LiveCity: A Secure Live Video-to-Video Interactive City Infrastructure

Joao Goncalves[1], Luis Cordeiro[1], Patricio Batista[1], and Edmundo Monteiro[2]

[1] OneSource, Consultoria Informatica, Lda.
{joagonca,cordeiro,pmbento}@onesource.pt
http://www.onesource.pt
[2] Centre for Informatics and Systems of the University of Coimbra
edmundo@dei.uc.pt
http://www.cisuc.uc.pt

Abstract. In typical video-to-video transmissions, security and confidentiality is becoming an issue of greater importance, but these features come at a cost. In mobile environments, where CPU time is a valuable resource, such features should be thoroughly thought over as they usually require heavy computational resources. In this paper a short analysis on existing streaming solutions, standardised and otherwise, is performed while taking into consideration the scope of the LiveCity project of developing applications destined to the end-user. An analysis of different transmission protocols and their specifications, as well as encryption protocols designed to work on top of streamed data, is performed as a means to access which specifications better fit LiveCity requirements.

Keywords: LiveCity, Video-to-Video, Mobile, Video encryption, Application, SRTP, ZRTP.

1 Introduction

The use of video-to-video applications on mobile environments has become a reality, from the fictional movies of the past, to the present day. A new level of interaction can be achieved through the use of video, superseding the traditional voice and SMS services. The use of simultaneous video and voice literally escalates interaction, increasing the capabilities of this resource to a superior level. With the expansion of resources and capabilities comes the threat of content tampering, and despite the existence of this threat, nowadays some of the video-to-video applications on mobile devices do not support video security mechanisms that ensure the confidentiality of the data being transmitted. Most of these applications are for personal use only, such as social networks and public web content video streaming, hence the video being transmitted has no security requirements. But institutions and companies are starting to use these devices to extend and improve their work, and security questions begin to arise.

LiveCity project addresses a number of communities where citizens of a city have specific challenges, from which can derive benefits through the use of live interactive

L. Iliadis et al. (Eds.): AIAI 2012 Workshops, IFIP AICT 382, pp. 260–267, 2012.

video-to-video. These communities include emergency ambulances, hospitals, doctors, patients, museum curators, city administrations and schools; institutions whose transmitted information could contain sensitive data. LiveCity implements a range of pilots for city communities across the public internet by building a wireline and 4G wireless network of 5 cities, including a right of way without interference from unwanted traffic, with authentication and media encryption so that each user can experience live and secure interactive video-to-video.

This article will start by giving an overview over existing video-to-video encryption solutions while describing some of its limitations, followed by a description of mobile applications while providing a general insight on video-to-video applications, and finally, concluding with a description of LiveCity project proposal.

2 Existing Video-to-Video Encryption Solutions

Video transmission technologies are widely used in the domestic sector, multimedia and communications in general. This provision usually occurs through transmission channels which are of public domain, therefore vulnerable to attacks from malicious sources. As a result, video security has become of paramount importance.

Commonly, the process of encoding video[1][2] is quite similar, usually by coding through Direct Cosine Transform (DCT) data, quantisation, and entropy coding. A video is composed by various frames, and one frame is composed of several macroblocks. Ultimately, a macroblock could be partitioned in several DCT blocks, which according to most of the available solutions, is the section of the video frame that goes through the process of encryption.

In current literature it exists several proposals and techniques aiming to provide encryption over video-to-video communication. The Secure Real-time Transport Protocol (SRTP)[3], which is a profile definition for the Real-time Transport Protocol (RTP)[4], provides encryption, message authentication and integrity in multimedia transmissions. The SRTP uses the Advanced Encryption Standard (AES) algorithm[5], together with its control protocol Secure Real-time Control Protocol (SRTCP)[3][6][7] to encrypt and decrypt the data flow, thus providing full confidentiality to the data transmitted.

The exchange of keys for encryption is not contemplated by any of the Requests For Comments (RFC) that define SRTP, having these values to be previously set on both communicating ends. As a key-agreement protocol to solve the issue with SRTP key negotiation, ZRTP appears. The ZRTP[8] protocol is widely used in Voice over IP (VoIP) communications, such as iCall[9], PJSIP[10], and Zfone[11], among others, and it does not require any Public Key Infrastructure (PKI) or certification authorities, executing instead a Diffie-Hellman or an Elliptic Curve Diffie-Hellman key exchange algorithm between peers, coupled with protection against man-in-the-middle attacks.

A new draft proposal, the Hypertext Transfer Protocol Live Streaming (HLS), is currently in the works[12]. This format is better suited for a server-client approach, since its main design is to transfer pre-recorded media files across the network rather than perform live streaming, although possible. The HLS protocol does not support

encryption, neither any form of authentication per se, relying on other forms of security. These security mechanisms can be implemented in the server[13] where HLS is installed, by configuring a Secure Sockets Layer (SSL) certificate, and depending on the technology used in the server, by enabling any authentication mechanisms available.

Prior to the proposal of HLS, only Hyper Text Transfer Protocol (HTTP) was used to transport video through TCP connections, but due to the growth of the mobile market, demand has augmented towards more resilient alternatives. The HLS protocol innovates by providing the same stream at different bitrates, enabling the user to switch streams according to its available bandwidth. To provide a secure HTTP stream, security methods similar to the ones employed in HLS are resorted to.

Currently undergoing further research and development, from the Internet Streaming Media Alliance (ISMA) comes the ISMA Encryption & Authentication specification. This specification, unofficially referred to as ISMACryp, presents itself as a framework for secure content delivery over IP networks[16] guaranteeing interoperability between encoders and streaming servers that respect standard-based technologies. Thought to work on top of RTP streams and ISO based media files, ISMA Encryption & Authentication abstracts from the codec used, shifting its focus towards content encryption and integrity assurance amongst devices. As of today, some of the Digital Video Broadcast (DVB) standards employ ISMA Encryption & Authentication specification on all of its MP4 ISO streams where a Digital Rights Management (DRM) application is mandatory[17].

Also, in the past years, research work that has not reached standard status has been performed regarding video encryption. Hung-Min Sun et al. proposed a selective video encryption using context-key control by modifying ElGamal Encryption control[18], a public key based encryption which is one of the best known cryptographic systems, proposed back in 1982 by ElGamal[19]. However, having worse efficiency than the deterministic encryption of algorithms like the Rivest-Shamir-Adleman (RSA) algorithm, the adoption of ElGamal based encryptions is proving difficult.

Video transport protocols whose implementations do not support any type of encryption nor authentication, such as Microsoft Media Server (MMS)[14] and the Real Time Streaming Protocol (RTSP) that superseded MMS[15], continue to be widely used in equipment's and streams across the Internet, despite the stated limitations. Secure communication using these protocols, although possible by resourcing to the establishment of secure tunnels between both ends, is beyond the scope of this project, therefore, not considered.

When in mobile environments, most video streams use the obsolete HTTP, or RTP when UDP transmission is available and preferred. Even though, none of the previous protocols are commonly used in live video-to-video transmission. Encountered limitations range from the server-client oriented approach used by HTTP and the new HLS draft proposal, not suitable for peer-to-peer communications, towards the fact that demand oriented protocols like RTP and ZRTP+SRTP exclusively towards VoIP implementations.

3 LiveCity: Mobile Solutions

More than ever, the use of mobile solutions to solve everyday problems rises as an efficient interaction method. Live high-quality video-to-video interaction is currently possible thanks to modern videoconferencing systems, but still lacking proper support in mobile environments. Limitations in bandwidth for mobile devices and the inability to assure proper quality of experience for these delay-sensitive applications stalled the development for such environments. The advantages of associating Live high-quality video-to-video with mobile environments are plain to see, from personal and social environments up to the corporate world. Distant friends and family can now remain closer and share special moments with the commodity of a mobile solution. Elderly and voice impaired people can also take significant advantages from a mobile video-to-video solution, facilitating communication and interaction with others. Telemedicine can take the advantages of a mobile video-to-video solution to extend the reach of consultations and on-the-go healthcare support to distant patients, presenting solutions and assistance right when needed. Video-to-video on tele-education can leverage the traditional online learning systems by allowing a new form of live interaction with the academic community, remotely and on a mobile environment. The use of a mobile video-to-video solution in the corporate world represents one of the most significant changes in remote interaction, allowing real-time videoconferencing possibilities in a mobile environment, effective commutation costs reduction and an improved business continuity.

By combining recent technologies and focusing in the development of a live high-quality video-to-video interaction system, LiveCity intends to provide a sophisticated communication infrastructure to use with a variety of applications, ranging from critical lifesaver uses to information providers. The interaction approach does not rely solely on high-quality video-to-video communications, but also on other forms of simultaneous data presentation mechanisms and feedback, defining a more immersive and rich experienced form of communication. Although video-to-video with mobility is a current possibility, it may not sustain the necessary requirements for certain use cases.

LiveCity stands on top of access technologies such as wirelines and 4G wireless networks, supported by a Virtual Path Slice (VPS)[20] controller derived from a FP6[21] project and owned by RedZinc[22]. The wireline and the 4G wireless network provide the necessary bandwidth requirements for the infrastructure, and the VPS controller gives a right of way without interference from unwanted traffic. By combining these two main characteristics with a video-to-video platform, LiveCity aims to deploy a sophisticated high-quality video-to-video platform pilot to over 3000 users in 5 european cities, with a variety of target users and scenarios.

When dealing with sensitive information or when privacy is a key aspect, secure measures must be in place. Being no exception, LiveCity is aiming to provide a secure transport layer for video, audio and any other form of data, when required. As referred in section 2, HTTP and RTP over UDP are the most used transports in mobile video stream scenarios, although not in a peer-to-peer basis. LiveCity intends to combine communication protocol standards such as RTP+SRTP+ZRTP for use in

secure, high-quality live video-to-video communications. The key challenge for this implementation resides in the capacity to ensure high-quality live video-to-video communications over an heterogeneous environment of equipment's.

It is a fact that mobile environments sustain rather different requirements and concerns, unlike the physically fixed ones. For example, mobile devices depend on battery life in order to operate; they must be portable, so they are usually smaller in size, which also leads to the necessity of a smaller battery, resulting in a more limiting battery life. They are also usually limited in terms of processing power, again, most often due to the need of a low power consumption and small size. Existing applications for mobile environments with video-to-video capabilities, such as Skype, offer live video calls but often without acceptable image quality, partly due to the lack of a right of way during network transmissions.

Nevertheless, the most important factor in LiveCity video-to-video communication resides in the ability to maintain an uninterrupted real-time video-to-video interaction over low bandwidth mobile scenarios while abiding with the security standards for privacy and integrity. Such low bandwidth mobile environments include the use case described further in section 4, where an ambulance uses a live video-to-video interaction with the hospital during the medicine's "golden hour", also known as the time period during which proper medical care can prevent future complications. The specified scenario involves frequent changes in signal reception and available bandwidth, provided by the nearest cell tower for example.

4 LiveCity Proposal

By providing a set of applications whose intention is to handle all previously stated limitations, LiveCity aim is to support the creation of peer-to-peer communications, aided by a server set to, at least, handle authentication and encryption negotiations between devices, as depicted in Figure 1. These applications comprise three different sections in LiveCity proposed architecture.

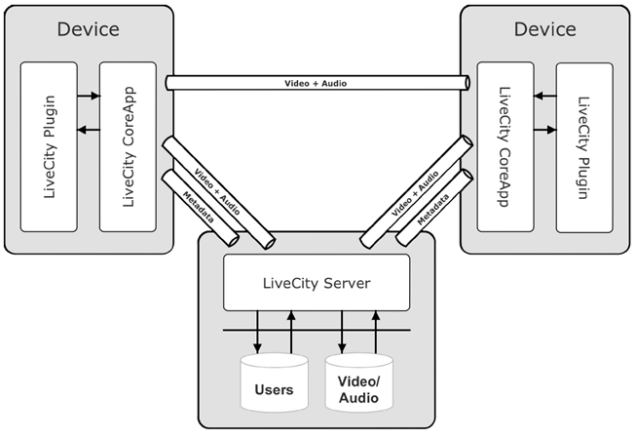

Fig. 1. LiveCity application architecture proposal

A core application for devices will be provided to aid in handling all communications, negotiations and encryption between devices, establishing and maintaining the video-to-video connections, as well as the connections to the server when required. Its initial design implies that it should receive via a specific interface, a video stream with an optional audio stream, and a secondary stream of metadata, packetizing this information and transmitting it to the other end. The application should be interoperable between different platforms and provide means for plugins to be developed and easily integrated, as the application will not provide means to acquire and display video/audio media and metadata, being these inputs fed through the use of the aforementioned plugins.

Being developed on top of the core application, intention is to create plugins capable of capturing video and audio from the device hardware, serving this media to the core application via a specific interface. Some flexibility will be given to the format of the captured media, being this handled through a signalling protocol between both applications. Metadata transmitted using LiveCity solution should be handled solely on the plugins end, being totally transparent to the remainder of the intervening parties.

Handling all the authentication procedures and encryption keys generation, LiveCity architecture comprises the installation of a central server holding the credentials of authorised users. Besides handling the negotiation between peers, the central server doubles as storage with the purpose to record the video-to-video communication when requested by both ends.

Table 1. Candidate streaming protocols

	Low latency	Defined standard	Supports encryption	Supports control messages	Open standard	Transport protocol
RTP	✓	✓	✓	✓	✓	UDP
HLS	X	X	✓	✓	✓	TCP
RTSP	✓	✓	X	✓	✓	UDP
MMS	✓	✓	X	✓	X	TCP/UDP
UDP	✓	✓	X	X	✓	UDP
HTTP	X	✓	✓	X	✓	TCP

To address all of the security concerns involved in such application, the proposed architecture that is currently in study, suggests an implementation of a custom RTP+SRTP+ZRTP protocol stack. The encryption process is intended to only process key elements (frames) from the RTP payload, lowering the need for constant processing and thus, reducing the communication overhead and permitting a sustainable live video-to-video interaction over low bandwidth mobile scenarios, since some of the video-to-video solutions for mobile environments present today are supporting this requirement, at a cost of a lower quality experienced[23][24][25]. It is also expected to decrease the impact on the equipment's battery life and overall performance.

Further tests are scheduled to assess if the limitations posed by the hardware do require a partial media encryption, possibly compromising full stream confidentiality, in favour of a full encryption scheme.

The implementation of a custom protocol stack based on RTP+SRTP+ZRTP was chosen mostly due to the low latency provided by the RTP protocol, caused by the use of UDP and reduced size control messages; also, to the fact that it is an open standard. Having a profile that supports MPEG-4 Part 12 is a surplus[26] as it eases its integration with current state-of-the-art systems. A comprehensive, yet simple, table (table 1) was depicted showing the main differences between all the analysed protocols. As for encryption, although ElGamal based encryption is not based on deterministic methods to generate its keys, it was abandoned in favour of the SRTP+ZRTP combination mostly because the latest has lighter processing requirements (table 2).

By combining a set of new technologies, while associating the proposed implementation to the larger bandwidth provided by the 4G wireless network and the VPS providing a right of way without interference from unwanted traffic, it is expected to obtain the platform requirements for live video-to-video interaction between all parties involved through the creation of a custom protocol stack interoperable between most types of devices.

Table 2. Candidate security protocols

	Payload encryption	Video encryption	Deterministic encryption	Key exchange
ISMACryp	✓	X	✓	✓
SRTP+ZRTP	X	✓	✓	✓
SSL	✓	X	✓	X
ElGamal	X	✓	X	✓

5 Conclusion

Although mobile video-to-video applications are available at the present, their use for a live, secure and uninterrupted user interaction it is still unfeasible. The main difficulties behind this matter are related to low the bandwidth provided by most operators and signal instability on mobile scenarios, but also, hardware limitations that hamper the implementation of stronger security measures.

Circumventing these limitations is part of LiveCity objectives. These objectives will be achieved through a complete solution that comprehends security, uninterrupted user interaction and live encryption. Next steps will be made towards a complete, stable and interoperable system architecture, leading up to an application being developed to support all the proposed scenarios.

References

1. Wang, Y., Ostermann, J., Zhang, Y.Q.: Video Processing and Communications. Prentice Hall (2002)
2. Richardson, I.E.G.: H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia. John Wiley & Sons (2003)

3. Baugher, M., McGrew, D., Naslund, M., Carrara, E., Norrman, K.: The Secure Real-time Transport Protocol (SRTP). RFC 3711 (Proposed Standard) (March 2004); Updated by RFC 5506
4. Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V.: RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (Standard) (July 2003); Updated by RFCs 5506, 5761, 6051, 6222
5. Paar, C., Pelzl, J., Preneel, B.: The Advanced Encryption Standard. In: Understanding Cryptography: A Textbook for Students and Practitioners, ch. 4. Springer (2009)
6. Huitema, C.: Real Time Control Protocol (RTCP) attribute in Session Description Protocol (SDP). RFC 3605 (Proposed Standard) (October 2003)
7. Friedman, T., Caceres, R., Clark, A.: RTP Control Protocol Extended Reports (RTCP XR). RFC 3611 (Proposed Standard) (November 2003)
8. Zimmermann, P., Johnston, A., Callas, J.: ZRTP: Media Path Key Agreement for Unicast Secure RTP. RFC 6189 (Informational) (April 2011)
9. iCall, http://www.icall.com/features/security
10. PJSIP, http://www.pjsip.org/
11. Zfone, http://zfoneproject.com/
12. Pantos, R. (ed.), Apple Inc.: HTTP Live Streaming. draft-pantos-http-livestreaming-08 (Informational) (March 2012)
13. Apache HTTP Server, http://httpd.apache.org/docs/2.0/ssl/ssl_faq.html
14. ISMA: INTERNET STREAMING MEDIA ALLIANCE Encryption and Authentication v2.0. External Proposed Specification (November 2007)
15. ITSecurity: ISMA End-to-End Video Encryption Specification Released after Extensive Interoperability Tests, http://www.itsecurity.com/press-releases/press-release-mobile-tv-crypto-101806/
16. Sun, H.M., Leu, M.C.: A real-time selective video encryption using context-key control. In: Fifth International Conference on Information Assurance and Security, IAS 2009, vol. 2, pp. 114–117 (August 2009)
17. Elgamal, T.: A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory 31(4), 469–472 (1985)
18. Microsoft Media Server (MMS) Protocol Specification, http://msdn.microsoft.com/en-us/library/cc234711%28v=prot.10%29.aspx
19. Microsoft protocol rollover, http://msdn.microsoft.com/en-gb/library/dd757582.aspx
20. RedZinc, http://www.redzinc.net/index.php?option=com_content&view=article&id=6&Itemid=18
21. Sixth Framework Programme, http://cordis.europa.eu/fp6/whatisfp6.html
22. RedZinc, http://www.redzinc.net/
23. ooVoo, http://support.oovoo.com/ics/support/TSList.asp?folderID=25&next=Continue+to+Step+2+-%3E&task=knowledge
24. Tango, http://support.tango.me/entries/20399717-is-tango-secure
25. Skype, https://support.skype.com/en/faq/FA31/Does-Skype-use-encryption
26. Schmidt, M., de Bont, F., Doehla, S., Kim, J.: RTP Payload Format for MPEG-4 Audio/Visual Streams. RFC 6416 (Proposed Standard) (October 2011)