Future Generation Computer Systems 42 (2015) 36–43

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

# Analysis of file download time in peer-to-peer networks with stochastic and time-varying service capacities

Keqin Li *

*Department of Computer Science, State University of New York, New Paltz, NY 12561, USA*

## HIGHLIGHTS

- Analyzing the file download time with stochastic and time-varying service capacities.
- Treating the service capacity of a source peer as a stochastic process.
- Validating analytical results by simulations.

ARTICLE INFO

*Article history:*
Received 6 February 2014
Received in revised form
19 August 2014
Accepted 1 September 2014
Available online 16 September 2014

*Keywords:*
Download time
File sharing system
Parallel downloading
Peer-to-peer network
Stochastic service capacity
Time-varying service capacity

ABSTRACT

The service capacities of a source peer at different times in a peer-to-peer (P2P) network exhibit temporal correlation. Unfortunately, there is no analytical result which clearly characterizes the expected download time from a source peer with stochastic and time-varying service capacity. The main contribution of this paper is to analyze the expected file download time in P2P networks with stochastic and time-varying service capacities. The service capacity of a source peer is treated as a stochastic process. Analytical results which characterize the expected download time from a source peer with stochastic and time-varying service capacity are derived for the autoregressive model of order 1. Simulation results are presented to validate our analytical results. Numerical data are given to show the impact of the degree of correlation and the strength of noise on the expected file download time. For any chunk allocation method, an analytical result of the expected parallel download time from a source peer with stochastic and time-varying service capacity is derived. It is shown that the algorithm which chooses chunk sizes proportional to the expected service capacities has better performance than the algorithm which chooses equal chunk sizes. It is also shown that multiple source peers do reduce the parallel download time significantly.


## 1. Introduction

### 1.1. Motivation

It has been well known that the analysis of file download time in a peer-to-peer (P2P) network has three challenges, i.e., randomness, spatial heterogeneity, and temporal correlation [1–4]. Randomness means that the service capacity of a source peer in a P2P network is a random variable, due to variable workload and unpredictable network traffic and congestion and delay encountered by a file transfer. Furthermore, service capacities in a P2P network exhibit both spatial heterogeneity and temporal correlation. Spatial heterogeneity means that the service capacities of different source peers have different probability distributions. Temporal correlation means that the service capacities of a source peer at different times are correlated. Randomness, spatial heterogeneity, and temporal correlation all have impact on file download times.

Analysis of file download time in a P2P network with heterogeneous source peers and random service capacities has been conducted. In [2], the problem of reducing download times in P2P file sharing systems with stochastic service capacities is addressed, and a chunk-based switching and peer selection algorithm using the method of probing high-capacity peers is proposed and the expected download time of the algorithm is analyzed. It is proved in [4] that for two or more heterogeneous source peers and sufficiently large file size, the expected file download time of the time-based switching algorithm is less than and can be arbitrarily less than the expected download time of the chunk-based switching algorithm and the expected download time of the permanent connection algorithm. Furthermore, it is

* Tel.: +1 845 257 3534; fax: +1 845 257 3996.
*E-mail address:* lik@newpaltz.edu.

http://dx.doi.org/10.1016/j.future.2014.09.003
0167-739X/

shown that the expected file download time of the time-based switching algorithm is in the range of the file size divided by the harmonic mean of service capacities and the file size divided by the arithmetic mean of service capacities.

In [1], the problem of minimizing file download time from source peers with time-varying service capacities is considered. A random chunk-based switching method and a random periodic switching method for single downloading is proposed, aiming to reduce the effect of spatial heterogeneity and temporal correlation and to achieve the harmonic mean of service capacities. A parallel downloading method which divides a file into chunks of equal sizes is also proposed. The performance of all these methods are evaluated by simulations. Despite the above effort, there is no analytical result which clearly characterizes the expected download time from a source peer with stochastic and time-varying service capacity.

### 1.2. Our contributions

The main contribution of this paper is to analyze the expected file download time in peer-to-peer networks with stochastic and time-varying service capacities. Our work in this paper has quantitative implication for real-world applications.

In Sections 2–4, we consider the expected time of a single download. We treat the service capacity of a source peer as a stochastic process (Section 2). For the autoregressive model of order 1, we derive analytical results which characterize the expected download time from a source peer with stochastic and time-varying service capacity (Section 3). In Section 4, we present simulation results to validate our analytical results. We also give numerical data to show the impact of the degree of correlation and the strength of noise on the expected file download time.

In Section 5, we consider the expected time of a parallel download. For any chunk allocation method, we derive an analytical result of the expected parallel download time from a source peer with stochastic and time-varying service capacity. We show that the algorithm which chooses chunk sizes proportional to the expected service capacities has better performance than the algorithm which chooses equal chunk sizes. We also show that multiple source peers do reduce the parallel download time significantly.

## 2. Time-varying service capacities

The service capacity of a source peer is treated as a stochastic process. Assume that the time of a P2P file sharing system is divided into time slots of equal length. Let $C_i(t)$ denote the random service capacity of source peer $i$ at time slot $t = 1, 2, 3, \ldots$, where $t = 1$ is the starting time of a file download.

Let $T_i(S)$ be the download time of a file of size $S > 0$ from source peer $i$, treated as a discrete (integer) random variable. Throughout the paper, we use $\boldsymbol{P}[e]$ to denote the probability of an event $e$, and $\boldsymbol{E}(X)$ the expectation of a random variable $X$.

The following theorem gives a general characterization of $\boldsymbol{E}(T_i(S))$.

**Theorem 1.** *The expected file download time is*

$$\boldsymbol{E}(T_i(S)) = \sum_{t=0}^{\infty} \boldsymbol{P}\left[\sum_{j=1}^{t} C_i(j) < S\right].$$

**Proof.** It is clear that $T_i(S) = t$ if and only if

$$C_i(1) + C_i(2) + \cdots + C_i(t-1) < S \le C_i(1) + C_i(2) + \cdots + C_i(t).$$

Notice that $T_i(S) > t$ if and only if $C_i(1) + C_i(2) + \cdots + C_i(t) < S$. Thus, we have the probability mass function (pmf) of $T_i(S)$,

$$\begin{aligned}\boldsymbol{P}[T_i(S) = t] &= \boldsymbol{P}[T_i(S) > t-1] - \boldsymbol{P}[T_i(S) > t] \\ &= \boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t-1) < S] \\ &\quad - \boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t) < S],\end{aligned}$$

for all $t = 1, 2, 3, \ldots$. The expectation of $T_i(S)$ is

$$\begin{aligned}\boldsymbol{E}(T_i(S)) &= \sum_{t=1}^{\infty} t\boldsymbol{P}[T_i(S) = t] \\ &= \sum_{t=1}^{\infty} \boldsymbol{P}[T_i(S) \ge t] \\ &= \sum_{t=1}^{\infty} \boldsymbol{P}[T_i(S) > t-1] \\ &= \sum_{t=1}^{\infty} \boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t-1) < S] \\ &= \sum_{t=0}^{\infty} \boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t) < S].\end{aligned}$$

The theorem is proven. ■

It remains to analyze $C_i(1) + C_i(2) + \cdots + C_i(t)$. In this paper, we will mainly use an autoregressive model to specify $C_i(t)$, which is an effective way to characterize stochastic and time-varying service capacities [1].

## 3. Analytical results

### 3.1. An autoregressive model

We consider a class of stochastic processes called the autoregressive model of order 1 (AR(1)) to study stochastic and time-varying service capacities. In an AR(1) stochastic process [5], we have

$$C_i(t) = \varphi_i C_i(t-1) + \varepsilon_i(t) + d_i,$$

for all $t \ge 2$. In the above equation, $C_i(1)$, i.e., the initial service capacity of source peer $i$, is a random variable with probability distribution function (pdf) $f_{C_i}(c)$, cumulative distribution function (cdf) $F_{C_i}(c)$, mean $\mu_{C_i}$, and variance $\sigma^2_{C_i}$. A constant $0 \le \varphi_i < 1$ indicates the degree of correlation between $C_i(t)$ and $C_i(t-1)$. The noise process $\varepsilon_i(2), \varepsilon_i(3), \varepsilon_i(4), \ldots$ is a sequence of independent and identically distributed (i.i.d.) random variables with pdf $f_{\varepsilon_i}(c)$, cdf $F_{\varepsilon_i}(c)$, zero mean $\mu_{\varepsilon_i} = 0$, and variance $\sigma^2_{\varepsilon_i}$. A constant $d_i = (1-\varphi_i)\mu_{C_i}$ is set in such a way that all the $C_i(t)$'s have the same expectation.

The following theorem shows that the summation $C_i(1) + C_i(2) + \cdots + C_i(t)$ of $t$ correlated random variables is actually a linear combination of $t$ independent random variables $C_i(1), \varepsilon_i(2), \varepsilon_i(3), \ldots, \varepsilon_i(t)$.

**Theorem 2.** *We have*

$$\begin{aligned}&C_i(1) + C_i(2) + \cdots + C_i(t) \\ &\quad = \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) C_i(1) + \sum_{j=2}^{t} \left(\frac{1-\varphi_i^{t-j+1}}{1-\varphi_i}\right) \varepsilon_i(j) \\ &\qquad + \left(t - \frac{1-\varphi_i^t}{1-\varphi_i}\right) \mu_{C_i}.\end{aligned}$$

**Proof.** Notice that by the recursive definition of $C_i(t)$, we have

$$C_i(1) = \varphi_i^0 C(1),$$
$$C_i(2) = \varphi_i^1 C(1) + \varepsilon_i(2) + d_i,$$
$$C_i(3) = \varphi_i^2 C(1) + (\varphi_i\varepsilon_i(2) + \varepsilon_i(3)) + (\varphi_i + 1)d_i,$$
$$\begin{aligned} C_i(4) &= \varphi_i^3 C(1) + (\varphi_i^2\varepsilon_i(2) + \varphi_i\varepsilon_i(3) + \varepsilon_i(4)) \\ &\quad + (\varphi_i^2 + \varphi_i + 1)d_i, \end{aligned}$$
$$\vdots$$
$$\begin{aligned} C_i(t) &= \varphi_i^{t-1} C(1) + (\varphi_i^{t-2}\varepsilon_i(2) + \varphi_i^{t-3}\varepsilon_i(3) + \cdots + \varepsilon_i(t)) \\ &\quad + (\varphi_i^{t-2} + \cdots + \varphi_i + 1)d_i. \end{aligned}$$

Hence, we get

$$\begin{aligned} &C_i(1) + C_i(2) + \cdots + C_i(t) \\ &= \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) C_i(1) + \left(\frac{1-\varphi_i^{t-1}}{1-\varphi_i}\right)\varepsilon_i(2) + \left(\frac{1-\varphi_i^{t-2}}{1-\varphi_i}\right)\varepsilon_i(3) \\ &\quad + \cdots + \left(\frac{1-\varphi_i^1}{1-\varphi_i}\right)\varepsilon_i(t) \\ &\quad + \left(\frac{1-\varphi_i^0}{1-\varphi_i} + \frac{1-\varphi_i^1}{1-\varphi_i} + \frac{1-\varphi_i^2}{1-\varphi_i} + \cdots + \frac{1-\varphi_i^{t-1}}{1-\varphi_i}\right) d_i \\ &= \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) C_i(1) + \left(\frac{1-\varphi_i^{t-1}}{1-\varphi_i}\right)\varepsilon_i(2) + \left(\frac{1-\varphi_i^{t-2}}{1-\varphi_i}\right)\varepsilon_i(3) \\ &\quad + \cdots + \left(\frac{1-\varphi_i^1}{1-\varphi_i}\right)\varepsilon_i(t) + \frac{1}{1-\varphi_i}\left(t - \frac{1-\varphi_i^t}{1-\varphi_i}\right) d_i \\ &= \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) C_i(1) + \left(\frac{1-\varphi_i^{t-1}}{1-\varphi_i}\right)\varepsilon_i(2) + \left(\frac{1-\varphi_i^{t-2}}{1-\varphi_i}\right)\varepsilon_i(3) \\ &\quad + \cdots + \left(\frac{1-\varphi_i^1}{1-\varphi_i}\right)\varepsilon_i(t) + \left(t - \frac{1-\varphi_i^t}{1-\varphi_i}\right)\mu_{C_i}. \end{aligned}$$

The theorem is proven. ■

It is clear that an AR(1) stochastic process $C_i(t)$ with $t \ge 1$ is determined by $f_{C_i}(c)$ and $f_{\varepsilon_i}(c)$. In this paper, we assume that the $\varepsilon_i(t)$'s are normal random variables with parameters $\mu_{\varepsilon_i} = 0$ and $\sigma_{\varepsilon_i}^2$, i.e.,

$$f_{\varepsilon_i}(c) = \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon_i}} e^{-c^2/(2\sigma_{\varepsilon_i}^2)},$$

in $(-\infty, +\infty)$.

The following subsections derive further results.

### 3.2. Arbitrary distribution of $f_{C_i}(c)$

In this section, we provide an analytical result on $\boldsymbol{E}(T_i(S))$ for an arbitrary distribution of $f_{C_i}(c)$. Let $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ be the pdf of a standard normal distribution.

**Theorem 3.** *The expected file download time is*

$$\begin{aligned} \boldsymbol{E}(T_i(S)) &= 2 + \int_0^\infty \left(\sum_{t=2}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right) f_{C_i}(c)dc \\ &= 2 + \int_0^\infty \left(F_{C_i}(c)\sum_{t=2}^\infty \frac{1}{\sigma_i(t)}\left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) \right. \\ &\quad \left. \times\ \phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right) dc, \end{aligned}$$

*for any pdf $f_{C_i}(c)$.*

**Proof.** Let $T_i(S,c)$ be the download time of a file of size $S$ from source peer $i$ when $C_i(1) = c$. In this case, by Theorem 2, we get

$$\begin{aligned} C_i(1) + C_i(2) + \cdots + C_i(t) &= \left(\frac{1-\varphi_i^{t-1}}{1-\varphi_i}\right)\varepsilon_i(2) \\ &\quad + \left(\frac{1-\varphi_i^{t-2}}{1-\varphi_i}\right)\varepsilon_i(3) + \cdots + \left(\frac{1-\varphi_i^1}{1-\varphi_i}\right)\varepsilon_i(t) \\ &\quad + \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) c + \left(t - \frac{1-\varphi_i^t}{1-\varphi_i}\right)\mu_{C_i}, \end{aligned}$$

that is, the sum $C_i(1)+C_i(2)+\cdots+C_i(t)$ is a linear combination of $t-1$ independent random variables $\varepsilon_i(2), \varepsilon_i(3), \ldots, \varepsilon_i(t)$. Since a linear combination of independent normal random variables is still a normal random variable [6], the sum $C_i(1) + C_i(2) + \cdots + C_i(t)$ is normally distributed with parameters $\mu_i(t,c)$ and $\sigma_i^2(t)$, where

$$\mu_i(t,c) = \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) c + \left(t - \frac{1-\varphi_i^t}{1-\varphi_i}\right)\mu_{C_i}, \quad t \ge 1,$$

with $\mu_i(0,c) = 0$, and

$$\sigma_i^2(t) = \frac{1}{(1-\varphi_i)^2}\left(t - 2\left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) + \left(\frac{1-\varphi_i^{2t}}{1-\varphi_i^2}\right)\right)\sigma_{\varepsilon_i}^2, \quad t \ge 2,$$

with $\sigma_i(0) = \sigma_i(1) = 0$. Therefore, we have

$$\begin{aligned} &\boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t) < S] \\ &= \int_0^S \frac{1}{\sqrt{2\pi}\sigma_i(t)} e^{-(c-\mu_i(t,c))^2/(2\sigma_i^2(t))} dc. \end{aligned}$$

Since

$$\frac{C_i(1) + C_i(2) + \cdots + C_i(t) - \mu_i(t,c)}{\sigma_i(t)}$$

is a standard normal random variable with mean 0 and variance 1, the above probability can also be represented as

$$\boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t) < S] = \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right),$$

where $\Phi(x)$ is the cdf of a standard normal distribution.

By Theorem 1, the expectation of $T_i(S,c)$ is

$$\begin{aligned} \boldsymbol{E}(T_i(S,c)) &= \sum_{t=0}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right) \\ &= 2 + \sum_{t=2}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right), \end{aligned}$$

where we notice that $\Phi(+\infty) = 1$ when $t = 0, 1$. By randomizing $c$ in $T_i(S,c)$, we get

$$\begin{aligned} \boldsymbol{E}(T_i(S)) &= \int_0^\infty \boldsymbol{E}(T_i(S,c)) f_{C_i}(c)dc \\ &= \int_0^\infty \left(2 + \sum_{t=2}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right) f_{C_i}(c)dc \\ &= 2 + \int_0^\infty \left(\sum_{t=2}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right) f_{C_i}(c)dc \\ &= 2 + \int_0^\infty \left(\sum_{t=2}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right) dF_{C_i}(c) \\ &= 2 + F_{C_i}(c)\left(\sum_{t=2}^\infty \Phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right)\Bigg|_0^\infty \end{aligned}$$

$$+\int_0^\infty \left(F_{C_i}(c)\sum_{t=2}^{\infty}\frac{1}{\sigma_i(t)}\left(\frac{1-\varphi_i^t}{1-\varphi_i}\right) \times \phi\left(\frac{S-\mu_i(t,c)}{\sigma_i(t)}\right)\right)dc.$$

The theorem is proven by noticing that $F_{C_i}(0) = 0$ and $\Phi(-\infty) = 0$. ■

Theorem 3 provides two different ways to calculate $\boldsymbol{E}(T_i(S))$ by using numerical integration.

### 3.3. Normal distribution of $f_{C_i}(c)$

The expressions in Theorem 3 can be simplified if more information about $f_{C_i}(c)$ is available. For instance, consider the case when $C_i(1)$ is a normal random variable with parameters $\mu_{C_i}$ and $\sigma_{C_i}^2$, i.e.,

$$f_{C_i}(c) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}}e^{-(c-\mu_{C_i})^2/(2\sigma_{C_i}^2)},$$

where we assume that $\mu_{C_i}$ is reasonably large while $\sigma_{C_i}$ is reasonably small such that the distribution of $f_{C_i}(c)$ in $(-\infty, 0]$ is negligible. A normal distribution can characterize the up and down fluctuation of stochastic service capacities.

**Theorem 4.** *For a normal distribution of $f_{C_i}(c)$, we have*

$$\boldsymbol{E}(T_i(S)) = \sum_{t=0}^{\infty}\Phi\left(\frac{S-\mu_i(t)}{\sigma_i(t)}\right).$$

**Proof.** Since a linear combination of independent normal random variables is still a normal random variable, the sum $C_i(1)+C_i(2)+\cdots+C_i(t)$ in Theorem 2 is normally distributed with parameters $\mu_i(t)$ and $\sigma_i^2(t)$, where

$$\mu_i(t) = t\mu_{C_i},$$

and

$$\sigma_i^2(t) = \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right)^2\sigma_{C_i}^2 + \left(\left(\frac{1-\varphi_i^{t-1}}{1-\varphi_i}\right)^2 + \left(\frac{1-\varphi_i^{t-2}}{1-\varphi_i}\right)^2 + \cdots + \left(\frac{1-\varphi_i^1}{1-\varphi_i}\right)^2\right)\sigma_{\varepsilon_i}^2.$$

Therefore, we have

$$\boldsymbol{P}[C_i(1)+C_i(2)+\cdots+C_i(t) < S] = \int_0^S \frac{1}{\sqrt{2\pi}\sigma_i(t)}e^{-(c-\mu_i(t))^2/(2\sigma_i^2(t))}dc.$$

Since

$$\frac{C_i(1)+C_i(2)+\cdots+C_i(t)-\mu_i(t)}{\sigma_i(t)}$$

is a standard normal random variable with mean 0 and variance 1, the above probability can also be represented as

$$\boldsymbol{P}[C_i(1)+C_i(2)+\cdots+C_i(t) < S] = \Phi\left(\frac{S-\mu_i(t)}{\sigma_i(t)}\right),$$

where $\Phi(x)$ is the cdf of a standard normal distribution.

Based on the above discussion, we know that the pmf of $T_i(S)$ is

$$\boldsymbol{P}[T_i(S) = t] = \Phi\left(\frac{S-\mu_i(t-1)}{\sigma_i(t-1)}\right) - \Phi\left(\frac{S-\mu_i(t)}{\sigma_i(t)}\right),$$

for all $t = 1, 2, 3, \ldots$. The expectation of $T_i(S)$ is

$$\boldsymbol{E}(T_i(S)) = \sum_{t=0}^{\infty}\Phi\left(\frac{S-\mu_i(t)}{\sigma_i(t)}\right).$$

The theorem is proven. ■

Notice that $\sigma_i^2(t)$ is an increasing function of $\sigma_{C_i}^2$ and $\sigma_{\varepsilon_i}^2$. Furthermore, $\sigma_i^2(t)$ is an increasing function of $\varphi_i$ and very sensitive to $\varphi_i$, i.e., the level of correlation. When $\varphi_i = 0$, $\sigma_i^2(t)$ gets its minimum value

$$\sigma_i^2(t) = \sigma_{C_i}^2 + (t-1)\sigma_{\varepsilon_i}^2.$$

When $\varphi_i \to 1$, $\sigma_i^2(t)$ can be arbitrarily close to

$$\begin{aligned}\sigma_i^2(t) &\to t^2\sigma_{C_i}^2 + (1^2+2^2+\cdots+(t-1)^2)\sigma_{\varepsilon_i}^2 \\ &= t^2\sigma_{C_i}^2 + \left(\frac{t(t-1)(2t-1)}{6}\right)\sigma_{\varepsilon_i}^2.\end{aligned}$$

The impact of the noise process increases as $\varphi_i$ increases. Since

$$\begin{aligned}\sum_{k=1}^{t-1}(1-\varphi_i^k)^2 &= \sum_{k=0}^{t-1}(1-\varphi_i^k)^2 \\ &= \sum_{k=0}^{t-1}(1-2\varphi_i^k+(\varphi_i^k)^2) \\ &= t-2\sum_{k=0}^{t-1}\varphi_i^k + \sum_{k=0}^{t-1}(\varphi_i^2)^k \\ &= t-2\left(\frac{1-\varphi_i^t}{1-\varphi_i}\right)+\left(\frac{1-\varphi_i^{2t}}{1-\varphi_i^2}\right),\end{aligned}$$

we obtain

$$\sigma_i^2(t) = \left(\frac{1-\varphi_i^t}{1-\varphi_i}\right)^2\sigma_{C_i}^2 + \frac{1}{(1-\varphi_i)^2}\left(t-2\left(\frac{1-\varphi_i^t}{1-\varphi_i}\right)+\left(\frac{1-\varphi_i^{2t}}{1-\varphi_i^2}\right)\right)\sigma_{\varepsilon_i}^2.$$

Notice that for fixed $\sigma_{C_i}$, $\sigma_{\varepsilon_i}$, and $\varphi_i$, $\sigma_i^2(t) = \Theta(t)$ as $t \to \infty$, where $f(x) = \Theta(g(x))$ means that $f(x)$ and $g(x)$ have the same growth rate. For a fixed $\varphi_i$, we have $\mu_i(t)/\sigma_i(t) = \Theta(\sqrt{t}) \to \infty$ as $t \to \infty$, i.e., the summation in Theorem 4 converges. However, as $\varphi_i \to 1$, we have $\sigma_i^2(t) = \Theta(t^3)$ and $\mu_i(t)/\sigma_i(t) = \Theta(1/\sqrt{t}) \to 0$ as $t \to \infty$. Since $\Phi(0) = 0.5$, the summation in Theorem 4 does not converge and $\boldsymbol{E}(T_i(S)) = \infty$.

### 3.4. Uncorrelated service capacities

As an extreme case, when $\varphi_i = 0$ and $\varepsilon_i(t) = C_i(1) - \mu_{C_i}$, the stochastic process $C_i(1), C_i(2), C_i(3), \ldots$ is actually a sequence of i.i.d. random variables with the same pdf $f_{C_i}(c)$ in $[0, \infty)$, i.e., we have uncorrelated service capacities. Let $\mu_i = \mu_{C_i}$ and $\sigma_i^2 = \sigma_{C_i}^2 = \sigma_{\varepsilon_i}^2$. Then, the $C_i(t)$'s are normal random variables with parameters $\mu_i$ and $\sigma_i^2$, i.e.,

$$f_{C_i}(c) = \frac{1}{\sqrt{2\pi}\sigma_i}e^{-(c-\mu_i)^2/(2\sigma_i^2)}.$$

**Theorem 5.** *For uncorrelated service capacities, we have*

$$\boldsymbol{E}(T_i(S)) = \sum_{t=0}^{\infty}\Phi\left(\frac{S-t\mu_i}{\sqrt{t}\sigma_i}\right).$$

**Proof.** It is clear that $C_i(1) + C_i(2) + \cdots + C_i(t)$ is also normally distributed with parameters $t\mu_i$ and $t\sigma_i^2$. Hence, we have

$$\boldsymbol{P}[C_i(1) + C_i(2) + \cdots + C_i(t) < S] = \Phi\left(\frac{S - t\mu_i}{\sqrt{t}\sigma_i}\right).$$

The pmf of $T_i(S)$ is

$$\boldsymbol{P}[T_i(S) = t] = \Phi\left(\frac{S - (t-1)\mu_i}{\sqrt{t-1}\sigma_i}\right) - \Phi\left(\frac{S - t\mu_i}{\sqrt{t}\sigma_i}\right),$$

for all $t = 1, 2, 3, \ldots$. The expectation of $T_i(S)$ is

$$\boldsymbol{E}(T_i(S)) = \sum_{t=0}^{\infty} \Phi\left(\frac{S - t\mu_i}{\sqrt{t}\sigma_i}\right).$$

The theorem is proven. ■

## 4. Simulation results and numerical data

In this section, we present our simulation results and numerical data. As in most P2P file sharing and exchange systems, the file sizes $S$ are in the range 10–1500 MB [7]. The service capacity of a source peer is in the range 50–1000 kbps, i.e., 0.375–7.5 MB/min.

### 4.1. Simulation results

To validate our analytical results in the last section, we have conducted simulations. We consider two distributions of $f_{C_i}(c)$. One is a normal distribution with $\mu_{C_i} = 5$ and $\sigma_{C_i} = 0.5$. Another is a Pareto distribution (commonly used for modeling network traffic) with pdf

$$f_{C_i}(c) = \frac{\alpha\beta^\alpha}{c^{\alpha+1}},$$

in the range $[\beta, \infty)$, where $\alpha$ is the shape parameter and $\beta$ is the scale parameter. We set $\alpha = 10$ and $\beta = 4.5$, such that

$$\mu_{C_i} = \frac{\alpha\beta}{\alpha - 1} = 5,$$

and

$$\sigma_{C_i} = \frac{\beta}{\alpha - 1}\sqrt{\frac{\alpha}{\alpha - 2}} = \frac{\sqrt{5}}{4} \approx 0.559.$$

Assume that $\sigma_{\varepsilon_i} = 0.5$.

In Table 1, we display our simulation results of $\boldsymbol{E}(T_i(S))$ and compare them with analytical results. For each $S = 10, 20, \ldots, 100$, we show analytical results calculated by using Theorem 3 for the Pareto distribution and Theorem 4 for the normal distribution. We also show simulation results, where each value is the average of 10,000 experiments. The maximum 99% confidence interval is $\pm 0.57667\%$. The relative difference of each simulation result compared with the corresponding analytical result is also given. We observe that each simulation result is very close to its corresponding analytical result, with relative difference no more than $\pm 0.5\%$. This validates the correctness of our analytical results.

### 4.2. Numerical data

In this section, we present numerical data to show the expected file download time on a source peer $i$ which has time-varying service capacity. Assume that $f_{C_i}(c)$ has a normal distribution.

The major parameters in our analytical model are the mean of the service capacity $\mu_{C_i}$, the variance of the service capacity $\sigma_{C_i}^2$, the degree of correlation $\varphi_i$, the strength of noise $\sigma_{\varepsilon_i}$, and the file size $S$. In Fig. 1, we demonstrate the expected file download time $\boldsymbol{E}(T_i(S))$ for $5 \le S \le 100$ and $\varphi_i = 0.91, 0.95, 0.99$, where $\mu_{C_i} = 5$, $\sigma_{C_i} = 0.5$, and $\sigma_{\varepsilon_i} = 0.5$. It is observed that $\boldsymbol{E}(T_i(S))$ is almost a linear function of $S$. It is also observed that $\boldsymbol{E}(T_i(S)) > S/\mu_{C_i}$. Furthermore, the difference between $\boldsymbol{E}(T_i(S))$ and $S/\mu_{C_i}$ gets more significant as $\varphi_i$ increases, i.e., increased correlation between $C_i(t-1)$ and $C_i(t)$ increases the expected file download time and reduces the performance of a source peer.

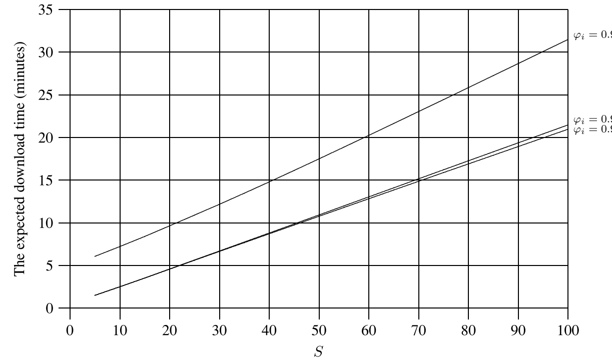


**Fig. 1.** The expected download time vs. file size (varying $\varphi_i$).

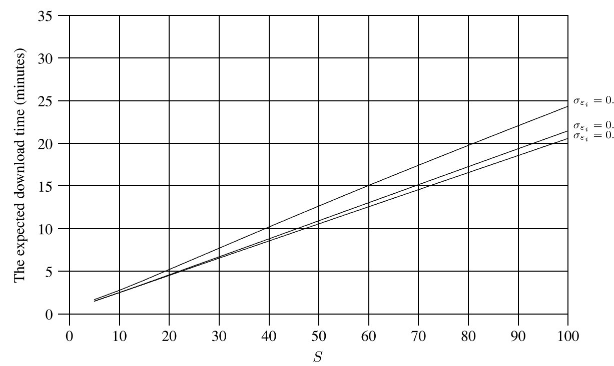


**Fig. 2.** The expected download time vs. file size (varying $\sigma_{\varepsilon_i}$).

In Fig. 2, we demonstrate the expected file download time $\boldsymbol{E}(T_i(S))$ for $5 \le S \le 100$ and $\sigma_{\varepsilon_i} = 0.1, 0.5, 0.9$, where $\mu_{C_i} = 5$, $\sigma_{C_i} = 0.5$, and $\varphi_i = 0.95$. It is observed that $\boldsymbol{E}(T_i(S))$ is almost a linear function of $S$. It is also observed that $\boldsymbol{E}(T_i(S)) > S/\mu_{C_i}$. Furthermore, the difference between $\boldsymbol{E}(T_i(S))$ and $S/\mu_{C_i}$ gets more significant as $\sigma_{\varepsilon_i}$ increases, i.e., increased noise increases the expected file download time and reduces the performance of a source peer.

The impact of $\varphi_i$ is further demonstrated in Fig. 3, where we show the expected file download time $\boldsymbol{E}(T_i(S))$ for $0.90 \le \varphi_i \le 0.99$ and $S = 20, 60, 100$, with $\mu_{C_i} = 5$, $\sigma_{C_i} = 0.5$, and $\sigma_{\varepsilon_i} = 0.5$. It is observed that as $\varphi_i$ increases, the expected download time also increases smoothly and slowly. Beyond 0.98, the expected download time increases dramatically, i.e., strong correlation degrades the performance of a source peer significantly.

The impact of $\sigma_{\varepsilon_i}$ is further demonstrated in Fig. 4, where we show the expected file download time $\boldsymbol{E}(T_i(S))$ for $0.1 \le \sigma_{\varepsilon_i} \le 0.9$ and $S = 20, 60, 100$, with $\mu_{C_i} = 5$, $\sigma_{C_i} = 0.5$, and $\varphi_i = 0.95$. It is observed that for reasonable noise level, as $\sigma_{\varepsilon_i}$ increases, the expected download time also increases smoothly and slowly.

## 5. Parallel download and chunk allocation

A file can be downloaded from $r$ source peers $1, 2, \ldots, r$ simultaneously. It is assumed that all source peers are stable and remain in a P2P network for significant amount of time. There is no effect of peer churn [8] for downloading the file of interest, i.e., all source peers are available during downloading of the file.

**Table 1**
Comparison of analytical and simulation results of $\boldsymbol{E}(T_i(S))$.

| $S$ | Normal distribution | | | Pareto distribution | | |
|---|---|---|---|---|---|---|
| | Analytical | Simulation | Relative differ. | Analytical | Simulation | Relative differ. |
| 10 | 2.50284 | 2.50930 | +0.25794% | 2.60386 | 2.59930 | −0.17507% |
| 20 | 4.58560 | 4.59390 | +0.18093% | 4.60198 | 4.59260 | −0.20374% |
| 30 | 6.69421 | 6.72100 | +0.40019% | 6.70408 | 6.69760 | −0.09668% |
| 40 | 8.80453 | 8.79400 | −0.11957% | 8.82034 | 8.80200 | −0.20788% |
| 50 | 10.92369 | 10.93020 | +0.05957% | 10.94329 | 10.94590 | +0.02384% |
| 60 | 13.04310 | 13.05000 | +0.05289% | 13.06634 | 13.05590 | −0.07988% |
| 70 | 15.15922 | 15.08620 | −0.48168% | 15.18546 | 15.19010 | +0.03052% |
| 80 | 17.27169 | 17.21880 | −0.30621% | 17.29847 | 17.37570 | +0.44647% |
| 90 | 19.37494 | 19.35900 | −0.08229% | 19.40427 | 19.34020 | −0.33016% |
| 100 | 21.47610 | 21.46150 | −0.06798% | 21.50247 | 21.42540 | −0.35841% |

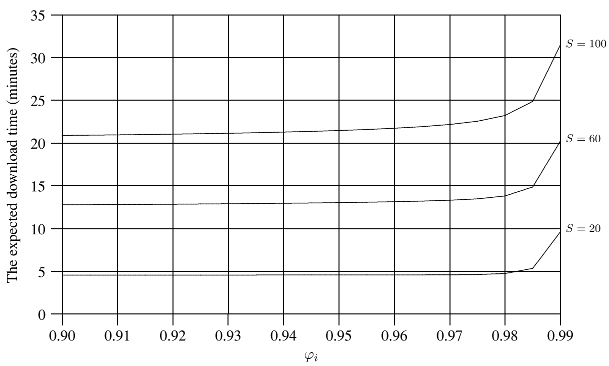

**Fig. 3.** The expected download time vs. degree of correlation (varying $S$).

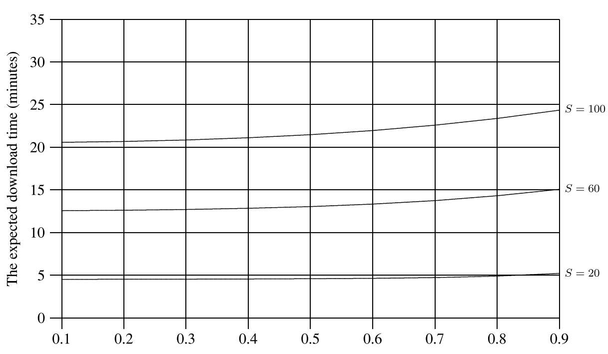

**Fig. 4.** The expected download time vs. noise strength (varying $S$).

Moreover, all the $r$ source peers are seed peers, i.e., they all hold a complete copy of a file, such that any chunk of the file can be obtained from any source peer.

In a parallel download, a file of size $S$ is divided into $r$ chunks of sizes $S_1, S_2, \ldots, S_r$, such that chunk $S_i$ is downloaded from source peer $i$, for all $1 \le i \le r$ in parallel. Different algorithms have different strategies in choosing the chunk sizes and have different parallel download times.

The following theorem analyzes the expected parallel download time.

**Theorem 6.** *If $f_{C_i}(c)$ is a normal distribution with parameters $\mu_{C_i}$ and $\sigma^2_{C_i}$, the expected parallel download time is*

$$\boldsymbol{E}(T_{\mathbf{PD}}(S,r)) = \sum_{t=0}^{\infty}\left(1 - \prod_{i=1}^{r}\left(1 - \Phi\left(\frac{S_i - \mu_i(t)}{\sigma_i(t)}\right)\right)\right).$$

**Proof.** The parallel download time for a file of size $S$ from $r$ source peers is

$$T_{\mathbf{PD}}(S,r) = \max\{T_1(S_1), T_2(S_2), \ldots, T_r(S_r)\}.$$

Notice that

$$\begin{aligned}\boldsymbol{P}[T_i(S_i) < t] &= 1 - \boldsymbol{P}[T_i(S_i) = t] - \boldsymbol{P}[T_i(S_i) > t] \\ &= 1 - (\boldsymbol{P}[T_i(S_i) > t-1] - \boldsymbol{P}[T_i(S_i) > t]) \\ &\quad - \boldsymbol{P}[T_i(S_i) > t] \\ &= 1 - \boldsymbol{P}[T_i(S_i) > t-1].\end{aligned}$$

Therefore, for a normal distribution of $f_{C_i}(c)$, we have

$$\begin{aligned}\boldsymbol{P}[T_{\mathbf{PD}}(S,r) < t] &= \prod_{i=1}^{r}\boldsymbol{P}[T_i(S_i) < t] \\ &= \prod_{i=1}^{r}(1 - \boldsymbol{P}[T_i(S_i) > t-1]) \\ &= \prod_{i=1}^{r}(1 - \boldsymbol{P}[C_i(1) + C_i(2) + \cdots \\ &\quad + C_i(t-1) < S_i]) \\ &= \prod_{i=1}^{r}\left(1 - \Phi\left(\frac{S_i - \mu_i(t-1)}{\sigma_i(t-1)}\right)\right) \\ &\quad \text{(by Theorem 4)}.\end{aligned}$$

The expected parallel download time is

$$\begin{aligned}\boldsymbol{E}(T_{\mathbf{PD}}(S,r)) &= \sum_{t=1}^{\infty}\boldsymbol{P}[T_{\mathbf{PD}}(S,r) \ge t] \\ &= \sum_{t=1}^{\infty}(1 - \boldsymbol{P}[T_{\mathbf{PD}}(S,r) < t]) \\ &= \sum_{t=1}^{\infty}\left(1 - \prod_{i=1}^{r}\left(1 - \Phi\left(\frac{S_i - \mu_i(t-1)}{\sigma_i(t-1)}\right)\right)\right) \\ &= \sum_{t=0}^{\infty}\left(1 - \prod_{i=1}^{r}\left(1 - \Phi\left(\frac{S_i - \mu_i(t)}{\sigma_i(t)}\right)\right)\right).\end{aligned}$$

The theorem is proven. ■

Our main problem here is to find chunk sizes $S_1, S_2, \ldots, S_r$, such that the expected parallel download time is minimized. It turns out that this is an extremely complicated multi-variable optimization problem and hard to solve. Instead, we will consider two heuristic solutions.

### *5.1. Chunk allocation algorithms*

We consider two parallel download and chunk allocation algorithms. In the naive parallel download algorithm $\mathbf{PD}_0$, a file of size $S$ is divided into $r$ chunks of equal size $S_1 = S_2 = \cdots = S_r = S/r$ [1].

Algorithm $\mathbf{PD}_1$ for parallel download and chunk allocation works as follows. Instead of dividing a file into chunks of equal

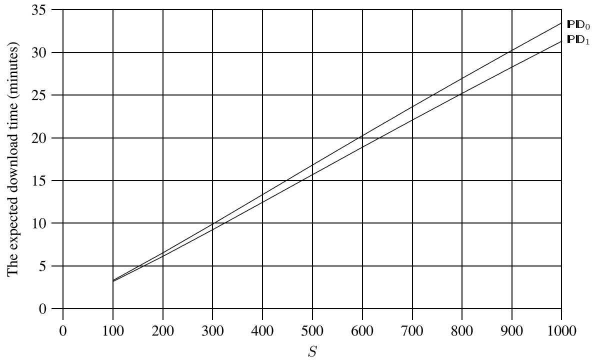


**Fig. 5.** The expected parallel download time vs. file size.

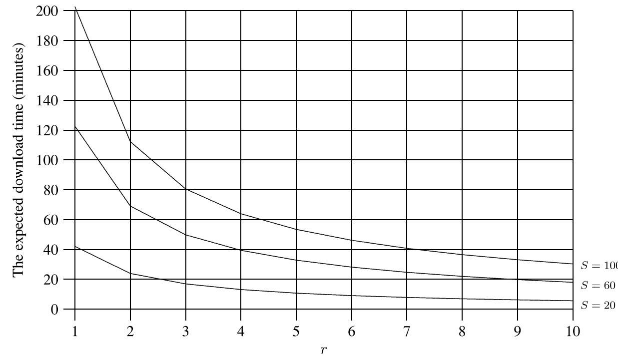


**Fig. 6.** The expected parallel download time vs. parallelism (varying $S$).

sizes, algorithm $\mathbf{PD}_1$ divides a file of size $S$ into chunks of sizes $S_1, S_2, \ldots, S_r$, such that chunk sizes are proportional to the expected service capacities, that is,

$$S_i = \left(\frac{\mu_{C_i}}{\mu_{C_1} + \mu_{C_2} + \cdots + \mu_{C_r}}\right) S,$$

for all $1 \le i \le r$. Algorithm $\mathbf{PD}_1$ has no knowledge of the current service capacities of the source peers. However, algorithm $\mathbf{PD}_1$ attempts to do chunk allocation based on the expected behavior of source peers.

### 5.2. Performance comparison

Let us consider a P2P file sharing system with $r = 10$ heterogeneous source peers. Assume that $C_i$ has time-varying service capacity, where the mean of the service capacity is $\mu_{C_i} = 4.1 + 0.2(i-1)$, and the variance of the service capacity is $\sigma_{C_i} = 0.5 + 0.05(i-1)$, and the degree of correlation is $\varphi_i = 0.95$, and the strength of noise is $\sigma_{\varepsilon_i} = 0.5$.

In Fig. 5, we demonstrate the expected parallel download time $\boldsymbol{E}(T_{\mathbf{PD}}(S, r))$ of algorithms $\mathbf{PD}_0$ and $\mathbf{PD}_1$ for file size $100 \le S \le 1000$. We observe that algorithm $\mathbf{PD}_1$ has better performance than $\mathbf{PD}_0$ due to a better chunk allocation strategy.

### 5.3. The impact of parallelism

We say that the $r$ source peers $1, 2, \ldots, r$ are homogeneous if their initial service capacities $C_1(1), C_2(1), \ldots, C_r(1)$ are identical random variables $C$ with the same pdf,

$$f_{C_1(1)}(c) = f_{C_2(1)}(c) = \cdots = f_{C_r(1)}(c) = f_C(c),$$

and their noise processes have the same pdf,

$$f_{\varepsilon_1}(c) = f_{\varepsilon_2}(c) = \cdots = f_{\varepsilon_r}(c) = f_\varepsilon(c),$$

and they have the same degree of correlation,

$$\varphi_1 = \varphi_2 = \cdots = \varphi_r = \varphi.$$

Notice that this does not mean that the $r$ source peers have the same service capacity. In fact, during any time slot, the service capacities of the $r$ source peers can be entirely and radically different as governed by $f_C(c)$. Even if they have the same initial service capacity, their subsequent service capacities can still change as governed by $f_\varepsilon(c)$.

In Fig. 6, we demonstrate the impact of parallelism for homogeneous source peers. Assume that $C$ is a normal random variable with parameters $\mu_C = 5$ and $\sigma_C = 0.5$. The common noise process has a normal distribution with mean 0 and $\sigma_\varepsilon = 0.5$. The degree of correlation is $\varphi = 0.95$. We demonstrate the expected parallel download time $\boldsymbol{E}(T_{\mathbf{PD}}(S, r))$ (notice that algorithms $\mathbf{PD}_0$ and $\mathbf{PD}_1$ are identical for homogeneous source peers) for $1 \le r \le 10$ and $S = 20, 60, 100$. We observe that although $r$ source peers do not reduce the expected parallel download time by a factor of $r$, multiple source peers do reduce the parallel download time significantly.

## 6. Related work

Extensive investigation has been performed by many researchers in the last few years for performance measurement, modeling, analysis, and optimization of file sharing in P2P networks. In this section, we review the related research.

Research in this area has been conducted at three different levels, i.e., system level, peer group level, and individual peer level. At the system level, research is focused on establishing models of P2P networks such as queueing models [9,10] and fluid models [11], so that overall system characterizations such as system throughput and average file download time can be obtained. At the peer group level, research is focused on distributing a file from a set of source peers to a set of user peers, so that the overall distribution time is minimized [12–17]. At the individual peer level, research is focused on analyzing and minimizing the file download time of a single peer [1,2,4].

It is clear that the vast majority of file downloads are performed by individual users. Therefore, P2P network performance optimization from a single peer's point of view has been an interesting and important issue. File download strategies for an individual user peer can be classified into two categories, namely, single download methods from one source peer and parallel download methods from several source peers simultaneously. The main concern in single download methods is the peer selection problem, namely, switching among source peers and finally settling on one, while keeping the total time of probing and downloading to a minimum [18–22,2,4].

It is well known that the method of parallel downloading can be used to reduce file download times. The main concern in parallel download methods is the chunk allocation problem, namely, how to divide a file to be downloaded into chunks which can be downloaded from several source peers simultaneously. In [1], it is proposed that a file is divided into chunks of equal sizes. In [23], it is observed that to achieve the maximum speedup, chunks should be allocated such that all servers finish their transmissions at the same time. It has been observed that performance improvement experienced by clients who perform parallel downloading comes at the expense of clients who simply go to a single server to retrieve files [24].

Performance measurement, modeling, analysis, and optimization of parallel document downloading in the Internet and file sharing in P2P networks have also been conducted at three different levels, i.e., system level, peer/client group level, and individual peer/client level. At the system level, research is focused on understanding the impact of large scale parallel downloading on the performance of a network [24–27]. At the peer/client group level, research is focused on parallel document/file downloading from multiple mirror sites and source peers such that duplicated transmissions are kept to a minimum by using efficient multicasting [28]. At the individual peer/client level, research is focused on minimizing the parallel file download time for a single peer/client [1,23]. Fine parallel downloading algorithms for an individual user peer are critical in competing for network resources. A comprehensive and analytical performance study of parallel download algorithms is given in [3] for source peers with random service capacities.

## 7. Conclusions

We have analyzed the expected file download time in peer-to-peer networks with stochastic and time-varying service capacities. We treated the service capacity of a source peer as a stochastic process. For the autoregressive model of order 1, we derived analytical results which characterize the expected download time from a source peer with stochastic and time-varying service capacity. We presented simulation results to validate our analytical results. We also gave numerical data to show the impact of the degree of correlation and the strength of noise on the expected file download time.

We have also considered the expected time of a parallel download. For any chunk allocation method, we derived an analytical result of the expected parallel download time from a source peer with stochastic and time-varying service capacity. We showed that the algorithm which chooses chunk sizes proportional to the expected service capacities has better performance than the algorithm which chooses equal chunk sizes. We also showed that multiple source peers do reduce the parallel download time significantly.

## Acknowledgments

The author would like to express his gratitude to the anonymous reviewers for their criticism and comments on the paper.

## References

[1] Y.-M. Chiu, D.Y. Eun, Minimizing file download time in stochastic peer-to-peer networks, IEEE/ACM Trans. Netw. 16 (2) (2008) 253–266.
[2] K. Li, Probing high-capacity peers to reduce download times in P2P file sharing systems with stochastic service capacities, Internat. J. Found Comput. Sci. 23 (6) (2012) 1341–1369.
[3] K. Li, Parallel file download in peer-to-peer networks with random service capacities, in: Proceedings of the 27th IEEE International Parallel and Distributed Processing Symposium Workshops, 15th Workshop on Advances in Parallel and Distributed Computational Models, Boston, Massachusetts, May 20–24, 2013, pp. 677–686.
[4] K. Li, On the expected file download time of the random time-based switching algorithm in P2P networks, Peer-to-Peer Netw. Appl. 7 (2) (2014) 147–158.
[5] http://en.wikipedia.org/wiki/Autoregressive.
[6] http://en.wikipedia.org/wiki/Normal_distribution.
[7] http://blogplots.blogspot.com/2008/02/p2p-file-size-distribution.html.
[8] D. Stutzbach, R. Rejaie, Understanding churn in peer-to-peer networks, in: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, 2006.
[9] Z. Ge, D.R. Figueiredo, S. Jaiswal, J. Kurose, D. Towsley, Modeling peer-peer file sharing systems, in: Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 3, 2003, pp. 2188–2198.
[10] S. Tewari, L. Kleinrock, On fairness, optimal download performance and proportional replication in peer-to-peer networks, in: Proceedings of the 4th International IFIP-TC6 Networking Conference, in: LNCS, vol. 3462, 2005, pp. 709–717.
[11] R. Gaeta, M. Gribaudo, D. Manini, M. Sereno, Analysis of resource transfers in peer-to-peer file sharing applications using fluid models, Perform. Eval. 63 (2006) 149–174.
[12] S.G.M. Koo, K. Kannan, C.S.G. Lee, On neighbor-selection strategy in hybrid peer-to-peer networks, Future Gener. Comput. Syst. 22 (2006) 732–741.
[13] R. Kumar, K.W. Ross, Peer-assisted file distribution: the minimum distribution time, in: Proceedings of the IEEE Workshop on Hot Topics in Web Systems and Technologies, 2006.
[14] M. Lingjun, K.-S. Lui, Scheduling in P2P file distribution—on reducing the average distribution time, in: Proceedings of the 5th IEEE Consumer Communications and Networking Conference, 2008, pp. 521–522.
[15] M. Lingjun, W. Xiaolei, K.-S. Lui, A novel peer grouping scheme for P2P file distribution networks, in: Proceedings of the IEEE International Conference on Communications, 2008, pp. 5598–5602.
[16] M. Teo, C. Carbunaru, B. Leong, Y. Nataraj, H.M.L. Vu, R. Tan, Y.M. Teo, Achieving high-bandwidth peer-to-peer file distribution, in: Proceedings of the 4th ACM International Conference on Emerging Networking Experiments and Technologies, 2008.
[17] X. Zheng, C. Cho, Y. Xia, Optimal peer-to-peer technique for massive content distribution, in: Proceedings of the 27th IEEE Conference on Computer Communications, 2008, pp. 151–155.
[18] M. Adler, R. Kumar, K. Ross, D. Rubenstein, T. Suel, D.D. Yao, Optimal peer selection for P2P downloading and streaming, in: Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 3, 2005, pp. 1538–1549.
[19] M. Adler, R. Kumar, K. Ross, D. Rubenstein, D. Turner, D.D. Yao, Optimal peer selection in a free-market peer-resource economy, in: Proceedings of the 2nd Workshop on the Economics of Peer-to-Peer Systems, 2004.
[20] D.S. Bernstein, Z. Feng, B.N. Levine, S. Zilberstein, Adaptive peer selection, in: Proceedings of the 2nd International Workshop on Peer-to-Peer Systems, 2003.
[21] R.L. Carter, M.E. Crovella, On the network impact of dynamic server selection, Comput. Netw. 31 (23–24) (1999) 2529–2558.
[22] S.G. Dykes, K.A. Robbins, C.L. Jeffery, An empirical evaluation of client-side server selection algorithms, in: Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 3, 2000, pp. 1361–1370.
[23] P. Rodriguez, E.W. Biersack, Dynamic parallel access to replicated content in the Internet, IEEE/ACM Trans. Netw. 10 (4) (2002) 455–465.
[24] C. Gkantsidis, M. Ammar, E. Zegura, On the effect of large-scale deployment of parallel downloading, in: Proceedings of the 3rd IEEE Workshop on Internet Applications, 2003, pp. 79–89.
[25] S.G.M. Koo, C. Rosenberg, D. Xu, Analysis of parallel downloading for large file distribution, in: Proceedings of the 9th IEEE Workshop on Future Trends of Distributed Computing Systems, 2003, pp. 128–135.
[26] Y. Liu, W. Gong, P. Shenoy, On the impact of concurrent downloads, in: Proceedings of the 33nd Winter Simulation Conference, 2001, pp. 1300–1305.
[27] D. Manini, M. Gribaudo, Modelling search, availability, and parallel download in P2P file sharing applications with fluid model, in: Proceedings of 14th International Conference on Advanced Computing and Communications, 2006, pp. 449–454.
[28] J.W. Byers, M. Luby, M. Mitzenmacher, Accessing multiple mirror sites in parallel: using Tornado codes to speed up downloads, in: Proceedings of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 1, 1999, pp. 275–283.

**Keqin Li** is a SUNY distinguished professor of computer science. His current research interests include parallel computing, distributed computing, energy-efficient computing and communication, heterogeneous computing systems, wireless communication, sensor networks, peer-to-peer file systems, cloud computing, big data computing, storage and file systems, CPU–GPU hybrid and cooperative computing, multicore computing, service computing, Internet of things and cyber–physical systems. He has published over 300 journal articles, book chapters, and research papers in refereed international conference proceedings. He is currently or has served on the editorial board of *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Computers*, *IEEE Transactions on Cloud Computing*.