第43卷第11期 中国科学技术大学学报 Vol. 43,No. 11
2013年11月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Nov. 2013

Article ID:0253-2778(2013)11-0899-08

SVC or not SVC, that is the question

LI Weiping

(*School of Information Science and Technology, University of Science and Technology of China, Hefei* 230027, *China*)

Abstract: Research on scalable video coding (SVC) has been motivated by the need of transmitting video over the Internet. There have been many good research results on SVC and they are included into international standards too. This paper is an attempt to analyze the likelihood of using SVC for network video applications. From the perspective of the game theory, the conclusion is that SVC is not likely to be used in the current Internet and the future network research has to set the game rules right in order to take advantage of SVC to achieve the global optimum of network video applications.

Key words: scalable video coding (SVC); network; game theory

CLC number: TN919.81 **Document code**: A doi:10.3969/j.issn.0253-2778.2013.11.005

Citation: Li Weiping. SVC or not SVC, that is the question[J]. Journal of University of Science and Technology of China, 2013,43(11):899-906.

是否使用可伸缩视频编码

李卫平

(中国科学技术大学信息科学技术学院,安徽合肥 230027)

摘要:可伸缩视频编码的研究来自于互联网上视频传输的需求.目前已经有了许多有关可伸缩视频编码的研

Received:2013-09-01;**Revised**:2013-10-09

Foundation item:Supported by USTC startup funding.

Biography:LI Weiping, PhD/Prof. He received his B. S. degree from University of Science and Technology of China (USTC) in 1982, and his M. S. and Ph. D. degrees from Stanford University in 1983 and 1988 respectively. In 1987, he joined the Faculty of Lehigh University as an Assistant Professor in the Department of Electrical Engineering and Computer Science. In 1993, he was promoted to Associate Professor with Tenure. In 1998, he was promoted to Full Professor. From 1998 to 2010, he worked in several high-tech companies in the Silicon Valley with technical and management responsibilities. In March of 2010, he was appointed to the position of the Dean for School of Information Science and Technology in USTC. He has been elected to Fellow of IEEE for contributions to image and video coding algorithms, standards, and implementations. He served as the Editor-in-Chief of IEEE Transactions on Circuits and Systems for Video Technology, a Guest Editor for a special issue of IEEE Proceedings, the Chair of several Technical Committees in IEEE Circuits and Systems Society and IEEE International Conferences, and the Chair of Best Student Paper Award Committee for SPIE Visual Communications and Image Processing Conference. He has made many contributions to International Standards. His inventions on Fine Granularity Scalable Video Coding and Shape Adaptive Wavelet Coding have been included into the MPEG-4 International Standard. He served as a member of MPEG (Moving Picture Experts Group) of ISO (International Standard Organization) and an Editor of MPEG-4 International Standard. He served as a founding member of the Board of Directors of MPEG-4 Industry Forum. As a technical advisor, he also made contributions to the Chinese Audio Video coding Standard (AVS) and its applications. He received Certificate of Appreciation from ISO/IEC as a Project Editor in development of International Standard in 2004, the Spira Award for Excellence in Teaching in 1992 at Lehigh University, and the first Guo Mo-Ruo Prize for Outstanding Student in 1980 at USTC. E-mail: wpli@ustc.edu.cn

究成果,并且包括在国际标准中了.本文试图分析在网络视频应用中使用可伸缩视频编码的可能性.从博弈论的角度来看,结论是可伸缩视频编码在现有的互联网中不太可能被用上,未来网络研究必须制定正确的游戏规则,从而能够使用可伸缩视频编码,以取得网络视频应用的全局优化.

关键词:可伸缩视频编码;网络;博弈论

0 Introduction

Scalable video coding (SVC) has been an active research topic for a while[1-2]. It encodes a video sequence into a bitstream that can be partially decoded and the reconstructed video quality from the partially decoded bitstream improves as the number of decoded bits increases. Fig. 1 illustrates the curves of non-scalable video coding (NSVC), layered SVC, and fine granularity scalable (FGS) video coding, as compared with the quality-rate bound. The horizontal axis in Fig. 1 is the channel capacity (not the encoded bit rate) of the channel to transmit the video sequence. The vertical axis of Fig. 1 is the received video quality (not the encoded video quality) through the transmission channel. An NSVC method would encode a video sequence at a given bit rate and result in a certain reconstructed video quality. In Fig. 1, there are three staircase curves illustrating NSVC at three different bit rates. NSVC results in a staircase curve in Fig. 1 because, once the encoded bit rate is set, the reconstructed video quality is determined and the received video quality is not going to improve even when the channel capacity is higher than the encoded bit rate but would become dramatically poor if the channel capacity is lower than the encoded bit rate. Layered SVC is illustrated in Fig. 1 as a staircase curve with two stairs indicating two layers (the number of stairs corresponds to the number of layers). When a video sequence is encoded into a layered SVC bitstream, the number of layers is determined first and the bit rates for the layers are then set. The first layer is called the base-layer and the other layers are called enhancement-layers. Decoding of the enhancement-layers depends on the base-layer and decoding of the higher enhancement-layers depends on the base-layer and the lower enhancement-layers. The reconstructed video quality depends on the number of layers decoded and keeps the same until a higher enhancement-layer is received and decoded. Similar to NSVC, if the channel capacity is lower than the base-layer bit rate, the received video quality becomes dramatically poor. An FGS video coding method is illustrated in Fig. 1 as a continuous curve (more precisely still a staircase curve with many small stairs), indicating that reconstructed video quality improves continuously as the number of bits received and decoded. When encoding a video sequence using an FGS video coding method, an interval of bit rate (instead of a point in NSVC or a set of points in layered SVC) is given. The base-layer of FGS is encoded at the lower bound of the bit rate interval and a single enhancement-layer (or many small incremental enhancement layers) is encoded with the bit rate equal to the length of the interval. The quality-rate bound is shown as a continuous curve in Fig. 1, but it indicates the bound only and does not imply it is achievable by a single bitstream.

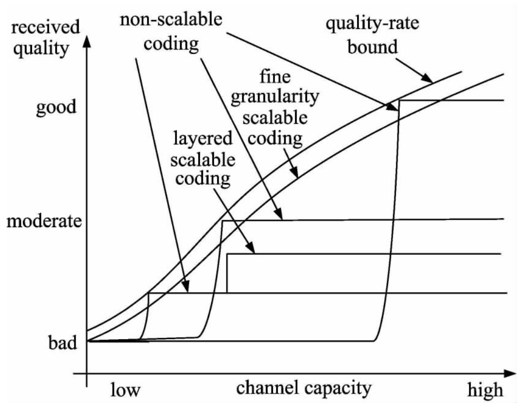

Fig. 1 Illustration of various video coding methods in the context of video transmission

The problem of finding the best SVC method can be formulated as an optimization problem. Let $Q(R)$ be the best quality achievable at the bit rate

R with or without the "partially decodable bitstream" constraint. $Q(R)$ may or may not be the quality-rate bound, depending on whether the bound is achievable or not. Let $Q_{\text{pdb}}(R)$ be the quality of a particular "partially decodable bitstream" (pdb) at the bit rate R. Then $Q(R)-Q_{\text{pdb}}(R)$ is a measure of how good this particular bitstream is at the bit rate R. Since we are concerned about a bit rate interval, we may set our objective as to find a pdb to minimize

$$\max_{R\in[R_l,R_h]}(Q(R)-Q_{\text{pdb}}(R))=D_{\max}$$

over $[R_l, R_h]$. We may also set the objective as to find a pdb to minimize

$$\int_{R_l}^{R_h}((Q(R)-Q_{\text{pdb}}(R))f(R)\mathrm{d}R=D_{\text{ave}},$$

where $f(R)$ is the probability density function of the bit rate distribution over $[R_l, R_h]$. The bit rate distribution may be measured across a set of users or a period of time for a single user or a combination of both.

SVC has become a main-stream research topic and many practical SVC methods have been developed. Some of them have been included in products. As stated in Ref. [1], one of the main applications for SVC was supposed to be video transmission over the Internet. However, until today, SVC is still not widely used in the applications of video delivery over the Internet. There may be many reasons for this. The most cited or debated reason is that the coding efficiency of SVC is still not as good as NSVC in the sense that, for a given bit rate, the video quality of SVC is not as good as NSVC, or put it in another way, for the same video quality, SVC needs a higher bit rate than NSVC. Another possible reason is that the complexity of SVC is higher than NSVC. Yet another possible reason is that the delivery of SVC bitstream over the Internet requires intelligent network devices that can detect channel capacity and determine how to best deliver the SVC bitstream based on the network conditions. In this paper, we try to analyze whether there are more fundamental reasons for using or not using SVC for network video applications. In our analysis, to put aside the above reasons, we assume (Ⅰ) the coding efficiency of SVC is the same as that of NSVC; (Ⅱ) the complexity of SVC is the same as that of NSVC; (Ⅲ) there may be intelligent network devices to detect channel capacity and best deliver SVC bitstreams based on the network conditions.

To focus on the essential discussions and not get side-tracked by any specific video codec or any particular video sequence, we define video quality at bit rates of 1 Mbps, 750 Kbps, 500 Kbps, and 250 Kbps as EXCELLENT, GOOD, FAIR, and MINIMUM, respectively. We also define received video quality as POOR if there are random packet losses due to network channel capacity lower than the video bit rate.

In the following section, we analyze a few much simplified cases and try to understand the likelihood of using SVC versus NSVC. In the analysis, we assume using layered SVC with four layers and each layer uses 250 Kbps bit rate. Since we assume the same coding efficiency for SVC and NSVC, the video quality of SVC at a given bit rate, e. g., at 250 Kbps (base-layer only), 500 Kbps (base-layer plus one enhancement-layer), 750 Kbps (base-layer plus two enhancement-layers), or 1 Mbps (base-layer plus three enhancement-layers), is the same as that of NSVC at the same bit rate. In Section 2, we draw some conclusions based on the analysis in Section 1.

1 Analysis of much simplified cases

In this section, we start with the simplest application of one source sending a video sequence to one receiver. Then we look at an application of one source sending a video sequence to many receivers, in which multicast (not necessarily IP multicast, maybe application layer multicast) is assumed, to save bandwidth. Finally, we consider, in a shared network environment, multiple applications with each having one source sending a video sequence to many receivers.

特约评述

1.1 Single application with one source and one receiver

The application is shown in Fig. 2. A video source is connected with a receiver through a Network Node. The channel capacity between the source and the Network Node is 1 Mbps and between the Network Node and the receiver is 500 Kbps.

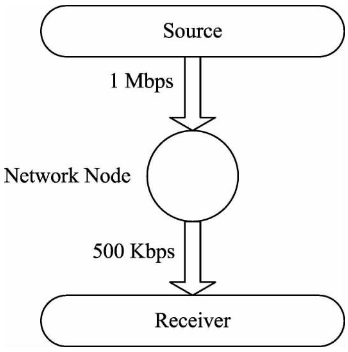

Fig. 2 A one-to-one application

The video source may use NSVC or SVC to code the video sequence. Let's consider the following cases:

(Ⅰ) Video sequence is encoded using NSVC at 1 Mbps

Encoded video at 1 Mbps rate is delivered from Source to the Network Node without any problems, but the Network Node randomly drops packets due to the output channel capacity being only 500 Kbps. Random packet drop results in POOR video quality at the receiver.

(Ⅱ) Video sequence is encoded using NSVC at 500 Kbps

The video Source may estimate the end-to-end channel capacity as 500 Kbps (probably through an end-to-end feedback mechanism) and encodes video to 500 Kbps. Encoded video at 500 Kbps rate is delivered from Source to the Network Node without any problems, and the Network Node forwards all packets to the receiver without any problems either. Received video quality is then FAIR. It does not matter if the Network Node has intelligence or not.

(Ⅲ) Video sequence is encoded using SVC at 1 Mbps, but the Network Node has no intelligence

Same as the first case in NSVC, the Network Node without intelligence randomly drops packets due to the output channel capacity being only 500 Kbps, resulting in POOR received video quality.

(Ⅳ) Video sequence is encoded using SVC at 1 Mbps, and the Network Node has intelligence

The Network Node estimates the output bandwidth as 500 Kbps (probably through a feedback mechanism from the receiver) and forwards only the base layer and the first enhancement layer of the SVC bitstream. Encoded video at 1 Mbps rate is delivered from Source to the Network Node without any problems, and the Network Node intelligently drops two enhancement layers due to output bandwidth being only 500 Kbps. The received video quality is FAIR.

Let's summarize the above discussion into Tab. 1.

Tab. 1 Summary of the discussion on single application with one source and one receiver

video encoding method	encoded bit rate	Network Node intelligence	received video quality
NSVC	1 Mbps	NO	POOR
	500 Kbps	YES or NO	FAIR
SVC	1 Mbps	NO	POOR
		YES	FAIR

In this case, it does not matter if the Network Node has intelligence or not, NSVC and SVC perform the same. Therefore, we may conclude that SVC is not useful and NSVC is good enough.

1.2 Single application with one source and many receivers

To simplify the discussion without loss of generality, let's assume a video source is to send a video sequence to two receivers as shown in Fig. 3.

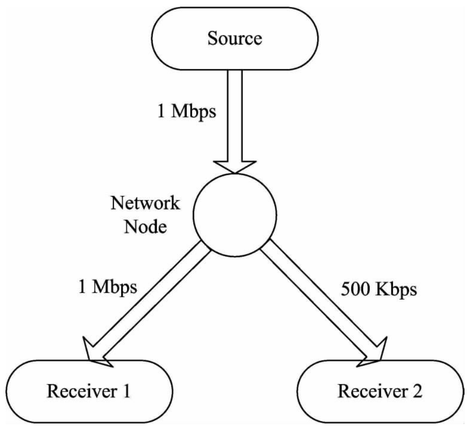

Fig. 3 One video sequence is to be sent to two receivers

The channel capacity between the Source and the Network Node is still 1 Mbps. The channel capacity from the Network Node to Receiver 1 is 1 Mbps and to Receiver 2 is 500 Kbps.

Let's consider the following cases:

(Ⅰ) Video sequence is encoded using NSVC at 1 Mbps

Assume that multicast (may not necessarily be IP multicast, could be application layer multicast) is used in delivering the encoded video bitstream from the Source to the two Receivers. The encoded video bitstream is delivered from the Source to the Network Node without any problems. The Network Node is to forward the video bitstream to the two Receivers. For Receiver 1, the bandwidth from the Network Node is 1 Mbps and the video bitstream is received without any packet losses. Therefore, Receiver 1 receives EXCELLENT video quality. On the other hand, there are packet losses from the Network Node to Receiver 2 due to its bandwidth being only 500 Kbps. Therefore, Receiver 2 receives POOR video quality.

(Ⅱ) Video sequence is encoded using NSVC at 500 Kbps

The Source may estimate the end-to-end bandwidth (probably through an end-to-end feedback mechanism) and encode the video sequence to 500 Kbps to avoid packet losses when delivering the video bitstream to Receiver 2. Since the same video bitstream at 500 Kbps is delivered to both Receivers, Receiver 1 also receives FAIR video quality as Receiver 2, although the bandwidth from the Network Node to Receiver 1 is 1 Mbps (network bandwidth resource is not fully utilized). Again, it does not matter if the Network Node has intelligence or not.

(Ⅲ) Video sequence is encoded using SVC at 1 Mbps, but the Network Node has no intelligence

Assume that the video is encoded into 4 layers using SVC with each layer using 250 Kbps so that the total bit rate is 1 Mbps. Same as in NSVC at 1 Mbps, Receiver 1 receives EXCELLENT video quality, but Receiver 2 receives POOR video quality due to random packet drops from the Network Node to Receiver 2 since the Network Node has no intelligence.

(Ⅳ) Video sequence is encoded using SVC at 1 Mbps, and the Network Node has intelligence

Since the Network Node has intelligence in this case, it may estimate the bandwidth to Receiver 2 as 500 Kbps and drops two higher enhancement layers. Therefore, Receiver 2 receives FAIR video quality while Receiver 1 receives EXCELLENT video quality.

Let's summarize the above discussion into Tab. 2.

Tab. 2 Summary of the discussion on single application with one source and multiple receivers

video encoding method	encoded bit rate	Network Node intelligence	Receiver 1 video quality	Receiver 2 video quality
NSVC	1 Mbps	NO	EXCELLENT	POOR
	500 Kbps	YES or NO	FAIR	FAIR
SVC	1 Mbps	NO	EXCELLENT	POOR
		YES	EXCELLENT	FAIR

In this case, if the Network Node has no intelligence, NSVC and SVC perform the same. However, if the Network Node has intelligence, SVC performs better than NSVC. Therefore, it seems that the conclusion should be to ALWAYS use SVC! This has been an argument for developing SVC all along.

1.3 Multiple applications with each having one source and many receivers

On the Internet, there are many applications running at the same time. Therefore, considering them together is closer to the real situation than considering only a single application. To simplify the discussion, let's consider two video applications with each having a video source to be delivered to two receivers through the same Network Node as shown in Fig. 4. The channel capacity between each source and the Network Node is still 1 Mbps. The channel capacity from the Network Node to Receiver 1s of the two applications is 2 Mbps and to Receiver 2s of the two applications is 1 Mbps.

Let's consider the following cases:

特约评述

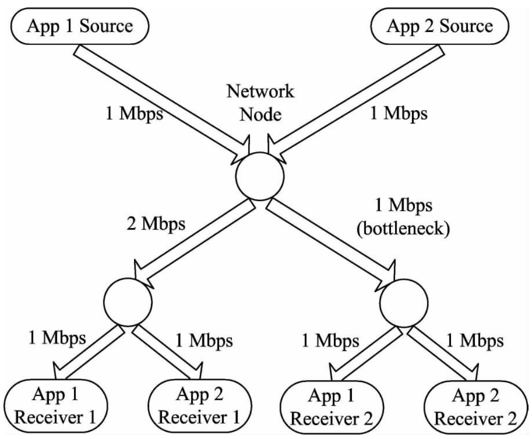

Fig. 4 Two applications with each having one source and two receivers connected through the same Network Node

(Ⅰ) Both applications use NSVC to encode video at 1 Mbps for each sequence

The two video sequences from both applications are delivered to the Network Node without any problems. The Network Node forwards the two video bitstreams to Receiver 1s of both applications without any problems either, since the bandwidth from the Network Node to the two Receivers is 2 Mbps. However, there are random packet drops when the Network Node forwards the two video bitstreams to Receiver 2s of both applications, since the total bandwidth from the Network Node to these two Receivers is only 1 Mbps. Therefore, in this case, Receiver 1s of both applications receive EXCELLENT video quality, but Receiver 2s of both applications receive POOR video quality.

(Ⅱ) Both applications use NSVC to encode video at 500 Kbps for each sequence

Assume that both applications may estimate the end-to-end bandwidth (again, possibly through an end-to-end feedback mechanism) and encode their video bitstrams to 500 Kbps each. Then, there are no problems of delivering both video bitstreams to both receivers of the two applications. All four receivers receive FAIR video quality.

(Ⅲ) Both applications use SVC to encode video at 1 Mbps for each sequence

Similar to the case of using NSVC at 1 Mbps, Receiver 1s of both applications receive EXCELLENT video quality. Assume that the Network Node has intelligence to drop two layers of each SVC bitstream when it forwards the SVC bitstreams to Receiver 2s of the both applications. Then Receiver 2s of both applications receive FAIR video quality.

(Ⅳ) Application 1 uses NSVC and application 2 uses SVC to encode video at 1 Mbps for each sequence

From the above three cases, it seems that all applications should use SVC without any doubt, since SVC results in (EXCELLENT, FAIR) video quality for the two receivers of both applications while NSVC results in either (EXCELLENT, POOR) or (FAIR, FAIR) video quality for the two receivers of both applications. However, in reality, each application makes its decision on which video encoding method to use independently of other applications and actually in competition with other applications for the network bandwidth resource. Therefore, it is interesting to look into what happens if some applications use NSVC and others use SVC. In this case, Receiver 1s of both applications receive EXCELLENT video quality. The question is what happens to Receiver 2s of both applications. Of course, if the Network Node has no intelligence or the video sources do not lower their bit rate, the random packet drops at the Network Node result in POOR video quality for Receiver 2s of both applications. To improve the situation, the SVC application requires the Network Node to have intelligence and to drop SVC enhancement layers, instead of randomly dropping packets. At the same time, the NSVC application requires the source video encoder to lower its bit rate to avoid random packet loss. NOW, let's think about what the most LIKELY strategy each application uses. For the SVC application, it would ask the Network Node to drop one enhancement layer first to see if the situation improves or not. For the NSVC application, it would ask the source video encoder to lower its bit rate a little bit (maybe by any small number, say 5Kbps) to see if the situation

特约评述

improves or not. In this case, as long as the total bit rate of both bitstreams is more than 1 Mbps, the situation would not improve. Therefore, the SVC application would continue dropping the enhancement layers, one at a time, until only the base layer is left at 250 Kbps. On the other hand, the NSVC application would continue adjusting its encoding bit rate until there is no packet drop anymore, resulting in 750 Kbps bit rate. Although it is hard to analyze such a situation in real Internet applications, the different philosophies of SVC and NSVC strategies would most likely result in what is shown in this simplified case. The SVC application behaves like a "good citizen", whenever there is a traffic congestion, it drops its enhancement layers one by one, until only the base layer is left. The NSVC application behaves like a "greedy gamer" who would not give in until absolutely necessary. In this case, the NSVC application would deliver 750 Kbps bitstream to both Receivers with GOOD video quality and the SVC application would deliver 1 Mbps bitstream to Receiver 1 with EXCELLENT video quality but only 250 Kbps bitstream to Receiver 2 with MINIMUM video quality.

Let's summarize the above discussion into Tab. 3.

We can put this into a simple game formulation as shown in Tab. 4. Since no application wants to have POOR quality for any of its receivers, the NSVC/NSVC case has received video quality of (FAIR, FAIR)/ (FAIR, FAIR) and the NSVC/SVC case has received video quality of (GOOD, GOOD)/ (EXCELLENT, MINIMUM).

Usually, the analysis of such a game looks at whether there are Nash equilibrium points by finding out whether each player has anything to gain by changing only their own strategy unilaterally. Let's look at the case of SVC/SVC first, in which both applications have the received video quality of (EXCELLENT, FAIR). If one of them changed to NSVC unilaterally, the received video quality would become (GOOD, GOOD) for the application, which means that one Receiver's video quality in the application is lowered from EXCELLENT to GOOD and the other Receiver's video quality in the application is increased from FAIR to GOOD. Therefore, it's hard to say that the application had anything to gain and we may consider the SVC/SVC case as a Nash equilibrium point. Now, let's look at the case of NSVC/NSVC in which both applications have the received video quality of (FAIR, FAIR). If one of them changed to SVC unilaterally, the received video quality would become (EXCELLENT, MINIMUM) for the application, which means that one Receiver's video quality in the application is lowered from

特约评述

Tab. 3　Summary of the discussion on multiple applications with each having one source and multiple receivers

video encoding method		encoded bit rate		received video quality	
App 1	App 2	App 1	App 2	App 1	App 2
NSVC	NSVC	1 Mbps	1 Mbps	(EXCELLENT, POOR)	(EXCELLENT, POOR)
NSVC	NSVC	500 Kbps	500 Kbps	(FAIR, FAIR)	(FAIR, FAIR)
SVC	SVC	1 Mbps	1 Mbps	(EXCELLENT, FAIR)	(EXCELLENT, FAIR)
NSVC	SVC	1 Mbps	1 Mbps	(EXCELLENT, POOR)	(EXCELLENT, POOR)
NSVC	SVC	750 Kbps	1 Mbps	(GOOD, GOOD)	(EXCELLENT, MINIMUM)

Tab. 4　A simple game formulation of the discussion on multiple applications with each having one source and multiple receivers

video quality (App 1 Receiver 1, App 1 Receiver 2), (App 2 Receiver 1, App 2 Receiver 2)		App 1 SVC	App 1 NSVC
App 2	SVC	(EXCELLENT, FAIR), (EXCELLENT, FAIR)	(GOOD, GOOD), (EXCELLENT, MINIMUM)
	NSVC	(EXCELLENT, MINIMUM), (GOOD, GOOD)	(FAIR, FAIR), (FAIR, FAIR)

Tab. 5 A change of the game rule results in global optimum for every video application

video quality (App 1 Receiver 1, App 1 Receiver 2), (App 2 Receiver 1, App 2 Receiver 2)		App 1 SVC	App 1 NSVC
App 2	SVC	(EXCELLENT, FAIR), (EXCELLENT, FAIR)	(FAIR, FAIR), (EXCELLENT, FAIR)
	NSVC	(EXCELLENT, FAIR), (FAIR, FAIR)	(FAIR, FAIR), (FAIR, FAIR)

FAIR to MINIMUM and the other Receiver's video quality in the application is increased from FAIR to EXCELLENT. Therefore, it seems that the application would have a gain since there are two-levels of increase from FAIR to EXCELLENT but only one-level of decrease from FAIR to MINIMUM. However, in reality, things are a little more complicated. The Receiver with the video quality increase may take it as granted and not appreciate the video quality increase, but the Receiver with the video quality decrease may complain about it. This would make the application provider think twice to change from NSVC to SVC unilaterally. More importantly, let's look at the off-diagonal boxes, in which the received video quality for the NSVC application is (GOOD, GOOD), but for the SVC application (EXCELLENT, MINIMUM). In real applications, SVC at EXCELLENT may not be too much better than NSVC at GOOD, but NSVC at GOOD is much better than SVC at MINIMUM. Therefore, it is very obvious that any application would choose NSVC, instead of SVC. Just like any game of this kind, every application settles on NSVC, resulting in (FAIR, FAIR) video quality for both applications, while we could have done better with (EXCELLENT, FAIR) video quality for both applications if both use SVC!

2 Conclusion

From the above discussions, we can draw a few conclusions:

(Ⅰ) Under the current Internet mechanism, even the Network Node has the ability to detect channel capacity, SVC is useless, since it would lose to NSVC in the game of competing for the shared channel capacity.

(Ⅱ) To achieve the global optimum of every video application using SVC, the Network Node has to be more intelligent than just detecting channel capacity. For example, in Fig. 4, if the Network Node has the ability to identify the bitstream associated with each application and set the maximum bandwidth usage to be equal for every application, then Receiver 2 of each application would be limited to 500 Kbps and the game would change to Tab. 5. In the off-diagonal boxes of the table, using SVC results in (EXCELLENT, FAIR) quality while using NSVC results in (FAIR, FAIR) quality. Therefore, every application would use SVC, resulting in global optimum. This example shows that, like any such a game, when the rules are set differently, player behaviors are different and global results are different.

(Ⅲ) There are many other types of applications on the Internet. The game between different types of applications in competing for the network resources is more complicated. The rules of the game will determine the behaviors of the applications that will, in turn, determine if the network resources are used to achieve the global optimum for every application. This will be an important part of the future network research.

References

[1] Li W P. Overview of fine granularity scalability in MPEG-4 video standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(3): 301-317.

[2] Schwarz H, Marpe D, Wiegand T. Overview of the scalable video coding extension of the H. 264/AVC standard,[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007,17(9): 1 103-1 120.