33).
- $I_W^{(1)}$: $C_W^{(2)}\mathbb{1}\left[C_W^{(3)} - 2 \geqslant P_W^{(\mathrm{tot})}\right]$.
- $\Psi_W^{(i+)}$: Upper bound on the potential change due to download of piece i in swarm-W without any departure of a peer. See (34).
- QV: See (35) and (36).
- $C^{(1)}$: $\max_{W \in \mathcal{W}} K_W C_W^{(1)}$. See (37).
- $\gamma_W^{(i)}$: $\sum_{S: W\backslash S = \{i\}} \frac{x_W^{(S)}}{|\mathbf{x}|_W}$.
- $\Psi_W^{(i)}$: See (42).
- $\widetilde{QV}_W$: Used in upper estimate of QV. See (43) and (44).
- D_W, $D_W^{(1)}$, and $D_W^{(2)}$. See (45).
- $\overline{T}_W$. Steady-state expected sojourn time of swarm-W peers (in the stability region).
- $\mathcal{L}_1$: $\mathcal{P}^{(stab)} \cap \{\lambda_W < 0.5K_W^{-2}LU\}$.
- $\mathcal{L}_2$: $\mathcal{P}^{(stab)} \cap \{\lambda_W \geqslant 0.5K_W^{-2}LU\}$.
- V_1: See (46).
- V_2: See (50).
- δ_W: Small number in $(0,1)$ used in Appendix E.
- $\mathcal{R}_W^{(kj)}$: Used in Appendix E.
- $N_W^{(1)}(\cdot)$: Used in Lemma 13.
- $\mathcal{A}_W$, B_W, N_W: Used in Lemma 14.
- $\mathcal{H}$, $\mathcal{S}_{\mathcal{H}}$, $\mathcal{A}_{\mathcal{H}}$, $\mathcal{A}$, B, $B_{\mathcal{W}}$: Used in Lemma 15.

Acknowledgment

This work was funded in part by NSF via EPCN1608361 (V. Subramanian) and EARS1516075, ECCS2038416, CCF2008130, CNS1955777, and CMMI2240981 (N. Khan and V. Subramanian), START grant from Michigan

Engineering (N. Khan), and the Rackham Predoctoral Fellowship from the University of Michigan (M. Moharrami). The bulk of the work was carried out when M. Moharrami was at the University of Michigan.

REFERENCES

[1] I. Norros, H. Reittu, and T. Eirola, "On the stability of two-chunk file-sharing systems," *Queueing Systems*, vol. 67, no. 3, pp. 183–206, March 2011.

[2] B. Cohen, "Incentives build robustness in bittorrent," *Workshop on Economics of Peer-to-Peer systems*, vol. 6, pp. 68–72, June 2003.

[3] Sandvine, "The global internet phenomena report covid-19 spotlight," 2020. [Online]. Available: https://www.sandvine.com/phenomena

[4] L. Massoulie and M. Vojnovic, "Coupon replication systems," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 603–616, June 2008.

[5] B. Hajek and J. Zhu, "The missing piece syndrome in peer-to-peer communication," *Stochastic Systems*, vol. 1, no. 2, pp. 246–273, 2011.

[6] J. Zhu and B. Hajek, "Stability of a peer-to-peer communication system," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4693–4713, July 2012.

[7] X. Zhou, S. Ioannidis, and L. Massoulie, "On the stability and optimality of universal swarms," in *ACM SIGMETRICS*, ser. SIGMETRICS '11. ACM, June 2011, pp. 341–352.

[8] J. Zhu, I. Stratis, N. Hegde, and L. Massoulié, "Stable and scalable universal swarms," *Distributed Computing*, vol. 28, no. 6, pp. 391–406, December 2015.

[9] B. Oğuz, V. Anantharam, and I. Norros, "Stable distributed p2p protocols based on random peer sampling," *IEEE/ACM Transactions on Networking*, vol. 23, no. 5, pp. 1444–1456, October 2015.

[10] O. Bilgen and A. B. Wagner, "A new stable peer-to-peer protocol with non-persistent peers," in *IEEE INFOCOM*, May 2017, pp. 1–9.

[11] V. Reddyvari, P. Parag, and S. Shakkottai, "Mode-suppression: A simple and provably stable chunk-sharing algorithm for p2p networks," in *IEEE INFOCOM*, April 2018, pp. 2573–2581.

[12] V. Reddyvari, S. C. Bobbili, P. Parag, and S. Shakkottai, "Mode-suppression: A simple, stable and scalable chunk-sharing algorithm for p2p networks," *IEEE/ACM Transactions on Networking*, pp. 1–12, 2021.

[13] D. Qiu and R. Srikant, "Modeling and performance analysis of bittorrent-like peer-to-peer networks," *SIGCOMM Computer Communication Review*, vol. 34, no. 4, pp. 367–378, August 2004.

[14] N. Khan, M. Moharrami, and V. Subramanian, "Stable and efficient piece-selection in multiple swarm bittorrent-like peer-to-peer networks," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 1153–1162.

[15] D. S. Menasche, A. A. d. A. Rocha, B. Li, D. Towsley, and A. Venkataramani, "Content availability and bundling in swarming systems," *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 580–593, April 2013.

[16] E. de Souza e Silva, R. M. Leão, D. S. Menasché, and D. Towsley, "On the scalability of p2p swarming systems," *Computer Networks*, vol. 151, pp. 93–113, Mar 2019. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2019.01.006

[17] D. X. Mendes, E. de Souza e Silva, D. Menasché, R. Leão, and D. Towsley, "An experimental reality check on the scaling laws of swarming systems," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.

[18] B. Hajek, *Random Processes for Engineers*. Cambridge University Press, 2015.

[19] R. Srikant and L. Ying, *Communication Networks: An Optimization, Control, and Stochastic Networks Perspective*. Cambridge University Press, 2014.

[20] A. Legout, G. Urvoy-Keller, and P. Michiardi, "Rarest first and choke algorithms are enough," in *ACM SIGCOMM*, ser. IMC '06. ACM, 2006, pp. 203–216.

Nouman Khan (Student Member, IEEE) is a Ph.D candidate in the Electrical Engineering and Computer Science (EECS) Department at the University of Michigan, Ann Arbor, MI, USA. He received the B.S. degree in electronic engineering from the GIK Institute of Engineering Sciences and Technology, Topi, KPK, Pakistan, in 2014, and the M.S. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA in 2019. His research interests include design and analysis of control and inference algorithms for stochastic systems and networks.

Mehrdad Moharrami is an Assistant Professor in the Computer Science (CS) department at the University of Iowa, Iowa City, IA, USA. He received B.S. degrees in mathematics and electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2014. He obtained his M.S. degrees in mathematics and electrical engineering, along with a Ph.D. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2017, 2019, and 2020, respectively. From 2020 through 2023, he was a TRIPODS Postdoctoral Research Fellow at the University of Illinois at Urbana Champaign, IL, USA. His research interests include Markov decision processes, reinforcement learning, and random graph models for economics, learning, and computation.

Vijay G. Subramanian (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1999. He was a researcher with Motorola Inc., and also with Hamilton Institute, Maynooth, Ireland, for a few years following which he was a Research Assistant Professor in the EECS Department at the Northwestern University, Evanston, IL, USA. In 2014, he joined the University of Michigan, Ann Arbor, MI, USA, where he is currently an Associate Professor with the EECS Department. His research interests are in stochastic analysis, random graphs, game theory, and mechanism and information design with applications to social, economic and technological networks.