Designing and Modeling for ITS

Yong Li

CONTENTS

I. INTRODUCTION

Intelligent Transportation Systems (ITS) consist of on-site embedded system with cameras for real-time video capturing/processing, transmission network to upload videos and data center for central management. Each component in the ITS should be careful designed and tuned to achieve high performance at low cost in terms of storage, network bandwidth as well as other hardware resources. This document presents an attempt to model our target ITS system performance as well as resource utilization and requirement as functions of a series of design parameters. The results of the modeling can be used to estimate the resource requirement and help in making design decisions in the early stage of the designing cycle.

II. Design and Resource Requirement Modeling for Terminals

Terminals are placed on site to capture real-time video and HD images for traffic monitoring. These video and images are pre-processed on site and transmitted, with other relevant information (e.g. image features, terminal status, maintenance information, location-specific info., etc.), to the data center through a dedicated network. This section presents the design choices and the modeling of resource utilizations for terminals under different situations.

A. General Design Considerations for Terminals (IP Cameras)

IP cameras usually offer rich parameters which can be changed to suit a specific application. Choosing appropriate camera model and parameters are crucial to meet the design objective. IP Cameras for surveillance purpose (e.g. Sony SNC-RZ50N, Cisco CIVS-IPC 4000 series, Axis P3301, etc.) also provide advanced features including motion detection, event trigger, customized objects tracking, etc. Leveraging these off-the-shelf features can drastically reduce the design cycle and cost. However, additional functions and video processing algorithms may have to be implemented to handle complicated situations in traffic surveillance.

Below are a number of key considerations for choosing and designing camera terminals for Intelligent Transportation Systems:

- **Bit Rate/Bandwidth Related Setting:** A camera's bit rate/bandwidth is determined by its resolution, frame rate, and compression standard used. The bit rate for commodity IP cameras varies from a few hundred Kbps to several Mbps, which requires the corresponding network bandwidth to transmit. Choosing the right bit rate is a tradeoff between the quality of the video and the required hardware resources.

  An important decision regarding bit rate is choosing between constant bit rate (CBR) and variable bit rate (VBR). With limited bandwidth the preferred mode is CBR because this mode generates a constant and predefined bit rate. The downside is that the image quality will vary depending on the amount of motion in the video scene. When there is a need for high quality images and the network infrastructure has a high capacity, VBR is the desired bit rate.

  The GOP structure defines the composition of the video stream (the interleave of I,B and P frames). Setting the GOP-length to a higher value results in a larger interval between any two I frames and saves considerable bandwidth but may have an adverse effect on image quality. In ITS the GOP length can be set to a large value (e.g. 128) due to the fact that the scene changes are not frequent for an IP camera at a given location.

  FPS (Frame per Second) is another factor that affecting bandwidth and video quality. FPS typically varies from $10 \sim 30$.
- **Power over Ethernet (PoE):** Camera with PoE lets you use the same Ethernet cable for supplying power as well as transferring data. The main benefit is the huge cost savings

associated with this. Hiring electricians to install separate power lines is unnecessary. In addition to that, PoE makes moving IP cameras to new locations and adding additional IP cameras simple.

- **Events Detection:** Interested events may include traffic accident, traffic jam, suspicious vehicle detection, speeding, other traffic violations, etc. Of course, these events can be detected at the data center after the video has been uploaded. Here are some general guidelines regarding which events should be detected by terminals: 1. Urgent or real-time events. 2. The events which can be detected using the built in features provided by the IP cameras. 3. Events which can be detected using inexpensive algorithms.
- **Events Handling:** Upon an interested event, certain action should be taken. The action could be recording that event to the local storage, uploading a video clip to the data center, triggering a HD camera for detailed information capturing, or sending an email to a specified address. What actions to be taken is dependent on the desired information (for the interested event) and the hardware configuration/availability. If there is enough network bandwidth, it is preferable that the relevant information is uploaded to the data center for management. There are two alternative ways to capture the detailed information upon an interested event. The first way is to use two cameras, one for recording/uploading SD video stream under normal condition, the other for capturing HD images upon interested events. This method is straightforward and the setting for each camera is simple. However, this requires a customized SoC design and the collaboration between the two cameras should be carefully designed. This method also has a low hardware resource utilization since the HD camera may not be used frequently. The alternative approach uses only one HD camera. The output video stream undergoes either a codecs channel with high compression rate under normal condition or a channels with low/zero compression rate (e.g. Motion JPEG) when detailed information is needed upon an interested event.
- **Video Upload:** The IP camera should support file transfer (FTP/TFTP) protocol and can serve as either client or server to ease the file transfer between terminals and the data center.
- **Other Features:** Other important features to consider including operating temperatures, waterproofness, lightning protection, night vision support, etc.

B. Network Bandwidth Requirement Modeling

Now we propose a more detailed modeling of network traffic and bandwidth requirement for H.264 streaming video in the context of ITS applications.

1) Modeling for VBR: Among the encoding schemes provided in H.264, VBR prevails since it provides better video quality for the same average bit rate without the need to adjust the quantization settings during the encoding as in CBR. The downside of VBR is that it results in a variable bit rate which requires a nondeterministic network bandwidth at a particular time. This also post a challenge for modeling the traffic characteristics and the bandwidth requirement.

Lombardo et al. [1] noticed that there is a strong correlation between the P/B-frame sizes and the I-frame size belonging to the same GOP, which is also called intra-GOP correlation. Motivated by their observation, we first assume the size (in bit) for any I frame is X_I (Under 528x384 resolution, the typical size of an I frame varies from tens to a few hundred Kbits, depending on its quantization settings. Under 2592x1944, the size of an I frame can reach a few Mbits. For the situation where the resolution, quantization setting and the background scene are fixed, such as a terminal in ITS, this value can be obtained experimentally). We further denote the number of frames in a GOP as N_{GOP} and the number of frame per seconds as N_{FPS}. We will use them as the baseline parameters to calculate the sizes of other B and P frames in the same GOP (for simplicity, we omit the differences between B and P frames). This is a reasonable assumption since the size of I frames are largely determined by the camera settings and the background images to be coded. In addition, camera settings in ITS applications are usually fixed and the background images usually do not show a huge variation. The size of B and P frames are dependent of the camera settings as well as the correlations with the reference and neighboring frames.

The Discrete Auto Regressive(1)(DAR(1)) model [2] is the most primitive yet an effective model that was proposed for modeling of video conference traffic. We modify it to model the video traffic in ITS:

$$X(n) = (1 - \alpha(t_n, l_n))X_I + \beta e(n) \tag{1}$$

where $X(n)$ is the size of the n_{th} B/P frame in a particular GOP, $e(n)$ is a Gaussian process with variance σ^2 and mean value of η ($0 < n < N_{GOP}$). β is a constant. $\alpha(t,l)$ is the correlation coefficient determined by the traffic condition being monitored at the time t_n and location l_n. Generally, $\alpha(t,l)$ is a function of time t and location l and derived based on the traffic condition. Particularly in our model, we use the number of vehicles passed the camera's view per unit of time (e.g. 1 minute) as the indication of traffic condition. The smaller the number is, the more correlation the intra-GOP frames exhibit, which indicates a more stationary video. At one extreme, when the number is zero (i.e. there is no vehicle passing or leaving), the produced video is almost stationary and the correlation is high . At the other extreme, when there is a large number of fast driving vehicles, every neighboring frames show significant differences in vehicles or their locations and the correlation in such a condition is low. An example of correlation coefficient function for a camera placed at location P over 24 hours is shown in Figure 1.

The total number of bits N_BG in the GOP thus can be calculated by:

$$N_{BG} = X_I(1 + \sum_{n=1}^{N_{GOP}-1}(1 - \alpha(t_n, l_n))) + \beta \sum_{n=1}^{N_{GOP}-1} e(n) \tag{2}$$

Since the duration of a GOP is not long (usually no more than a few seconds), the correlation coefficient function $\alpha(t_n, l_n)$ can be approximately viewed as a constant $\alpha(T, L)$ for a camera at

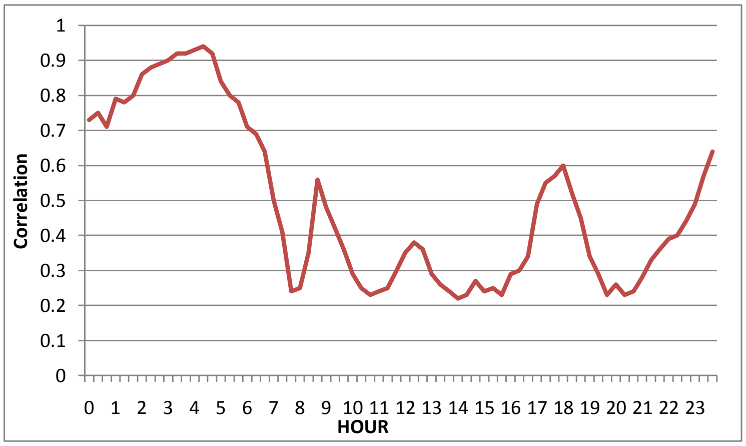

Fig. 1: A synthetic correlation coefficient function based on 24-hour traffic condition

the given location L, where T is the time when the GOP begins to be processed. Therefore, the above equation can be simplified as:

$$N_{BG} = X_I(1 + (N_{GOP} - 1)(1 - \alpha(T, L))) + \beta \sum_{n=1}^{N_{GOP}-1} e(n) \tag{3}$$

N_{BG} reaches its maximum value, denoted as N_{BGMax}, when $\alpha(T, L) = \alpha_{Min}$ where α_{Min} is the minimum value of the α coefficient function. Considering the factor of the random variable $e(n)$ and the 3-σ rule, the maximum number of bits in a GOP thus is:

$$N_{BGMax} = X_I(1 + (N_{GOP} - 1)(1 - \alpha_{Min})) + \beta(N_{GOP} - 1)\eta + 3\beta\sigma \tag{4}$$

Similarly, the mean number of bits in a GOP is:

$$N_{BGMean} = X_I(1 + (N_{GOP} - 1)(1 - \alpha_{Mean})) + \beta(N_{GOP} - 1)\eta \tag{5}$$

We can calculate max/mean bit rate or network bandwidth necessary for transmitting the VBR video:

$$BR_{VMax} = N_{BGMax} \frac{N_{FPS}}{N_{GOP}} \tag{6}$$

$$BR_{VMean} = N_{BGMean} \frac{N_{FPS}}{N_{GOP}} \tag{7}$$

In a network environment where the bandwidth is limited, the parameter *maximum bit rate* (BR_{Max}) in the VBR mode can be used to ensure the encoded video bit rate does not exceed the bandwidth limit at the expense of dropped frames. In a situation in which the BR_{Max} is used, the maximum bit rate is calculated by $Min(BR_{VMax}, BR_{Max})$.

2) Traffic Characteristics for CBR: As opposed to VBR, the CBR encoder maintains the constant bit rate by changing the quantization level for macro-blocks to compensate the variances among frames in the encoded video, at the expense of non-constant visual quality and low coding efficiency. When CBR is chosen as the encoding method, the bit rate is more deterministic than that of the VBR. For every GOP in CBR, the average bit rate is a constant value BR_C, which is preset by the administrator. However, the network designer should be aware that the bit rate within a particular GOP is not a constant. The network experiences bursts during the transmission of reference frames, or I-frames. The burst bit rate can be calculated by:

$$BR_{CBurst} = X_I N_{FPS} \tag{8}$$

An example of network traffic characteristic for H.264 video encoded in CBR mode is illustrated in Figure 2.

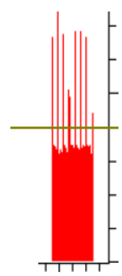

Fig. 2: Illustration of network traffic characteristic for H.264 video encoded in CBR mode

3) Other Network Traffic Sources in ITS: In addition to the real-time H.264 video steaming, the network bandwidth resources could also be consumed by other traffic sources such as information of interested events, maintenance/control traffic, software updates, etc. These traffic sources are totally nondeterministic at the early design stage of the ITS and the network resources required for these traffic sources are dependent of a variety of factors including the what kind of information is required upon interested events, the frequency of interested events, maintenance/control operations, size and frequency for software updates, etc. Comparing to the real-time video traffic however, these network traffic sources do not pressure network resource significantly. Additionally, the network traffic due to maintenance/control operation and software updates primarily flows the opposite direction against the video traffic. Thus, we only consider the additional bandwidth required for uploading the information of the interested events.

We shall assume that the information generated upon an interested event are N_{Img} HD images and a series of features (e.g. time, event type, vehicle information, etc). Among these, the HD images take up the most bandwidth and thus is the focus for our modeling.

Take 5-megapixel cameras(2,592x1,944) for instance, the sizes of produced images are determined by the resolution, the number of bits used to represent one pixel and the compression standard used. Table I presents typical image sizes for different formats.

TABLE I: File Sizes for Images with Different Formats (5Mega Pixels)

Image Format	File Size	Description
RAW12	7.6MB	(uncompressed 12 bits/pixel - Bayer mask)
RAW8	5.0MB	(uncompressed 8 bits/pixel - Bayer mask)
BMP	15.1 MB	(uncompressed RGB 24bit/pixel)
JPG100	1.4M	(100% quality - 24bit/pixel)
GIF	1.7M	(compressed - 8bit/pixel)
PNG	3.7M	(lossless compressed - 24bit/pixel)

To calculate the bit rate contributed by interested events for a particular terminal, we denote the number of interested events per hour as N_{IE}. For PNG format, the bit rate can be represented as

$$BR_{IE} = \frac{3.7 \times 8 \times N_{IE} N_{Img}}{3600} \tag{9}$$

4) A Case Study: Given the bit rate and network bandwidth modeling above, we shall study an example with the modeling parameters being replaced by specific values.

In our case study the HD camera's resolution is 5M pixels (2,592x1,944) and we estimate the size of an I frame as 2Mbits (we can adjust the quantization settings Q to reach this value). We also choose $N_{GOP} = 128$, $N_{FPS} = 15$, $\beta = 128Kbits$ and $e(n)$ to be a standard normal distribution ($\sigma = 1$ and $\eta = 0$). Based on Eq. (4) and Eq. (5), the maximum and mean number of bits in GOP are $N_{BGMax} = 2Mbits \times (1 + 127 * (1 - 0.23)) + 3 \times 128Kbits \times 1 = 197.58Mbits + 0.384Mbits = 197.964Mbits$ and $N_{BGMean} = 2Mbits \times (1 + 127 * (1 - 0.49)) = 131.54Mbits$, respectively. The maximum and mean bit rates can be calculated by Eq. (6) and Eq. (7): $BR_{VMax} = (197.964Mbits \times 15)/128 = 23.19Mbps$; $BR_{VMean} = (131.54Mbits \times 15)/128 = 15.4Mbps$.

To take the interested events into consideration, we shall assume that $N_{IE} = 10$, $N_{Img} = 3$ and the files to be uploaded is PNG. Using Eq. (9), the bit rate due to the interested events for a camera is $BR_{IE} = (3.7 \times 8 \times 10 \times 3)Mbits/3600 = 0.24Mbps$.

To summarize all the modeled parameters above, we list them in Table II. The + - symbols indicate whether the corresponding modeled parameter has positive or negative effect on bit rate (EoB) or network bandwidth. + means the bit rate increases as the parameter increases while - means the bit rate decrease as the parameter increases. Effecting factors list the H.264 encoding parameters that affect the modeled parameters.

The above table provides a clear picture of how various parameters affect the model parameters and a convenient way of adjusting parameters under specific network bandwidth constraints. For example, reducing the resolution to 1920 x 1080 (still HD) could approximately save the bandwidth requirement by $(2,592x1,944 - 1920x1080)/2,592x1,944 = 58.84\%$. The maximum and mean bandwidth requirements after saving are $23.19 \times (1 - 58.84\%) = 9.54Mbps$ and $15.4 \times (1 - 58.84\%) = 6.34Mbps$. This requirement can be fulfilled by a 10 / 100 Mbps ethernet switch.

TABLE II: Summary of modeled parameters

Modeled Parameter	Description	EoB	Typical Value	Effecting Factors
CBR/VBR	Constant/Variable bitrate		CBR (for constant bitrate)/VBR(for constant video quality)	None
X_I	Size of I frame	+	100Kbits $\sim$ 10Mbits	resolution, quantization level
N_{GOP}	Frames in GOP	-	10$\sim$few hundred (use larger value in ITS)	None
N_{FPS}	Frames per second	+	10 $\sim$ 30	None
BR_C	Constant bit rate	+	100Kbps $\sim$ 10Mbps	None
N_{Img}	Number of Images to upload upon an interested event	+	1$\sim$10	None
N_{IE}	Number of interested events per hour	+	0$\sim$3600	None

We will use these parameters for the network bandwidth modeling for data center, as will be elaborated in Section III-B.

C. Network Designing for Terminals

After modeling the network bandwidth requirement, this section lists accessing networking design considerations to meet the required bandwidth as well as other metrics such as QoS and latency:

1) First, estimate the network bandwidth requirement using the introduced modeling for each terminal. Make sure the accessing switches (the local LAN switch for terminals) provide enough bandwidth for the terminals. When designing the accessing network, the traverse distance limitations should also be considered: 100BASE-TX (100 Mbps over two-pair Cat5/ Twisted-pair cablingCAT-6, CAT-7) 100 meters; 1000BASE-SX Multi-mode fiber 220 meters; 100BASE-FX SFP multimode fiber-optic (MMF) 2 kilometers; 100BASE-LX10 SFP single-mode fiber-optic (SMF) 10 kilometers;
2) Number of supported switch ports (e.g. 24, 48, etc.) on a single switch should be co-considered with the aggregate bit rates from all the connected terminals. Uplink ports should satisfy the bandwidth requirement as well. The switches should support IEEE 802.3af, the standard for Power over Ethernet (for IP Cameras);
3) Data and video network QoS usually is required to allocate resources to transport video with low loss. Packet loss in the network will be noticeable in the video quality of MPEG-4 and H.264 video feeds. QoS marking or policies can be enforced by terminals, switches or routers;
4) H.264 transported in TCP is not tolerant of high latency. IP cameras with two-way need low latency. H.264 in UDP/RTP is tolerant of high latency.

III. DATA CENTER MODELING

The data center, which consists of a large number of servers, provides a central platform for monitoring the real-time traffic condition, management of terminals and databases, storing large video files and other useful information. The target platform hosted by the servers in the data center is a distributed system where the storage requests (upload video files) are served by large number of *datanodes* servers using HDFS (Hadoop Distributed File System), the directory namespace service is hosted by the *namenode* servers and the computation tasks are processed using map-reduce servers (*jobtracker* and *tasktrackers*). An overview of the data center structure in the ITS is shown in Figure 3

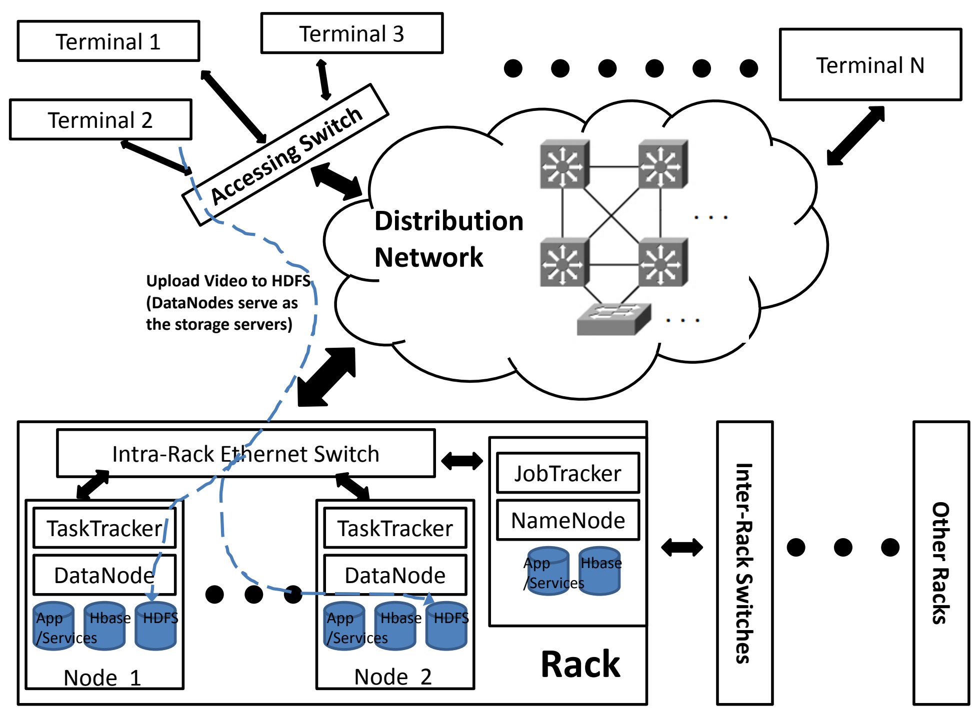

Fig. 3: Data center and networks in the ITS

A. Configuration for Data Center

A datacenter consists of physical hosts, storage and networking infrastructure, power/cooling hardware and necessary software frameworks. The hosts are organized into racks. Power is routed to each rack, and each rack also includes some amount of storage and network hardware.

The following is the configuration at one site of VMWare: “Racks in a datacenter come with a variety of power specifications that support a varying number of hosts. We assume configurations ranging from 10 hosts in a rack, representing the lower end of the power spectrum, to a very

TABLE III: Common configuration for server racks

Configuration	Number of Hosts/Rack	Cores/Rack
Small	10	320
Medium	20	640
Large	30	960
Future	40	1280

dense configuration of 40 hosts in a rack. For our default server building blocks, we use quad-socket servers with 8 cores per socket, for a total of 32 cores per host." [3]

Table III lists the common configuration (suggested by VMWare):

To calculate the exact number needed in the data center, we first estimate the network bandwidth and storage requirement, as will be discussed in the following sections.

B. Network Bandwidth

From servers' perspective, a single networking interface is used to multiplex both inter-rack/intra-rack storage-based traffic (result from block split and replication in HDFS) as well as the standard network traffic (video uploading, maintenance, other network services). The amount of traffic resulting from HDFS is affected by the frequency of file operations, running applications and important hadoop parameters such as *dfs.replication*.

Particularly when $dfs.replication = N_{DR}$, an uploaded video file will be split into small blocks and each of them is replicated to N_{DR} datanodes (including the datanode that provides the ftp service for the uploaded video file from a terminal). This indicates that each datanode should provide at least N_{DR} times as much bandwidth as required for uploading all the video files onto this datanode.

On the other hand, the frequency of file operation and running applications are entirely nondeterministic and could exhibit great variations. To design network in data center to meet specific bandwidth requirement, one may ask the following questions: What is bandwidth requirement observed from a synthetic benchmark? What bandwidth is observed in a production cluster with a mix of user jobs? What bandwidth can be obtained by the most carefully constructed large-scale user application? A research group from Yahoo [4] answered the above questions. Briefly, they draw the following conclusions:

1) They use DFSIO, a benchmark which measures average throughput for read, write and append operations, to test how much bandwidth a file processing benchmark could take. This MapReduce program reads/writes/appends random data from/to large files. It is designed to measure performance only during data transfer, and excludes the overheads of task scheduling, startup, and the reduce task. Their results shows that DFSIO reads files at $66MB/s$ per node and writes $40MB/s$ per node.

2) For a production cluster, the number of bytes read and written is reported to a metrics collection system. These averages are taken over a few weeks and represent the utilization of the cluster by jobs from hundreds of individual users. On average each node was occupied by one or two application tasks at any moment (fewer than the number of processor cores available). The reported result indicates that busy cluster read $1.02MB/s$ per node and writes $1.09MB/s$ per node.
3) Yahoo participated in the Gray Sort competition. The nature of this task stresses the systems ability to move data from and to the file system. The result shows a bandwidth consumption of $22MB/s$ per node.

Given the above three different scenarios, one can observe that both the running applications and user's behaviors utilizes network bandwidth quiet differently, from a few MBps to less than 100 MBps. We shall denote this part of bandwidth requirement as B_UA, total bandwidth support by the NIC on a server as B_S. To be conservative, we also assume an overhead of 30% of all the available bandwidth for interacting with namenode, exchanging metadata, maintaining the distributed database, synchronizations, other maintenance/management traffic, etc. Therefore, the rest bandwidth is used to support video file uploading from terminals. The number of terminals that one datanode/storage server can support is:

$$\frac{0.7 \times B_S - 8 \times B_{UA}}{(BR_{VMax} + BR_{IE})N_{DR}} \tag{10}$$

As a concrete example, we study a typical cluster node from Yahoo with the following configurations:

- 2 quad core Xeon processors @ 2.5ghz
- Red Hat Enterprise Linux Server Release 5.1 Sun Java JDK 1.6.0_13-b03
- 4 directly attached SATA drives (one terabyte each)
- 16G RAM
- 1-gigabit Ethernet

Let $B_UA = 22MB/s = 176Mbps$, $BR_{VMax} = 9.54Mbps$, $BR_{IE} = 0.24Mbps$ and $N_{DR} = 3$, then the number of terminals supported by one server node is $(0.7 \times 1G - 176Mbps)/((9.54Mbps + 0.24Mpbs) \times 3)$ = 18.

C. Storage Cost Modeling

The total storage cost (TSC, in GB) in data center can be calculated based on the mean bit rate of each camera/terminal (BR_{VMax} for VBR or BR_C for CBR), the number of terminals (N_T), the data retention time (T_{DR}, in days) and the number of copies/replications (N_{DR}, same as in the bandwidth modeling):

$$TSC = 3600 \times 24 \times T_{DR} N_T \frac{(BR_{VMax} + BR_{IE})}{8 \times 10^9} \tag{11}$$

D. Performance Modeling

Please refer to [5] for a detailed performance modeling of map-reduce tasks on hadoop-based distributed system.

REFERENCES

[1] G. M. A. Lombardo and G. Schembra, "An accurate and treatable markov model of mpeg-video traffic," in *Proc. of INFOCOM*, 2008.

[2] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Transactions on Commununications*, vol. 36, no. 7, pp. 834–843, 1988.

[3] V. Soundararajan and J. M. Anderson, "The impact of management operations on the virtualized datacenter," *SIGARCH Comput. Archit. News*, vol. 38, pp. 326–337, June 2010. [Online]. Available: http://doi.acm.org/10.1145/1816038.1816003

[4] S. R. Konstantin Shvachko, Hairong Kuang and R. Chansler, "The hadoop distributed file system," in *IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 2010.

[5] H. Herodotou, "Hadoop performance models," in *Technical Report, Computer Science Department Duke University*, 2011.