

# Deadline-aware Data Plane for Internet Video

Junchen Jiang, and Hui Zhang  
Computer Science Department, Carnegie Mellon University  
5000 Forbes Avenue, Pittsburgh, PA, USA  
junchenj@cs.cmu.edu

## ABSTRACT

Internet video requires large-scale content-centric multicast and is highly latency-sensitive. The recently proposed Named Data Network (a.k.a., NDN) is promising in supporting content-centric multicast applications, but is lack of deadline-aware processing in which packets are processed differently according to its processing deadline. In this paper, we extend NDN data plane to support deadline-aware packet processing. In particular, we propose to add a new attribute called “deadline” in the NDN data name, give concrete forwarding strategies that differentiate packets by their deadline attribute, and show its potentials in reducing latency.

## Categories and Subject Descriptors

C.2.1 [Computer-Communication Networks]: Network Architecture and Design—*Network communications*

## Keywords

Internet video, Named Data Network

## 1. INTRODUCTION

In today’s Internet, video traffic is becoming dominant and the trend would not be slowed as videos with higher resolution are viewed on more devices more frequently through Internet. The uniqueness of Internet video lies in the high requirement of content-centric multicast and latency-sensitivity. There has been many approaches to Internet video, including application-layer (e.g., [8]) and transport-layer solutions, service-aware routing, and more recently, clean-slate designs (e.g., NDN [6]). However, it is still unclear how to meet multicast and latency-sensitivity at the same time.

On the one hand, scalable multicast will only be more challenging with the increasing number of devices and viewers. The emerging technique of HTTP-based video (a.k.a., DASH) which leverages existing CDN infrastructure is an IP-based approach that uses CDN caching infrastructure to realize efficient multicast. However, despite of the deployability, DASH uses TCP/IP as the underlying protocol, which is not content-centric. In contrast, NDN provides content-centric routing and forwarding by replacing IP with named data so that any end points requesting for the same data name or

any machine hosting the data could be identified and aggregated by any node in the network.

On the other hand, video is highly latency-sensitive, which means literally, every packet has a processing “deadline” after which the penalty could be greatly increased. For example, Live video has relatively small buffer length (about 4-8 seconds), so any delay of a packet could cause re-buffering issues. Thus, each packet of the next video frame must arrive before the buffer is drained. NDN reduces the retrieval latency by flexible forwarding strategies that send named data requests (i.e., Interest packets) selectively via multiple interfaces in the hope of reaching the content as early as possible. But it does not differentiate packets by their deadlines. For example, when multiple interfaces are available but all with low bandwidth, a latency-sensitive request should be sent on multiple interfaces simultaneously while others should try one interface after another to save bandwidth. Therefore, the forwarding strategy must first be aware of the deadline of each packet, and treat them differently according to deadline.

In this paper, we extend NDN data plane to support deadline-aware packet processing while preserving content-centric multicast. In particular, we add a new attribute, “deadline”, into NDN data name, and show the potentials of deadline-aware forwarding strategies. For space limit, readers may refer to [6] for details of NDN.

## 2. VIDEO TRAFFIC IS DEADLINE-AWARE

The challenge of latency-sensitive application is the nonlinear penalty function of latency, which increases sharply after a turning point of latency. For video, the turning point could be the buffer length, since the viewing quality will be greatly lowered if the next frame comes after buffer is drained. Not only video content has a deadline, so does other traffic of a video session. Taking DASH player as an example, it also downloads a metadata file, called manifest file, from server before fetching the first frame. The manifest file must be fetched within a certain deadline, as this fetch time is included in the join time (i.e., the time between click and play starts) of a video, which has negative impact on the viewer engagement [2].

However, not every packet in video traffic is latency-sensitive. For example, it would be better to forward the packet with same priority as other packets when buffer is full, for purpose of minimizing impact on other traffics and estimating real network condition.

## 3. WHERE TO CHANGE IN NETWORK

This section crystallizes the difference between our approach and some alternatives.

**Application-layer approaches:** HTTP-based [8] and RTMP-based [1] video are representative solutions. They are more readily deployable, but the major drawback is that they have no control over la-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CoNEXT Student’12, December 10, 2012, Nice, France.

Copyright 2012 ACM 978-1-4503-1779-5/12/12 ...\$15.00.

tency, and delegate the improvement of latency to transport layer. The early work of realtime services on NDN (e.g., VoCCN [5]) only achieves comparable performance with IP-based services.

**Application-aware network:** [10] exploits static knowledge of traffic pattern of an application to improve performance compared to general purpose routing algorithms. However, all packets in video traffic are not with equal latency-sensitivity as shown in last section, so being application-aware is not flexible enough.

**Transport-layer approaches:** Examples include  $D^2$ CTCP [9] and  $D^3$  [11] that shortens the latency in data center environment and [3] that reduces the web service latency. Its major drawback is the lack of multicast support in IP-based data plane.

**Deadline-aware data plane:** Our approach is to express the deadline in every packet header and make NDN packet forward strategy deadline-aware. It has three advantages.

First, NDN offers content-centric multicast by replacing location-dependent IP address with content-dependent data name as a new narrow waist. Moreover, this approach can be used together with application-layer approaches like DASH by transporting HTTP protocol on top of NDN as demonstrated in [6]. Second, by expressing deadline in data name, we also enable fine granular packet processing for different deadlines, and avoid the use of static application knowledge as in application-aware routing. Finally, NDN enables flexibility in forwarding strategy of a router, offering opportunities to differentiate packet with various deadlines.

#### 4. DEADLINE-AWARE DATA PLANE

This section briefly describes the deadline-aware design of forwarding strategy in NDN.

**Extended name and deadline update:** First, we extend the data name by adding a new prefix called deadline to the head of a data name. Let  $N$  denote the extended name,  $D(N)$  be its deadline field, and  $S(N)$  be original data name without deadline (Figure 1). The extended prefix is a number indicating how long it can wait for being satisfied. For example, when a video player sends an interest packet for part of the next frame, the deadline value attached to data name could be initially the buffer length. Of course, when traversing the network, the deadline must be updated just like TTL (time to live) field in IP header. When an interest packet is sent through an interface, the deadline value is decreased by the time spent in the node. Accordingly, the FIB (Forward Information Base) should also be keyed by the extended names of which the deadline field refers to the expected time to satisfy the request if the corresponding forwarding strategy is used (Figure 1 shows an example).

**Longest prefix matching:** Second, since NDN forwarding is based on longest prefix matching, an immediate problem is how to support the longest prefix matching of two extended names. Consider  $N_1$  in interest packet and  $N_2$  in FIB. On one hand, longer the common prefix of  $S(N_1)$  and  $S(N_2)$  is, larger the similarity they have is. On the other hand, the closer (but also smaller)  $D(N_2)$  to  $D(N_1)$  is, the better the forwarding strategy is fit to the deadline. So with  $S(\cdot)$  and  $D(\cdot)$ , we have two sorted lists of FIB keys. We consider the case when they are not conflicting, i.e., there is one best choice for both. It is viable if the  $D(N)$  values of any  $S(N)$  in the FIB keys are the same (e.g., in Figure 1, the deadline values associated for each original name are always 10sec, 1sec and 100ms). The figure also shows the best choice of two example requests).

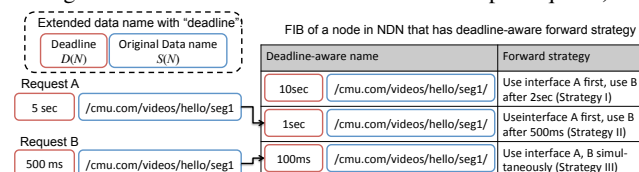


Figure 1: Extended name, and longest prefix matching

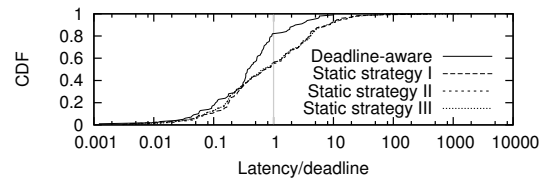


Figure 2: Deadline-aware forwarding strategy has higher deadline-hit rate compared to static forward strategies.

**Deadline-aware forwarding strategy:** We use simulation to show the potentials of deadline-aware forwarding strategy. We consider one node with the FIB in Figure 1. The node has 2 outgoing interfaces A and B, each with a link of average throughput of 1kbps to a server. The node receives randomly generated interest packets (100ms interval) each with a deadline drawn from a Poisson distribution with an average of 1sec. Each packet is forwarded by looking up the FIB and then following the forwarding strategy. The fetching delay is link throughput divided by number of packets on the link (i.e., fair sharing). The performance metric is the fetching delay divided by the deadline: above one means the deadline is missed. We compare the performance of deadline-aware forwarding strategy with static strategies. Figure 2 shows the deadline-aware strategy provides a much lower deadline-miss (82%) rate than any static strategy (55%).

#### 5. CONCLUSION AND FUTURE WORK

Motivated by the challenge of multicast and latency-sensitivity of Internet video, this paper proposes a deadline-aware design of data plane in NDN that improves latency-sensitive services like Internet video while preserving content-centric multicast.

The proposed approach also raises challenges. By extending NDN data name, it also complicates the narrow waist for which we already have many candidates, including IP, named data [6] and HTTP [7]. This can be potentially addressed by the recently proposed XIA [4] platform. Also, it will be interesting to see the impact to other non-deadline-aware traffic and the overhead of implementation. Finally, the deployment issue can be partly addressed by only deploying such data plane to where deadline is most likely to be missed (e.g., last mile).

#### 6. ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation under awards CNS-1040757, CNS-1040800, and CNS-1040801.

#### 7. REFERENCES

- [1] Real-time messaging protocol. [www.adobe.com/devnet/rtmp.html](http://www.adobe.com/devnet/rtmp.html).
- [2] F. Dobrian, A. Awan, D. Joseph, A. Ganjam, J. Zhan, V. Sekar, I. Stoica, and H. Zhang. Understanding the impact of video quality on user engagement. *SIGCOMM-Computer Communication Review*, 41(4):362, 2011.
- [3] N. Dukkipati, T. Refice, Y. Cheng, J. Chu, T. Herbert, A. Agarwal, A. Jain, and N. Sutin. An argument for increasing tcp's initial congestion window. *ACM SIGCOMM Computer Communication Review*, 40(3):27–33, 2010.
- [4] D. Han, A. Anand, F. Dogar, B. Li, H. Lim, M. Machado, A. Mukundan, W. Wu, A. Akella, D. Andersen, et al. Xia: Efficient support for evolvable internetworking. *Proceedings of NSDI 2012*, 2012.
- [5] V. Jacobson, D. Smetters, N. Briggs, M. Plass, P. Stewart, J. Thornton, and R. Braynard. VoCCN: voice-over content-centric networks. In *Proceedings of the 2009 workshop on Re-architecting the internet*, pages 1–6. ACM, 2009.
- [6] V. Jacobson, D. Smetters, J. Thornton, M. Plass, N. Briggs, and R. Braynard. Networking named content. In *Proc. CoNext*, pages 1–12. ACM, 2009.
- [7] L. Popa, A. Ghodsi, and I. Stoica. Http as the narrow waist of the future internet. In *Proc. HotNet*, page 6. ACM, 2010.
- [8] I. Sodagar. The mpeg-dash standard for multimedia streaming over the internet. *MultiMedia, IEEE*, 18(4):62–67, 2011.
- [9] B. Vamanan, J. Hasan, and T. Vijaykumar. Deadline-aware datacenter tcp (d2tcp). In *Proc. SIGCOMM*, pages 115–126. ACM, 2012.
- [10] M. Veeraraghavan, P. Pancha, and K. Eng. Application-aware routing protocol. In *Proc. IEEE ISCC*, pages 442–448. IEEE, 1997.
- [11] C. Wilson, H. Ballani, T. Karagiannis, and A. Rowtron. Better never than late: Meeting deadlines in datacenter networks. *SIGCOMM-CCR*, 41(4):50, 2011.