Analyzing the Resilience-Complexity Tradeoff of Network Coding in Dynamic P2P Networks

Di Niu, Baochun Li, *Senior Member, IEEE*

Abstract—Most current-generation P2P content distribution protocols use fine-granularity blocks to distribute content to all the peers in a decentralized fashion. Such protocols often suffer from a significant degree of imbalance in block distributions, especially when the users are highly dynamic. As certain blocks become rare or even unavailable, content availability and download efficiency are adversely affected. Randomized network coding may improve block diversity and availability in P2P networks, as coded blocks are equally innovative and useful to peers. However, the computational complexity of network coding mandates that, in reality, network coding needs to be performed within segments, each containing a subset of blocks. In this paper, we quantitatively evaluate how network coding may improve content availability, block diversity and download performance in the presence of churn, as the number of blocks in each segment for coding varies. Based on stochastic models and a differential equation approach, we explore the fundamental tradeoff between the resilience gain of network coding to peer dynamics and its inherent coding complexity. We conclude that a small number of blocks in each segment is sufficient to realize the major benefits of network coding, with acceptable coding cost.

Index Terms—Peer-to-peer content distribution, generation-based network coding, peer dynamics, content availability, resilience.

1 Introduction

Peer-to-peer (P2P) communication paradigm has become the *de facto* standard in current-generation content distribution protocols. The basic idea of P2P content distribution is to break the content of interest into fine-granularity blocks, and distribute these blocks in a decentralized manner by letting peers exchange the blocks on an overlay network.

In reality, however, such protocols often suffer from the *rare block problem*. There might exist a significant degree of variation with respect to the availability of different blocks, adversely affecting content availability and download efficiency. Such a phenomenon is exacerbated by the presence of peer dynamics, a characteristic inherent of end hosts in the Internet. It is not only hard to maintain a reasonable degree of block diversity in large-scale dynamic networks to guarantee download efficiency, but unexpected peer departures may also cause rare blocks to disappear from the network, undermining the integrity of the content.

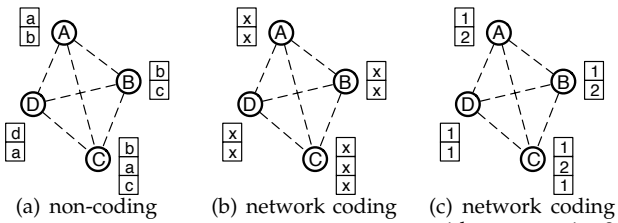
Randomized network coding [1], first proposed in information theory, has been introduced into P2P content distribution to improve system performance [2]. With randomized network coding, each peer linearly encodes all the blocks it possesses using random coefficients and transmits the encoded blocks to its downstream peers. This simplifies protocol design by avoiding block scheduling. In this paper, we focus on another more important benefit of network coding, which is its resilience to peer dynamics. Intuitively, the problem of locating rare blocks in a non-coding protocol does not exist in a network coding based protocol, since all coded blocks are equally innovative and useful to any peers with high probability.

Nevertheless, network coding may not realize its benefits without introducing significant computational complexity at users. In order to reduce coding cost, Chou et al. [3] has proposed the concept of generation-based or segment-based network coding, which divides the content of interest into segments, each containing a prescribed (and arguably small) number of blocks, and performs coding only across the blocks within the same segment. Though coding within segments reduces computational cost, it is yet to be understood how segmentation affects the resilience of network coding to peer dynamics — the major advantage of using network coding in the first place.

This paper aims to theoretically analyze the resilience gain of network coding in highly dynamic P2P networks when the number of blocks in each segment for coding (referred to as the segment size) varies. Let us consider two extremes of P2P protocol design, as illustrated in Fig. 1(a) and Fig. 1(b). The first one does not use coding at all, whereas the second one applies network coding across all existing blocks. Intuitively, it is much more difficult for a non-coding protocol to ensure a uniform distribution of all blocks in dynamic networks, e.g., in Fig. 1(a), if peer D leaves, block d is missing. In contrast, when randomized network coding is applied (Fig. 1(b)), all coded blocks x's are equally useful with high probability. Even if both peer C and D leave, the content is still decodable, as peer A and B collectively possess 4 coded blocks. The use of network coding also leads to a higher download efficiency, since peers will not be delayed in locating rare blocks.

If we consider segment-based network coding, we

D. Niu and B. Li are affiliated with the Department of Electrical and Computer Engineering, University of Toronto. Their email addresses are {dniu, bli}@eecg.toronto.edu.



(a) non-coding (b) network coding (c) network coding with segment size 2

Fig. 1. A snapshot of a network of 4 peers, A, B, C, and D, distributing 4 original blocks, a, b, c and d.

may observe that Fig. 1(a) corresponds to a segment size of one, with as many segments as blocks, while Fig. 1(b) corresponds to the case of grouping all blocks into the same segment. If we vary the segment size when network coding is used, we are fundamentally moving from one extreme to another, making our choice in the challenge of *resilience-complexity tradeoff*. For example, Fig. 1(c) shows network coding performed with a segment size of 2, so that there are two types of coded blocks "1" and "2" in the network. We ask the question — what is an appropriate segment size to choose so that it is sufficient for network coding to yield its resilience advantage, yet with acceptable computational cost?

We consider dynamic network models, where peers have random lifetimes and coded blocks are propagated by gossip-like protocols. We quantify the block availability of different segments and content persistence using appropriate metrics, as the segment size varies. The major technique used in the analysis is approximating density dependent jump Markov processes with differential equations. Extensive simulations corroborate the analysis and shed light on the problem in a wide range of settings. We find that the variation in the availability of different segments decreases as the segment size increases, and the download performance is closely related to such variation. Furthermore, there is a sweet spot of segment size, beyond which network coding can hardly bring further benefits in terms of resilience.

The remainder of the paper is organized as follows. In Sec. 2, we outline the system models for analysis. In Sec. 3, we characterize the asymptotic system behavior with differential equations, based on which the steady-state segment availability is analyzed in Sec. 4. In Sec. 5, we study content persistence in the absence of original content sources. Sec. 6 presents simulation results to corroborate and extend our theoretical findings. We review related work in Sec. 7 and conclude the paper in Sec. 8.

2 Background and System Models

We consider a BitTorrent-like file distribution system, in which a large file of size F bytes (usually in the order of hundreds of Megabytes or several Gigabytes) is broadcast to every participating peer. The content is broken into M blocks, each of size $k = F/M$ bytes. Participating peers are organized into a randomized overlay mesh exchanging these blocks. Each peer maintains TCP connections with a number of other peers (e.g., around 40 in BitTorrent), which are called its *neighbors*. The neighbors of each peer are assumed to be uniform samples of the entire network and can be changing over time. Data transmission can only occur between a peer and its neighbors. Normally, a peer only uploads to a small number (e.g., less than 5) of its neighbors at the same time, which are called its *downstream peers*. The goal is to expedite the transfer of individual blocks from a peer to its downstream peers, given its limited upload bandwidth.

A peer is a *seed* if it has obtained a complete copy of the content, otherwise it remains as a *downloader*. Let N and N_s denote the numbers of online downloaders and online seeds, respectively. We further assume that each peer has an average upload capacity of μ bytes per unit time, or $\tilde{\mu} = \mu/k$ blocks per unit time, and a separate downlink of sufficiently large capacity.

When segment-based network coding [3] is applied, all the M data blocks are grouped into G segments, each containing m blocks, referred to as the *segment size*. A random linear code (RLC) is applied to each of these segments. Assume segment i has original blocks $\mathbf{B^{(i)}} = [B_1^i, B_2^i, \ldots, B_m^i]$, then a coded block b from segment i is a linear combination of $[B_1^i, B_2^i, \ldots, B_m^i]$ in the Galois field $GF(2^q)$. Coding operations are not limited to the source: if a peer (including the source) has $l \leq m$ coded blocks $[b_1^i, b_2^i, \ldots, b_l^i]$ from segment i, when serving another peer p, it independently and randomly chooses a set of coding coefficients $[c_1^p, c_2^p, \ldots, c_l^p]$ in $GF(2^q)$, and encodes all the blocks it possesses from segment i, and produces a coded block x of k bytes: $x = \sum_{j=1}^{l} c_j^p \cdot b_j^i$.

The coding coefficients used to encode *original blocks* to x are embedded in the header of each coded block. As soon as a peer has received a total of m linearly independent coded blocks $\mathbf{x} = [x_1^i, x_2^i, \ldots, x_m^i]$ from segment i, it will be able to decode segment i. To decode segment i, we first need to compute the inverse of the $m \times m$ coefficient matrix $\mathbf{A_i}$ using Gaussian elimination, which requires $O(m^3)$ operations in total, or $O(m^2)$ operations per original block. To obtain the original m blocks $\mathbf{B^{(i)}}$, it then needs to multiply $\mathbf{A_i}^{-1}$ and $\mathbf{x}$, which takes $m^2 \cdot k$ multiplications of two bytes in $GF(2^q)$ and requires $O(m^2 k)$ operations in total, or $O(mk)$ operations per original block. It turns out the latter cost dominates the overall decoding time, because the cost of the latter phase also depends on the block size k, which is usually in the order of kilobytes. Apparently, the overall decoding complexity increases as the segment size m increases.

We introduce a framework that can model both non-coding and network-coded distribution systems by tuning the segment size. We consider *random downstream peer selection* and *random useful segment encoding*, which are typical in network-coded content distribution systems such as Avalanche [2].

Let $P(A)$ denote the set of segments for which peer A has received at least one block, and $C(A)$ denote the set of segments that peer A has completely received.

Assume the time to upload a block is exponentially distributed with mean $1/\tilde{\mu}$. At rate $\tilde{\mu}$, each peer A serves a downstream peer B randomly chosen from all A's neighbors that are not seeds. Peer A then randomly chooses a segment in $P(A)\backslash C(B)$ and encodes all the blocks it possesses for that segment, and transmits the encoded block to peer B. When $m = 1$, this model morphs into a non-coding protocol where each peer transmits random useful blocks that the downstream peer needs. By varying the segment size m from 1 to a large value, we are essentially moving from a non-coding protocol to a network-coded protocol with variable computational cost.

Our theoretical analysis accommodates minor variations of the above protocol. Since the exact characterization of network coding in gossiping systems is extremely difficult if not impossible as shown in [4], [5], we trade the accuracy of analysis for the cleanness of the results. Specifically, we apply approximated analysis onto such protocols based on several heuristics, which proves to be an accurate approximation of the real system, as verified by the simulations.

We consider two models for peer dynamics:

Replacement model. Assume there are always N downloaders and N_s seeds simultaneously online. Each downloader has an i.i.d. lifetime $\boldsymbol{L}$. Every departed downloader is replaced by a new empty peer, as shown in Fig. 2. The N_s seeds are always online.

Poisson arrival model. Downloaders enter the system in a Poisson process with rate λ. Each downloader has an i.i.d. lifetime $\boldsymbol{L}$. There are always N_s seeds online.

In both models, $\boldsymbol{L}$ follows a general distribution $F_L(x)$ with mean $\overline{L}$ and variance σ^2. As a starting point, the replacement model allows us to focus on the effect of peer churn on data availability and download efficiency, decoupling the impact of the change in online peer number. Even if there are always a large number of peers simultaneously online, the dynamic nature of these peers can induce a significant variation in the availability of different segments, which causes the *rare block problem*. Such a model for churn has previously been used in [6] to study the reliability of unstructured P2P networks. Poisson arrival model, on the other hand, represents a more common peer joining process in the real world. We show both models yield similar results with regard to block-level segment availability in asymptotic systems.

We evaluate the resilience of network coding to peer churn from multiple aspects. Let random variable $\boldsymbol{I}$ denote the number of (coded) blocks each segment has in all the downloaders in the steady state. Our first goal is to determine the ***block availability distribution*** of different segments, i.e., $\overline{p}_i := \Pr\{\boldsymbol{I} = i\}$ for $\boldsymbol{I}$ in steady-state networks. We can show that $\boldsymbol{I}$ follows the *negative binomial distribution* under certain conditions.

To quantitatively evaluate the *rare block problem*, we define ***block variation*** as

$$\gamma_I^2 = \frac{\mathbf{Var}(\boldsymbol{I})}{\mathbf{E}^2\{\boldsymbol{I}\}} = \frac{\sum_{i=0}^{\infty} i^2 \overline{p}_i - (\sum_{i=0}^{\infty} i \cdot \overline{p}_i)^2}{(\sum_{i=0}^{\infty} i \cdot \overline{p}_i)^2}, \qquad (1)$$

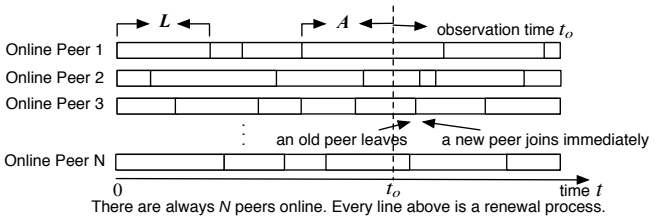


Fig. 2. The replacement model for peer dynamics. $\boldsymbol{L}$ denotes a peer's lifetime. $\boldsymbol{A}$ denotes a peer's age at a given observation time t_o. Every departed downloader is replaced by a new empty peer.

which is inspired by a typical fairness measure in resource allocation [7]. γ_I^2 always lies between 0 and ∞, and is 0 only if all the segments have the same number of (coded) blocks in the network.

As a major theoretical contribution, we show that, under both the replacement and Poisson arrival models, γ_I^2 is inversely proportional to the segment size m in asymptotic networks. We find through experiments that the expected time needed for a long-lived peer to download the entire content, or the ***download time***, is closely related to block variation. We also evaluate the block diversity in the steady-state network using the Shannon entropy $H(\boldsymbol{I})$ of $\boldsymbol{I}$, which proves to be a logarithmic function of m.

To study the transient behavior of the system, we assume all the seeds depart in the steady-state and see how long the content can be kept complete by the unreliable downloaders in a distributed fashion. Under such a scenario without seeds, we also study the content loss and the evolution of block availability distribution over time.

3 System Characterization via ODEs

In this section, we derive a set of ordinary differential equations (ODEs) to asymptotically approximate the block availability evolution under the replacement model and Poisson arrival model. The solutions to these ODEs are used to study the steady-state block distribution and variation in Sec. 4, and the transient behavior of the system in Sec. 5.

For a segment s, we use $\deg(s)$ (degree of s) to denote the number of (coded) blocks s has in all the downloaders. To characterize the distribution of $\boldsymbol{I}$, we introduce the following notations, concerning the block statistics in downloaders:

- $\triangleright$ $n_i(t)$: the number of segments which have a degree of i. $\sum_{i=0}^{\infty} n_i(t) = G$.
- $\triangleright$ $p_i(t)$: the fraction of segments which have a degree of i. $p_i(t) = n_i(t)/G$.
- $\triangleright$ $r_i(t)$: the number of segments with a degree of at least i. $n_i(t) = r_i(t) - r_{i+1}(t)$.
- $\triangleright$ $s_i(t)$: the fraction of segments which have a degree of at least i. $s_i(t) = r_i(t)/G$.

These notations cover both non-coding and coding cases as m varies from 1 to M. Apparently, as the

segment number $G \to \infty$, $p_i(t)$ will approach the probability mass function (PMF) of $\boldsymbol{I}$. For simplicity, we may omit t in the following context. We can also get the simple facts that at time t, the total number of blocks in downloaders is $Y(t) = \sum_{i=1}^{\infty} r_i(t)$, and the average degree of a segment is $\sum_{i=1}^{\infty} s_i(t) = Y(t)/G$.

The system state could be represented by the vector $\boldsymbol{R}(t) := (r_0(t), r_1(t), \ldots)$, or $\boldsymbol{S}(t) := (s_0(t), s_1(t), \ldots)$ or $\boldsymbol{P}(t) := (p_0(t), p_1(t), \ldots)$. Clearly, $\boldsymbol{S}(t+h)$, for any $h > 0$, not only depends on $\boldsymbol{S}(t)$, but also depends on the particular subset of blocks possessed by each peer at time t. Thus, $\boldsymbol{S}(t)$ is a complicated process which has a large state space and is extremely hard to characterize. However, under certain assumptions, $\boldsymbol{S}(t)$ will become a Markov process, whose limiting behavior converges to a deterministic process that can be represented by a set of ODEs. It is through such a method that the block availability evolution can be characterized. A similar approach has been used by Massoulie et al. [8] to analyze a coupon collection system that models P2P content distribution without coding. The validity of the assumptions made is verified by the simulations in Sec. 6.

Our characterization for the replacement model follows the following path. First, we consider the "peer process" and show the total number of blocks in the network $\boldsymbol{Y}(t)$ tends to concentrate around a certain value after the network has evolved for sufficiently long time. We then focus on the "block exchange process" and make certain approximations so that $\boldsymbol{S}(t)$ forms a density dependent jump Markov process, which can be asymptotically characterized by a set of differential equations with arbitrarily small error.

We first consider the "peer process". At a given observation time t_o, we use $\boldsymbol{A}$ to denote a peer's *age*, which can be given by the following lemma. (Refer to [9] for the proof.)

Lemma 1: Let $\boldsymbol{L}$ denote a peer's lifetime with mean $\overline{L}$ and variance σ^2. Under the replacement model, if N is large, as $t_o \to \infty$, we have $\overline{A} := \mathbf{E}[\boldsymbol{A}] = (\overline{L}^2 + \sigma^2)/2\overline{L}$.

We further establish in the following lemma that $\boldsymbol{Y}(t)/(N+N_s)$ almost always remains constant as $t \to \infty$, the proof of which is given in Appendix A.

Lemma 2: Let $\boldsymbol{Y}(t)$ denote the total number of blocks in downloaders from all the segments at a given time t. Under the replacement model, as $N \to \infty$, $M \to \infty$, $N = \alpha M$, $N_s = \alpha_s M$, where α, α_s are finite constants,

$$\lim_{N\to\infty} \boldsymbol{Y}(t)/(N+N_s) = \tilde{\mu}\overline{A} \quad \text{with probability 1}, \tag{2}$$

for t sufficiently large.

From Lemma 2, we can see that if G is large enough, the average number of blocks each segment has in all the downloaders is

$$\sum_{i=1}^{\infty} s_i(t) = \boldsymbol{Y}(t)/G = (N+N_s)\tilde{\mu}\overline{A}/G = (\alpha+\alpha_s)m\tilde{\mu}\overline{A}.$$

Now we consider the "data exchange process" and characterize the evolution of $\boldsymbol{S}(t)$ as $t \to \infty$. We make appropriate linear approximations on upload behavior and content loss patterns that convert $\boldsymbol{S}(t)$ into a tractable Markov process. From a network perspective, we assume that the upload of a block by a seed has the same effect as choosing from all the segments a random segment s and increasing deg(s) by 1, whereas the upload of a block by a downloader has the same effect as choosing from all the segments a random segment s with probability $\deg(s)/\sum_s \deg(s)$ and increasing $\deg(s)$ by 1. Further, the number of blocks a segment loses due to peer departures is assumed to be linear in the number of blocks it has in the downloaders. These approximations are formally described in Appendix B and validated by simulations in Sec. 6.

We can thus show that **as $t \to \infty$, $M \to \infty$, $N \to \infty$, $N_s \to \infty$, $N/M \to \alpha$, $N_s/M \to \alpha_s$, while m remains finite, $\boldsymbol{S}(t)$ in replacement model converges to the following system of ODEs with an arbitrarily small error:**

$$\begin{cases} \overline{A}\frac{\mathrm{d}s_i}{\mathrm{d}t} = \alpha_s m\tilde{\mu}\overline{A}p_{i-1} + \frac{\alpha}{\alpha+\alpha_s}(i-1)p_{i-1} - ip_i, \ \forall i \geq 1, \\ s_0 \quad = 1. \end{cases} \tag{3}$$

The derivation of (3) can be found in Appendix B.

Similarly, a system of ODEs can be established for the Poisson arrival model. **As $t \to \infty$, $M \to \infty$, $\lambda \to \infty$, $N_s \to \infty$, $\lambda\overline{L}/M \to \alpha'$, $N_s/M \to \alpha_s$, while m remains finite, $\boldsymbol{S}(t)$ under the Poisson arrival model converges to the following system of ODEs:**

$$\begin{cases} \overline{A}\frac{\mathrm{d}s_i}{\mathrm{d}t} = \alpha_s m\tilde{\mu}\overline{A}p_{i-1} + \frac{\alpha'}{\alpha'+\alpha_s}(i-1)p_{i-1} - ip_i, \ \forall i \geq 1, \\ s_0 \quad = 1. \end{cases} \tag{4}$$

The derivation of (4) can be found in Appendix C.

In addition, we have also considered the effect of the block size k on the system evolution. Intuitively speaking, if the block size k is large, when a peer departs while still uploading a block, more data will be lost due to the lack of granularity in the upload process. However, it turns out that the block size will almost not affect the differential equations (3) and (4) at all, as long as $\mu\overline{A} \gg k$, which is satisfied in usual cases ($\mu\overline{A} \gg k$ does not conflict with the finiteness of $\tilde{\mu}\overline{A}$).

4 Segment Availability in Steady States

Based on the ODEs derived above, in this section we determine the *block availability distribution*, *block variation* and *block distribution entropy* in steady-state networks for both the replacement and Poisson arrival models, and discuss their implications.

Let us first consider the block availability distribution for the replacement model in its steady state. Denote the steady-state solutions to the ODEs (3) by $\overline{p}_0, \overline{p}_1, \ldots, \overline{p}_G$. By setting $\frac{\mathrm{d}s_i}{\mathrm{d}t} = 0$ in (3), we can obtain the steady-state block distribution as a function of $\overline{p}_0$:

$$\overline{p}_i = \overline{p}_0 \cdot \prod_{j=1}^{i}(\beta + \frac{B-\beta}{j}), \quad \text{for } i \geq 1, \tag{5}$$

with $\sum_{i=0}^{G} \overline{p}_i = 1$, where $B = \alpha_s m \tilde{\mu} \overline{A}$, $\beta = \frac{\alpha}{\alpha+\alpha_s}$. It is in general very difficult to obtain closed form solutions for $\overline{p}_i$. However, under certain mild conditions appropriate for engineering purposes, we can convert (5) into its closed form as follows:

Theorem 3: **(Steady-State Block Distribution for Replacement Model)** Let $B = \alpha_s m \tilde{\mu} \overline{A}$ and $\beta = \frac{\alpha}{\alpha+\alpha_s}$. If $\beta > 0$ is rational and $\frac{B}{\beta} \in \{1, 2, \ldots\}$, then $\boldsymbol{I}$ follows the negative binomial distribution, i.e., for $i = 0, 1, \ldots$,

$$\Pr\{\boldsymbol{I} = i\} = \overline{p}_i = \binom{i + \frac{B}{\beta} - 1}{i} \beta^i (1-\beta)^{\frac{B}{\beta}}. \quad (6)$$

Interested readers are referred to Appendix D for a complete proof. In practice, β is rational for sure, since both the seed number and the downloader number in the network are integers. $\overline{p}_i$ for all valid β and B can thus be approximated by using the nearest β and B satisfying the above conditions.

The steady-state block distribution for the Poisson arrival model can be derived in a similar way. Denote the steady-state solutions to (4) by $\overline{p}'_0, \overline{p}'_1, \ldots, \overline{p}'_G$. We have the following theorem:

Theorem 4: **(Steady-State Block Distribution for Poisson Arrival Model)** Let $B = \alpha_s m \tilde{\mu} \overline{A}$ and $\beta' = \frac{\alpha'}{\alpha'+\alpha_s}$. If $\beta' > 0$ is rational and $\frac{B}{\beta'} \in \{1, 2, \ldots\}$, then in the steady state, for $i = 0, 1, \ldots$,

$$\Pr\{\boldsymbol{I} = i\} = \overline{p}'_i = \binom{i + \frac{B}{\beta'} - 1}{i} \beta'^i (1-\beta')^{\frac{B}{\beta'}}. \quad (7)$$

For the replacement model, we plot the CDF of $\boldsymbol{I}$ in Fig. 3 according to Theorem 3, with $N = 6000$, $N_s = 100$, $F = 768$ MB, $k = 256$ KB, $\overline{A} = 5$, and the average upload rate $\tilde{\mu} = \mu/k = 3.7$. In Sec. 6, we also plot the CDF of $\boldsymbol{I}$ from simulation in Fig. 7 under the same parameters. The match between Fig. 7 and Fig. 3 verifies the validity of the approaches in our theoretical analysis.

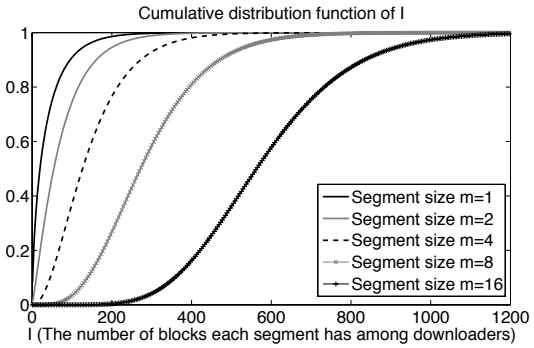


Fig. 3. Steady-state block distribution (CDF of $\boldsymbol{I}$).

One implication of Fig. 3 is that as m increases, there will be fewer rare blocks in the network, and the risk that the content becomes incomplete upon seed departures will be lower. For example, for a non-coding scheme ($m = 1$), 20% of all the segments do not have any blocks in downloaders in the steady state. If all the seeds leave, these segments will become undecodable immediately. As the segment size m increases, we observe an increasing trend in the mean as well as the minimum number of blocks each segment has in downloaders. When segment size $m = 16$, the fraction of segments with less than 200 blocks in downloaders is close to 0. This means even if seeds leave altogether, the content will not become incomplete immediately. In addition, the variation in block distribution is subdued as m increases, which we will demonstrate subsequently.

Now let us quantitatively characterize the block variation γ_I^2 defined in Sec. 2. For the replacement model, from (6) we have $\gamma_I^2 = \sigma_I^2/\mu_I^2 = 1/\alpha_s m \tilde{\mu} \overline{A} = F/N_s m \mu \overline{A}$. For the Poisson arrival model, the same result is obtained from (7). Thus, we have the following theorem:

Theorem 5: **(Steady-State Block Variation)** Under both the replacement and Poisson arrival models,

$$\gamma_I^2 := \sigma_I^2/\mu_I^2 = F/N_s m \mu \overline{A}, \quad (8)$$

where $\overline{A}$ is the average age of a downloader given by Lemma 1. All the parameters are assumed to satisfy the conditions set in Theorem 3 or Theorem 4.

Theorem 5 shows that as more blocks in each segment are used for coding, the variation of the availability of different segments decreases and the rare block problem is mitigated. Furthermore, the block distribution variation is inversely proportional to the segment size m. As a result, there exists a sweet spot in the curve of resilience-complexity tradeoff, at which network coding suffices to yield its major benefit with acceptable coding cost. The block diversity can also be evaluated by the Shannon entropy $H(\boldsymbol{I})$ of $\boldsymbol{I}$. In fact, in Appendix E, we show that $H(\boldsymbol{I}) = \frac{1}{2} \ln m + o(m)$. The logarithm (concave) trend of the entropy of block availability distribution confirms that there can be a sweet spot in the resilience offered by network coding as m varies. We discuss the choice of this sweet spot in Sec. 6.

It is quite counter-intuitive that the block size k does not affect γ_I^2 or $H(\boldsymbol{I})$ at all. One might expect that with a larger k, the content is broken into fewer blocks, and thus the block distribution variation may decrease. However, with a larger k, blocks will be disseminated throughout the network at a lower rate $\tilde{\mu} = \mu/k$, given limited upload bandwidth at each peer. Both effects counteract, resulting in the same γ_I^2 or $H(\boldsymbol{I})$.

The content size F and the number of online seeds N_s are also critical parameters that affect block variation. A lack of seeds will increase block variation, not because seeds hold more blocks than downloaders, but because the upload behavior of seeds are fundamentally different from that of downloaders. As we have pointed out through the analysis in Appendix B, seeds can choose both prevalent and rare segments equally likely when uploading, whereas downloaders tend to choose the segments that are already prevalent in the network. Also note that the number of online downloaders N does not affect the block variation.

5 Transient Behavior without Seeds

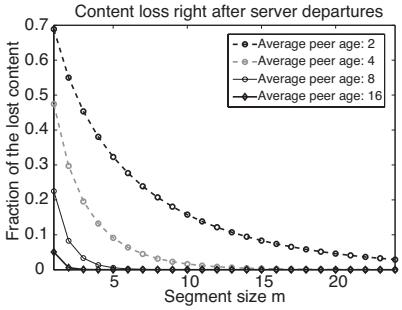


Fig. 4. The fraction of content loss immediately upon seeds departure.

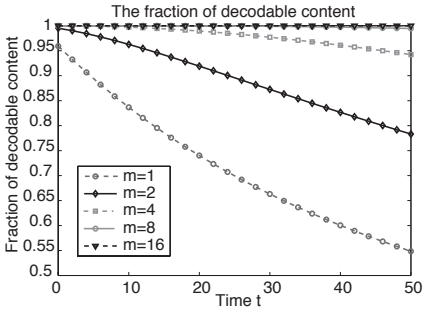


Fig. 5. Decodable content over time after seeds departure.

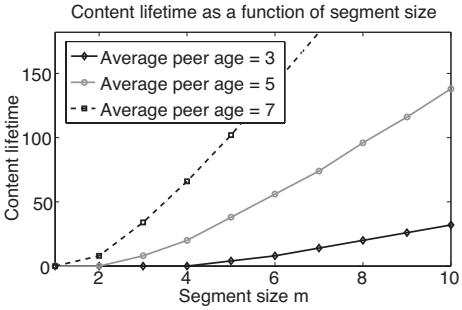


Fig. 6. Content lifetime after seeds departure as m and $\overline{A}$ vary.

In a dynamic P2P network, the original seeds that provide the content may leave the session. In the absence of seeds, if a certain segment has less than m blocks in the network, the content will become incomplete henceforth. The hope is that the content can be kept complete merely by downloaders for a sufficiently long time, before certain downloaders become seeds. Therefore, it is important to analyze how network coding can enhance content persistence in a dynamic network, when seeds are absent.

We only consider the replacement model in this section. We assume the seeds leave altogether in the steady state and set $t = 0$ on their departure. By letting $N_s = 0$ or $\alpha_s = 0$, the system after seed departures is characterized by

$$\begin{cases} \overline{A} \cdot \frac{\mathrm{d}s_i(t)}{\mathrm{d}t} = (i-1)p_{i-1}(t) - ip_i(t), & \text{for } i \geq 1, \\ s_0(t) \qquad\qquad\quad = 1, \end{cases} \tag{9}$$

with initial conditions $p_i(0) = \overline{p}_i$, $i = 0, 1, \ldots$, given by (6). When $N_s = 0$, the network does not have a steady state and the content will become incomplete eventually. However, we are still interested in certain transient behavior of the system.

First, by (9), the fraction of undecodable segments upon seed departures is $1 - s_m(0) = \sum_{i=0}^{m-1} p_i(0) = \sum_{i=0}^{m-1} \overline{p}_i$, the values of which are numerically plotted in Fig. 4 with $N = 1000$, $N_s = 10$, $M = 1000$, $\tilde{\mu} = 4$. As segment size m increases, there is a decreasing trend in the content loss. Moreover, there exists some knee in the curve, beyond which the content loss upon seed departures will not be further reduced by increasing m. Fig. 5 plots the fraction of decodable content at time t after seeds departure, which equals to $s_m(t)$, the fraction of segments with at least m blocks in the network. The parameters are $N = 1000$, $N_s = 50$, $M = 1000$, $\tilde{\mu} = 4$, $\overline{A} = 5$. We can see that only a small increment in m leads to a salient decrease in the content loss rate.

Furthermore, we numerically plot content lifetime after seeds departure in Fig. 6, given that the content still remains complete upon seeds departure. The parameters are the same as in Fig. 5. Clearly, the content becomes incomplete when there is a segment with less than m (coded) blocks in the network, and its lifetime equals to the first hitting time of $1 - s_m(t)$ to $1/G$ starting from the initial conditions $p_i(0) = \overline{p}_i$, $i = 0, 1, \ldots$. In general, the content persistence offered by network coding enhances as the segment size m increases.

6 Simulation Results

We have developed a simulating environment to experimentally evaluate the behavior of P2P content distribution systems with peer churn. The initial input to the simulator is a set of empty downloaders with i.i.d. lifetimes drawn from a general distribution, and a set of seeds which hold complete copies of the source content. Each downloader will leave the network once its lifetime has expired or it has finished downloading the content. The simulator has a "tracking server", which connects each peer to at least 40 other randomly chosen peers (in accordance to BitTorrent), which form its neighborhood. All other tasks, including downstream peer selection and data exchanges, are performed locally at peers.

The simulator runs in rounds. In each round, a peer first randomly chooses 4 of its neighbors as its downstream peers. We also require each peer to have at most 6 upstream peers to avoid load imbalance. With this policy, most peers turn out to have at least one upstream peer in each round. Each peer then uploads blocks to each of its downstream peers independently using random useful segment encoding described in Sec. 2. To accommodate peer heterogeneity, we assume there are an equal number of peers with upload bandwidth 2 MB/round, 512 KB/round and 256 KB/round. A peer with 256 KB/round connectivity can upload one block of 256 KB in a round. It is easy to check that the average upload rate is $\tilde{\mu} = \mu/k = 3.7$ blocks/round. Coding operations are done in $GF(2^8)$.

We first evaluate the block distribution and download performance in steady-state networks. Fig. 7 shows the steady-state block distribution for the replacement model. The parameters are set to be the same as in Fig. 3. We can see the experimental results of Fig. 7 matches the analytical results of Fig. 3, substantiating the correctness of Theorem 3 and the validity of the differential equations derived in Sec. 3.

The average block variation in steady state is plotted in Fig. 8 and Fig. 9 for the replacement model and Poisson arrival model, respectively, with N_s = 100, F = 768 MB, k = 256 KB, and $N = 6000$ in steady states. (For Poisson

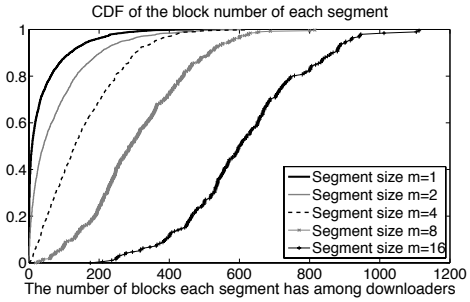


Fig. 7. Steady-state CDF of the number of blocks of each segment in all the downloaders.

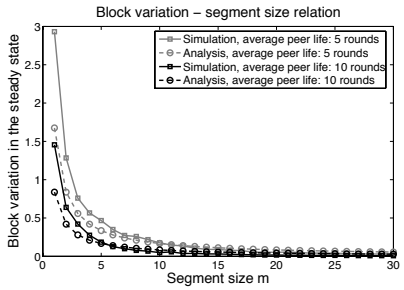


Fig. 8. Steady-state block variation as a function of segment size m for the replacement model.

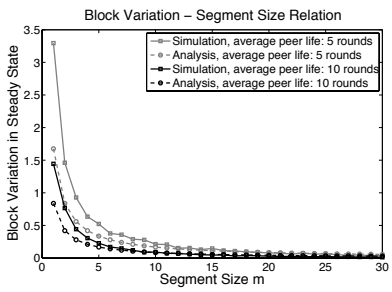


Fig. 9. Steady-state block variation as a function of segment size m for the Poisson arrival model.

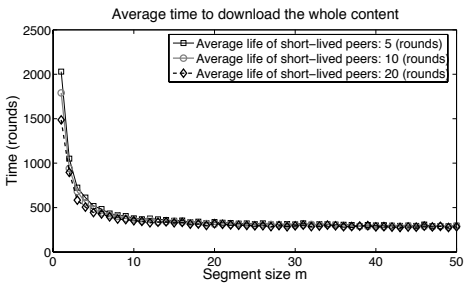


Fig. 10. The average time for a long-lived peer to download the entire content in steady state for the replacement model. $N = 2000$.

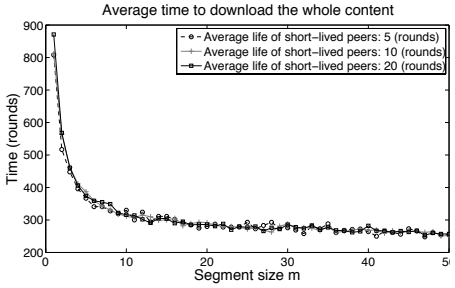


Fig. 11. The average time for a long-lived peer to download the entire content in steady state for the Poisson arrival model. $\lambda = 100$.

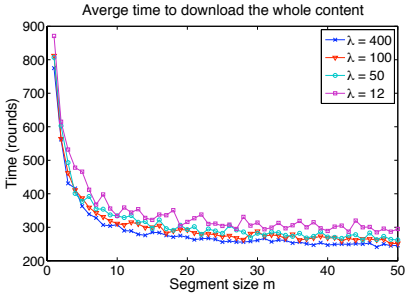


Fig. 12. The average time for a long-lived peer to download the entire content in the steady state for the Poisson arrival model.

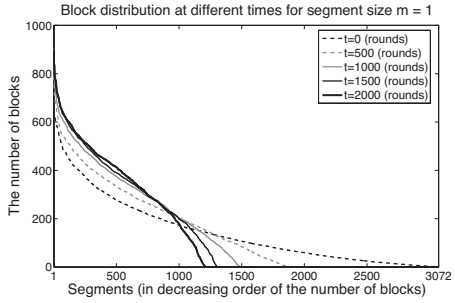
arrival model, the expected number of online downloaders is $\lambda\overline{L}$ in the steady state.) Even for such a heterogeneous environment, simulation results are quite close to the conclusion of Theorem 5. The simulation results differ from the analysis by a small constant coefficient, possibly due to the difference between the round-based simulation and the continuous-time analytical model. Simulation confirms that the steady-state block variation γ_I^2 is inversely proportional to the segment size m in both models, and when peer churn is more severe, or when the average peer age is lower, the network suffers from greater block variation.

We now evaluate the average time required for a long-lived peer to download the entire content as the segment size m varies. The results are plotted in Fig. 10 and Fig. 11 for the replacement model and Poisson arrival model, respectively. We set N_s = 30, F = 256 MB, k =256 KB. 1% of all the downloaders are long-lived ones which will not leave until they finish downloading. The other peers are short-lived and have lifetimes exponentially distributed. The download time exhibits a similar inversely proportional relationship with m, just as the block variation does. The download time is closely related to block variation because when there is lower block variation, peers will not be hindered in obtaining rare blocks, leading to lower delay in the download process. More importantly, we note that *there is a sweet spot of segment size, beyond which download time can hardly be reduced.* Thus, the use of a small segment size, e.g., 10-20, suffices to optimize the download efficiency, with only a moderate computational cost incurred.

The average download time is also plotted as λ varies in Fig. 12, where we set F = 256MB, N_s = 30, k = 256 KB. 1% of all the downloaders to be long-lived. The average lifetime of short-lived peers is 10 rounds. We can see that as the joining rate λ increases, the download time is reduced. In fact, we find through simulation that with a bigger λ and thus a larger network size, the steady-state block variation decreases and approaches its theoretical value, as asymptotic behavior comes into effect.

Let us now consider the extreme case that seeds leave in the steady state (at round 400) altogether under the replacement model, and let $t = 0$ upon seeds departure. We set $N = 6000$, $N_s = 100$, $F = 768$ MB, $k = 256$ KB, and $\overline{L} = 20$, and experimentally evaluate the content persistence in the absence of seeds. First, we show the block distribution evolution over time after seeds departure in Fig. 13. In the non-coding case ($m = 1$), 50% of all the blocks are missing at round 1000, while for network coding with $m = 4$, the block distribution stays almost unchanged within 2000 rounds, with only a small fraction of content loss. The figure clearly demonstrates that even network coding with small segment sizes can significantly enhance content persistence in a dynamic network without seeds. We also plot the fraction of decodable content over time after seeds departure in Fig. 14, which agrees with the analysis in Sec. 5 in trend.

7 Related Work



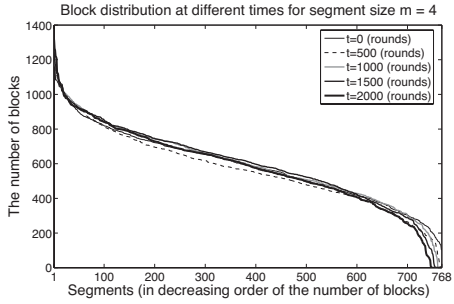


Fig. 13. The number of blocks that each segment has at different times t (rounds). Left: non-coding ($m = 1$); Right: network coding with $m = 4$.

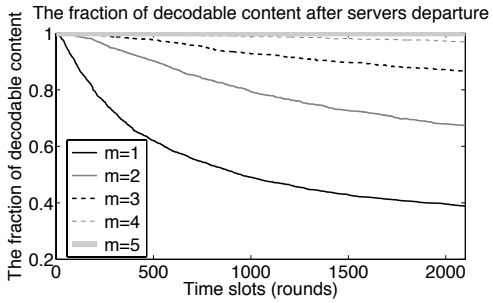


Fig. 14. The fraction of decodable content after seeds departure.

Randomized network coding has been applied to BitTorrent-like P2P content distribution by Gkantsidis et al. [2]. They show that network coding speeds up download over non-coding random block selection. To reduce the computational cost of coding across all the content, generation-based or segment-based network coding has been proposed by Chou et al. [3] and subsequently applied to P2P content distribution systems [10].

Despite the experimental studies, there exists a limited amount of work that theoretically analyzes the benefits of network coding in P2P networks. Yeung [11] shows in a time-slotted model that network coding achieves the optimal time to distribute k blocks given any transmission schedule in P2P networks. Deb et al. [4] show in a time-slotted model that network coding can achieve a shorter broadcast delay of k blocks in complete graphs, as compared to a sequential dissemination.

From a distinctly different perspective than throughput benefits, this paper analyzes how network coding can preserve block diversity and availability in large-scale dynamic P2P networks as the segment size varies. We quantitatively show that such a resilience gain of network coding is the major reason why network coding provides performance improvements with regard to download time and content persistence.

8 Conclusions

In this paper, we study the resilience gain of randomized network coding in terms of enhancing block diversity and availability in dynamic P2P networks. Using differential equations to approximate large deviations, we quantify the resilience gain of network coding, as different numbers of blocks are used for coding in a segment. We evaluate a wide range of performance metrics, including block availability distribution and download performance in steady-state networks, and content lifetime and content loss in transient networks without seeds.

We show that there is an inverse proportional relationship between the segment availability variation and segment size. We further find that the time needed for a peer to download the entire content from a dynamic P2P network is closely related to such segment availability variation. Our studies reveal that small segment sizes — around 20-30 even with high peer volatility — suffice to realize the major benefit of network coding in terms of preserving block diversity and reducing download time in steady-state networks. In the absence of seeds, content lifetime increases and content loss rate decreases in trend, as the segment size for coding increases. Our insights into the resilience-complexity tradeoff of segment-based network coding in dynamic P2P networks contributes to understanding the benefits and concerns of applying network coding in practical P2P file sharing and content dissemination systems.

References

[1] T. Ho, R. Koetter, M. Medard, D. R. Karger, and M. Effros, "The Benefits of Coding over Routing in a Randomized Setting," in *Proc. of IEEE International Symposium on Information Theory*, 2003.

[2] C. Gkantsidis and P. Rodriguez, "Network Coding for Large Scale Content Distribution," in *Proc. of IEEE INFOCOM 2005*, March 2005.

[3] P. A. Chou, Y. Wu, and K. Jain, "Practical Network Coding," in *Proc. of 41th Annual Allerton Conference on Communication, Control and Computing*, October 2003.

[4] S. Deb, M. Médard, and C. Choute, "Algebraic Gossip: A Network Coding Approach to Optimal Multiple Rumor Mongering," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2486–2507, June 2006.

[5] D.Mosk-Aoyama and D.Shah, "Information Dissemination via Network Coding," in *Proc. of IEEE International Symposium on Information Theory (ISIT'06)*, Seattle, WA, October 2006.

[6] D. Leonard, V. Rai, and D. Loguinov, "On Lifetime-Based Node Failure and Stochastic Resilience of Decentralized Peer-to-Peer Networks," in *Proc. of ACM SIGMETRICS'05*, Banff, Alberta, Canada, June 2005.

[7] K. K. Ramakrishnan and R. Jain, "A binary feedback scheme for congestion avoidance in computer networks," *ACM Transactions on Computer Systems*, vol. 8, no. 2, pp. 158–181, May 1990.

[8] L. Massoulie and M. Vojnovic, "Coupon Replication Systems," in *Proc. of ACM SIGMETRICS'05*, Banff, Alberta, Canada, June 2005.

[9] S. Resnick, *Adventures in Stochastic Processes*. Birkhauser, Boston, 2002.

[10] C. Gkantsidis, J. Miller, and P. Rodriguez, "Anatomy of a P2P Content Distribution System with Network Coding," in *Proceedings of 5th International Workshop on Peer-to-Peer Systems (IPTPS'06)*, 2006.

[11] R. W. Yeung, "Avalanche: A Network Coding Analysis," *Communications in Information and Systems*, vol. 7, no. 4, pp. 353–358, 2007.

[12] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed. Academic Press, Inc., 1975.

[13] T. G. Kurtz, "Approximation of Population Processes," in *CBMS-NSF Regional Conference Series in Applied Math*. SIAM, 1981.

[14] G. Tan and S. A. Jarvis, "Stochastic Analysis and Improvement of the Reliability of DHT-based Multicast," in *Proc. of IEEE INFOCOM 2007*, Anchorage, Alaska, USA, May 2007.

[15] P. Jacquet and W. Szpankowski, "Entropy Computations Via Analytic Depoissonization," *IEEE Transactions on Information Theory*, vol. 45, pp. 1072–1081, 1999.

APPENDIX A
PROOF OF LEMMA 2

Let i.i.d. random variables $A_1, A_2, \ldots, A_N$ denote the ages of online downloaders. We consider the challenging case when the network consists of short-lived peers, and thus assume a peer's mean lifetime $\overline{L}$ is finite. Since $\tilde{\mu}$ and $\overline{L}$ are finite, as $N \to \infty$, $M \to \infty$, with very small probability will two online peers have large sets of overlapped blocks in their buffers. Thus, a peer can almost always transmit a useful block to another peer. As a result, the total download bandwidth in the network always equals to the total upload bandwidth $(N+N_s)\tilde{\mu}$. Moreover, we can use identical and mutually independent random processes $\tilde{\mu}_1(t), \tilde{\mu}_2(t), \ldots, \tilde{\mu}_N(t)$ to denote the actual download bandwidth (blocks/unit time) of online downloaders at time t. Assume $\tilde{\mu}_i(t)$ are wide-sense stationary and mean-ergodic. We also assume that all $\tilde{\mu}_i(t)$ have the same mean function $\mathbf{E}\{\tilde{\mu}_i(t)\}$. Now we are ready to establish Lemma 2 using the strong law of large numbers.

At any given time t, we have

$$\boldsymbol{Y}(t) = \sum_{i=1}^{N} \int_0^{A_i} \tilde{\mu}_i(\tau) \mathrm{d}\tau.$$

According to the strong law of large numbers,

$$\lim_{N\to\infty} \frac{\boldsymbol{Y}(t)}{N} = \mathbf{E}\{\int_0^{A_i} \tilde{\mu}_i(\tau)\mathrm{d}\tau\} \quad \text{with probability 1.}$$

We have

$$\begin{aligned}
&\mathbf{E}\{\int_0^{A_i} \tilde{\mu}_i(\tau)\mathrm{d}\tau\} = \mathbf{E}\{\mathbf{E}\{\int_0^{A_i} \tilde{\mu}_i(\tau)\mathrm{d}\tau | A_i\}\} \\
&= \mathbf{E}\{\int_0^{A_i} \mathbf{E}\{\tilde{\mu}_i(\tau)\}\mathrm{d}\tau | A_i\} = \mathbf{E}\{A_i \cdot \mathbf{E}\{\tilde{\mu}_i(\tau)\}\} \\
&= \overline{A} \cdot \mathbf{E}\{\tilde{\mu}_i(\tau)\} = \overline{A} \cdot \frac{\sum_{j=1}^{N} \tilde{\mu}_j(\tau)}{N} = \frac{(N+N_s)\tilde{\mu}\overline{A}}{N},
\end{aligned}$$

where the second equality holds because $\tilde{\mu}_i(\tau)$ is mean-ergodic. The second last equality holds with probability 1 as $N \to \infty$. The last equality holds because $\tilde{\mu}_i(\tau)$ are wide-sense stationary and the total download bandwidth in the network always equals to the total upload bandwidth $(N+N_s)\tilde{\mu}$. Therefore, we have shown $\lim_{N\to\infty} \frac{\boldsymbol{Y}(t)}{(N+N_s)} = \tilde{\mu}\overline{A}$ w.p. 1 for large t. □

APPENDIX B
DERIVING THE DIFFERENTIAL EQUATIONS FOR THE REPLACEMENT MODEL

Note that there are three factors that contribute to the change of $\boldsymbol{S}(t)$, namely seed uploads, downloader uploads, and downloader departures. No matter what distribution the time for a peer to upload a block follows, the aggregate uploads from all peers form a Poisson process of rate $(N+N_s)\tilde{\mu}$, as $N \to \infty$, $N_s \to \infty$ (refer to [12] pp. 221). At each upload of the aggregate upload process, it is a random peer in the network that uploads. Similarly, no matter what distribution the lifetime of a peer $\boldsymbol{L}$ follows, the aggregate departures of downloaders form a Poisson process with rate $N/\overline{L}$. Even with the Markovian property of uploads and download departures, it is still difficult to write the transition rates for $\boldsymbol{S}(t)$ due to the dependency among the actions of different peers. To further introduce analytical tractability that makes $\boldsymbol{S}(t)$ a Markov process with linear intensities, we make the following linear approximations on the effect of random downstream peer selection and random useful block selection:

Assumption 1 (Linear Approximation for Uploads): Whenever a seed uploads a useful block, the effect is approximated by choosing *uniformly* from all the segments a random segment s and increasing deg(s) by 1; whenever a downloader uploads a useful block, the effect is approximated by choosing from all the segments a random segment s with probability $\deg(s)/\sum_s \deg(s) = \deg(s)/\boldsymbol{Y}(t)$ and increasing $\deg(s)$ by 1.

Remarks: Since blocks of all segments are distributed in seeds in a balanced way, and blocks of all segments are largely needed by downloaders due to a finite $\overline{L}$, each segment has an equal chance of being chosen when a seed uploads. However, when a downloader uploads, each segment does not enjoy an equal chance of being chosen, as different segments have different availability. Instead, the more blocks a segment has in downloaders, the more frequently it will be chosen and encoded for transfer. The accuracy of this approximation is verified in Sec. 6 by simulations.

Let $\boldsymbol{e}_i$ denote a unit vector of the same dimension as $\boldsymbol{R}(t)$ with its ith element being 1, and all other elements being 0. Now it is easy to write the intensities $q^{(G)}_{R,R+e_i}$ for the transitions on $\boldsymbol{R}$ due to uploads, when the total number of segments is G:

$$\begin{aligned}
\boldsymbol{R} \to \boldsymbol{R} + \boldsymbol{e}_i,\ q^{(G)}_{R,R+e_i} &= N_s\tilde{\mu}p_{i-1} + N\tilde{\mu} \cdot \frac{(i-1)n_{i-1}}{\boldsymbol{Y}} \\
&= N_s\tilde{\mu}p_{i-1} + \frac{\alpha}{\alpha+\alpha_s}\frac{(i-1)n_{i-1}}{\overline{A}}
\end{aligned} \tag{10}$$

for all finite i, where $N/M = \alpha$, $N_s/M = \alpha_s$, and the second equality is due to Lemma 2. The rationale behind is that seeds upload a block at a total rate of $N_s\tilde{\mu}$ with all segments being chosen for encoding equally likely, while downloaders upload a block at a total rate of $N\tilde{\mu}$ with segments being chosen according to Approximation 1. Note that in (10), it is implicitly assumed that no linear dependency will occur when applying network coding, because by Lemma 2.1 in [4], a random linear combination of all the blocks from the same segment at a peer p is useful to another randomly chosen peer in the network with probability at least $1 - 1/q$ if network coding is done in $GF(2^q)$. And this argument is true regardless of whether peer p is a seed or a downloader.

Let us consider block loss due to downloader departures. The total number of downloader departures in a small interval Δt is $N\Delta t/\overline{L}$. Each downloader downloads at a rate of $\frac{(N+N_s)\tilde{\mu}}{N}$ on average. Upon departure, a downloader takes away $\overline{L}\cdot\frac{(N+N_s)\tilde{\mu}}{N}$ blocks on average. Therefore, the network loses $\frac{N\Delta t}{\overline{L}}\cdot\overline{L}\cdot\frac{(N+N_s)\tilde{\mu}}{N}=(N+N_s)\tilde{\mu}\Delta t$ blocks in Δt in total. We make a similar linear approximation to block loss:

Assumption 2 (Linear Approximation for Loss): As $N\to\infty$, $N_s\to\infty$, $M\to\infty$, among all the blocks lost due to downloader departures in a small time interval $\Delta t=O(1/\sqrt{N})$, the proportion of the blocks that segments of degree i lose is $in_i/\sum_i in_i=in_i/\boldsymbol{Y}(t)$.

Remarks: This essentially means that the more blocks a segment has in downloaders, the more it loses due to peer departures. The choice of $\Delta t=O(1/\sqrt{N})$ ensures each segment can lose at most one block in Δt. The accuracy of this approximation is verified in Sec. 6 through simulations.

Thus, in Δt, the total block loss of degree-i segments is

$$(N_s+N)\tilde{\mu}\Delta t\cdot\frac{i\cdot n_i}{\boldsymbol{Y}}=\frac{i\cdot n_i}{\overline{A}}\cdot\Delta t.$$

As $N\to\infty$ while $\overline{L}$ remains finite, such a block loss pattern will be equivalent to having individual blocks of different segments get lost according to the following transitions:

$$\boldsymbol{R}\to\boldsymbol{R}-\boldsymbol{e}_i,\quad q^{(G)}_{R,R-e_i}=\frac{i\cdot n_i}{\overline{A}},\quad\forall i<\infty. \tag{11}$$

Let $\boldsymbol{S}^{(G)}(t)=\{s_0^{(G)},s_1^{(G)},s_2^{(G)}\ldots\}:=\boldsymbol{R}^{(G)}(t)/G$ denote the normalized process when there are G segments. Note that the intensities for $\boldsymbol{R}$ can be rewritten in the form:

$$q^{(G)}_{R,R+l}=G\cdot\beta_l(\frac{\boldsymbol{R}}{G})=G\cdot\beta_l(\boldsymbol{S}),\quad l=+e_i,-e_i\quad\forall i<\infty,$$

where

$$\begin{aligned}\beta_{+e_i}(\boldsymbol{S})&=\frac{N_s\tilde{\mu}p_{i-1}}{G}+\frac{\alpha}{\alpha+\alpha_s}\cdot\frac{(i-1)n_{i-1}}{\overline{A}G}\\&=\alpha_s m\tilde{\mu}p_{i-1}+\frac{\alpha}{\alpha+\alpha_s}\cdot\frac{(i-1)p_{i-1}}{\overline{A}},\end{aligned}\tag{12}$$

$$\beta_{-e_i}(\boldsymbol{S})=\frac{i\cdot n_i}{\overline{A}G}=\frac{i\cdot p_i}{\overline{A}}. \tag{13}$$

Thus, $\boldsymbol{S}^{(G)}(t)$ forms a *density dependent jump Markov process* (refer to [13], pp.51). We set

$$F(x)=\sum_l l\beta_l(x)=\sum_i\boldsymbol{e}_i(\beta_{+e_i}(x)-\beta_{-e_i}(x)). \tag{14}$$

By Kurtz Theorem (Theorem 8.1 in [13]), for $\boldsymbol{S}^{(G)}(t)$ to converge to a deterministic fluid, we need that $\boldsymbol{S}^{(G)}(0)$ converges to a certain value $\boldsymbol{S}(0)$ that does not depend on G. Since $\boldsymbol{S}^{(G)}(0)=(1,0,0,\ldots)$, this condition is satisfied. Moreover, we need the boundedness and Lipschitz continuity of $F(x)$. These conditions are guaranteed in our model by the finiteness of $\sum_{i=1}^{\infty}s_i(t)\to(\alpha+\alpha_s)m\tilde{\mu}\overline{A}$, which remains finite due to a finite $\overline{L}$. Therefore, by Kurtz Theorem, $\boldsymbol{S}^{(G)}(t)$ converges almost surely to a deterministic fluid $\boldsymbol{S}(t)=\{s_0,s_1,s_2\ldots\}$ for large G:

$$\boldsymbol{S}(t)=\boldsymbol{S}(0)+\int_0^t F(\boldsymbol{S}(u))du,\quad t\geq 0, \tag{15}$$

Substituting (12), (13), and (14) into (15), we have shown that as $t\to\infty$, $M\to\infty$, $N\to\infty$, $N_s\to\infty$, $N/M\to\alpha$, $N_s/M\to\alpha_s$, while m remains finite, $\boldsymbol{S}(t)$ in replacement model converges to the system of ODEs (3) with an arbitrarily small error.

APPENDIX C
DERIVING THE DIFFERENTIAL EQUATIONS FOR THE POISSON ARRIVAL MODEL

In Poisson arrival model, it turns out the distribution of $\boldsymbol{A}$ can be determined in the same formula as in Lemma 1. We state this result in the following lemma, the proof of which can be found in Lemma 3 in [14].

Lemma 6: Under the Poisson arrival model with joining rate λ, as the observation time $t_o\to\infty$, we have $\mathbf{E}[\boldsymbol{A}]=\overline{A}=(\overline{L}^2+\sigma^2)/2\overline{L}$, regardless of λ.

Similarly, if the joining rate λ is large enough, the network will reach a stationary stage, where $\boldsymbol{Y}(t)/(\lambda\overline{L}+N_s)$ almost always remains the same:

Lemma 7: Let $\boldsymbol{Y}(t)$ denote the total number of blocks in downloaders from all the segments at a given time t. Under the Poisson arrival model, as $\lambda\to\infty$,$M\to\infty$, $\lambda\overline{L}=\alpha' M$, $N_s=\alpha_s M$, where α', α_s are finite constants,

$$\lim_{\lambda\to\infty}\frac{\boldsymbol{Y}(t)}{\lambda\overline{L}+N_s}=\tilde{\mu}\overline{A}\quad\text{with probability 1}, \tag{16}$$

for t sufficiently large.

Proof: We first show $\lim_{\lambda\to\infty}N(t)/\lambda\overline{L}=1$ and then $\frac{\boldsymbol{Y}(t)}{N(t)+N_s}\to\tilde{\mu}\overline{A}$ to prove the theorem. Since the arrival process is Poisson, it can be split into n i.i.d. Poisson processes, each with rate $\lambda_0=\lambda/n$, and the system can be split into n sub-systems, each being a $M/G/\infty$ queue. Let $N_i(t)$ be the number of peers in the ith subsystem. To prove $\lim_{\lambda\to\infty}N(t)/\lambda\overline{L}=1$ for large t, we only need to show that $\lim_{n\to\infty}N(t)/n\lambda_0\overline{L}=1$ with probability 1 for any positive and finite λ_0. This is true because

$$\begin{aligned}&\lim_{n\to\infty}\lim_{t\to\infty}\frac{N(t)}{n\lambda_0\overline{L}}=\lim_{n\to\infty}\lim_{t\to\infty}\frac{\sum_{i=1}^n N_i(t)}{n}\cdot\frac{1}{\lambda_0\overline{L}}\\&=\lim_{t\to\infty}\frac{\mathbf{E}\{N_i(t)\}}{\lambda_0\overline{L}}=\frac{\lambda_0\overline{L}}{\lambda_0\overline{L}}=1,\end{aligned}$$

where the second equality holds w.p. 1 by the strong law of large numbers, and the third equality holds due to Little's Theorem. Hence, we have

$$\begin{aligned}&\lim_{t\to\infty}\lim_{\lambda\to\infty}\frac{\boldsymbol{Y}(t)}{\lambda\overline{L}+N_s}\\&=\lim_{t\to\infty}\lim_{N(t)\to\infty}\frac{\boldsymbol{Y}(t)}{N(t)+N_s}\cdot\lim_{t\to\infty}\lim_{\lambda\to\infty}\frac{N(t)+N_s}{\lambda\overline{L}+N_s}=\tilde{\mu}\overline{A}\end{aligned}$$

with probability 1, where $\lim_{t\to\infty}\lim_{N(t)\to\infty}\frac{\boldsymbol{Y}(t)}{N(t)+N_s}$ can be shown to approach $\tilde{\mu}\overline{A}$ following the same argument as in the proof of Lemma 2 by replacing N with $N(t)$. Thus, we have shown that $\lim_{\lambda\to\infty}\frac{\boldsymbol{Y}(t)}{\lambda\overline{L}+N_s}=\tilde{\mu}\overline{A}$ with probability 1 for large t. □

Given the above two lemmas, the derivation of (4) follows the same approach as in the derivation of (3) for the replacement model. The only difference is that we replace N in the replacement model with $\lambda\overline{L}$ and apply Lemma 7 instead of Lemma 2 when we derive (4).

APPENDIX D
PROOF OF THEOREM 3

Clearly, $\Pr\{\boldsymbol{I}=i\}=\overline{p}_i$ as $M\to\infty$, while m remains finite. Based on the steady-state solutions to (3), let us derive $\overline{p}_i$ in a closed form under the conditions set for B and β. Let $C=B/\beta-1$. We first prove $\overline{p}_i=\beta^i\binom{i+C}{i}\cdot\overline{p}_0$, for $i\geq 1$, and then determine $\overline{p}_0$ from $\sum_{i=0}^{\infty}\overline{p}_i=1$. Let $\beta=\frac{l}{n}<1$, where $l,n\in\mathbb{N}$. Because $C\in\mathbb{Z}^+$, we have

$$\overline{p}_i=\overline{p}_0\prod_{j=1}^{i}(\frac{l}{n}+\frac{B-\beta}{j}),$$

from which we can get

$$\overline{p}_i=\overline{p}_0(\frac{l}{n})^i\cdot\frac{(i+C)!}{i!C!}=\beta^i\binom{i+C}{i}\cdot\overline{p}_0. \qquad (17)$$

We now determine $\overline{p}_0$. Let $a_i=\overline{p}_i/\overline{p}_0=\beta_i\binom{i+C}{i}$, $i=0,1,2,\ldots$. Then we have $1/\overline{p}_0=\sum_{i=0}^{\infty}a_i$. Because

$$\binom{i+C}{i}=\binom{C+i-1}{i}+\binom{C+i-1}{i-1},\quad i\geq 1,$$

we have

$$\sum_{i=0}^{\infty}a_i=\sum_{i=1}^{\infty}\beta^i\binom{(C-1)+i}{i}+\beta\sum_{i-1=0}^{\infty}\beta^{i-1}\binom{C+i-1}{i-1}+a_0$$

If we view a_i as a function of C and let

$$f(C)=\sum_{i=1}^{\infty}a_i(C)=\sum_{i=1}^{\infty}\beta^i\binom{i+C}{i},$$

we have

$$\begin{cases}(\beta-1)f(C)+f(C-1)+\beta=0, & \text{for } C\geq 1\\ f(0)=\sum_{i=1}^{\infty}\beta^i=\frac{\beta}{1-\beta}\end{cases}$$

By induction, it is not hard to get

$$f(C)=(1-\beta)^{-(C+1)}-1,\quad C=0,1,2,\ldots.$$

Hence,

$$\sum_{i=0}^{\infty}a_i=f(C)+a_0=(1-\beta)^{-(C+1)}$$

Thus, we have $\overline{p}_0=(1-\beta)^{\frac{B}{\beta}}$. Substituting $\overline{p}_0$ into (17) will yield $\overline{p}_i$. □

APPENDIX E
STEADY-STATE ENTROPY OF $\boldsymbol{I}$

Theorem 8: Suppose all the parameters satisfy the conditions set in Theorem 3 or Theorem 4. Under the replacement model, for big $\alpha_s m\tilde{\mu}\overline{A}$, the entropy of $\boldsymbol{I}$ in the steady state is given by

$$H(\boldsymbol{I})=\frac{1}{2}\ln m+\frac{1}{2}(1+\ln\frac{2\pi\tilde{\mu}\overline{A}(\alpha+\alpha_s)^2}{\alpha_s})+o(1), \quad (18)$$

where $\overline{A}$ is given by Lemma 1, $\alpha=N/M$, $\alpha=N_s/M$. Under the Poisson arrival model, for big $\alpha_s m\tilde{\mu}\overline{A}$, the entropy of $\boldsymbol{I}$ in the steady state is given by

$$H(\boldsymbol{I})=\frac{1}{2}\ln m+\frac{1}{2}(1+\ln\frac{2\pi\tilde{\mu}\overline{A}(\alpha'+\alpha_s)^2}{\alpha_s})+o(1), \quad (19)$$

where $\alpha'=\lambda\overline{L}/M$, $\alpha=N_s/M$.

Proof: We derive $H(\boldsymbol{I})$ using Theorem 1 in [15]. For the replacement model, since $\boldsymbol{I}$ follows the negative binomial distribution (6), $\boldsymbol{I}$ can be represented as a sum of B/β i.i.d. geometrically distributed random variables. Thus, the distribution of $\boldsymbol{I}$ in (6) satisfies the required assumptions (A1)-(A2) in [15]. By Theorem 1 in [15], for a big B/β, the Shannon entropy of $\boldsymbol{I}$ is

$$H(\boldsymbol{I})=\frac{1}{2}\ln\frac{B}{\beta}+\frac{1}{2}(1+\ln\frac{2\pi\beta}{(1-\beta)^2})+o(1). \qquad (20)$$

Plugging in $B=\alpha_s m\tilde{\mu}\overline{A}$ and $\beta=\frac{\alpha}{\alpha+\alpha_s}$ gives (18). (19) can be derived in a similar way based on (7). □

Di Niu. Di Niu received his B.Engr. degree in 2005 from the Department of Electronics and Communication Engineering, Sun Yat-sen (Zhongshan) University, Guangzhou, Guangdong, China, and his M.A.Sc. degree in 2009 from the Department of Electrical and Computer Engineering, University of Toronto. Since 2008, he has been a PhD candidate in the Department of Electrical and Computer Engineering at University of Toronto. His research interests include measurement, data mining and implementation of large-scale multimedia systems, peer-to-peer networks, applications of network coding.

Baochun Li. Baochun Li received his B.Engr. degree in 1995 from Department of Computer Science and Technology, Tsinghua University, Beijing, China, and his M.S. and Ph.D. degrees in 1997 and 2000 from the Department of Computer Science, University of Illinois at Urbana-Champaign. Since 2000, he has been with the Department of Electrical and Computer Engineering at the University of Toronto, where he is currently a Full Professor. He holds the Nortel Networks Junior Chair in Network Architecture and Services since October 2003, and the Bell University Laboratories Endowed Chair in Computer Engineering since August 2005. In 2000, he was the recipient of the IEEE Communications Society Leonard G. Abraham Award in the Field of Communications Systems. In 2005, he won the Best Paper Award at the Thirteenth IEEE International Workshop on Quality of Service (IWQoS). His research interests include large-scale multimedia systems, peer-to-peer networks, applications of network coding, and wireless networks.