Simulation Modelling Practice and Theory 18 (2010) 1214–1224

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

BitTorrent packet traffic features over IPv6 and IPv4

Cebrail Çiflikli [a], Ali Gezer [a,*], A. Tuncay Özşahin [a], Öznur Özkasap [b]

[a] Erciyes University, Kayseri Vocational High School, 38020 Kayseri, Turkey
[b] Koc University, Department of Computer Engineering, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 10 December 2009
Received in revised form 12 April 2010
Accepted 26 April 2010
Available online 18 May 2010

Keywords:
BitTorrent
Peer-to-peer traffic
IPv6
IPv4
Traffic characteristics

ABSTRACT

At present, BitTorrent protocol packets constitute a large part of peer-to-peer application traffic on the Internet. Due to the increasing amount of BitTorrent traffic, it has become inevitable to take into account its effects on network management. Generally, studies on BitTorrent traffic measurement have involved analysis with packets transmitted via IPv4 protocol. However, with several facilities provided by IPv6 protocol, its traffic volume in operational networks is increasing day by day. New features of IPv6 enhance packet processing speeds over routers, switches and end systems. We consider that traffic features and packet traffic characteristics are likely to be affected with increasing amount of IPv6 protocol traffic. Therefore, it becomes significant to explore IPv6 packet traffic characteristics and application traffic features over IPv6. In this study, we investigate the IPv6 BitTorrent packet traffic characteristics in terms of autocorrelation, power spectral density and self similarity of packet size and packet interarrival time. We also perform distribution modeling for IPv4 and IPv6 BitTorrent packet traffic. With these models, efficient packet traffic traces are generated for network simulation studies. A detailed comparison is performed to determine differences between IPv4 and IPv6 BitTorrent packet traffic.

1. Introduction

Peer-to-peer (P2P) protocols, applications and their resulting network traffic have formed a significant part of the Internet [1–5]. A study performed in five different regions between August and September 2007 by IPOQUE [6] demonstrates that P2P file sharing produces majority of the Internet traffic, and its share varies between 48% (in the Middle East) and 80% (in Eastern Europe). P2P users cooperate through an overlay network that allows applications to share resources by acting as both clients and servers [7]. This approach was first popularized by the Napster system mostly for sharing music, video, and software files. Different download and file sharing strategies are adopted by several P2P applications. Some of them such as Napster use a centralized server to index files. Gnutella uses a fully distributed approach where queries are flooded to neighboring peers [8]. Some P2P applications such as BitTorrent, Gnutella 2, and Kazaa integrate centralized and distributed mechanisms. Supernodes in Kazaa, ultrapeers in Gnutella 2, and trackers in BitTorrent are responsible for handling index files for peers.

BitTorrent (BT) has been the most popular P2P file sharing protocol attracting millions of users since its introduction in 2001 [9–13]. BitTorrent system scales fairly well and is now widely used for various purposes, such as data distribution [15], media streaming, media on demand [16], and even to launch DDos attacks [17]. In recent years, there is a significant increase in P2P traffic amount, in particular, due to the popularity of BitTorrent protocol [13,14].

* Corresponding author. Tel.: +90 505 7795686.
E-mail addresses: cebrailc@erciyes.edu.tr (C. Çiflikli), aligezerr@hotmail.com (A. Gezer), atozsahin@erciyes.edu.tr (A.T. Özşahin), oozkasap@ku.edu.tr (Ö. Özkasap).

1569-190X/$ - see front matter
doi:10.1016/j.simpat.2010.04.009

For efficient resource utilization in networks, the nature of the traffic must be determined accurately. Due to the increasing amount of BT traffic on backbones, determination of its traffic characteristic would be invaluable. Most of the studies related to BT traffic are based on IPv4 protocol. With the increasing volume of IPv6 traffic in total, the characteristics of BT traffic via IPv6 must be determined elaborately. In this study, we investigate the characteristics of IPv6 BT packet traffic and the differences between IPv6 and IPv4 BT traffic in terms of spectral density, autocorrelation, distribution and self similarity of packet interarrival time and packet size.

BT divides a single large file into small pieces. Peers query a centralized server to retrieve a peer list, and can simultaneously download different pieces of the file from multiple peers. Several studies have been performed to model BT networks, BT peers, and BT traffic. For instance, modeling of BT-like P2P networks was studied in [18], and an analytical model of a BT peer was proposed in [19]. Furthermore, many studies have been conducted to investigate and model BitTorrent traffic. In [20], BT session characteristics and message characteristics were modeled using distributions. The self-similarity of BT traffic was estimated using variance time, Rescaled Range (R/S), and a periodogram estimator [21]. According to that study, the packet size of BT traffic gives Hurst degrees of 0.82, 0.77, and 0.83, respectively. Generally, in measurement studies of BT traffic, packets are transmitted over the IPv4 internet protocol.

Due to the restrictions of IPv4, it was essential to design a new internet protocol. Therefore, IPv6 was proposed in the middle of the 1990s, providing solutions to a number of shortcomings. 128 bit IP address size of the IPv6 protocol has solved the problem of limited address space of IPv4. This variation brings an advantage of a much greater number of addressable nodes on the Internet. IPv6 provides a simpler and more extensible header structure. Some IPv4 header fields have been dropped or made optional to reduce the common-case processing cost of packet handling and to limit the bandwidth cost of IPv6 headers [22]. Flow labeling capability is added to address the packets to a particular traffic flow with the new internet protocol. Despite the many facilities provided with the IPv6 protocol, its widespread adoption has been delayed for many reasons. It also seems unlikely IPv4 will be abandoned in the very near future. Nowadays, the Internet is in transition from IPv4 to IPv6. Transition from one to the other is widely achieved with IPv6-over-IPv4 tunnels. IPv6 traffic load in total is increasing day by day with the provided facilities such as tunneling and pilot programs. Many internet service providers use IPv6 pilot programs to provide IPv6 connectivity to their customers. Free IPv6 tunnel brokers make it possible for anyone to obtain an IPv6 connection [23]. The growth in facilities to provide IPv6 internet connections results in a traffic load much greater than that in the near past.

In this study, our aim is to make a characterization of IPv6 BT packet traffic. Our work comprises analysis of self-similarity, autocorrelation, and power spectral density for packet interarrival time and packet size. Moreover, detailed analyses are performed to determine characteristic differences between IPv6 and IPv4 BT packet traffic in terms of the statistics mentioned. Network traffic exhibits non-deterministic behavior. Therefore, distributions are more suitable to characterize packet traffic of computer networks. We also perform distribution modeling for IPv6 and IPv4 BT packets it terms of packet interarrival time and packet size. With correct distribution models, efficient packet traffic would be generated for network simulation studies.

This paper is organized as follows. Background information about BT protocol, IPv4 and IPv6 are presented in Section 2. Measurement details of inspected Internet traces and analysis perspective are provided in the next section. Packet interarrival time, packet size comparisons of IPv6 and IPv4 traffic in terms of autocorrelation, power spectral density, and cumulative distribution function are described in Section 4 together with self-similarity analysis and distribution modeling of the obtained time series. Finally, Section 5 states our conclusions.

2. Background

2.1. BitTorrent protocol

BitTorrent is a P2P content distribution system comprised of a set of network protocols for realizing communication among the participating peers. The idea behind BT is organizing the peers in such a way that the load of peer is distributed to the entire system [24]. It organizes peers into an overlay network to distribute files. Peers can connect several other peers simultaneously and download different blocks of the file in parallel.

For sharing files via BT, at first a torrent file is created which specifies the tracker and describes how the file is partitioned into small blocks. The tracker is the server responsible for keeping track of the registered peers and helps peers find each other. The initial distributer posts the .torrent file to the tracker. Distributed files are divided into small pieces and pieces are subdivided into smaller data units called blocks, typically 16 KB in size. A block is a data exchange unit in BT. To download a file, a peer first downloads the .torrent file associated to the file of interest, then contacts the tracker to obtain the list of peers sharing the file. Peers use this subset to connect to other peers to exchange the pieces of the file. The peer then requests blocks of the file.

BT employs rarest first policy to select pieces for downloading. Each peer tries to download the least replicated pieces among its neighbors. Therefore, the distribution of the file among the peers could be realized faster. A peer can only upload to a limited number of peers due to the choking procedure. Choking is the temporary refusal of uploading to some neighbors. If permission is given, unchoke message is sent. Choked neighbors are chosen according to tit-for-tat policy. In particular a peer uploads to the peers providing the best download rates. BT employs a tit-for-tat policy to penalize free-riding [24].

There are two types of peers in BT, namely seeds and leechers. Seeds have the entire file while leechers have a part of the file. Blocks can be downloaded directly from seeds or leechers. When a new peer joins the swarm, it contacts the trackers to obtain a random list of active peers, then it tries to establish connections to these peers and finds out what pieces are there. Once it has received an unchoke message from a neighbor peer, it requests pieces and start downloading.

In this study, we investigate the packet level characteristics of BT traffic over IPv6 and IPv4. Using a packet sniffer application, BT packets are filtered from real IPv6 and IPv4 traffic traces. Port based filtering is used to obtain BT packets. Due to the increasing amount of BT traffic on the Internet, we determine BT packet traffic characteristics over IPv6 and perform a comparison with IPv4.

2.2. IPv4 versus IPv6

IPv6 (i.e. next generation IP) [25–27] is the recent version of the Internet protocol that provides a number of enhancements over the existing IPv4 [28]. The most significant problem with the IPv4 is its limited addressing capacity. IPv4 can address 2^{32} nodes on the Internet due to 32 bits addressing structure. IPv6 increases the address length from 32 bits to 128 bits to enhance addressing capability. A new type of address, namely anycast, is introduced and used to send a packet to any one of a group of nodes. IPv4 header varies between 20 octets and 60 octets. It consists mandatory 20 octets and maximum 60 octets option and padding fields. On the other hand IPv6 consists mandatory 40 octets for IPv6 header. While IPv4 header consists of 13 different fields, IPv6 header only consists of 8 fields. Packet fragmentation and reassembly fields have been dropped from IPv6 packet header. Packet fragmentation and reassembly are performed only by source and destination nodes respectively. Therefore, in IPv6 protocol, packet processing speeds increase in intermediate routers and header format simplification helps decreasing the bandwidth cost. On the other hand, IPv4 datagrams must be fragmented according to link and node maximum packet size by border routers.

Flow labeling capability is another significant property offered by IPv6. It labels the packets which request special handling by the intermediate routers. A flow is identified by the combination of a source address, destination address and flow label. Thus, all packets to be part of the same flow are assigned the same flow label by the source. Hop limit shows the remaining number of allowable hops for a packet in its routing. It is decremented by 1 at each node that forwards the packet. Time to Live field in IPv6 perform the similar job but it works via time.

Extension headers are used to enhance packet processing capability in the IPv6. Hop by hop options header, routing header, fragment header, authentication header, encapsulating security payload header and destination option header are extension headers used in IPv6. Each header includes a next header field except the encapsulating security payload header. Next header identifies the type immediately following the header. It may be the identifier of another extension header or an upper layer protocol header such as TCP or UDP.

Novel properties of IPv6 enhances packet processing speeds over routers, switches and end systems. We consider that traffic features and packet traffic characteristics are also affected with increasing amount of IPv6 protocol traffic. Therefore, our study focuses on a detailed investigation of IPv6 packet traffic characteristics and application traffic features over IPv6, as well as its comparison with IPv4 traffic.

3. Measurement details and perspective

3.1. Internet traces

We use the MAWI Working Group traffic archive [29] and analyze packet-level internet traces for 18 May – 22 June 2008 on an IPv6 line connected to a WIDE-6Bone. The traffic is logged in tcpdump format. All traces consist of different kinds of application traffic including P2P, HTTP, FTP, and SMTP. A packet sniffer application is used to filter BT packets to analyze packet-level traffic traces and extract information for IPv6 BT packet traffic. Port-based filtering is used to obtain BT packets. Each day trace comprises 2 million packets and nearly 4–8 thousand IPv6 BT packets are obtained for every day of traffic. Nearly 2 million packets are captured in 8 h every day. Capturing starts at 6 p.m. each day and lasts about 8 h for every day trace. To determine differences between IPv4 and IPv6 BT traffic, we also present the traffic characteristics of IPv4 BT traffic. For IPv4 traffic, we analyze the traffic traces between 1 and 31 August 2008 on a 150 Mbps transpacific link between the US and Japan from the MAWI Working Group traffic archive [29]. Each trace covers 15 minutes of traffic for every day. Capturing starts at 2 p.m. and finishes at 2:15 p.m. Every day 3–8 thousand IPv4 BT packets are obtained. Comparisons are performed in terms of packet size and packet interarrival time.

3.2. Analysis perspective

Simulations and network traffic modeling studies must be based on true traffic models to give accurate performance estimation results. Most of the network traffic types have shown self-similarity in terms of packet interarrival time and packet size. Spectral density, autocorrelation, and cumulative distribution analyses give significant details about the self-similar behavior of network traffic. Many self-similarity estimation methods take into consideration the results of these analyses. We perform comparative power spectral density analysis for IPv6 and IPv4 BT traffic to observe 1/f spectrum behavior. Recent studies have demonstrated that most network traffic types show long-range dependency. Autocorrelation does not decay so quickly with increasing time lag. We perform detailed analyses to compare IPv6 and IPv4 BT packet traffic in terms of autocorrelations.

Another significant field in network traffic characterization is determination of distribution models. According to previous studies, several network traffic types show heavy tailed behavior in their distributions. In order to simulate self-similar traffic, distribution functions that represent heavy tailed behavior are used. To compare IPv6 and IPv4 BT traffic in terms of distributions, distribution fitting is performed for IPv4 and IPv6 BT packet interarrival time and packet size.

4. Analysis results

4.1. Time series analysis

Time series plots are given for two day traces of IPv6 and IPv4 BT packets to demonstrate how the bursty nature is preserved in different time scales. Different time scales are selected for IPv6 and IPv4 BT packet traffic. To compare the time length of total traces 1-s and 10-s windows are given in both IPv6 and IPv4 traffic. As stated before, IPv4 traces cover 15 min but IPv6 traces cover nearly 8 h each day. Fig. 1 represents the IPv6 BT packet traffic of 1 June 2008 in 1-, 10-, 60-, and 100-s scales.

Fig. 2 represents the IPv4 BT packet traffic of 1 August 2008 in 0.01-, 0.1-, 1-, and 10-s scales.

An important point is that the bursty nature has been preserved in different time scales for both types of protocol traffic. This result is consistent with [21].

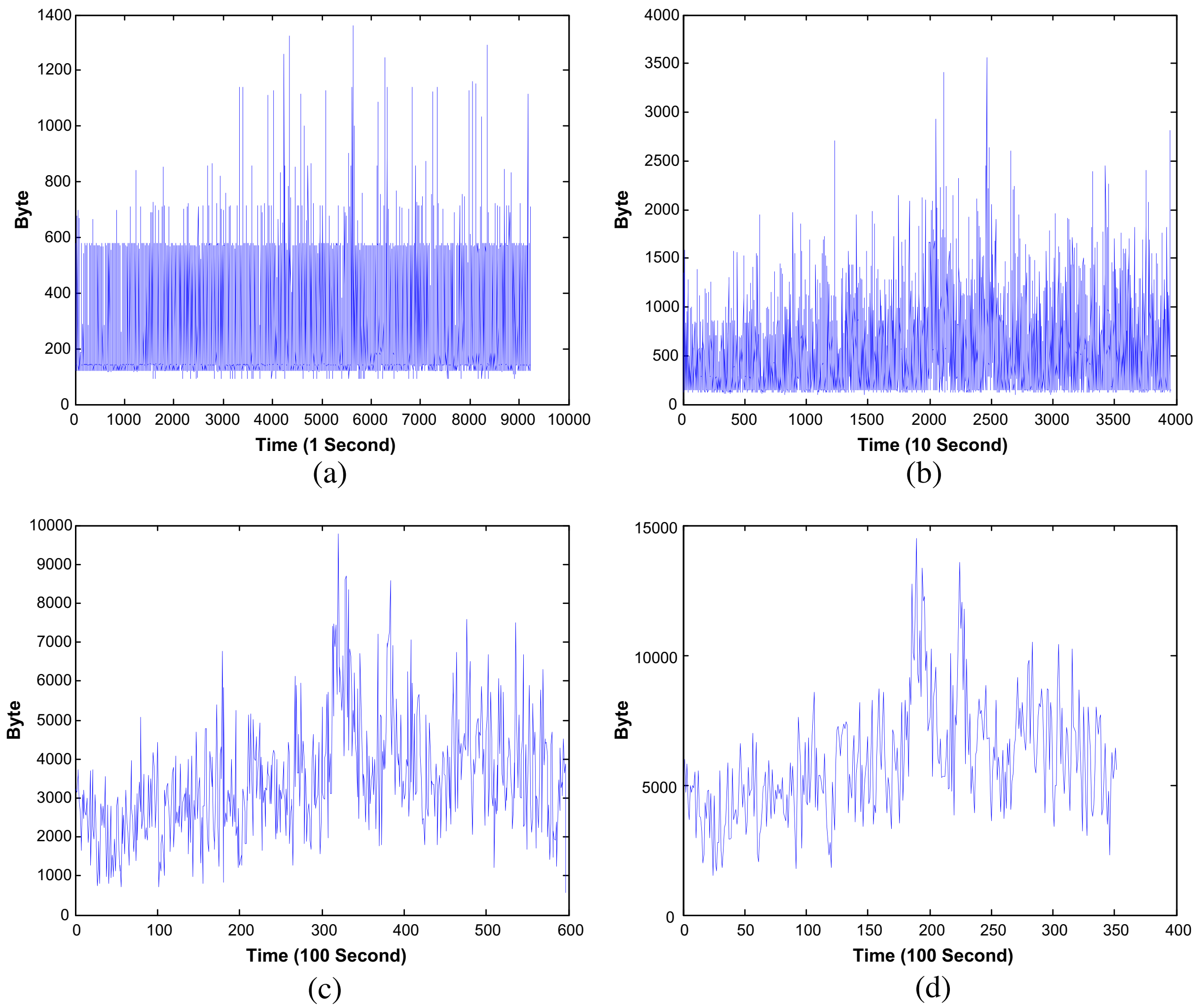

Fig. 1. Time series of 1 June 2008 IPv6 BT packets. Every point shows (a) 1-s, (b)10-s, (c) 60-s, (d) 100-s.

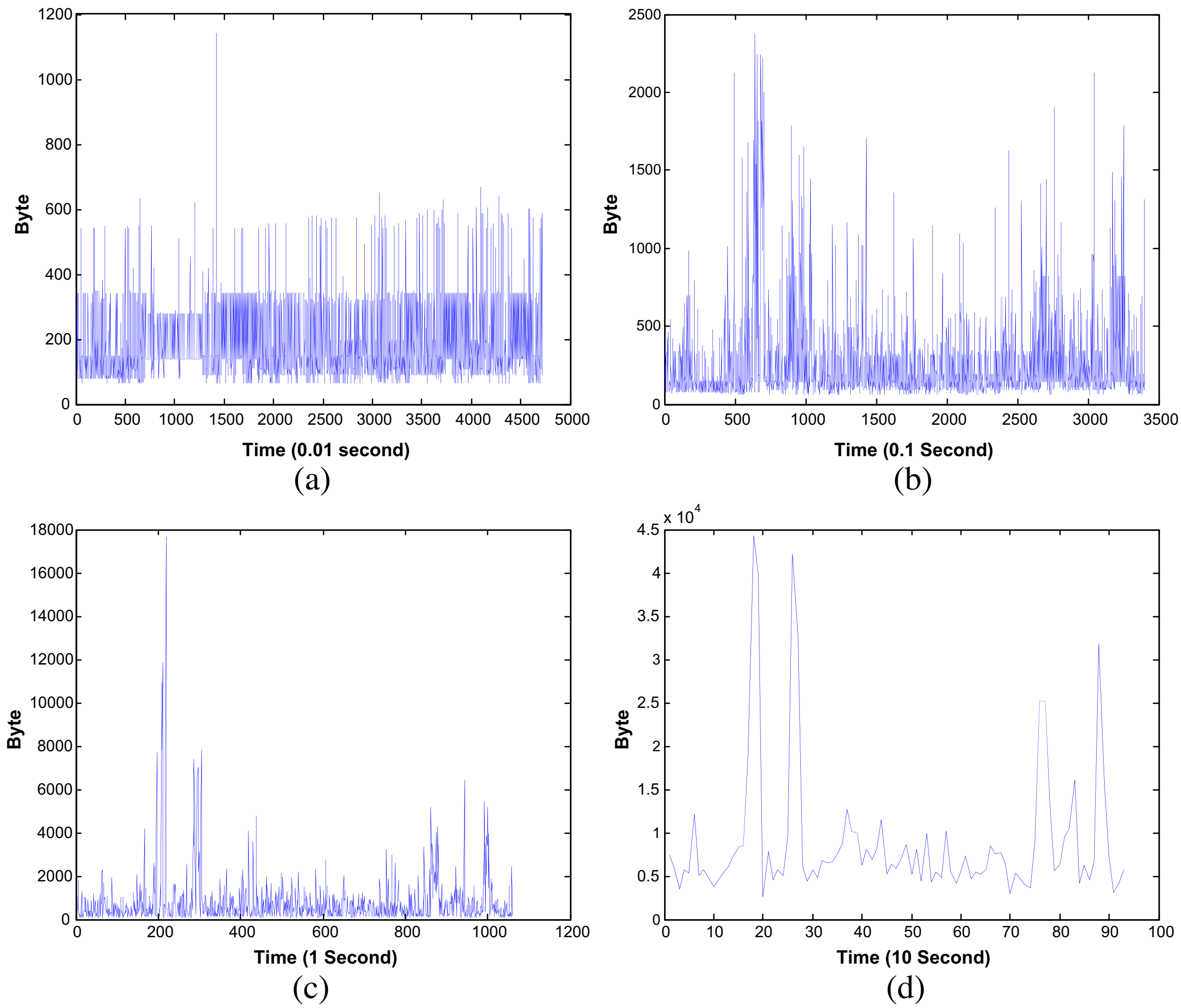

Fig. 2. Time series of 1 August 2008 IPv4 BT packets. Every point shows (a) 0.01-s, (b) 0.1-s, (c) 1-s, (d) 10-s.

4.2. Autocorrelation analysis

Long range dependent packet traffic has a serious impact on the performance of computer networks. Packet traffic is relevant during long time lags. That is, the autocorrelation of packets does not decay exponentially. For long range dependent time series the autocorrelation function follows:

$$r_x(k) \approx |k|^{-\beta} \text{as } k \to \infty, 0 < \beta < 1, \tag{1}$$

where k represents time lag and β relates closely to the Hurst parameter. Another important characteristic of the autocorrelation function is that it is non-summable.

$$\sum_k r_x(k) = \infty. \tag{2}$$

Both IPv6 and IPv4 BT traffic analyses cover a one-month period. Traffic is captured in tcpdump format. We extract the interarrival time and size of each packet via packet sniffer software. According to the autocorrelation values obtained, IPv6 and IPv4 BT traffic shows similar correlations in terms of packet size and interrarrival time. They both represent low autocorrelation values during long lags. As we mentioned, analysis covers one month for each traffic type. One representative day for IPv6 BT traces and one for IPv4 BT traces are shown in Figs. 3 and 4. The autocorrelation figures obtained obviously show that IPv6 and IPv4 BT packet size and packet interarrival time do not give a high degree of long range dependent behavior. This result gives some sense of the degrees of the Hurst exponent of the metrics mentioned. Similar types of autocorrelation plot are obtained for the other day traces.

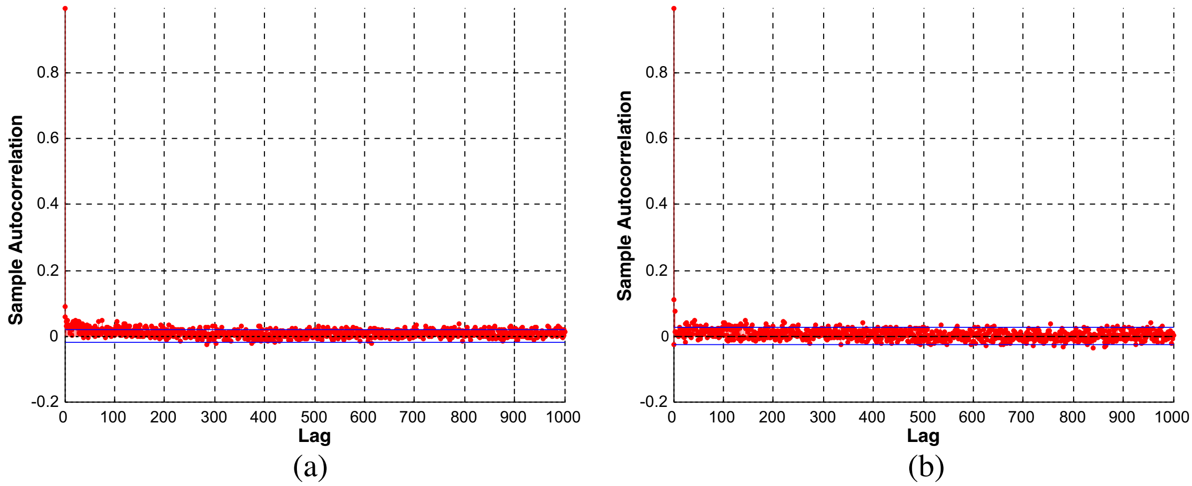

Fig. 3. Packet size autocorrelation: (a) 3 June 2008 IPv6 BitTorrent traffic, (b) 4 August 2008 IPv4 BitTorrent traffic.

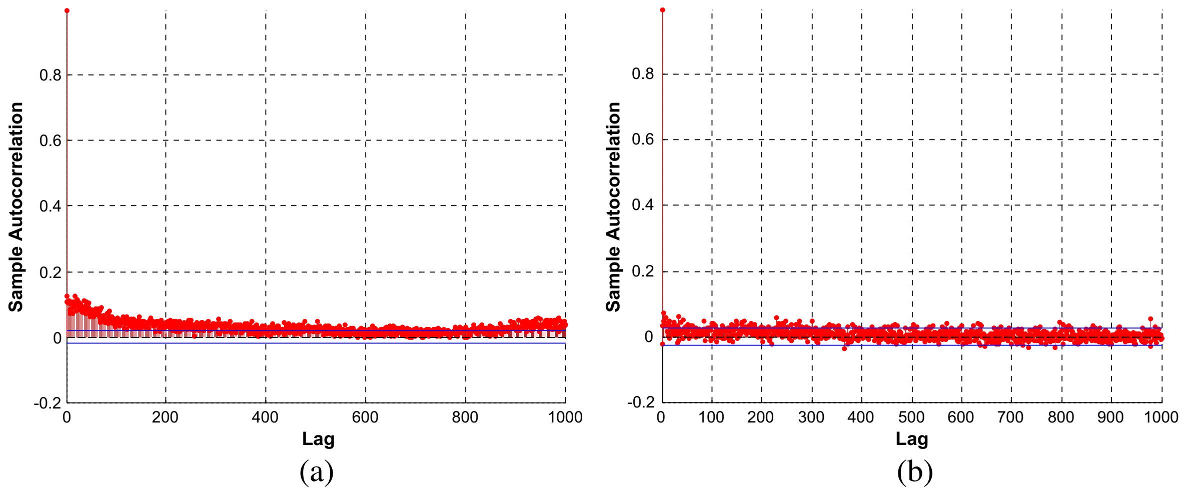

Fig. 4. Packet interarrival time autocorrelation: (a) 3 June 2008 IPv6 BitTorrent traffic, (b) 4 August 2008 IPv4 BitTorrent traffic.

4.3. Cumulative distribution analysis

Since network traffic represents non-deterministic behavior, it cannot be modeled using definite formulations. Statistical analysis is more suitable to extract significant information for such time series. Many performance-related calculations of network traffic take into consideration the results of statistical analysis. Distribution functions are also very important in network traffic studies. Modeling of network traffic via a known distribution function is an important step to determine the properties of network traffic. If a time series could be defined with a known distribution function, similar time series could be generated easily for network simulation studies. Therefore, we would like to observe cumulative distribution differences between IPv6 and IPv4 BT packet traffic in terms of packet size. Network traffic generally represents heavy tailed distribution functions. That is, very large values also show significant probabilities. If a time series is heavy tailed, it shows:

$$P[X > x \approx x^{-\alpha}] \text{ as } x \to \infty, 0 < \alpha < 2, \quad (3)$$

where α is related with Hurst parameter.

Plotted cumulative distribution functions (CDFs) for interarrival time and packet size show some differences between IPv4 and IPv6 BT protocol traffic. We assert that this difference is caused by the packet structure difference between the two protocols. In the IPv4 protocol, there are some optional headers that are not present in IPv6. According to Fig. 5, IPv6 BT packets aggregated into some levels. However, IPv4 BT packet sizes represent a distributed packet size distribution structure. Based on this observation, IPv6 BT packet sizes cause some peaks at some levels in their probability densities. These differences between packet size CDFs indicate that modeling must be done with different distribution functions.

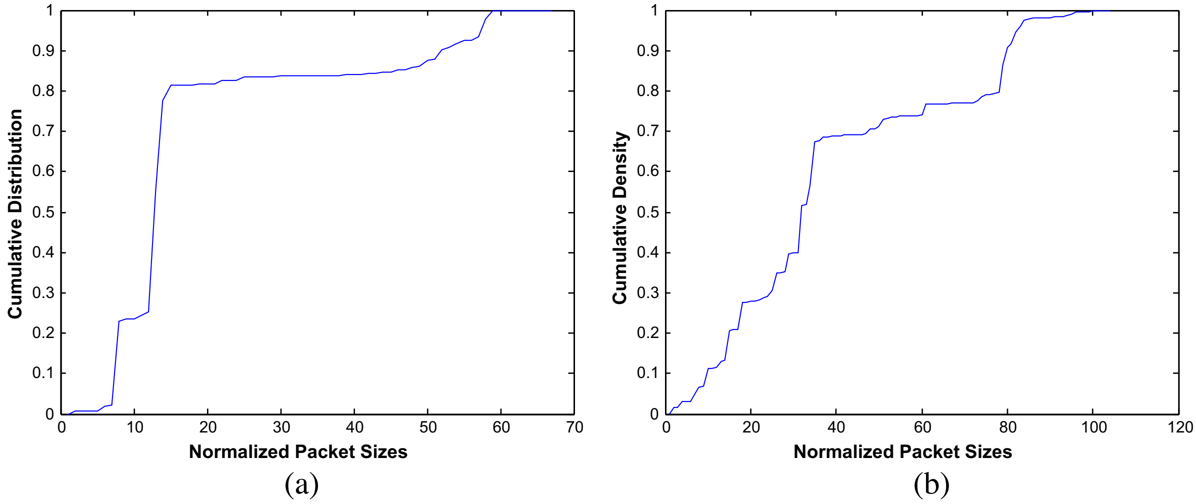

Fig. 5. Packet size CDF of both protocols' traffic: (a) 3 June 2008 IPv6 BitTorrent traffic, (b) 4 August 2008 IPv4 BitTorrent traffic.

4.4. Power spectral density analysis

Another important characteristic for self-similar packet traffic is spectral density. If a time series is long range dependent, its spectral density follows a power law near the origin:

$$s_x(w) \approx |w|^{-\gamma} \text{as } w \to \infty, 0 < \gamma < 1, \quad (4)$$

where w is the frequency, $s_x(w)$ is the spectral density and

$$\gamma = H - 1. \quad (5)$$

Power spectral density (PSD) plots of both IPv4 and IPv6 BT packets represent similar characteristics and both of them represent a little 1/f type power spectrum behavior as shown in Figs. 6 and 7. Moreover, their spectra are more like Gaussian type power spectra. However, the self-similarity of IPv4 packet interarrival times and packet sizes look slightly greater than their IPv4 counterparts. We report only one representative day PSD plots. We also obtain similar plots for other day traces.

4.5. Self similarity analysis

Self-similarity has attracted attention in the network traffic community. It is not only a simple phenomenon related to correlation. It has changed the way of looking into some basic artifacts in many disciplines. For example, the modeling of

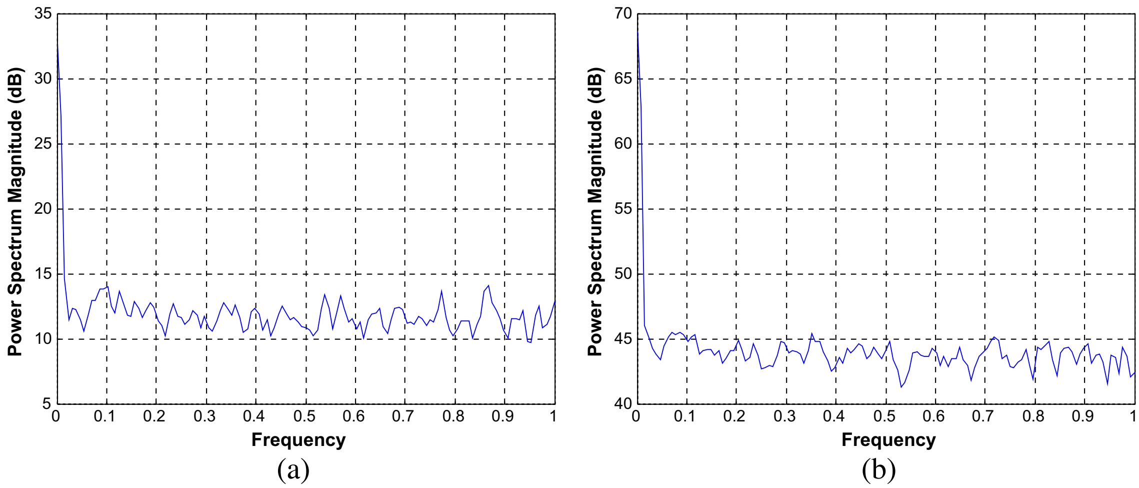

Fig. 6. PSD of 3 June 2008 WIDE-6Bone IPv6 traffic: (a) packet size, (b) interarrival time.

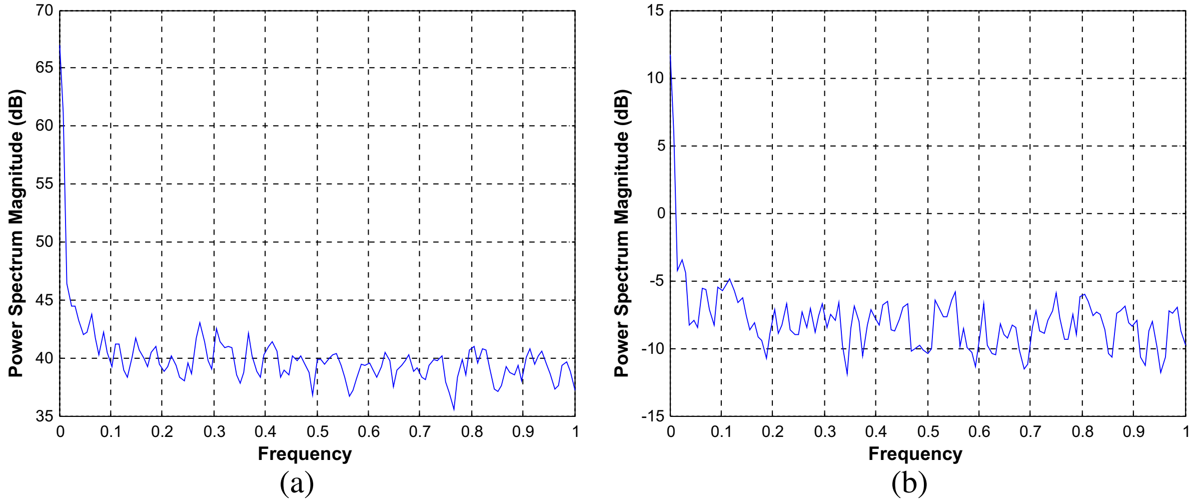

Fig. 7. PSD of 11 August 2008 150 Mbps trans-pacific line IPv4 traffic: (a) packet size, (b) interarrival time.

network traffic used to be done differently. Then Hurst discovered a parameter related to self-similarity while investigating discharge time series of the Nile in the design of a reservoir [30]. Performance-related calculations for computer networks such as resource sharing, efficient queue management, and routing management have been studied as self-similarity artifacts occurred in most teletraffic modeling.

As self-similarity is significant in several disciplines, its efficient estimation is vital. The Hurst parameter is a numerical measure of self-similarity. It indicates whether a stochastic process has long range dependency or not. Moreover, the process is randomly scattered when the Hurst value is 0.5. A continuous time stochastic process $\{X(t), t \in R\}$ is strictly self-similar with the Hurst parameter $\{H, 0 < H < 1\}$ if the following equation is supplied:

$$X(at) \stackrel{d}{=} a^H X(t). \tag{6}$$

In Eq. (6), $X(at)$ is a new process scaled by factor a and $\stackrel{d}{=}$ means equal in finite dimensional distributions. The R/S method, variance time estimator, the estimator based on absolute moments, and the estimator based on variance of residuals are commonly used time-based estimators. Frequency-based Hurst estimation methods take into consideration the power law behavior of power spectral density. The Daniell PB estimator, Whittle Maximum Likelihood estimator, and Local Whittle ML estimator are some examples of frequency-based estimators.

In this subsection, self-similarity analysis of IPv6 and IPv4 BT traffic in terms of packet interarrival time and packet size is given. In the Hurst estimation procedure, various Hurst estimation methods in the literature are used to obtain comparable results. The results are given in Tables 1 and 2. We compare the self-similarity of packet interarrival time and packet size of both protocols' traffic. Self-similarity degrees of packet size and interarrival times give very close values for each traffic type. The only difference is that all estimation methods give slightly greater Hurst values for IPv4 BT packet traffic in terms of interarrival time and packet size. Average values during one month are given in the tables.

An important conclusion that we have drawn from the self-similarity analysis is that packet interarrival time and packet size of both types of BT traffic give a Hurst degree of about 0.7. The results obtained show that the network management of the two protocols could be handled in the same fashion.

Table 1
Average Hurst values for interarrival time of both protocol.

Protocol	Absolute value	Aggregated variance	Rescaled range	Wavelet method
IPv4	0.789925	0.71942	0.78450	0.733095
IPv6	0.76870	0.77089	0.75714	0.683119

Table 2
Average Hurst values for packet size of both protocol.

Protocol	Absolute value	Aggregate variance	Rescaled range	Wavelet method
IPv4	0.770096	0.773168	0.79379	0.729898
IPv6	0.729126	0.731504	0.748181	0.705134

4.6. Distribution analysis

In this section, we aim to obtain a suitable distribution function for packet size and interarrival time of IPv6 and IPv4 BT packet traffic. Since network traffic exhibits non-deterministic behavior, it could be well characterized using distribution functions. With efficient modeling, realistic traffic traces could be generated for both IPv6 and IPv4 BT packet traffic in network simulation studies. In the distribution modeling procedure, we use 48 different distribution types. In selection criteria for goodness of fitting, we use Anderson-Darling and chi square test results and determine the suitable distribution type for packet interarrival time and packet size of IPv6 and IPv4 BT traffic.

Log Logistic with 3 parameters gives the best result for IPv4 BT packet size modeling. Its probability density function is as follows:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-y}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x-y}{\beta}\right)^{\alpha}\right)^{-2}, \tag{7}$$

and CDF function is given as

$$F(x) = \left(1+\left(\frac{\beta}{x-y}\right)^{\alpha}\right)^{-1}. \tag{8}$$

According to the test results, the span for the beta distribution parameters α, β, and y vary in the ranges [1.2075–2.425], [101.14–120.44], and [21.368–39.797] for the analyzed traces, respectively. Weibull distribution is selected for IPv4 interarrival time fitting according to the Anderson–Darling and chi square test results. The PSD function and CDF of the Weibull distribution are as follows:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right), \tag{9}$$

$$F(x) = 1-\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right), \tag{10}$$

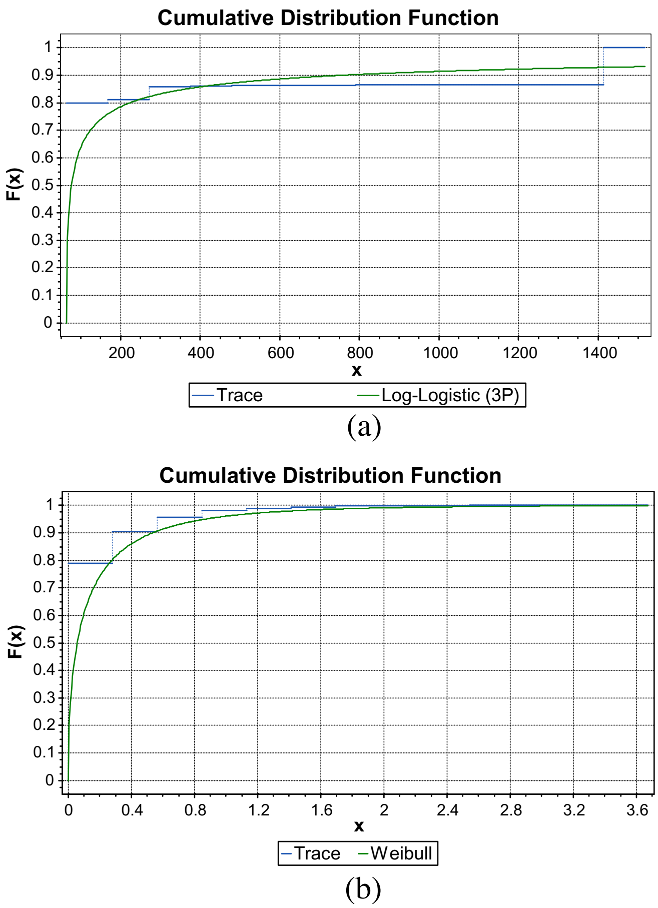

Fig. 8. (a) Packet size CDF of IPv4 BT traffic. (b) Interarrival time CDF of IPv4 BT traffic.

The α and β parameters ranges are [0.43661–0.639] and [0.11532–0.15817] for interarrival time, respectively. Fig. 8 shows the cumulative distribution and fitted distribution for BT packet size and interarrival time on IPv6.

Generalized Gamma with 3 parameters is the best model for IPv6 BT packet interarrival time according to the test statistics. Its PSD distribution is given as

$$f(x) = \frac{kx^{k\alpha-1}}{\beta^{k\alpha}\Gamma(\alpha)} \exp\left(-(x/\beta)^k\right), \tag{11}$$

and CDF distribution of IPv6 is as follows:

$$F(x) = \frac{\Gamma_{(x/\beta)^k}(\alpha)}{\Gamma(\alpha)}. \tag{12}$$

The k, α, and β parameters' ranges for all day traces are [0.66868–0.98909], [0.64778–0.90336], and [4.5007–6.3001], respectively. Lastly the Pareto distribution gives the best fitting result for IPv6 BT packet sizes. The PSD and CDF of the Pareto distribution are as follows respectively in Eq. (13) and Eq. (14):

$$f(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}, \tag{13}$$
$$F(x) = 1 - (\beta/x)^{\alpha}, \tag{14}$$

The α and β parameters' ranges for all day traces are [1.5024–1.7287] and [78–94], respectively. The distribution results obtained show that IPv6 and IPv4 packet size and packet interarrival time have different probability statistics. Fig. 9 shows the cumulative distribution and fitted distribution for BT packet size and interarrival time on the IPv4 protocol for one day randomly selected. Similar distribution plots are obtained for other day traces.

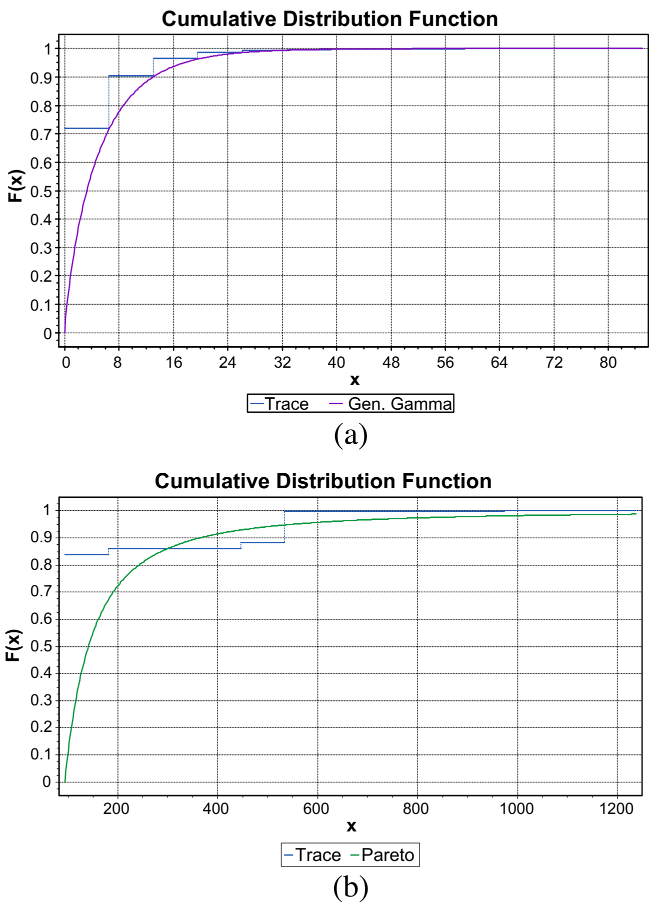

Fig. 9. (a) Packet size distribution of IPv6 BT traffic. (b) Interarrival time distribution of IPv6 BT traffic.

5. Conclusions

In this study, we analyze characteristics of IPv6 and IPv4 BitTorrent traffic obtained through real network traces. We analyze IPv6 traces of a month length obtained on a link connected to a WIDE-6Bone node of the MAWI Working Group traffic archive. Every day trace consists of nearly 4–8 thousand BitTorrent packets. To analyze differences between IPv6 and IPv4 BitTorrent traffic, we also use a month of IPv4 traffic traces, obtained on a 150 Mbps transpacific link between Japan and the US from the MAWI Working Group traffic archive. Detailed analyses are performed in terms of power spectral density, probability density, autocorrelation, self-similarity, and cumulative distribution. Our study indicates that there are more similarities than differences in the mentioned autocorrelations and power spectral densities for IPv6 and IPv4 BitTorrent traffic in terms of packet interarrival time and packet size. However, the distribution fitting results prove that their modeling must be handled differently. According to the Anderson–Darling and chi square test statistics, Log logistic with 3 parameters and Weibull distribution could be used to model IPv4 BT packet size and packet interarrival time. However, the Generated Gamma and Pareto distribution give the best fitting results for IPv6 BT packet interarrival time and packet size. The self-similarity results obtained for interarrival time and packet size series are very similar for both protocols. They give Hurst degrees between 0.7 and 0.8 according to different Hurst estimation methods. The results indicate that the network management of the two traffic types could be handled in the same fashion. We will also consider other P2P application traffic over the IPv6 protocol in our future research.

Acknowledgements

The authors would like to thank the editors and anonymous reviewers for their valuable suggestions and comments. The fourth author's research was supported by TUBITAK (The Scientific and Technical Research Council of Turkey) under CAREER Award Grant 104E064.

References

[1] B. Liu, Y. Cao, Y. Cui, Y. Lu, Y. Xue, Locality analysis of BitTorrent-like peer-to-peer systems, in: 7th IEEE Consumer Communications and Networking Conference (CCNC), 2010, pp. 1–5.
[2] Morgan Stanley Report. Available at: <http://www.morganstanley.com/institutional/techresearch/webtwopto2006.html>.
[3] J.S.K. Chan, V.O.K. Li, K.S. Lui, Performance comparison of scheduling algorithms for peer-to-peer collaborative file distribution, IEEE Journal on Selected Areas in Communications 25 (1) (2007) 146–154.
[4] Top Applications (bytes) for Subinterface: SD-NAP Traffic: <http://www.caida.org/analysis/workload/byapplicationlsdnap>.
[5] F.R. Santos, L.P. Gaspary, M.P. Barcellos, Towards a robust pollution control mechanism for P2P file sharing systems, in: IEEE INFOCOM Workshops, 2009, pp. 1–2.
[6] IPOQUE, Internet Study 2007: The Impact of P2P File Sharing, Voice over IP, Skype, Joost, Instant Messaging, One-Click Hosting and Media Streaming such as YouTube on the Internet. <http://www.ipoque.com/media/internet_studies/internet_study_2007>, 2007.
[7] D.S. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, Z. Xu, Peer-to-peer Computing, HP Laboratories, Berlin, 2002. pp. 11–12.
[8] Gnutella, <http://www.gnutella.com>, 2005.
[9] J.A. Pouwelse, P. Garbacki, D.H. Epema, H. Sips, The BitTorrent P2P file-sharing system: measurements and analysis, in: Proceedings of the IPTPS, 2005, pp. 1–13.
[10] K. Eger, U. Killat, Bandwidth trading in BitTorrent-like P2P networks for content distribution, Computer Communication 31 (2008) 201–211.
[11] N. Andrade, E.S. Neto, F. Brasileiro, M. Ripeanu, Resource demand and supply in BitTorrent content-sharing communities, Computer Networks 53 (2009) 515–527.
[12] Z. Zhang, Y. Lin, Y. Chen, Y. Xiong, J. Shen, H. Liu, B. Deng, X. Li, Experimental study of broadcatching in BitTorrent, in: 6th IEEE Consumer Communications and Networking Conference, 2009, pp. 1–5.
[13] L. Ye, H. Zhang, W. Zhang, J. Tan, Measurement and analysis of BitTorrent availability, in: 15th International Conference on Parallel and Distributed Systems (ICPADS), 2009, pp. 787–792.
[14] Q.H. Li, C.S. Lui, On modeling clustering indexes of BT-like systems, in: IEEE International Conference on Communications, 2009, pp. 1–6.
[15] W. Baohua, F. Gilles, C. Franck, Collaborative data distribution with BitTorrent for computational desktop grids, in: Proceedings of the 4th International Symposium on Parallel and Distributed Computing, IEEE Computer Society, 2005.
[16] P. Nadim, W. Carey, M. Anirban, C. Niklas, Analysis of BitTorrent-like protocols for on-demand stored media streaming, in: Proceedings of the 2008 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, 2008.
[17] D. Karim El, G. Minas, M. Athina, BitTorrent: misusing BitTorrent to launch ddos attacks, in: Proceedings of the 3rd USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet, USENIX Association, 2007.
[18] D. Qiu, R. Srikant, Modeling and performance analysis of BitTorrent-like peer-to-peer networks, in: ACM SIGCOMM'04, 2004, pp. 11–12.
[19] M. Barbera, A. Lombardo, G. Schembra, M. Tribastone, An analytical model of a BitTorrent peer, in: 15th EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing IEEE, 2007.
[20] D. Erman, D. Ilie, A. Popescu, BitTorrent traffic characteristics, in: Proceedings of the International Multi-Conference on Computing in the Global Information Technology (ICCGI'06), IEEE, 2006.
[21] G. Liu, M. Hu, B. Fang. H. Zhang, Explaining BitTorrent traffic self similarity, in: PDCAT 2004, LNCS, vol. 3320, 2004, pp. 839–843.
[22] D. Minoli, Voice over IPv6: Architectures for Next Generation VoIP Networks, Elsevier, Burlington, 2006.
[23] B.J. Nikkel, An introduction to investigating IPv6 networks, The International Journal of Digital Forensics and Incident Response (2007) 4.
[24] O. Ozkasap, M. Caglar, A. Alagöz, Principles and performance analysis of second: a system for epidemic peer to peer content distribution, Journal of Network and Computer Applications 32 (2009) 666–683.
[25] S. Deering, R. Hinden, Internet Protocol, version 6 (IPv6), Specification RFC 1883, Internet Engineering Task Force, 1995.
[26] M. Goncalves, K. Niles, IPv6 Networks, McGraw-Hill, 1998.
[27] C. Huitema, IPv6: The New Internet Protocol, second ed., Prentice Hall, 1997.
[28] Information Sciences Institute, USC, Internet Protocol, RFC 791, IETF, 1981.
[29] Wide Project. <http://mawi.wide.ad.jp/mawi/>.
[30] H.E. Hurst, Long term storage of reservoirs, Trans. Am. Soc. Civ. Eng. 116 (1950) 770–808.