Secured Video Streaming Development On Smartphones With Android Platform

Danang Tri Massandy[#1], Dr. Ir. Rinaldi Munir, M.T.[*2]

Informatics Engineering, Bandung Institute of Technology
Jl. Ganesha 10, Bandung, Indonesia

[1]danangmassandy@gmail.com

[2]rinaldi@informatika.org

Abstract—**In this paper, we proposed an application for video streaming on Android smartphones that can capture video from camera smartphone and then send it to the computer in real-time. On the computer, the video is played using a video player such as VideoLAN VLC. The rapid development of information cause the information exchanged is very sensitive and important, as well as the information on the video. With cryptography, information on the video can be secured by selective encryption of critical data in the video. Selective encryption only selects important part of the video which will be encrypted. H.264 video format is most used today in various multimedia applications. In selective encryption on H.264 video format, slice data on slice with type I is selected as part of the I-Frame to be encrypted. On proposed system, selective encryption process is performed on Android smartphones while selective decryption is performed on the computer. Based on the test results, the video can be selectively encrypted by selecting slice data on slice I so that video image become broken and difficult to see. However, in certain circumstances there are images on the video that are not broken or damaged even though all of slice I have been encrypted.**

Keyword— **video streaming, H.264 video, RTP, selective encryption, Android smartphone**

I. INTRODUCTION

Video streaming allows a person to see or witness an event without having to be in the same place. This technology can be used for various communication such as between families, friends, the enterprises, or governments. Video streaming requires devices that connect to network. Network is necessary to transmit multimedia data from a sender to a receiver through Internet.

Devices can be a computer, tablet or smartphone. These devices must be connected to the appropriate network. The use of smartphones is easier to be carried everywhere than a computer. One popular type of platform for smartphones is Android. Android platform is a mobile operating system first developed by Android Inc., and then was acquired by Google[1].

The use of video streaming technology in particular smartphones with Android platform makes it easier to report an event or share an important event to others. We can share our moments or events that occur around us with friends, families, or coworkers using this technology anytime and anywhere because it uses smartphones that we carry everywhere we go.

This video streaming technology makes it possible to transmit video in real-time from Android smartphones to a computer. Data transmission can stream through the Internet so that it can be viewed by anyone. On the computer, the multimedia stream which is consists of both video and audio can be played by a video player such as VLC VideoLAN.

An example of the use of this technology is to send a video of the concert-going so that the others can see the concert without having to come to the concert. To view the transmitted video in live or real-time, one has to pay a certain fee if the concert is closed to the public.

Video content that can be streamed can vary, ranging from day-to-day activities, a unique event, or private/confidential that concerns company business, government or military activities. A paid streaming video content or a private streaming video content is highly confidential and require security guarantees when transmitted on the Internet.

However, the security of the data that is transmitted through network (Internet) can be compromised. The confidentiality of data needs to be considered. This is because the data on the Internet is spread around to the world that is possible for data theft. If a third party successfully got the data from the video stream, then that person can monitor all activities that occur include important information in it. Therefore, we need a method of securing data sent to the Internet.

Data security method can use any existing encryption algorithms in cryptography such as Advanced Encryption Standard (AES). AES is a symmetric key encryption standard that was proposed by J. Daemen and V. Rijmen[2].

The use of encryption for data on video streaming should pay attention to resources on the smartphone device which is very limited. Efficiency was assessed by two criteria: a small computational time in order to process video data and memory usage according to the resources on the smartphone.

In this paper, we proposed an application for encrypting video streaming on smartphones with Android platform. The rest of the paper is organized as follows: Section II briefly discusses some basic theories; Section III describes the implementation of proposed system; Section IV presents experiment and testing for the application; Section V discusses the analysis of the test results; and finally Section VI concludes the paper.

978-1-4673-4550-7/12/$31.00 ©2012 IEEE

II. Literature Study

Some basic theories will be discussed in this paper, namely, digital video, RTP protocol, and AES algorithm.

A. Digital Video

Digital video is a type of video recording using a digital video signal. Digital video was introduced in 1990. In that year, the change from analog video to digital video requires a prohibitive cost on hardware for compression [3]. Therefore, from year to year for digital video compression methods is growing as MPEG-1, MPEG-2, MPEG-4, H.261, H.263, H.264 or MPEG-4 AVC, and VP8.

Some of the characteristics of digital video are owned by frame dimensions, pixel depth, and framerate. Frame dimensions are the resolution of the video frame, it is measured in pixels. Pixel depth is the unit used to measure the number of colours shown in each image pixel. Frame rate is frame speed which are displayed in one second, measured in fps (frames per second).

MPEG is a compression format for digital video. Digital video compression is used for saving the size of a video especially for video transmitted in the network. MPEG compression format itself is the most commonly used. The algorithm used for compression in this format is JPEG algorithm.

There are four sequences of images in MPEG architectures, namely I-Frame, P-Frame, B-Frame, and D-Frame. The four frames are distinguished by compression. I-Frame is compressed independently of other images with intraframe compression method. P-frame motion compensation compressed by the method of I-Frame or P-Frame before. Meanwhile, B-Frame is compressed using bi-directional prediction based on interpolation of the I-Frame or P-Frame previously [4].

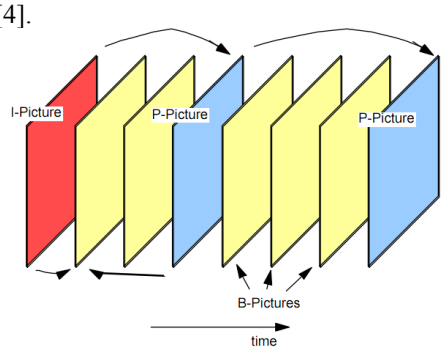

Figure 1: Group of Pictures (GOP) in MPEG format [5]

Set of frames in MPEG format called a Group of Pictures (GOP), which is shown in Figure 1. In the GOP there are usually 8-24 frames. GOP shows the movement of the frame with another frame, such as P-Frame only points to the previous frame and B-frame only point to the previous and next frame.

One type or version of MPEG is H.264. H.264 NAL unit is based on a collection of data as a wrapper layer underneath. NAL unit types are divided into two groups based video coding,the VCL (Video Coding Layer) and Non-VCL. VCL NAL unit is a NAL unit that contains video data, while the Non-VCL NAL unit contains other information about the video.

Overall structure shown in Figure 2. From NAL units wrap slice layer. In this layer, the slice is divided into two parts, namely the slice header and slice data. Then, in the slice of data divided into a collection of macroblock. Actually video frame data contained in the macroblock layer.

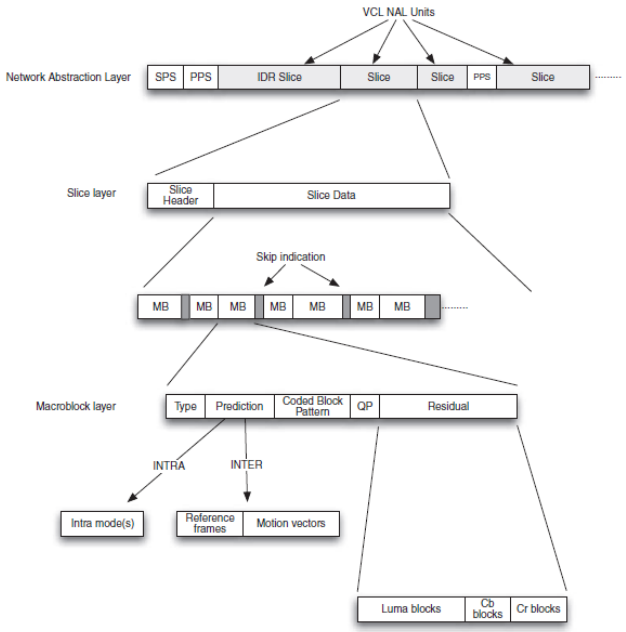

Figure 2: Structure of Byte Stream H.264 [6]

B. Real-Time Transport Protocol (RTP)

Real-Time Transport Protocol (RTP) is a standard protocol for sending packets in real-time, ie audio and video.RTP are in user space and running over UDP protocol. RTP is made to complement the features of the UDP but it runs in real time. Real-time applications that require a specific protocol requirements specification due time for data transmission is very significant impact on the application. This protocol is used in a variety of applications, such as internet radio, mobile internet, video conferencing, music-on-demand, video-on-demand [7].

RTP is a transport protocol, but RTP is implemented on the application layer. From this design, illustrate that a transport protocol (RTP) running over the UDP transport protocol other.audio or video sources from multimedia applications which then formed into RTP packets. After the RTP packet is formed, UDP packets are generated and attached to the IP packet. IP packets are sent over the network (subnet) connected.

Format header in the RTP protocol can be seen in Figure 3. Minimum length of the header is 12 bytes. Explanation of the header sections as follows,

1) Version (V): the version of the protocol used.
2) Padding (P): indicate whether there are additional bytes at the end of the RTP packet.

3) Extension (X): indicates whether there is a standard extension header between RTP payload and header.
4) CRSCCount (CC): contains the number of CSRC identifiers.
5) Marker (M): used in application-level and is determined by the application that makes the RTP packet.
6) Payload Type (PT): shows the format of the data payload.
7) Sequence Number: a number of RTP packets sent.
8) Timestamp: used by the receiver to restart the data received at certain intervals.
9) SSRC identifier: used to identify the source of the stream.
10) CRSC identifier: as a contributor ID sources of data if the data is sent from many sources.
11) Extension headers: an optional field.
12) RTP payload: contains data to be transmitted and is specified by the application.

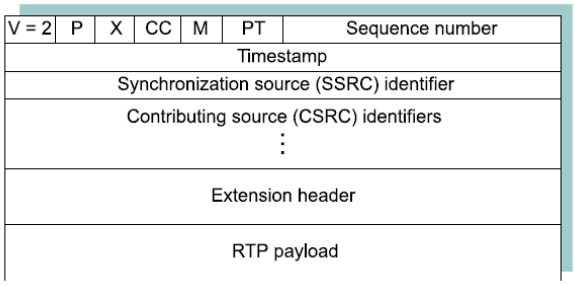

Figure 3: Format of RTP header [8]

C. Algoritma Advanced Encryption Standard (AES)

Standard encryption algorithm at first (since 1973) is the DES (Data Encryption Standard). However, the security of DES has been questioned by the U.S. around 1993. This is because the development of computer technology can not be matched by the design of the DES algorithm itself [9]. Therefore, the National Institute of Standards and Technology (NITS) held a contest to find a new standard encryption algorithm in 1997. Standard new algorithm is named AES (Advanced Encryption Standard).

AES is a symmetric algorithm based chipper block. There are two variants of the AES-128 AES (128-bit block size with 128-bit key length) and AES-256 (128-bit block size with 256-bit key length). Also supports AES with 192-bit key, but AES with 192-bit key length is seldom used.

This adjusts the size of 16 bytes of data block size and a key size of 128-bits to be inserted into the array. AES operates on a 4x4 matrix of bytes (for 128-bit block of data) are called state. Initialized state two-dimensional array with plaintext and modified each time step for computation. Broadly speaking, the process of the Rijndael algorithm can be seen in Figure 4.

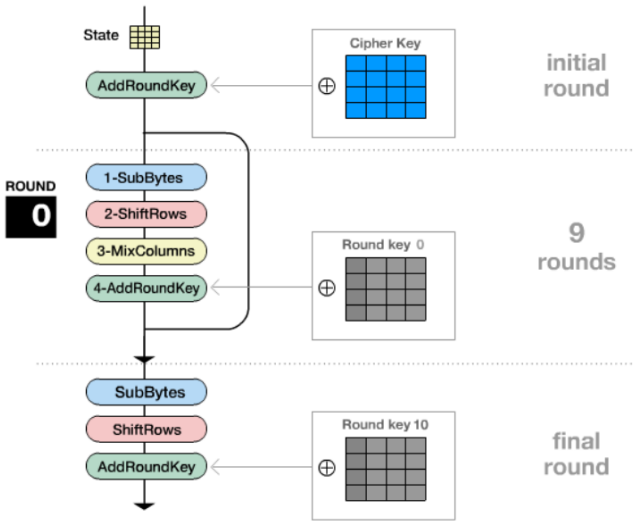

Figure 4: Process Flow AES Encryption [10]

III. IMPLEMENTATION

A. General System Architecture

Video streaming application with selective encryption is named *S-Streaming*, which is a continuation of the Secured Streaming. *S-Streaming* consists of two applications, application on Android and Java application on a computer.

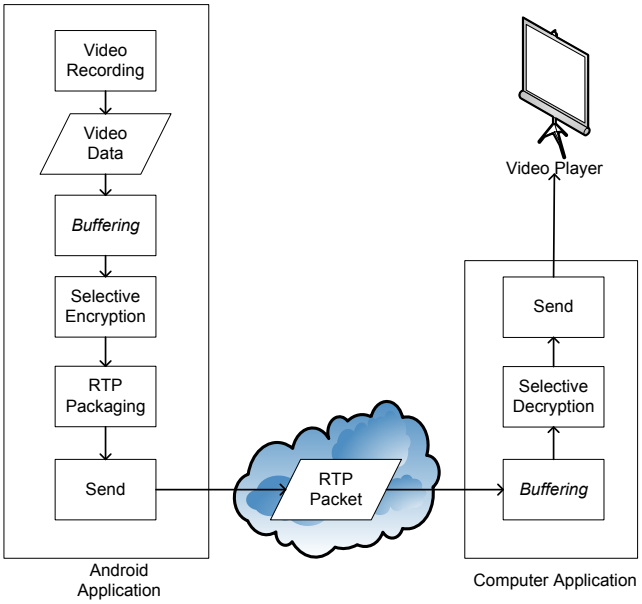

Figure 5: General Architecture of *S-Streaming*

S-Streaming on Android is an application that records video, and sends it to the application on the computer. Then from application on the computer, the video was passed on to the video player for playback. *S-Streaming* architecture as a whole can be seen in Figure 5. As in Figure 5, the encryption is done on an Android application, while the decryption process is done on the application on the computer. The interface of *S-Streaming* on Android is shown in Figure 6.

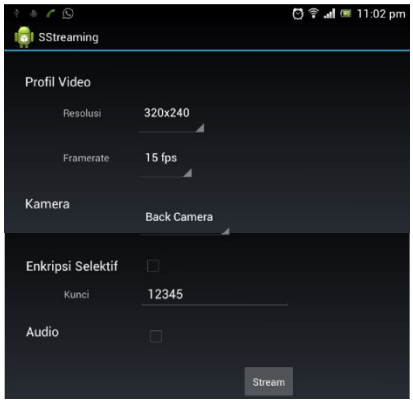

Figure 6: The Interface of Android Application

Some restrictions on the implementation of applications that are built are as follows,

1) Video streaming is done using Wi-Fi connections in the local network.
2) Target applications on the Android platform with minimal 4.0.3 API 15.
3) Decryption applications running simultaneously on a computer with VLC video player application.
4) Application decryption on the computer does not address the issue of sequencing of RTP packets received.

B. Selective Encryption Process

The process of encryption and decryption of selective attention to the structure of H.264 video streams, namely by looking for where is the I-frame. In the H.264 video data packets that stores video streams called the NAL unit (Network Abstraction Layer unit). Each NAL unit can save part of the slice. In the slice itself are parts of a video frame.

Selective encryption for H.264 video done by finding NAL unit containing slices of type I. These slices are at a particular NAL unit is marked with the type of NAL IDR (Instantaneous Decoding Refresh). NAL type IDR is characterized by a value of type 5.

After getting the IDR NAL units, the next step is to read the header of the slice contained in the NAL unit. The reading is done using VLC (variable length coding). Then the encryption can be performed on the data that follows the slice header. Selective encryption process can be seen in Figure 7.

Slice data that have been obtained are encrypted using AES. Block cipher modes that can be used is the OFB which yield the same length as the plaintext cipher text. However, the use of AES OFB mode requires IV (initialization vector) which is unique according to the recommendations of NIST 800-38A [11].

IV which will be used throughout the 16 byte consists of 8 byte random IV and 8 byte counter. Sequence value in RTP packet header can be used as a counter because its value always increases. Thus, each RTP packet will use a different IV values for encryption.

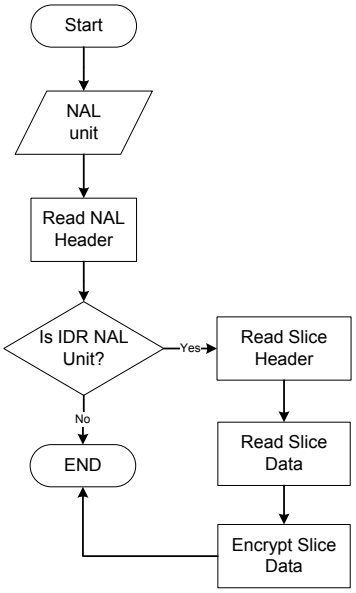

Figure 7: Selective Encryption Process

IV. EXPERIMENT

Experiment for the application was done on purpose as follows,

1) Determine whether all the functional and non-functional requirements are met.
2) Testing the reliability/performance of the application (delay and packaging computation).
3) Testing whether the streaming video data has been successfully selectively encrypted when transmitted.

For the use of selective encryption, the test results were obtained as shown in Figure 8. The average delay obtained from the test are shown in Table 1.

Figure 8: Test results of selective encryption

Table 1: Average Delay

Test Case	Average Delay (ms)
Without encryption	211.7
With encryption	219.4

Performance testing measurements in addition to measuring the amounts of delay is the total package that is produced and delivered. The video profile for this testing is using 176x144 pixel of video resolution with variable values of frame rate. Video streaming is taken within 30 seconds. The test results can be seen in Table 2.

Table 2: Packaging Computation

Type	10 fps	15 fps	20 fps
Total Packet	677	805	863
Total *Slice*	322	476	629
Total *I-Slice*	30	30	30
Total *P-Slice*	292	446	599

Security testing is done to check whether the RTP packets that was sent is encrypted correctly. The test case is to compare the NAL Unit data before encryption with the data after encryption. The encrypted NAL Unit data is taken from the network using wireshark application.

Part of the data before it is encrypted is shown in Figure 9, while Figure 10 shows the data after it is encrypted.

```
7c85b8000206fc66221806df07400040af16a60ef097d341d406a2
95c000a970403048a50f499a11e2f4b884879b8997587262e7919e
3a110a1a1e649 …
```

Figure 9: NAL Unit before encryption

```
0030  1e b0 f0 86 7c d3 7c 85  b8 00 02 06 fc 99 7d 47
0040  ce ea 4c 84 18 a3 a0 1e  52 13 a0 4b d2 93 7b d5
0050  dd c7 8f 66 77 dc 63 33  31 02 be 20 7e 12 41 9d
0060  ce 57 9d 32 90 d3 03 fc  e5 11 a4 99 dc 52 8a d3
0070  7f 9b 21 7d 37 b0 f9 db  3c 7d b1 31 f4 11 07 31
0080  1a 4f 34 0e 1b fa 7c 07  c1 6e 7a db 87 2f f2 4e
0090  f9 b4 a0 83 9f 7c 16 be  8f 57 c7 f3 dc 52 19 9c
00a0  fc 88 80 e5 5f b2 38 f2  ba 62 4c d6 7a e4 4a 63
00b0  33 cd 15 a1 1d 28 b9 ec  3a cb 46 9d ba 88 4c bb
```

Figure 10: NAL Unit after encryption

Some initial byte is still the same, namely 0x7c85b8000206fc. These bytes indicate 2-byte header and indicator of FU-A and the rest is the header of the slice[12]. After these bytes, the data is not the same which shows that the slice data is successfully encrypted.

V. Analysis

On the experiments for selective encryption/decryption of video, there are some images that are still partly visible. For example, as shown in Figure 11. In this figure, it still can be seen that the object inside the image can be recognized as a portable computer. From this image, the quality of video encryption is very unsafe because objects in the image can still be seen and recognized.

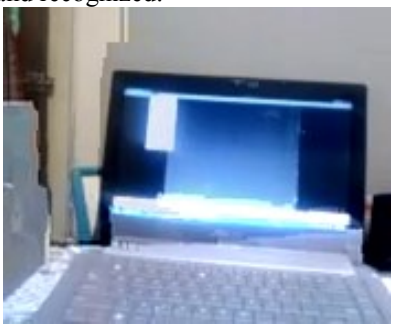

Figure 11: Picture that is not perfectly encrypted

Based on encryption security testing that is performed by comparing NAL units which have not been encrypted with NAL units which have been encryption, it can be ensured that all the NAL units which contain slice I has been encrypted. This security testing indicates that the encryption process has been performed correctly on data slice. Encryption has been done carefully in order to avoid the slice header. It is intended that the encrypted video can still be played by the video player.

Although all of NAL units which contain slice I have been encrypted, the Figure 11 shows that the encryption is still not perfect. Based on packaging computation testing, it is known that the number of slice I is much smaller than the number of slice P. Slice I has fixed frequency of occurrence which is 1 slice/second while slice frequency of occurrence of slice P depends on the value of frame rate.

The possible causes in Figure 11 that is still can be seen is because there is a part from slice P that is not encrypted. These images are often occurred when there is a movement of the object or movement of the camera smartphone.

According to H.264 standard document, it says that every slice P is possible to contains macroblocks with type I or P[13]. If slice P contains some macroblock with type I, then these macroblocks does not require certain parameters (eg. motion vectors) in the decoding process. In addition, with the fixed frequency of occurrence of slice I is just 1 slice/second, the H.264 encoder will put macroblock with type I into slice P. Thus, these slice P can still be decoded into almost perfect frame without the need for another frame in the decoding process.

Delay measurement is done by calculating the time difference between the time of decryption program will send RTP packets to video player with the time of streaming program will send RTP packets to decryption program. Delay measurement is calculated with consideration of synchronization time between Android smartphones with the computer.

If encryption is used, then the delay is calculated based on the difference between the time of decryption program finish the decryption process of RTP packets with the time of streaming program start the encryption process of NAL units. From the test results, it is obtained an average delay if no encryption is used at 211.7 ms, while the average delay of 219.4 ms when encryption is used. This delay can be tolerated because according to standards that specified by the ITU-T, tolerated delay is less than 10 seconds for video streaming [14]. With a maximum delay of 219.4 ms, the application can be run interactively.

If seen from these results, the use of encryption or no encryption process is not produced very significant delay. Delay that has been measured is a combination of the time of RTP packets packaging, encryption, transmission over a network, as well as the decryption time. Because testing is used during local Wi-Fi network, the delivery time can be assumed to be very small.

However, there are other types of time which need to be considered for the calculation of delay, such as the time of the video data stored in the buffer, the time of video data processed by recorder, the encoded time of video data by video encoder. In addition, the video player, VLC, is possible

to has buffer storing video data so that the time video data stored in this buffer can also be added to delay.

VI. CONCLUSION

Based on the analysis of the test results, the system *S-Streaming* software can run well. The software can perform video streaming of videos captured using the camera on a smartphone with the Android platform and sends it to the computer. On the computer, video/audio data can be played well using VLC video player.

Software system S-Streams can also do selective encryption/decryption of video data. The selection process is done by selecting slice with type I only. Then the slice data is encrypted.

By encrypting data only in slice I, time for parsing H.264 structure can be reduced. The results of selective encryption is pretty good, although there are some pictures that still can be recognized. This is because the slice P contains unencrypted data frames in the video which can be decoded without relying on slice I.

Delay of selective encryption or decryption process is tolerated, namely the value of 219 ms so that the resulting video streaming smoothly enough to be seen.

For further system development, selective encryption can be continued to the macroblock layer so that video is more guaranteed its safety. Selective Encryption H.264 parser has to be effective and efficient because of limited resources on Android so that the computing time of the parser can still be tolerated.

REFERENCE

[1] Jeyaganesh. (2010), "Introduction to Android Operating System". [Online]. Available: http://devlup.com/mobile/what-is-android/348/.

[2] Daemen, J.,danRijmen, V. (1999), "Rijndael AES Proposal". *AES Algorithm Submission*.

[3] Pctechguide. (2012), "The History of Digital Video". [Online]. Available: http://www.pctechguide.com/uncategorized/the-history-of-digital-video.

[4] Basith, S.A. and Done, S.R. (1996), "Digital Video, MPEG and Associated Artifacts". [Online]. Available: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/sab/report.html.

[5] Girod, Bernd. (2007), "Video Coding Standards". [Online]. Available: http://www.stanford.edu/class/ee398b/handouts/lectures/04-StandardsMPEG124.pdf.

[6] Richardson, Iain E. G. (2010), "The H.264 advanced video compression standard 2nd Ed". Chichester, West Sussex, United Kingdom: John Wiley & Sons, Ltd.

[7] Tanenbaum, Andrew S. (2003), "Computer Networks". New Jersey : Prentice Hall PTR.

[8] Peterson, Larry L. dan Davie, Bruce S. (2007), "Computer Networks A Systems Approach". San Fransisco, CA : Elsevier, Inc.

[9] Anderson, Benjamin. (2007), "Advanced Encryption Standard". [Online]. Available: http://home.eng.iastate.edu/~hawklan/aes.pdf.

[10] Munir, Rinaldi. (2011), "Advanced Encryption Standard (AES)". [Online]. Available: http://www.informatika.org/~rinaldi/Kriptografi/2010-2011/Advanced%20Encryption%20Standard%20(AES).ppt.

[11] Dworkin, Morris. (2001), "Recommendation for Block Cipher Block Operation, Methods and Techniques". *National Institute of Standards and Technology Special Publication 800-38A*. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-38a/sp800-38a.pdf.

[12] Wenger, S., Hannuksela, M.M., Stockhammer, T., Westerlund, M., Singer, D. (2005), "RTP Payload Format For H.264 Video, RFC 3984". [Online]. Available: http://www.ietf.org/rfc/rfc3984.txt.

[13] ITU-T. (2005), "Advanced Video Coding for Generic Audiovisual Services : Recommendation ITU-T for H.264". [Online]. Available: http://www.itu.int/rec/T-REC-H.264-201201-I/en.

[14] ITU-T. (2001), "End-User Multimedia QoS Categories : Recommendation ITU-T for G.1010". [Online]. Available: http://www.itu.int/rec/T-REC-G.1010-200111-I.