Real-Time Video Quality Control For Multimedia Network

Biao Jiang *Tarek Saadawi*
Department of Electrical Engineering
The City University of New York, City College
New York, U.S.A

{bjiang&saadawi}@ccny.cuny.edu

***Abstract*—In this paper, we propose a new approach for video quality control for multimedia networks. Our new approach is based on video quality measure that combines both the network quality of service (QoS) as well as the user quality of experience (QoE). The proposed approach improves the end-to-end traditional video quality control for multimedia network by including the human perception of video data, which is major concern for the video client, along with the network quality of service (QoS) measurements. We will show that the proposed QoS-QoE based video quality control algorithm can reflect both the condition of the network environment and the human perception of the received networked video data stream. According to both QoS and QoE parameters, rather than using only QoS parameter, video quality control action will satisfy the user needs more than relying only on the network conditions. Since our proposed QoE parameter Self-Reference Complex Wavelet Video Structural Similarity Index (SRCW-VSSIM) can be obtained with no reference (NF) video data, it satisfies the requirement of real-time video transmission.**

***Keywords*—QoS, QoE, Video Quality Control, SRCW-VSSIM, No-Reference, Networks**

II. INTRODUCTION

Nowadays, huge amount of video is streamed over IP-Based multimedia networks, such as the Internet. However, both service providers and end users still suffer from unreliability of packet transmission. Ensuring the quality of video streaming becomes a major concern of general public. In order to meet user satisfaction, there is a need to monitor and control video quality. Currently, applications, such as video conference and video streaming, require a guaranteed Quality of Service (QoS) to work properly. Therefore current real-time video quality control (VQC) algorithms attempt to adapt streaming rate to avoid severe frame delay, frame distortion and frame loss.

The main video quality control approaches can be classified into two types: formula-based approach and measurement-based approach. Formula-based approach attempts to describe traffic, analyze and predict network condition based on mathematical models. Measurement-based method gathers path resource information, such as available bandwidth, packet loss, delay, and applies these statistics to control the source sending rate in order to satisfy the QoS requirement.

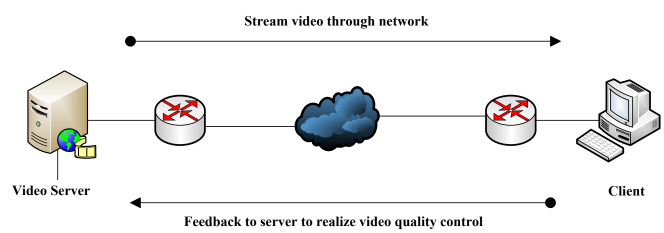

Figure 1. A networked video transmission system

Figure 1 shows the diagram of a networked video transmission system. However, both methods of video quality control attempt to adjust the video sending rate to adapt to the network condition only, and don't include the human perception. We concede that network condition has significant influence on video transmission, especially when severe congestion happens, but human perception, or referred to as quality of experience (QoE) [1], represents the major consideration for networked video data, and it should be properly included in video quality control algorithm to trigger proper actions to meet the needs of both QoS and QoE.

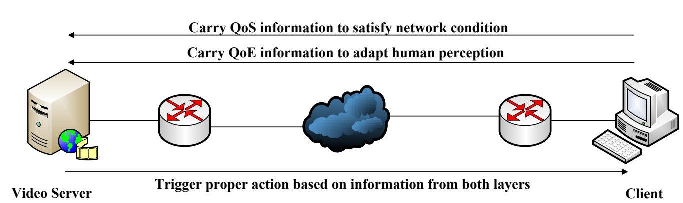

Figure 2. Cross-layer video quality control technique

As shown in Figure 2, our key contribution to video quality control technique is that we bring our innovated real time QoE parameter to correlate with QoS parameter, and introduce a

MOBILWARE 2013, November 11-12, Bologna, Italy

DOI 10.4108/icst.mobilware.2013.254287

cross-layer based real time video quality control indicator. Compared with traditional QoS only video quality control technique, our cross layer design has the ability to trigger proper action to satisfy the needs of both human perception and network condition. To better design our cross-layer video quality control indicator, we should choose appropriate QoE parameter to represent human vision perception in real time, and proper QoS parameter to represent network condition.

Currently, quality of experience (QoE) is an intense research area, and different QoE assessment methods have been proposed to describe subjective video service experience. Generally, most of prior research on QoE assessment is divided into three categories: subjective assessment, objective assessment, and hybrid respectively.

Subjective assessment is considered the most accurate approach to assess perceived quality, since it is the indicator given directly by humans. Mean opinion score (MOS) [2] is the output of subjective assessment, and it rates the perceived quality using 5 grades: Excellent, Good, Fair, Poor and Bad. But its high cost of manpower limits the use of MOS.

Current objective QoE assessment is classified into three types: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR). FR and RR methods, such as Mean Square Error (MSE) and Peak Signal Noise Ratio (PSNR), need additional network resources to access to full or portion of the original video signal [3], [4]. This can be considered as a serious drawback when it comes to real-time multimedia communication, because of the limited bandwidth and unavailability of reference data. At the same time, most of current FR or RR methods are error sensitive methods, and this can be considered as another main drawback, since error sensitive methods may not properly reflect the real human vision perception. We will discuss in details this part in the following section.

NR method becomes a very suitable option for real-time QoE assessment, since it does not require the original video [5]. However, with no reference video signal, NR method suffers from less accuracy when compared with FR and RR methods. NR methods can be divided into three types: No-Reference Pixel (NR-P), No-Reference Bit-stream (NR-B), and hybrid of NR-P and NR-B. Current NR methods mainly focus on video coding and transmission. However statistics from video coding and transmission system, such as coding rate and packet loss rate, hardly linear correlate with human vision perception. For example, with certain packet loss rates due to transmission errors, human vision perception may still be regarded as acceptable. In essence, human vision perception should be the core concern of QoE assessment.

As part of our QoS-QoE cross-layer design video quality control indicator, our QoE parameter, Self-Reference Complex Wavelet Video Structural Similarity Index (SRCW-VSSIM), can evaluate human perception in real-time. SRCW-VSSIM is a No-Reference QoE assessment method, since there is no need of original video data, it satisfies the requirement of real-time measurement. Also, instead of tracking error statistics, SRCW-VSSIM directly evaluates human vision perception.

Meanwhile, on QoS side, packet loss is an ideal parameter to apply to our proposed approach, since networked video is packetized, and transmitted through packet switched network. Packet loss can show the network condition clearly, so packet loss is chosen as the QoS parameter for our algorithm.

Based on the chosen QoE and QoS parameters, we propose a QoS-QoE cross-layer based video quality control approach. Such approach is an objective video quality control algorithm that works in real-time and reflects both human perception and network condition.

The remainder of this thesis is organized as follows. Section 3 presents the related work. We develop a QoS based video quality software application and discuss the limitations of current video quality control technique in section 4. In section 5, we introduce a novel video QoE measurement technique. Section 6 applies our new QoE measurement to QoS-QoE based cross-layer video quality control approach, and section 7 evaluates the performance of our video quality measurement approach and real-time video quality control algorithm by simulation. In section 8, we conclude our work and discuss the future research.

III. Related work

With rapid development of communication networks, high volume of video streaming is possible to be transmitted by various consumer applications. Meanwhile, quality of networked video becomes the key concern of both video service providers and video service receivers. Since video transmission weight a very heavy portion in total network data flow, control the quality of the video is intensively researched. Current video quality control technique only takes statistics from network layer, QoS, into account, and service provider can adjust video streaming rate based on the network condition.

Most of prior research on this problem is divided into two categories. One category is formula-based approach. Formula-based approach attempts to describe traffic, analyze and predict network condition based on mathematical models. Floyd et al. [6] propose an equation-based congestion control approach for unicast application. The approach lays on the TCP-Friendly rate control (TFRC) protocol. In [7], Suki et al. discuss the relationship of TFRC congestion control protocol to video rate control optimization. In [8], Huang et al. develop a feedback control system model for video streaming systems, which takes into account the interactions among video rate control, RED active queue management, and received video quality. The authors also derived a P controller that stabilizes both homogeneous video and heterogeneous video system.

The other category is measurement-based approach. Measurement-based method gathers path resource information, such as available bandwidth, packet loss [9], delay, and applies these statistics to control the source sending rate in order to satisfy the QoS requirement, which is believed to ultimately contribute to the user's QoE [10]. A multi-layer active queue management method is proposed by Kang et al [11], the authors allow the video to mark their own packets with different priority, and use the proposed queue management method to control the router to drop the less-important in order to stable the video quality when congestion happen. In [12], Kim et al. propose video quality control system which can control video service quality through the monitoring of end-to-end available bandwidth for video streaming service like IPTV in NGN convergence network. In [13], Jammeh et al. propose a delay-based congestion avoidance approach for video communication with fuzzy logic control, and this approach use delay and computational intelligence to replace packet loss and throughput modeling as input to proposed algorithm.

Although major measurement-based approaches focus on QoS parameter, such as bandwidth, delay and packet loss, a few researchers make their efforts to build mathematical model of QoE based on network measurement. In [14], Kim et al. propose a QoE assessment model for video streaming service using QoS parameters. Kitamura et al. [15] consider the relationship between the QoE of Video streaming and QoS, and propose mirco-second resolution to capture the precise behavior which effects the codec system's performance. In [16], Suzuki et al. estimate QoE from MAC-level QoS in audio-video transmission with IEEE 802.11e EDCA. Even though, all these methods suffer from uncertain relationship between QoE and QoS.

IV. QOS BASED VIDEO QUALITY CONTROL

In this section, we develop our own application to better understand the QoS only video quality control technique. During testing, we identified some limitations of current QoS only video quality control technique, and we will prove our cross-layer real time video quality indicator can overcome these limitations.

In recent years, the design of video distribution system is an intensive research area. Well-designed streaming application has to face two main challenges:

1) How to adapt the needs from users with different heterogeneous capabilities such as buffers size, processing speed rate, reception rate.

2) How to adapt the dynamic network condition, such as transmission delay, packet loss rate.

Thus a successful video streaming application should keep tuning the streaming rate to prevent network congestion and avoid overwhelming the client buffer. Therefore, a proper choice to solve the problem is that client side detects its network condition and asks to tune streaming rate to achieve better human perception.

Our software implementation [25] focuses on two scenarios:

1) End-to-End unicast streaming. As shown in Figure 3, receiver keeps monitoring and collecting network quality of service (QoS) parameter, such as packet loss rate. Server side will receive the update of network condition from receiver side, and will adjust the source sending rate once the pre-determined threshold is reached.

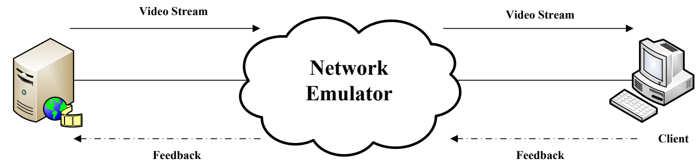

Figure 3. Unicast rate control technique

2) Multicast streaming. As shown in Figure 4, service provider creates multiple multicast streams with different rates, and allows users to switch between different multicast groups. Receivers dynamically joining and leaving the multicast groups. The same as the unicast scenario, receiver monitors and collects packet loss rate information, and choose the multicast group with proper stream rate to join.

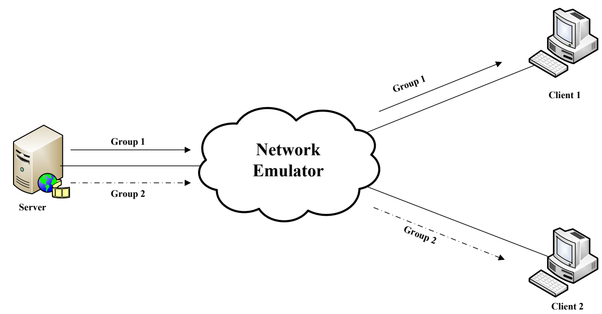

Figure 4. Multicast video quality control technique

We setup a complex wireless scenario using 18 mobile wireless routers, and these routers can move randomly. Another 4 fixed wireless routers connected with cross-traffic generator, which is composed by 4 virtual workstations. Two OPNET System In The Loop (SITL) ports are connected to two fixed wireless routers to allow actual laptops to run our software.

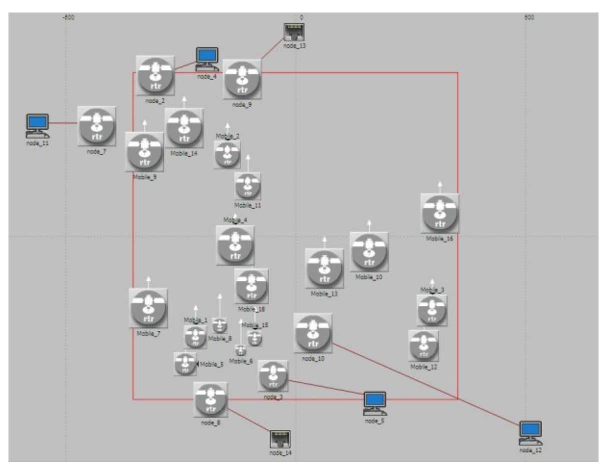

Figure 5. Topology to test software using build-in QoS measurement

In Figure 6, client side detected the network congestion, and notify video server to tune the stream sending rate. We can see from Figure 7, when network condition is good, video distribution system increased the stream rate and try to achieve a better video quality. When network congestion happen, our system can downgrade the video sending rate and release the network congestion.

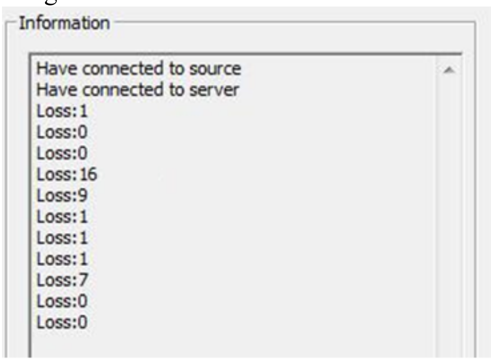

Figure 6. Client side network condition detection

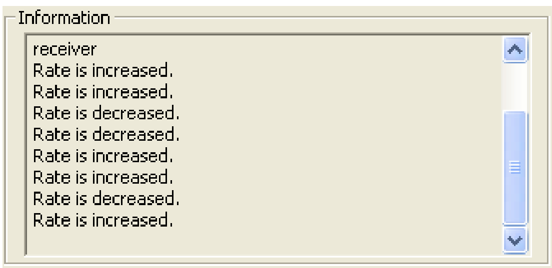

Figure 7. Server side tune stream sending rate

Our software implementation well illustrates the basic idea of QoS only video quality control technique. During the period of video transmission, software measurement tool is gathering the network statistics using different tools, and then feedback to video service provider to trigger proper actions. Actions such as sending rate increasing, sending rate decreasing and group switching, can help video stream go through the network in real time without suffering from network congestion.

Although QoS based technique is widely considered as the ideal method to avoid sever quality loss of video transmission, we identify some scenarios that triggered action can hardly help relief network congestion or even make the network congestion getting worse. For instance, current video transmission service suffer from medium packet loss rate, however, client might consider the video quality as acceptable. This is because lost packets may belong to less important frames, or codec technique may recover the lost packets using its own algorithm. Traditional QoS based technique will trigger the action to lower the sending rate while it is not necessary at all. Another drawback is that QoS based technique cannot properly reach the needs of human perception when network condition allows to do so.

The nature of QoS only based video quality control technique makes it only sensitive to network condition not to human perception. However, video service provider do cares about user perception rather than network condition, so the introduction of QoE to video quality control technique becoming more and more desirable.

The introduction of QoE, along with QoS parameter, will enable the video quality indicator to trigger proper action, which takes both QoE and QoS into account. Once we choose proper QoE parameter, and correlate with QoS parameter, and design a new real time video quality indicator, video server can follow the indicator and make the right action.

The following section will introduce a novel QoE parameter, and then use this QoE parameter to correlate with chosen QoS parameter, so that the new indicator can overcome the drawback of QoS based video quality control technique.

V. SELF-REFERENCE COMPLEX WAVELET VIDEO STRUCTURAL SIMILARITY

A. Structural Similarity

In [17], the assumption that human visual perception is highly adapted for extracting structural information from a scene helps find a new direction to evaluate the image quality assessment (IQA), referred to as Structural Similarity Index (SSIM), which is proved much closer to human vision perception and simpler than traditional error sensitive methods.

In [17], two images in spatial domain can be represented as $x=\{x_i \mid i=1,...M\}$ and $y=\{y_i \mid i=1,...M\}$, and SSIM between image x and image y is defined as

$$S(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}, \quad (1)$$

Where C_1 and C_2 are two small positive constants.

$$\mu_x=\frac{1}{N}\sum_{i=1}^{N}x_i, \quad (2)$$

$$\mu_y=\frac{1}{N}\sum_{i=1}^{N}y_i, \quad (3)$$

are used to construct the luminance comparison function.

$$\sigma_x=\left(\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\mu_x\right)^2\right)^{\frac{1}{2}}, \quad (4)$$

$$\sigma_y=\left(\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i-\mu_y\right)^2\right)^{\frac{1}{2}}, \quad (5)$$

are used to construct the contrast comparison function.

$$\sigma_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\mu_x\right)\left(y_i-\mu_y\right), \quad (6)$$

is used to construct the structure comparison.

Compared with traditional IQA and VQA method, spatial domain SSIM has the advantage of analyzing the structural information, making it more sensitive to human vision system (HVS) than the error itself, while we cannot overlook the drawback of SSIM. SSIM is highly sensitive to translation, scaling, and rotate which normally occur during coding, decoding and transmission. However in certain scenario these changes to image will not influence the structural information from the image. For example, comparing to the reference image, current image is shifted to the right by two pixels, SSIM will rate the image as an image with poor quality, but human vision system still recognize this image as an acceptable quality one, since human vision perception ranks it as acceptable. Similar argument can be discussed with regard to scaling and rotation.

In [18], in order to avoid the drawback of SSIM in spatial domain, a modified version of SSIM is proposed, Complex Wavelet Structural Similarity (CW-SSIM), and Complex Wavelet Transform (CWT) successfully overcomes the drawback of the original method.

If two images can be represented as two sets of coefficients extracted at the same spatial location in complex wavelet transform domain, $c_x = \{c_{x,i} \mid i = 1, \ldots N\}$ and $c_y = \{c_{y,i} \mid i = 1, \ldots N\}$, SSIM now can be written in this domain as:

$$\bar{S}(c_x, c_y) = \frac{2\left|\sum_{i=1}^{N} c_{x,i} c^{*}_{y,i}\right| + K}{\sum_{i=1}^{N}\left|c_{x,i}\right|^2 + \sum_{i=1}^{N}\left|c_{y,i}\right|^2 + K}, \qquad (7)$$

Where $\bar{S}(c_x, c_y) \in (0,1)$, and greater value means image y has closer human vision perception to reference image x.

As discussed in [18], *translation, scaling and rotation factor* in 2-D spatial domain can be defined as:

$$\begin{pmatrix} \Delta t_1 \\ \Delta t_2 \end{pmatrix}, \qquad (8)$$

$$\begin{pmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{pmatrix}, \qquad (9)$$

$$\begin{pmatrix} 1+\Delta s_1 & 0 \\ 0 & 1+\Delta s_2 \end{pmatrix}, \qquad (10)$$

Where $\Delta\theta$ is small, so $\cos\Delta\theta \approx 1$ and $\sin\Delta\theta \approx \Delta\theta$, and therefore

$$\begin{aligned} c_y\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= c_x\left(\begin{pmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{pmatrix}\begin{pmatrix} 1+\Delta s_1 & 0 \\ 0 & 1+\Delta s_2 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}\right. \\ &\left. + \begin{pmatrix} \Delta t_1 \\ \Delta t_2 \end{pmatrix}\right) \approx c_x\begin{pmatrix} u_1 + (u_1\Delta s_1 - u_2\Delta\theta + \Delta t_1 - u_2\Delta s_2\Delta\theta) \\ u_2 + (u_2\Delta s_2 + u_1\Delta\theta + \Delta t_2 + u_1\Delta s_1\Delta\theta) \end{pmatrix} \\ &= c_x\begin{pmatrix} u_1 + \Delta u_1 \\ u_2 + \Delta u_2 \end{pmatrix} \end{aligned}, \qquad (11)$$

We can consider this is only the linear phase shift in Fourier domain, and $\bar{S}(c_x, c_y) \approx 1$.

Translation, scaling, and rotation are proved to be not sensitive to CW-SSIM, and CW-SSIM can still perform well as reference free image quality assessment. This is the reason why we choose CW-SSIM, not SSIM, to evaluate the quality of continuous video streaming. Only change of structural information, not regular motion of video stream, can sharply change CW-SSIM of back-to-back frames, while both changes of structural information and regular motion can influence SSIM of back-to-back frames.

B. Self-Reference Complex Wavelet Video Structural Similarity

Many factors can affect and/or impair the video quality. Due to current motion compensation blocked-based coding technique, networked video always suffers from blockiness, blurriness, color bleeding, ringing, false edges, jagged motion [19]. To detect these video quality issues, many techniques are introduced, but most of these techniques are FR or RR methods, which need full or portion of the original video stream. Obtaining the original transmitted video stream at the receiving end for comparison purposes with the actual received stream is difficult due to the highly needed bandwidth. At the same time, many real-time video communication applications, such as video conference, battle field real-time video communications, can never provide the original video data. In other words, it's essential to develop a technique that is based on no reference frame approach.

Since there are no reference video data techniques available, current NR methods are mainly based on coding rate and QoS parameter, however both of them can not reflect the video quality directly to human vision perception.

At the receiving end, video streaming can be recognized as a set of frames. Continuous and clear video requires no random change between back-to-back frames. The differences between back-to-back frames are translation, scaling, and rotation. Scene shifting can be considered as translation of previous frame, and zoom-in or zoom-out can be treated as the scaling of previous frame, while scene rotation can be recognized as the rotation of the previous frame. For the special need of real-time video, especially the efficiency of information transmission, we only need to distinguish severe distortion, which leads to bad human perception and will influence the client to access the information carried by video. While slight distortion, which can be detected and probably corrected by error sensitive FR method, is not that crucial, since human perception in many cases may still ranks it as acceptable, and most of the information will be transmitted efficiently. We should not waste limited computational and network resource to detect and attempt to correct these distortions.

Back-to-back frames from video with slow and regular motion have high CW-SSIM (as discussed earlier in Section II), while those frames from video with fast motion tend to have relative low CW-SSIM. This can be explained since slow motion means slight changes between frames, and fast motion means large changes. However, for any continuous video (no sudden scene switch), the set of CW-SSIM for all back-to-back frames should be continuous, since the motion is continuous, whether it is slow motion, fast motion or mixture of these two. If the set of CW-SSIM is not continuous, or in other words, if the discrete degree of the set of CW-SSIM is large, that means some substantial unpredicted changes being introduced to the video, such as blockiness, blurriness, or false edges, In this case, we can conclude that a severe distortion has occurred. If the set of CW-SSIM is continuous, or with very small discrete degree, we can consider the video has good human perception, even there might be slight distortion.

We assume real-time video can be represented as a set of frames, $v = \{v_i \mid i = 1,...M\}$, v_n and v_{n+1} are back-to-back frames. CW-SSIM between v_n and v_{n+1} by using equation (7) is $\overline{S}(v_n, v_{n+1})$. We set up a slide window, and the width of the window is M frames, then we describe the discrete degree by standard deviation:

$$SRCW-VSSIM = \sqrt{\frac{1}{M-1}\sum_{n=1}^{M-1}\left(\overline{S}(v_n,v_{n+1}) - \frac{1}{M-1}\sum_{n=1}^{M-1}\overline{S}(v_n,v_{n+1})\right)^2}\ , \quad (8)$$

When real-time video quality is good, and we experience a fluent and clear video, SRCW-VSSIM is relatively small and close to 0, otherwise, video with poor human vision perception will lead SRCW-VSSIM relatively large. The simulation results for this section are presented in Section IV.

VI. QoS-QoE BASED VIDEO QUALITY CONTROL APPROACH

Most of the video quality control approaches, both formula-based and measurement-based, try to contain the network congestion and may adjust the source sending rate according to quality of service (QoS) rather than the quality of experience (QoE). While QoS can provide many useful information about the network conditions, and sometimes can also tell the quality of the real-time video service, QoS can hardly provide any information about QoE. For example, although we experience packet loss, less important frame lost won't affect the overall quality of the video, so it is unnecessary in this case to attempt to mitigate the network conditions and attempt to tune the source sending rate dramatically. Besides, even when network condition is good, proper action can hardly be decided without the information from QoE side.

On the other hand, using QoE alone to tune the video service is also unacceptable. Since we try to control the video quality by adapting the source sending rate, we have to pay attention to the network condition. If the original video quality is poor from the source side, QoE parameter will show that tuning is needed to be done, while the truth is that tuning is useless.

In summary, parameters generated from a single layer cannot determine accurately when and how to tune the network to meet the needs of user's satisfaction. Hence, we propose here a new indicator including both QoS parameter (here we will use packet loss as the QoS parameter) and QoE parameter (here we will use SRCW-VSSIM as the QoE parameter), to trigger the proper action.

A. Framework

We plan to design an end to end cross-layer video quality control system as shown in Figure 8. Video service client collects statistics of packet loss rate and SRCW-VSSIM as input to our QoS-QoE based video quality control indicator (QQVQCI). We define the packet loss rate as the ratio between the number of the lost packets and the number of transported packets during each interval.

$$pktlossrate = \frac{P_{Loss}}{P_{Loss} + P_{received}} \quad (9)$$

P_{Loss} is the number of packet loss per time interval, and $P_{received}$ is the number of received packets per time interval. Indicator is sent to video service source, and video service source determines whether tuning is needed, basically increasing or decreasing the sent video rate according to the QQVQCI value feedback from the client.

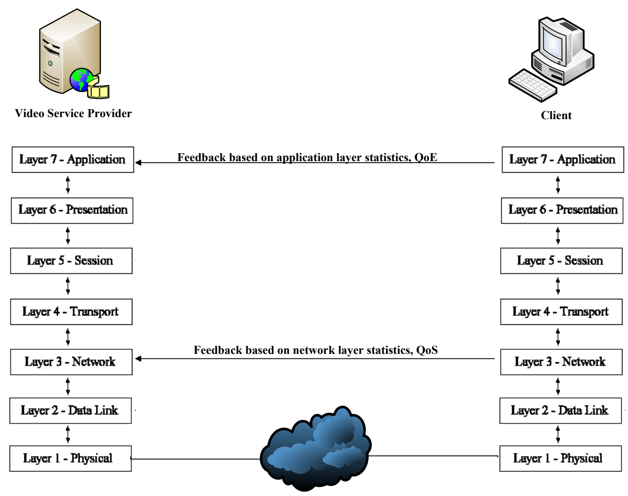

Figure 8. Proposed QoS-QoE based video quality control technique

Our QoS-QoE based video quality control indicator (QQVQCI) is defined as：

$$\text{QQVQCI} = (pktlossrate + \delta) \times (SRCW\_VSSIM - \alpha) \quad (10)$$

where δ is a small positive constant, and α is the threshold for SRCW-VSSIM.

	Good QoE	**Bad QoE**
Good QoS	Scenario 1 No Action is needed	Scenario 3 Action (sender increases the sending rate ect.)
Bad QoS	Scenario 2 No Action is needed	Scenario 4 Action (Network tuning, e.g. relief congestion, reduce source.)

Table 1. Scenarios for different QoS and QoE conditions

To demonstrate our cross-layer design video quality control system, we discuss the following scenarios as stated in Table 1:

1) When both network condition and human perception are good, packet loss rate is small and SRCW-VSSIM is less than the threshold α , so we expect our indicator be negative, and its absolute value is small . No further action should be done.

2) Another possible scenario is as follows: network condition may experience slight congestion, so some packets may be lost. Traditional QoS based video quality control approach considers this as the reason to lower the sending rate. However, this action may n't be necessary, since limited number of lost packets may not degrade the human perception of the video, for example, key frames may n't experience

packet losses, and in this case human perception may still be acceptable. Instead of enhancing the quality of experience, the action hurts the human perception of the video data. QQVQCI successfully avoids this exception due to the introduction of QoE as a factor of our indicator. Our indicator is a negative value since QoE index is less than α , and video sender should not change the sending rate.

3) Sometimes, low sending rate leads a poor video quality, however, the network condition is good enough to allow a higher sending rate. Under this situation, our indicator is a small positive number. The reason is simple: one factor of our indicator, SRCW-VSSIM, is much greater than threshold α , but another factor, packet loss rate, stays low. Once the video service provider receives such indicator, video service provider should increase the sending rate to enhance the user's quality of experience.

4) If both network condition and human perception are bad, packet loss rate and SRCW-VSSIM are large at the same time. QQVQCI turns out to be a large value. This indicates the video source or network should take action to relieve congestion and achieve a better perceived video quality.

Our contribution is that we can accurately distinguish Scenario 2 and Scenario 3 from the traditional Scenario 1 and Scenario 4, and make proper actions based on statistics from both layers to tune network or source server to reach the user's needs correctly.

Generally speaking, the introduction of QoE parameter along with QoS parameter helps video sending rate tuning adapt not only to network conditions but also to human perception. The proper actions according to our new indicator, QQVQCI, are as followed:

- When QQVQCI is a large positive number, source sending rate should be reduced to adapt to the network condition.
- When QQVQCI is negative number, no further action should be done since human perception is within the tolerated range.
- When QQVQCI is a small positive number, source sending rate should be increased to achieve a better human perception, since current network condition allows to do so.

Proper thresholds of QQVQCI to trigger reasonable action are decided in the experiment section.

The simulation results for this section are presented in Section VII.

VII. Simulation Results

In this section, part A to part D prove that our new QoE index, SRCW-VSSIM, is a sensitive real-time video quality measurement. Part E demonstrates that QQVQCI can be used to trigger proper actions in order to satisfy the needs for both the network and the human perception.

A. Implementation of Experiment

In order to accelerate the computational speed, we take every other frame as back-to-back frames, and use the same implementation of CW-SSIM in [21] to calculate the CW-SSIM between back-to-back frames. We choose window size as 10 frames, so the first ten frames are used to initialize our index. We will show the CW-SSIM, SRCW-VSSIM, and PSNR for each video sample. PSNR is calculated by MSU Video Quality Measurement Tool, [22].

For video with slight distortion, we use LIVE Video Quality Database, [23] [24] to show our results. Meanwhile, we use Sirannon, [20] to simulate heavy packet loss to video stream during transmission, and the sample video is provided by [20].

B. Experiment: Video with slight distortion

As shown in Figure 9 and Figure 10, original video without distortion has continuous CW-SSIM for back-to-back frames, and our SRCW-VSSIM stays at a very low level.

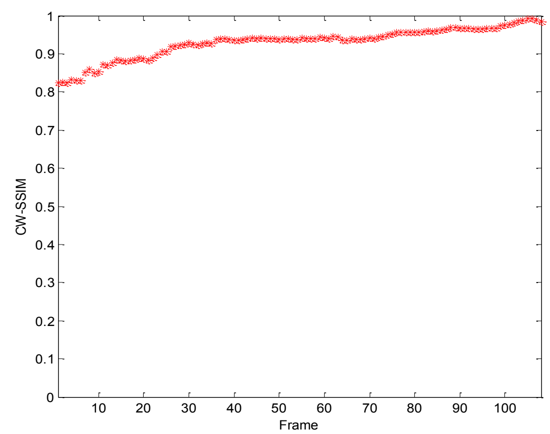

Figure 9. CW-SSIM of back-to-back frames for video sample 1 without distortion

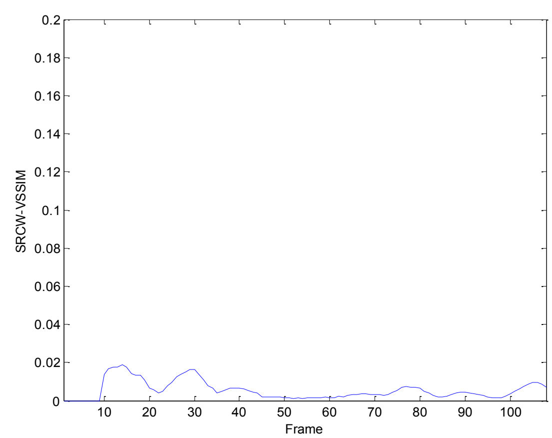

Figure 10. SRCW-VSSIM of video sample 1 without distortion

Figure 11 and Figure 12 show the CW-SSIM and SRCW-VSSIM for video with slight distortion. We can see some discontinuity points which represent worse quality video segment. SRCW-VSSIM is relatively higher when compared with the previous data. However, SRCW-VSSIM still stays at a very low level, less than 0.05. Human perception is still good even with the existence of some slight distortion. PSNR in Figure 12 can also tell that overall quality of the distorted video is accepted, and PSNR is generally above 25 dB.

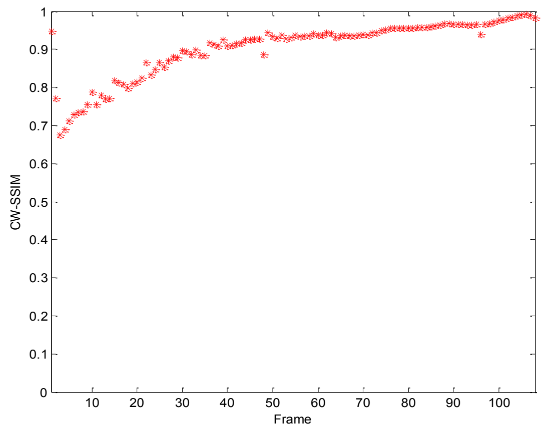

Figure 11. CW-SSIM of back-to-back frames for video sample 1 with slight distortion

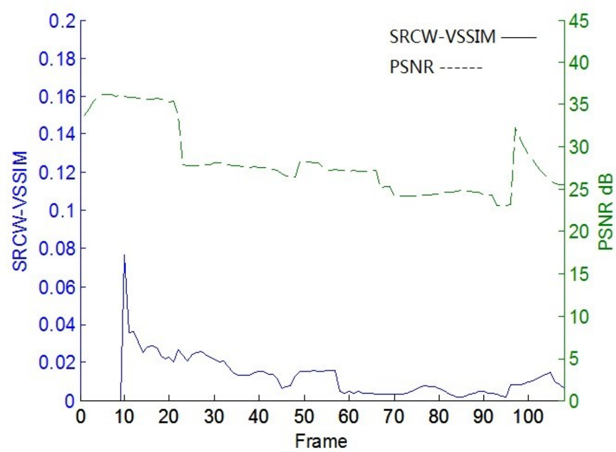

Figure 12. SRCW-VSSIM and PSNR of video sample 1 with slight distortion

C. Experiment: Video with heavy distortion

We can see from Figure 13 and Figure 14, for the original video, CW-SSIM is continuous and SRCW-VSSIM stays at a low level, less than or around 0.05. When heavy distortion is introduced, as shown in Figure 15 and Figure 16, CW-SSIM for video sample 3 becomes discontinuous. Meanwhile, SRCW-VSSIM increases to a very high level, greater than 0.05, when heavy distortion happens. At certain instant, the human vision system (HVS) can easily detect video quality changes, and information carried by video stream can hardly be accepted. When comparing with PSNR shown in Figure 16, all peak value of SRCW-VSSIM can match the PSNR less than 25 dB, which means unacceptable video quality.

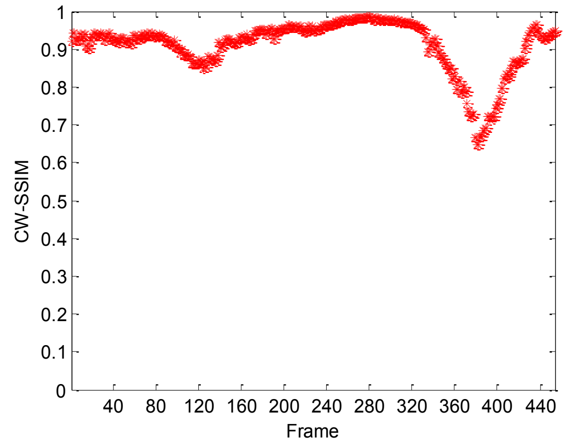

Figure 13. CW-SSIM of back-to-back frames for video sample 3 without distortion

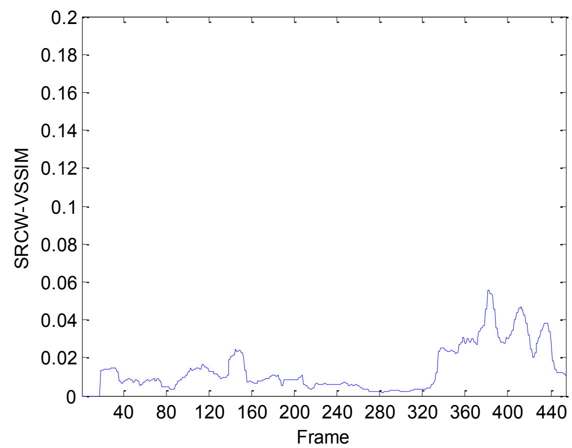

Figure 14. SRCW-VSSIM of video sample 3 without distortion

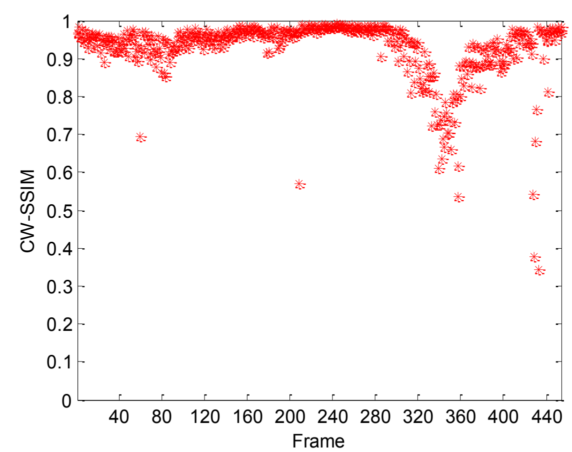

Figure 15. CW-SSIM of back-to-back frames for video sample 3 with heavy distortion

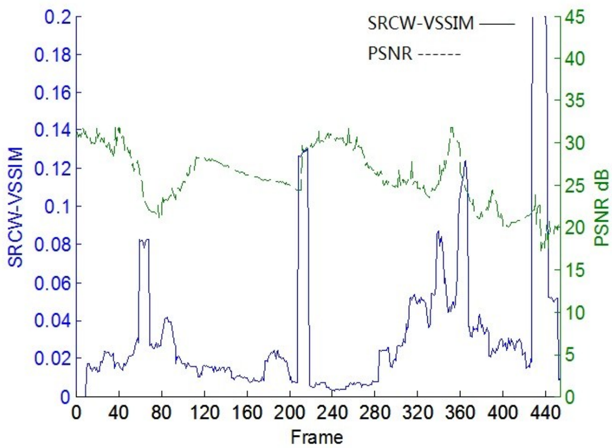

Figure 16. SRCW-VSSIM and PSNR of video sample 3 with heavy distortion

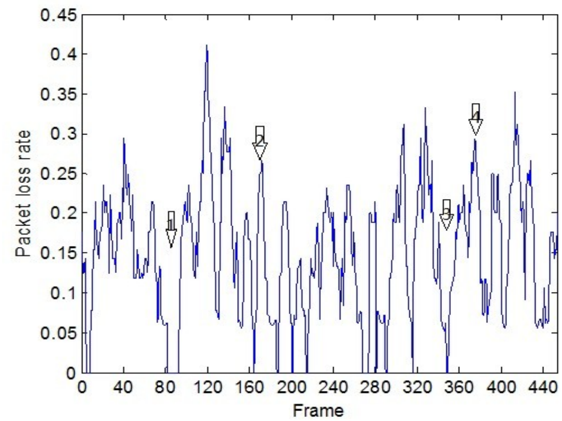

Figure 17. Packet loss rate of video sample 3 with heavy distortion

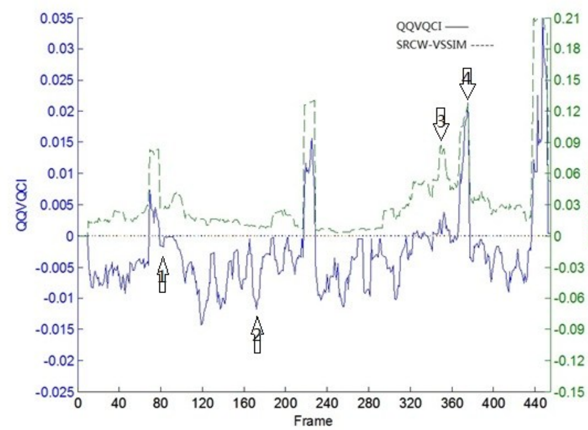

Figure 18. QoS-QoE based Video Quality Control Indicator and SRCW-VSSIM of Video sample 3 with heavy distortion

D. *Experiment Summary*

Conclusion can be reached based on the above experiment. Compared with PSNR, when SRCW-VSSIM is less than or around 0.05, human vision percetion is acceptable, even though the video is suffering from slight distortion. When SRCW-VSSIM changes sharply from a relatively low value (less than or around 0.05) to a relatively high value (around 0.1), human vision percetion becomes worse, and video quality is heavily decreased. Our algorithm is proved to be a sensitive no-reference video quality measurment technique.

E. QoS-QoE based video quality control indicator (QQVQCI)

The parameter δ and α in (10) are set to be 0.01 and 0.05 (according to the simulation result of part A to part D). We can see from Figure 17 and Figure 18, our indicator can detect network condition and human perception of quality of video clearly, and trigger the proper action to enhance the overall video transmission service. Indicator with value greater than 0.005 indicates poor network condition and human perception (Scenario 4 in Table 1), and reducing sending rate is needed. For instance, QQVQCI is greater than 0.005 near frame 370 (Arrow 4), hence server need to lower the sending rate to realize the optimization of perceived video quality. When the value of the indicator is around 0.005, for example near frame 350 (Arrow 3), network condition is good enough to increase the sending rate to enhance the video transmission service (Scenario 3 in Table 1). All other negative values of QQVQCI demonstrate that even when the network experience packet loss, since human perception of video quality still satisfies the user's requirement (Scenario 1 and Scenario 2 in Table 1), no further action is needed to tune the video sending rate (Arrow 1 and Arrow 2).

VIII. Conclusion

This paper introduces a novel end-to-end video quality control approach, which is based on QoS-QoE based cross-layer design. Our video quality control indicator, QQVQCI, gives more accurate information of the video transmission service, and successfully avoid the drawbacks of QoS-based or QoE-based video quality control algorithm. Compared with traditional single layer based video quality control algorithm, QQVQCI can look into both network condition and human perception, and trigger proper actions to balance the satisfaction of both layer's requirement. Especially, our newly introduced QoE index, SRCW-VSSIM, is reference free and closer to human perception of video quality, and this makes our video quality control indicator works properly under real-time video transmission environment.

References

[1] "Definition of Quality of Experience (QoE)," Reference: TD 109rev2 (PLEN/12), ITU - International Telecommunication Union.

[2] "ITU-T Recommendation P.910. Subjective video quality assessment methods for multimedia applications, 2000."

[3] Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference. ITU-T Rec. J.144, Geneva, Switzerland, Mar. 2004.

[4] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol. 50, no. 3, pp.312-322, Sep. 2004

[5] J. Caviedes and F. Oberti, "No-reference quality metric for degraded and enhanced video," in Proc. SPIE, vol. 5150. Jul. 2003, pp. 621–632.

[6] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based congestion control for unicast applications," in Proc. ACM SIGCOMM '00, Stockholm, Sweden, Sept. 2000.

[7] A. Suki M. Arif, Suhaidi Hassan, Osman Ghazali, Shahrudin Awang Nor, "The Relationship of TFRC Congestion Control to Video Rate Control Optimization," in Proc. 2nd Int. Conf. Network App. Protocols and Service (NETAPPS), 2010, pp. 31-36.

[8] Y. Huang, S. Mao, and S. Midkiff, "A Control-Theoretic Approach to Rate Control for Streaming Videos," IEEE Transactions on Multimedia, Vol. 11, No. 6, October 2009.

[9] G. Van der Auwera and M. Reisslein, "Implications of smoothing on statistical multiplexing of H.264/AVC and SVC video streams," IEEE Trans. Broadcast., vol. 55, no. 3, pp. 541–558, Sep. 2009.

[10] N. Staelens, S. Moens, W. Van den Broeck, I. Marien, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "Assessing quality of experience of IPTV and video on demand services in real-life environments," IEEE Trans. Broadcast., vol. 56, no. 4, pp. 458–466, Sep. 2010.

[11] S. Kang, Y. Zhang, M. Dai, D. Loguinov, "Multi-layer Active Queue Management and Congestion Control for Scalable Video Streaming," in Proc. 24th Intl. Conf. Distributed Computing Systems, 2004, pp. 768-777.

[12] Hyun long Kim, Seong Gon Choi, "Design of video streaming service quality control system using available bandwidth information," in Proc. 13th Intl. Conf. Advanced Communication Technology (ICACT), 2011, pp. 1540-1544.

[13] E. Jammeh, M. Fleury, and M. Ghanbari, "Delay-based congestion avoidance for video communication with fuzzy logic control," in Packet Video 2007, pp. 8-17.

[14] Hyun Jong Kim, Dong Geun Yun, Hwa-Suk Kim, Kee Seong Cho, Seong Gon Choi, "QoE Assessment Model for Video Streaming Service using QoS Parameters in Wired-Wireless Network," in Proc. 14th Intl. Conf. Advanced Communication Technology (ICACT), 2012, pp. 459-464.

[15] [15] Masahiko Kitamura, Takahiro Murooka, Mitsuhiro Teshima, Hisaya Hadama, Atsushi Takahara, "A Study on the Correlation between QoE of 4K Super High Definition Video Streamings and QoS of Network," in Proc. 18th Intl. Conf. Comp. Commu. and Network, 2009, pp. 1-6.

[16] Takahiro Suzukit, Tasuku Kutsunatt, and Shuji Tasakatt, "QoE Estimation from MAC-Level 008 in Audio-Video Transmission with IEEE SOZ.lle EDCA," in Proc IEEE 19th Intl. Symposium PIMRC, 2008, pp. 1-6.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13, pp. 600–612, Apr. 2004.

[18] Zhou Wang and Eero P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," Proc. IEEE Inter. Conf. Acoustic, Speech & Signal Processing Volume II, Pages 573-576, March 2005

[19] Shyamprasad Chikkerur, Vijay Sundaram, Martin Reisslein, and Lina J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," IEEE TRANSACTIONS ON BROADCASTING, VOL. 57, NO. 2, JUNE 2011

[20] UGent-IBCN, "Sirannon 1.0.0: Modular Multimedia Streaming," [Online]. Available: http://xstreamer.atlantis.ugent.be

[21] "Complex Wavelet Structural Similarity: A New Image Similarity Index", M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, M. K. Markey. IEEE Transactions on Image Processing, 18(11), 2385-401, 2009.

[22] MSU Graphics & Media Lab, "MSU Video Quality Measurement Tool," [Online]. Available: http:// compression.ru/video/quality_measure

[23] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video",IEEE Transactions on Image Processing, vol.19, no.6, pp.1427-1441, June 2010.

[24] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms", SPIE Proceedings Human Vision and Electronic Imaging, Jan. 2010.

[25] Qihua Yang, Tarek Saadawi, Ahmed Abdelal, Mitesh Patel, Biao Jiang, "Multicast Multi-Streaming and Real-Time Stream Rate Control Demo," in Globecom 2010, Miami, USA, Dec 2010