

Time-Dimensional Traffic Engineering with Storage Aware Routing

著者	Yamashita Shigeyuki, Yagyu Tomohiko, Yamamoto Miki
journal or publication title	IEICE Transactions on Communications
volume	E97-B
number	12
page range	2641-2649
year	2014-12
権利	Copyright (C) 2014 The Institute of Electronics, Information and Communication Engineers https://search.ieice.org/ The publisher's version is also available at https://www.jstage.jst.go.jp/article/transcom/E97.B/12/E97.B_2641/_article
URL	http://hdl.handle.net/10112/11052

doi: 10.1587/transcom.E97.B.2641

Time-Dimensional Traffic Engineering with Storage Aware Routing*

Shigeyuki YAMASHITA[†], Student Member, Tomohiko YAGYU^{††a)}, Member, and Miki YAMAMOTO[†], Fellow

SUMMARY Because of the popularity of rich content, such as video files, the amount of traffic on the Internet continues to grow every year. Not only is the overall traffic increasing, but also the temporal fluctuations in traffic are increasing, and differences in the amounts of traffic between peak and off-peak periods are becoming very large. Consequently, efficient use of link bandwidth is becoming more challenging. In this paper, we propose a new system for content distribution: storage aware routing (SAR). With SAR, routers having large storage capacities can exploit those links that are underutilized. Our performance evaluations show that SAR can smooth the fluctuations in link utilization.

key words: storage, content distribution, new generation network, traffic peak shift

1. Introduction

The volume of traffic across the Internet continues to increase steadily, in large part because of continuous growth in video traffic. According to a Cisco white paper [1], traffic is increasing eight fold every five years, and will approach 1.3 zettabytes by the end of 2016. To effectively accommodate such an increasing demand, sophisticated management of networks will be indispensable. Content delivery networks (CDNs) [2] and traffic engineering [3] are practical examples of effective network management. These techniques aim to use adequate routing to distribute traffic throughout all spatial dimensions; i.e., these conventional techniques are limited to managing traffic only in space.

Besides traffic volume, another impact aspect of current Internet traffic use is the difference between peak and minimum traffic volume on a daily basis. According to a Japanese government report [4], the difference between the daily peak and minimum traffic in day-time scale has recently increased significantly. In current network planning, the bandwidths of links are designed to accommodate peak traffic. Since most Internet traffic propagates in bursts, average link utilization should be small; this means that network resources cannot be utilized effectively if only spatial traffic management techniques are applied.

In-network caching can be used to avoid network congestion caused by content flooding [5]. Paul et al. proposed a cache-and-forward architecture [6], in which routers in the networks cache content. Further, the concept of data centric networking has been proposed [7], [8]. Large storage equipment in routers could be utilized for time-dimensional traffic management. When the network is congested, content could be stored in the network, i.e., at the routers, and the stored content could be transmitted during off-peak periods. A delay tolerant network (DTN) [9] is a cache-and-forward technology with storage in routers. It enables discontinuous forwarding; however, the main purpose of DTN is to provide reliable communications in networks having intermittent connections. Therefore, DTN is not a technology for traffic management. Other methods to control transmission times by using node storage have been proposed [10], [11]. However, because DTN and these other methods do not consider deadlines for data delivery, they cannot guarantee the timely arrival of priority data. This is the essential difference between DTN and the method proposed here. Another method for asynchronous forwarding, NetStitcher, has been proposed for bulk transfers between datacenters [12]. NetStitcher schedules bulk data transfers to smooth resource fluctuations into diurnal patterns. NetStitcher is designed to be applied to worldwide datacenters; so, bulk data transfers are scheduled to occur during off-peak nocturnal fluctuations. Though the purpose of NetStitcher is similar to our proposal, i.e., traffic management in the time dimension, it can only be applied to worldwide datacenter systems because it adapts to resource fluctuations by taking advantage of differences in local times.

In the near future, one possibility for video services could be a deadline-aware scenario, in which a user reserves their preferred video service for their desired time. For example, a user on a commuter train in the morning could reserve a video file to be available at home in the evening. In current on-demand video services, such as YouTube, the peak traffic period is exactly the same as the peak demand period, and those peak periods often occur in the evening at primetime. In contrast, in a deadline-aware video service, there is leeway between the time a video file is reserved and the time it is received. Another example of deadline-aware delivery is a backup service for critical data between distant datacenters for business continuity planning (BCP). When a router has a large storage space and can store content when the forwarding link is congested during peak traffic loads, on-peak traffic can be shifted to an off-peak period.

Manuscript received March 16, 2014.

Manuscript revised July 2, 2014.

[†]The authors are with the Faculty of Engineering Science, Kansai University, Suita-shi, 564-8680 Japan.

^{††}The author is with the Cloud System Research Laboratories, NEC Corporation, Kawasaki-shi, 211-8666 Japan.

*This paper is an extended paper of "Evaluation of Time Dimensional Traffic Engineering with Storage Aware Routing" [13], by S. Yamashita, T. Yagyu, and M. Yamamoto which appeared in The Fifth International Conference on Advances in Future Internet.

a) E-mail: yagyu@cp.jp.nec.com

DOI: 10.1587/transcom.E97.B.2641

In this paper, we propose storage aware routing (SAR), a novel traffic engineering technique that spatially and temporally manages traffic [13]. SAR aims to smooth fluctuations during link utilization throughout a network. In SAR, a router stores traffic content and shifts on-peak traffic to off-peak periods. We assume that storage in routers is large enough to enable time shifts of on-peak traffic; the required storage is much larger than the traffic of queues in current routers. In this paper, we present the algorithm for SAR traffic management in space-time. We investigate the feasibility and fundamental benefits of deadline-aware content delivery with SAR. Our simulation results demonstrate that SAR can effectively adapt traffic delivery to smooth fluctuations in link utilization.

This paper is structured as follows. In Sect. 2, we explain our proposed new traffic engineering technique (SAR) in detail. In Sect. 3, we evaluate the performance of SAR, and in Sect. 4, we draw conclusions.

2. Storage Aware Routing

2.1 Design Concept of SAR

In the current Internet, the sharp increase in traffic is one of the most important technical issues. In particular, the gap between peak and off-peak traffic is increasing significantly. If some on-peak traffic can be shifted to off-peak periods, the smoother use of links will be able to support the increasing traffic. Therefore, we propose SAR that will enable traffic to be shifted in time while providing timely delivery of deadline-aware content.

According to traffic statistics reported by the Ministry of Internal Affairs and Communications [4], Internet traffic peaks daily between 1900 and 2300 hrs. One possible reason for this peak is concentrated requests for online video files; i.e., large numbers of users tend to watch videos during the evening hours. One target of SAR is video services with delivery reservations, in which users can reserve their videos during a convenient time, such as lunchtime or during the morning commute. Then, SAR can exploit the time difference between the request (morning or lunchtime) and the deadline for video delivery, which is typically during the primetime between 1900 and 2300. SAR assumes a granularity of time for traffic management that is from tens of minutes to several hours. This means that SAR does not affect the quality of video content because all data segments have been delivered before a user starts viewing the content. In SAR traffic management, storage read/write latency is on the order of several μs , or at most 1 ms, and therefore is negligible. The end user is interested only in the content; they do not care about content retrieval time so long as the content is available at their convenience. SAR tactically utilizes the difference between the time of a request and the time of content consumption. For a deadline-aware video service, content can be delivered between these two times. SAR schedules its hop-by-hop transmission times for each session (end-to-end session for one content retrieval) so that

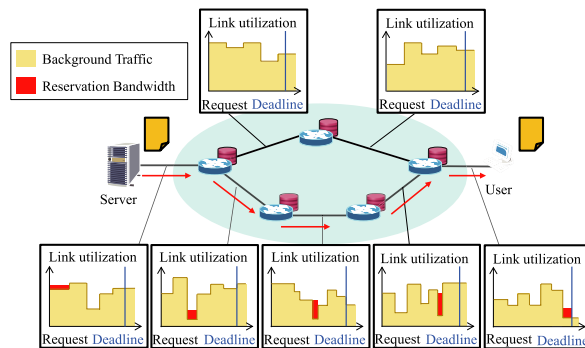


Fig. 1 Storage aware routing.

fluctuations in link utilization are minimized across the entire transmission path. With this smoothed link utilization, the network can preserve its resources for later requests.

2.2 Storage Aware Routing

Figure 1 shows an overview of SAR. The framework for SAR is composed of the following three elements:

1. Request by end user,
2. Scheduling of transmission at the server, and
3. Routing and scheduling in the network.

SAR is a time-dimensional traffic engineering technique that makes use of the time difference between request and content consumption. Requests for content by end users trigger SAR. A content server and the associated network schedule transmission times for the corresponding content files so that fluctuations in link utilization are minimized. The content server schedules its transmission time, followed by the network choosing a path for transmitting content files and scheduling hop-by-hop transmission times.

As shown in Fig. 1, SAR chooses the route and transmission time that minimizes disparities in link utilization. In the figure, a graph for each link depicts link utilization over time. At the first link in Fig. 1, the content file is not transferred during its lowest utilization period because that period is out of sync with low utilization periods at subsequent links. If the content file is transferred during the lowest utilization period at the first link, then the content must be transferred during high utilization periods at the following links so as to reach the destination before the deadline (law of causality). In this way, SAR optimizes not only routing but also transmission times at each hop on the route, i.e., SAR addresses both spatial routing and temporal scheduling.

2.3 Algorithm for Traffic Management in Space-Time

SAR can calculate a minimum cost route from a space-time diagram. Figure 3 shows a space-time diagram for the tandem topology model shown in Fig. 2. The horizontal axis shows time and the vertical axis represents space. In Fig. 3, the space dimension corresponds to the tandem network in



Fig. 2 Tandem topology model.

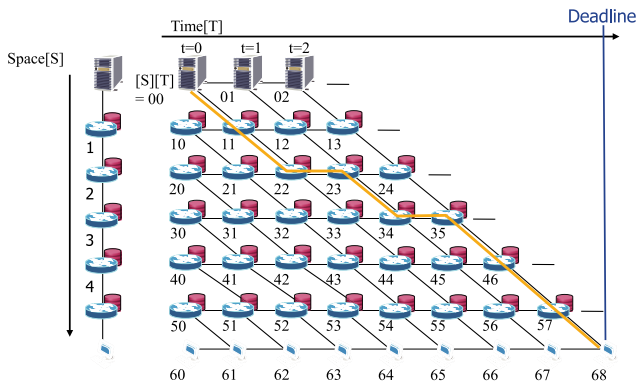


Fig. 3 Space-time diagram.

Fig. 2. Time zero is the time at which a user makes a request for content delivery, and time 8 is the deadline for delivery, as requested by the user. The origin $([S][T] = 00)$ of this space-time diagram shows the content server at time 0. Here, time is normalized by the file transmission time at a link (we simply assume that a content file is transmitted at a link in one time unit). In a sample path in Fig. 3, the server transmits the content file to router 1 at time 0. Instead of storing the file after router 1 receives the whole file at time 1, it immediately forwards the file to router 2. At routers 2 and 3, the content file is stored for one unit time to avoid congestion; this is temporal scheduling. In this example, the deadline for content receipt is time 8 and the user receives the file just at the deadline.

As explained above, SAR is designed to minimize the cost of a space-time path. We define a space-time path as a path from the source at the time of content requested to the receiver at the deadline (like the sample path in Fig. 3). To heavily penalize the choice for a highly utilized link, we set the cost of the link as the square of the utilization ratio. We set any storage cost to zero, under the assumption that storage capacity in routers can be as large as required. We also assume that future link resources can be reserved as needed.

We propose to use a centralized controller such as OpenFlow [14]. A centralized controller has the potential to manage link bandwidths in the network and to dictate scheduled transmission times to routers. The scalability of SAR for network size can be resolved by implementing several research proposals for OpenFlow, such as a hierarchical structure for the OpenFlow controller and flow aggregation [15].

The objective is to find a route and the transmission times that minimize fluctuations in link utilization under the restriction that content must be delivered on or before the deadline. We use the Dijkstra algorithm to calculate the minimum cost path on a space-time diagram. Because we

Table 1 Notation used in Algorithm 1.

Notation	Meaning
$src(i)$	Sender node of content i
$dst(i)$	Receiver node of content i
$rqtime(i)$	Request generation time of content i
$dl(i)$	Deadline period of content distribution
G	Graph of network topology
$G'(i)$	space-time diagram for G
	Origin in $(scr(i), rqtime(i))$
E'_{pl}	Set of physical links in G' (diagonal links in Fig. 3)
E'_s	Set of time dimension links (storage : horizontal links in Fig. 3) in G'
$src_{rqtime}(i)$	$src(i)$ at $rqtime(i)$
$dst_{dl}(i)$	$dst(i)$ at $dl(i)$

Algorithm 1 Calculations performed in SAR for optimizing the path for routing and for temporal scheduling.

GIVEN:

$BackgroundTraffic(l), \forall l \in E'_{pl}$

Initialize:

$AccumulatedSARtraffic(l) = 0, \forall l \in E'_{pl}$

Calc SAR Route for request for content delivery i :

$G'(i) = ExtractGraph$

$ToTimeDimension(G, src(i), dst(i), rqtime(i), dl(i))$

$SetCost(l \in E'_{pl}(i))$

$Path(i) = CalcMinimumCostPath(G'(i), src_{rqtime}(i), dst_{dl}(i))$

$AccumulatedSARtraffic(l) += traffic(i), \forall l \in Path(i)$

Subroutine SetCost:

$Cost(l) = ((BackgroundTraffic(l) + AccumulatedSARtraffic(l) + traffic(i)) / Capacity(l))^2$

use the square of the link utilization ratio as the link cost, the Dijkstra algorithm can find the best path that smoothens link utilization. Whenever the controller receives a request, it calculates the best path and sends transmission schedules to the content server and all routers on the path. The route is calculated when the request arrives at the service. Therefore, the controller does not need to have full knowledge of all requests beforehand. It only has to know the accumulated reserved bandwidth for SAR traffic. This means that the controller does not incur a burden for calculating the route.

Algorithm 1 shows our Dijkstra-based SAR algorithm (the Dijkstra algorithm is applied at $CalcMinimumCostPath$). Table 1 shows the notation used in this algorithm. Background traffic is the amount of non-SAR traffic and is estimated statistically in advance. In $SetCost$, link costs in E'_{pl} are calculated for the sum of background traffic, the current accumulated SAR traffic, and the corresponding content i . After the minimum cost path has been identified, this designed traffic is added to the accumulated SAR traffic, which is used in calculating the path for the next content request. If the requested content cannot be accommodated due to limitations on link capacity, the content request is rejected.

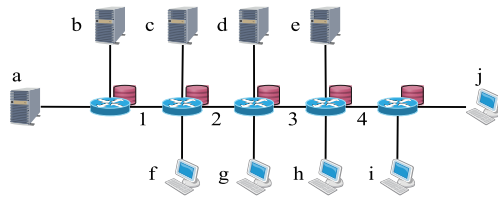


Fig. 4 Mesh network model.

3. Evaluation

3.1 Evaluation in Time

3.1.1 Effect of SAR on Smoothing Link Utilization

In this section, we evaluate the effectiveness of time-dimensional traffic engineering by SAR using a simple mesh model as an accumulated session model. We compare SAR under two design policies.

1. SAR-edge

In SAR-edge, the sender, i.e., the content server, can schedule its transmission time to minimize the total cost of link use and storage on the path. In SAR, all routers on the path can store content files; however, in SAR-edge, no router can store content files.

2. Fastest Reservation (FR)

At the time a request is generated, transmission of the content file is scheduled as fast as possible on condition that link usage of the scheduled traffic is below link capacity.

Figure 4 shows a mesh model with five senders ($a - e$), five receivers ($f - j$), and five tandem routers. In this model, the capacity of each link is 1, and link usage for each session is simply given by the fixed value 0.03. Sessions are set up for five server-receiver pairs: (a, j) , (b, f) , (c, g) , (d, h) , and (e, i) . The last four pairs were used to design the traffic before considering the first pair, so the designed traffic for these four pairs becomes the accumulated SAR traffic for the first pair. First, 1200 sessions (300 sessions for each pair) with short time periods between requests and deadlines were designed for the last four pairs. Based on the Internet traffic report by the Ministry of Internal Affairs and Communications [4], we assumed that the duration of online video viewing by users had a shape similar to the normal distribution. Specifically, deadlines were normally distributed with a mean value given in Table 2 and a standard deviation of 7.5. The average value of the request time was 20 time units earlier than the deadline. Requests were generated according to an exponential distribution. In Table 2, server e has the earliest mean time for its requests for pattern α and the latest one for pattern β .

After the accumulated data for SAR traffic were generated, node j transmitted a request to Server a to start its

Table 2 Mean value of normal distribution.

server-receiver pair	pattern α mean value	pattern β mean value
$b - f$	80	20
$c - g$	60	40
$d - h$	40	60
$e - i$	20	80

1200 sessions. The deadlines of these 1200 sessions were normally distributed with mean 80 and standard deviation of 7.5. Request times were exponentially distributed between 0 and the deadline. Figures 5 and 6 show link utilization for the initially accumulated SAR traffic (1200 sessions: red lines) and for a total of 2400 sessions (blue lines).

In this evaluation, the initially accumulated SAR traffic was generated by 1200 prior sessions of SAR traffic with tightly assigned deadlines. The red lines in the top rows of panels in Fig. 6 (Panels (a) SAR) shows that SAR could not smoothly schedule the 1200 traffic sessions after the first 1200. This was because the times between requests and deadlines for the initial 1200 sessions, which served as accumulated SAR traffic, were quite short. The request times and deadlines among nodes $f - i$ were also cooperatively dependent. [†]The tight deadline assignment in this evaluation is just for investigating SAR behavior when extreme constraints were imposed on the accumulated SAR traffic.

One time unit is defined as the time needed to transmit the maximum number of content files through a (one-hop) link. Suppose that a 0.5 Gbps bandwidth is assigned for SAR traffic and 33 content files of 1 GB each are transmitted in parallel, then the time unit is about 7.2 min. So, in Figs. 5 and 6, 100 time units on the horizontal axis correspond to 12 hours. As shown in Fig. 5(a), for the initially accumulated SAR traffic in pattern α , SAR can schedule the later 1200 sessions so as to smooth link utilization. In SAR-edge (Fig. 5(b)), the high utilization periods used to obtain the initially accumulated SAR traffic for each link remain high-utilization periods after the last 1200 sessions were accommodated. This was because, in SAR-edge, no router can time shift traffic; hence, traffic peaks occurring at different times at each link cannot be avoided by temporal scheduling only at the sender side. The panels in Fig. 5(c) show that since FR starts transmitting as soon as possible, several time periods occur with extremely high utilization of links.

As shown in Figs. 6(b) and 6(c), SAR-edge and FR perform poorly in terms of smoothing link utilization. In pattern β , SAR (Fig. 6(a)) also shows poor performance. This is because, even with temporal scheduling inside a network, periods of high utilization cannot always be avoided. Even when a session can be scheduled to avoid the peak traffic period at one router, at other subsequent routers, high utilization might not be avoidable. Figure 7 simply explains

[†]In real situations using SAR, request times are around lunchtime or morning commute times and deadlines are around primetime, e.g., after 1900. So there would be plenty of time for SAR requests to be made. Further, in real situations, end users independently reserve content and set deadlines.

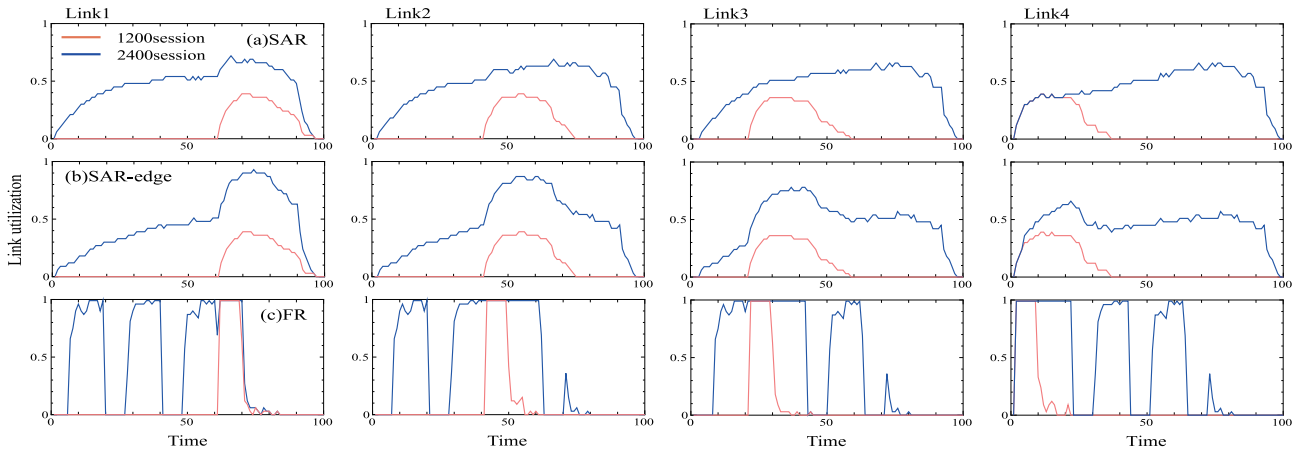


Fig. 5 Link utilization - pattern α .

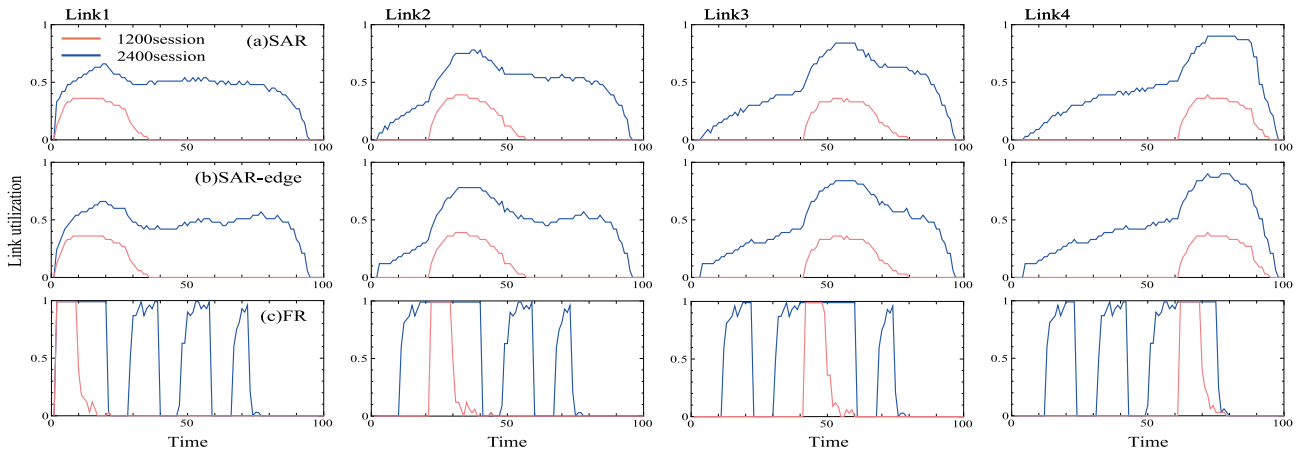


Fig. 6 Link utilization - pattern β .

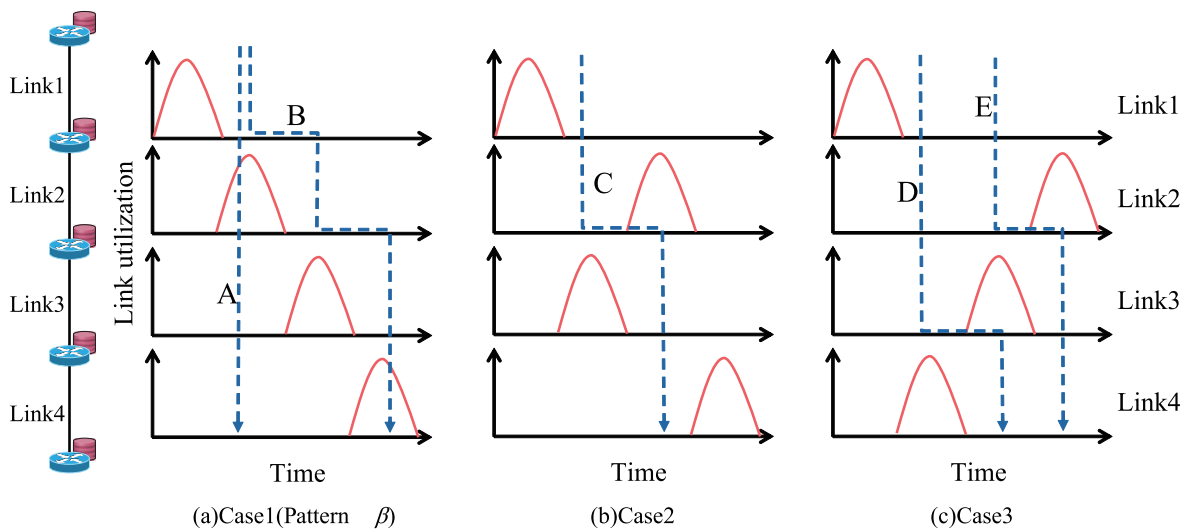


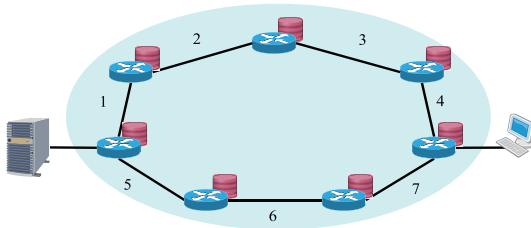
Fig. 7 Effective scheduling of SAR in several peak traffic patterns.

this situation. As shown in Fig. 7(a), when peak traffic periods at different routers form a linear temporal sequence, as in pattern β , no path in a space-time diagram can be found

that will avoid peak periods at all links. However, this can happen only when peak periods form strict linear sequences in time. If such a sequence of peak periods is broken by even

Table 3 Number of rejected sessions.

	pattern α	pattern β
SAR	4	224
SAR-edge	255	243
FR	657	669

**Fig. 8** Ring model.

one pair of links (see Fig. 7(b)), a path exists that will avoid use of all routers during their peak periods. This means that, generally, SAR will find a good path on a space-time diagram. Even when a path has a linear temporal sequence of peak periods, changing the routing in space might find a better path.

3.1.2 Numbers of Rejected Requests

In this section, we evaluate SAR in terms of the number of rejected requests. After accommodating the 2400 sessions described in Sect. 3.1.1, another 1200 sessions were injected. These 1200 sessions were designed between the pair $[a, j]$ using the same distributions of request times and deadlines as in Sect. 3.1.1. Table 3 shows the number of rejected requests for session generation and deadline distributions of patterns α and β . In both cases, SAR had the lowest number of rejected requests. So, SAR can design file transmission times so as to leave more room for upcoming sessions. This means that SAR can allow networks to accommodate traffic generated by users who want to watch videos immediately.

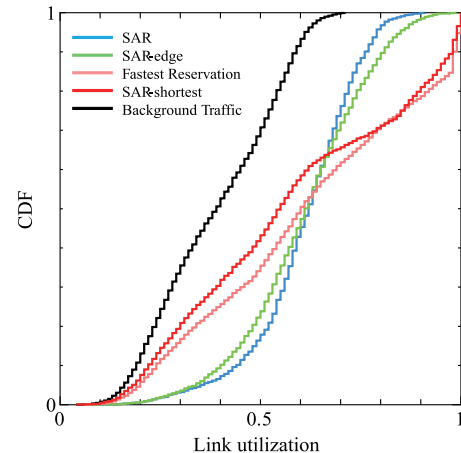
3.2 Evaluation in Time and Space

3.2.1 Homogeneous Link Capacity Model

In this section, we evaluate the efficiency of SAR in both time and space. We used a ring model with one sender, one receiver, and seven routers, as shown in Fig. 8. For SAR-edge and FR, the minimum-cost path to the receiver was first selected[†] and then scheduling by SAR-edge and FR was applied to this path. In this evaluation, we add SAR-shortest as a comparable design policy. In SAR-shortest, the minimum hop route is selected first, and SAR scheduling in time is applied to this route.

In this model, background traffic was preloaded on each link. In Sect. 3.1, we generated background traffic

[†]Definition of link cost is the same as in the previous subsection.

**Fig. 9** CDF of unit-time link utilization (ring model).

using artificially generated deadlines to investigate the potential and limitations of SAR. In this section, we evaluate SAR using more general and realistic background traffic; *BackgroundTraffic* in Algorithm 1 was generated as described below (the form of *BackgroundTraffic* is defined in (1)). As reported by Fraleigh et al. [16], fluctuations in network traffic generally have two different time granularities: wavy fluctuations over long times and bumpy fluctuations over short times. To emulate both, background traffic was taken to be the superposition of a sine curve and normally distributed random variables generated in each unit of time. The standard deviation for this normal distribution was 0.05. Wave-shaped traffic fluctuations on a daily scale have been generally reported [4], so the mean value of the normal distribution was varied along a sine curve as follows:

$$0.2 \cdot \sin\left(\frac{2\pi}{T}t - \varphi\right) + 0.4 \quad (1)$$

The mean of background traffic was set to 0.4, and the maximum fluctuation in background traffic, i.e., the amplitude of the sine curve, was set to 0.2. The period T of the sine curve was 100 time units. Peak traffic periods in business and residential areas generally differ because of differences in the day and night populations. So, the phase of the sine curve, φ , was randomly assigned to each link to avoid loss of generality with regard to variations in population. For the deadline and request distributions, we used the same assumptions as above, i.e., deadlines were normally distributed with mean 80 and standard deviation 7.5, and requests were exponentially distributed between 0 and the deadline.

Figure 9 shows cumulative distribution functions (CDFs) for unit-time link utilization in each method. The CDF for background traffic is also shown in the figure. The figure shows that the maximum link utilization is 1 for both FR and SAR-shortest; this means that both these methods rejected requests in their traffic designs. Further, the link utilizations of FR and SAR-shortest have flatter CDFs than SAR-edge and SAR. This means SAR-edge and SAR both enable smooth link utilization distributions.

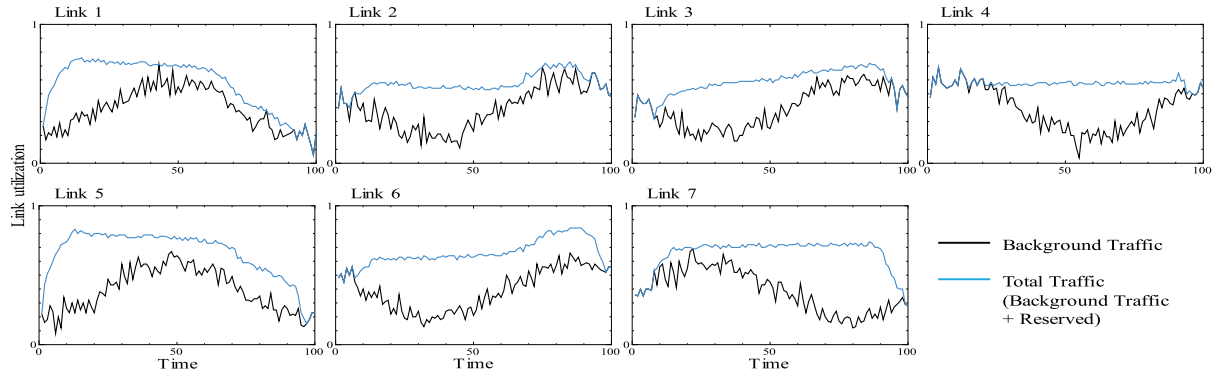


Fig. 10 Link utilization - SAR.

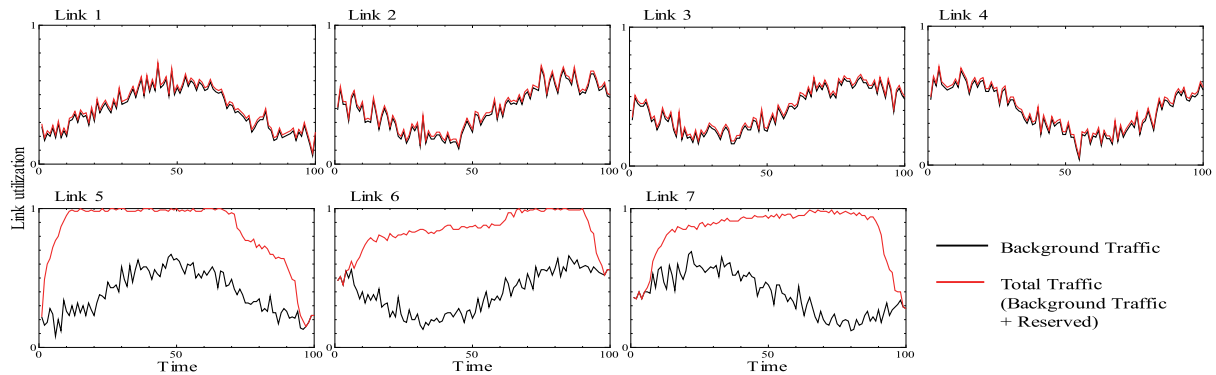


Fig. 11 Link utilization - SAR-shortest.

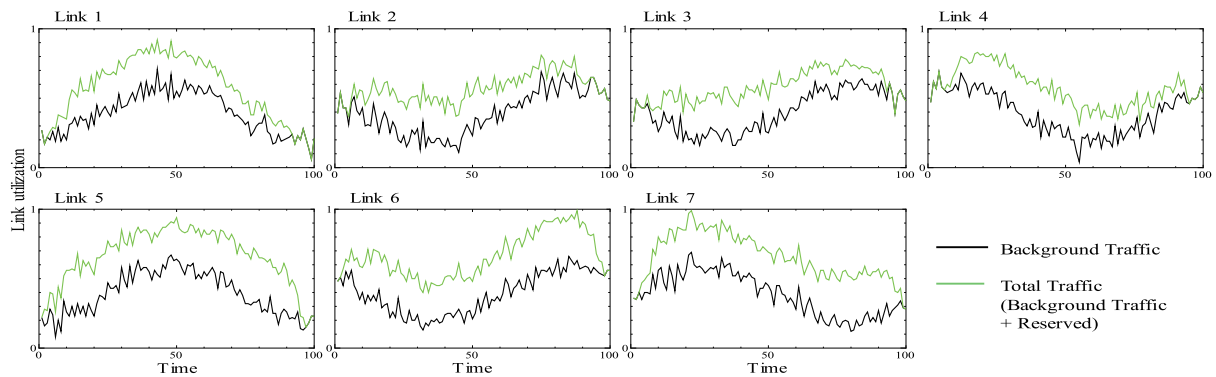


Fig. 12 Link utilization - SAR-edge.

Figures 10–13 show temporal evolutions of link utilization at each link for SAR, SAR-shortest, SAR-edge, and FR, respectively. In each figure, variations in link utilization for background traffic are shown as black curves. In SAR-shortest (Fig. 11), only the shortest path was selected in the space dimension, then temporal scheduling by SAR was applied. Therefore, effective traffic distributions in space cannot be realized using SAR-shortest; this caused the high link utilizations appearing for links 5, 6, and 7 in Fig. 11. For FR in Fig. 13, the fastest transmission time was selected in the time dimension, meaning that traffic could be distributed effectively in space but not in time.

When peak periods on each link differ, as in Fig. 7, SAR-edge cannot avoid using peak periods for requests. So,

SAR-edge cannot accommodate traffic smoothly (Fig. 12). In contrast, in many cases, SAR can select an effective spatio-temporal path that can smoothly absorb not only long-term fluctuations (sine curve fluctuation) but also short-term fluctuations, producing the flat traffic distribution shown in Fig. 10.

3.2.2 Heterogeneous Link Capacity Model

Thus far, we have evaluated our proposed SAR method using a homogeneous link capacity model. In this section, SAR is evaluated using a heterogeneous link capacity model. Figure 14 shows temporal evolution of link utilization at each link when SAR is applied to the model in Fig. 8

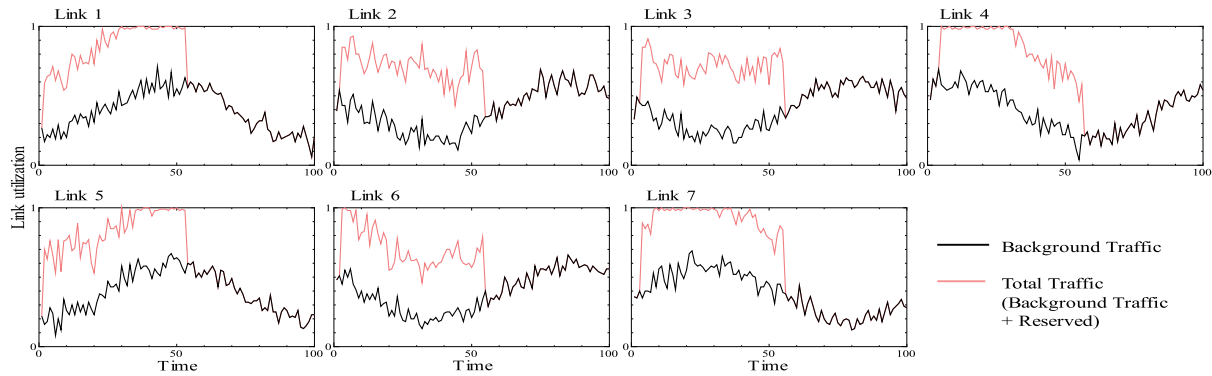


Fig. 13 Link utilization - FR.

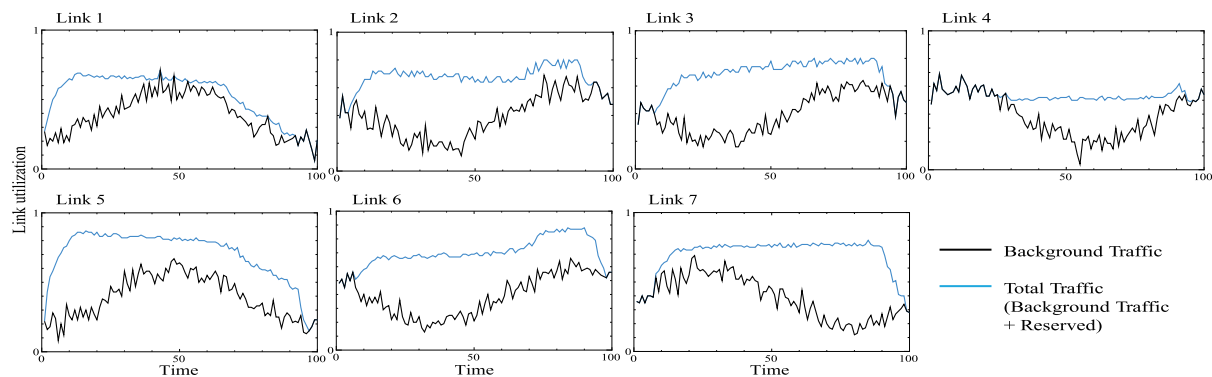


Fig. 14 Link utilization - SAR (heterogeneous link capacity).

with the capacities of links 2 and 3 each reduced by half. Background traffic for these two links was also halved. Figure 14 shows that utilization of links 2 and 3 increased because of their reductions in capacity. However, the increases were not critical because traffic was adequately distributed along other paths (via links 1, 2, 3, and 4), so link utilization remained moderate. These results show that SAR can distribute network traffic adequately in both time and space even when capacities of network links differ.

4. Conclusions

In this paper, we proposed a new traffic engineering technique, SAR, for content distribution. When there is a time difference between a user request for content and user consumption of the content, the user is satisfied so long as content is received before the time of intended consumption. Routers with large storage capacities can schedule their own transmission times. SAR exploits this temporal traffic control to schedule the file transmission time at each router; SAR also uses conventional spatial traffic control; i.e., it selects an optimal path for routing. The numerical examples discussed here show that SAR can, in general, accommodate traffic to smooth link utilization at all links throughout a network. However, there is one limiting condition in which SAR cannot avoid routing content during high utilization periods. This limiting condition occurs when high link utilization periods form a linear temporal sequence along

the path. Nevertheless, such a situation rarely occurs; thus, SAR generally improves overall transmission times. Even if this situation does occur along a certain path, the problem is purely temporal, so spatial re-routing can usually find a better path. Thus, the combination of spatial routing and temporal scheduling is more effective than either one alone. Our future work will focus on avoiding congestion at router storage devices, and on implementation of multicast addressing to those storage devices.

Acknowledgment

This research is partly supported by the National Institute of Information and Communications Technology, Japan.

References

- [1] Cisco Visual Networking Index: Forecast and Methodology, 2012-2017, May. 2013. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html [retrieved: Jan. 2014]
- [2] I. Lazar and W. Terill, "Exploring content delivery network," *IEEE IT Professional*, vol.3, no.4, pp.47-49, 2001.
- [3] Y. Wang and Z. Wang, "Explicit routing algorithms for Internet traffic engineering," *Proc. IEEE ICCCN*, pp.582-588, Oct. 1999.
- [4] Grasp of total traffic amount on the Internet in Japan, Ministry of Internal Affairs and Communications, http://www.soumu.go.jp/main_content/000211328.pdf
- [5] G. Tyson, S. Kaune, S. Miles, Y. El-khatib, A. Mauthe, and A.

- Taweel, "A trace-driven analysis of caching in content centric networks," Proc. IEEE ICCCN, pp.1–7, July 2012.
- [6] S. Paul, R. Yates, D. Raychaudhuri, and J. Kurose, "The cache-and-forward network architecture for efficient mobile content delivery services in the future internet," First ITU-T Kaleidoscope Academic Conference, pp.367–374, May 2008.
- [7] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K.H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," Proc. ACM SIGCOMM, pp.181–192, Aug. 2007.
- [8] V. Jacobson, D. Smetters, J. Thornton, M. Plass, N. Briggs, and R. Braynard, "Networking named content," Proc. ACM CoNEXT, pp.1–12, Dec. 2009.
- [9] K. Fall, "A delay-tolerant network architecture for challenged Internets," Proc. ACM SIGCOMM, pp.27–34, Aug. 2003.
- [10] S. Jain, S. Gopinath, and D. Raychaudhuri, "STAR: Storage aware routing protocol for generalized Delay Tolerant Networks," Proc. IEEE WoWMoM, pp.1–4, Aug. 2011.
- [11] N. Somani, A. Chanda, S.C. Nelson, and D. Raychaudhuri, "Storage aware routing protocol for robust and efficient services in the future mobile Internet," Proc. IEEE ICC, pp.5849–5853, June 2012.
- [12] N. Laoutaris, M. Sirivianos, X. Yang, and P. Rodriguez, "Inter-datacenter bulk transfers with NetStitcher," Proc. ACM SIGCOMM, pp.74–85, Aug. 2011.
- [13] S. Yamashita, T. Yagyū, and M. Yamamoto, "Evaluation of time dimensional traffic engineering with storage aware routing," Proc. Fifth International Conference on Advances in Future Internet, pp.21–26, Aug. 2013.
- [14] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling innovation in campus networks," ACM SIGCOMM Computer Communication Review, vol.38, no.2, pp.69–74, 2008.
- [15] K. Suzuki, K. Sonoda, N. Tomizawa, Y. Yakuwa, T. Uchida, Y. Higuchi, T. Tonouchi, and H. Shimonishi, "A survey on OpenFlow technologies," IEICE Trans. Commun., vol.E97-B, no.2, pp.375–386, Feb. 2014.
- [16] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level traffic measurements from the sprint IP backbone," IEEE Communications Society, vol.17, no.6, pp.6–16, Nov. 2003.



Miki Yamamoto received his B.E., M.E., and Ph.D. in communications engineering from Osaka University in 1983, 1985, and 1988. He joined the Department of Communications Engineering at Osaka University in 1988. He moved to the Department of Electrical Engineering and Computer Science of Kansai University in 2005, where he is a professor. He visited the University of Massachusetts at Amherst in 1995 and 1996 as a visiting professor. His research interests include multicast communications, high-speed networks, wireless networks, and the evaluation of performance of these systems. He is a member of IEEE, ACM, and IPSJ.



Shigeyuki Yamashita received B.E. from Kansai University in 2012 in Electrical and Electronic Engineering. He is currently a student of Graduate school, Kansai University. His research interests include content distribution in new generation networks.



Tomohiko Yagyū received his B.E. and M.E. from Osaka University in 1993 and 1995, respectively. He received his Ph.D. from Tsukuba University in 2009. He is a principal researcher at Cloud System Research Laboratories in NEC Corporation. His research interests include ad hoc network, delay tolerant network and information centric network. He is a member of IEEE.