6000 series video stream user data format

Version 1.0

2005/5/12

MPEG-4 Simple Profile Video Encoder on Equator

MAPCA Technical User's Guide

Vivotek may make changes to specifications and product descriptions at any time, without notice.

The following is trademarks of Vivotek Inc., and may be used to identify Vivotek products only: Vivotek. Other product and company names contained herein may be trademarks of their respective owners.

Revision History

version	Issue date	author	comment
0.9	2005/5/11	Albus	First draft from internal document by Yun
1.0	2005/05/12	Joe	Update some fields definitions

TABLE of CONTENTS

OVERVIEW 5
DATA PACKET 6
NETWORK TRANSMISSION 9

1.1 Overview

Vivotek embedded some useful information in the video stream so that the developer can use them for advanced features in their software. The information includes digital input states, digital output states, motion detection, etc. This document describes the media data format used in 4000/5000/6000 series product.

Data Packet

Data Packet contains one or multiple frames of one media stream. For video stream, one Data Packet contains one frame. For audio stream, one Data Packet contains one or multiple frames.

To achieve 32-bits alignment, the size of user data must be multiple of 32-bits. The unit of "User Data Length" is 32-bits. Pad zero bits to achieve 32-bits alignment.

The layout of Data Packet:

Packet Size	Stream Type (FourCC)	Frame Type	DI Alert flag	Reserved	DO flag	Time Modified flag	Audio DI flag	Reserved
32 bits	32 bits	4 bits	4 bits	8 bits	2 bits	1 bit	1 bit	20 bits
	←---------- Data Packet size ---							

Second	Millisecond	User Data Length	User Data	Bitstream
32 bits	16 bits	8 bits	UserDataLength*4 bytes	Packet size – 16 – UserDataLength*4 bytes
----------→				

◆ Field descriptions:

Packet Size	32 bits	The size in bytes of Data Packet
Stream Type	32 bits	The unique ID for the codec type (FourCC) Currently support stream types are listed below: 1. "MP4V" => MPEG4 simple profile video 2. "JPEG" => Motion JPEG 3. "H263" => H.263 or MPEG4-SHM video 4. "AAC " => MPEG4 AAC audio
Frame Type	4 bits	Indicate the type of frame, intra, prediction or bi-direction prediction. For audio stream, the field is always zero. 0x0: Intra frame, (for MPEG-4, it means intra frame with VO header) 0x1: Prediction frame (P) 0x2: Bi-direction prediction frame (B)
DI alert flag	4 bits	Each bit is used to indicate the DI alert triggered by user defined (H/L). It will add the flexibility except the three alert windows to motion detection with extra

		devices. It supports four digital input sources in the present. The LSB indicates the first digital input source.
Reserved	8 bits	This field is reserved and should be set to 0.
DO flag	2 bits	Each bit is to indicate the digital output status. The LSB indicates the first digital output.
Time modified flag	1 bit	The time is modified according to timezone. This bit is always 1 in 4000/5000/6000 series. The “Second” field below represents the relative to 0h local time on 1 January 1970.
Audio DI flag	1 bit	Audio packets take the DI Alert information. This bit is always 1 in 4000/5000/6000 series.
Reserved	20 bits	This field is reserved and should be set to 0.
Second	32 bits	The second that the first frame in the Data Packet belongs to. It is relative to 0h UTC on 1 January 1970.
Millisecond	16 bits	The millisecond that the first frame in the Data Packet belongs to. The range is (0, 999).
User Data Length	8 bits	The number of 4-bytes in the User Data field
User Data	Varies	The content of user data is stream-dependent. The detail description is in the next section.
Bitstream	varies	The media bitstream. The size of bitstream can calculate from the packet size and header size.

◆ User Data in video stream

Motion Detect W0	Alert Flag	Percent	Axis	Motion Detect W2	Alert Flag	Percent	Axis	No Signal
1 bit	1 bit	7 bits	4*10 bits	1 bit	1 bit	7 bits	4*10 bits	1 bit
1	0	20	(1,1,10,10)	1	1	50	(50,50,60,55)	1

◆ Field descriptions

Motion Detect W#	1 bit	Indicate the on/off of motion detection. 0: disable motion detection 1: enable motion detection If the motion detection was enabled, alert flag, percent and axis fields follows. If the motion detection was disabled, there are no these three fields.
Alert Flag	1 bit	Indicates the movement is over the threshold the user

		defined.
Percent	7 bits	The percent of motion detection. It ranges from 0 to 100.
Axis	4*10 bits	The rectangle of window for motion detection. Each element is encoded in 10 bits. ✧ 1st element (the top-left x-axis): the real x1-axis plus one (+1). The range is (1, 1023) ✧ 2nd element (the top-left y-axis): the real y1-axis plus one (+1). The range is (1, 1023) ✧ 3rd element: the width of rectangle ✧ 4th element: the height of rectangle
No Signal	1 bit	Indicates the loss of video signal

◆ User Data in audio stream

In the present, there is no security information for audio stream. If the size of bitstream is fixed (mostly for speech), store the total number of access unit and the fixed size of access unit. If the size of bitstream is variable, store the total number of access unit and the size of each access unit in this Data Packet. Make the size of user data is the multiple of 32-bits for alignment. System needs to control the times of calling audio decoder. Calculate the time difference from sampling frequency. If there are multiple selections of sampling frequency and channel number in an audio stream type, the user data contains the sampling frequency and channel number.

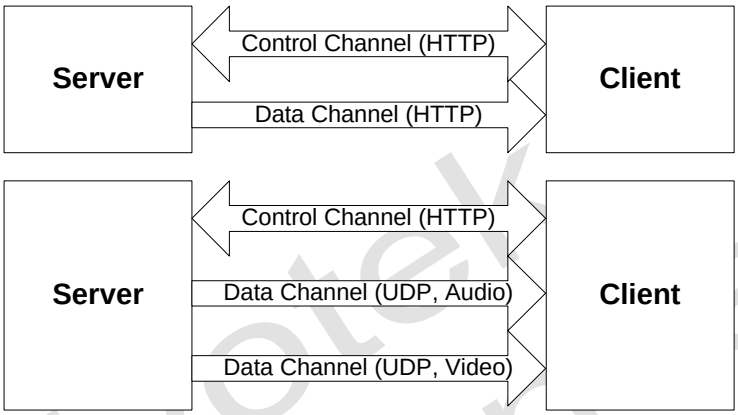
Fixed unit size flag	Sampling freq & channel number field flag	Reserved	Access unit number (n)	Access unit size (1)	…	Access unit size (n)	Sampling frequency	Channel number
1 bit	1 bit	6 bits	8 bits	16 bits		16 bits	24 bits	8 bits

● Field descriptions

Fixed unit size flag	1 bit	Indicating the size of access unit is fixed or not
Sampling freq & channel number field flag	1 bit	Indicating the existence of sampling frequency and channel number fields
Reserved	6 bits	This field is reserved and should be set to 0.
Access unit number	8 bits	The number of access units in this Data Packet
Access unit size	16 bits	The size in bytes of access units
Sampling Frequency	24 bits	The audio sampling frequency
Channel number	8 bits	The audio output channel configuration

Network Transmission

Between server and client, there are two channels for transmission when using HTTP. One is control channel. The other is data channel. When using UDP, there are three channels: two media channels and one control channel.



The server uses the control channel sending information to the client about the machine type, location, audio sampling frequency and channel number and if there is alert event (motion detection) or digital input alert happened, it sends the alert event to client. The control channel uses a reliable channel, HTTP. When data channel uses UDP, if there is any packet loss in client side, it can use control channel to inform server. The server can stop sending P frame and waiting next Intra frame to send. The data channel sends the Data Packet of media.

When using UDP, the packetizer is necessary. If the length of a UDP packet (i.e. the length of Data Packet in this case) is larger than the length of the network packet, the bottom layer will partition it into several network packets; in client side, it returned a UDP packet until all network packets belongs a UDP packet arrived. If there is some network packet lost, the timeout of waiting all network packets belongs a UDP packet arrived is not certain. Therefore, the behavior of receiving UDP packets becomes unpredictable. To prevent this situation, the Data Packet is partitioned into several packets using packetizer. Let the length of each UDP packets be smaller the maximum size of a network packet. In order to restore packet sequence and detect packet loss, additional information is added in front of a UDP packet.

For TCP and HTTP, we don't need the packetizer basically. But if the length of a Data Packet is larger than the size of TCP/HTTP buffer, additional procedure is needed to send a Data Packet (calculate the position of Data Packet sent). Therefore, we use

the packetizer for TCP/HTTP, too.

The layout of packet after packetizer:

Length	Sequence number	Media type	Packet flag	Data Packet

- Field descriptions:

Length	16 bits	The size in bytes of a packet, including the size of length
Sequence number	16 bits	The sequence number increments by one for each packet sent
Media type	4 bits	Indicate the type of media 0x1: audio 0x2: video
Packet flag	4 bits	Indicate the situation of Data Packet in this packet 0x0: a whole Data Packet in this packet 0x1: the first packet of a Data Packet 0x2: the last packet of a Data Packet 0x3: the rest packets of a Data Packet

The sequence number of every media stream is independent.