Network Working Group
Internet-Draft
Intended status: BCP
Expires: January 7, 2012

C. Perkins
University of Glasgow
JM. Valin
Octasic Inc.

July 6, 2011

Guidelines for the use of Variable Bit Rate Audio with Secure RTP
draft-ietf-avtcore-srtp-vbr-audio-03.txt

Abstract

This memo discusses potential security issues that arise when using variable bit rate audio with the secure RTP profile. Guidelines to mitigate these issues are suggested.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 7, 2012.

Copyright Notice

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 3
2. Scenario-Dependent Risk 3
3. Guidelines for use of VBR Audio with SRTP 4
4. Guidelines for use of Voice Activity Detection with SRTP . . . 4
5. Padding the output of VBR codecs 5
6. Security Considerations 6
7. IANA Considerations 6
8. Acknowledgements 6
9. References 6
9.1. Normative References 6
9.2. Informative References 6
Authors' Addresses 6

1. Introduction

The secure RTP framework (SRTP) [RFC3711] is a widely used framework for securing RTP sessions. SRTP provides the ability to encrypt the payload of an RTP packet, and optionally add an authentication tag, while leaving the RTP header and any header extension in the clear. A range of encryption transforms can be used with SRTP, but none of the pre-defined encryption transforms use any padding; the RTP and SRTP payload sizes match exactly.

When using SRTP with voice streams compressed using variable bit rate (VBR) codecs, the length of the compressed packets will therefore depend on the characteristics of the speech signal. This variation in packet size will leak a small amount of information about the contents of the speech signal. For example [spot-me] shows that known phrases in an encrypted call using the Speex codec in VBR mode can be recognised with high accuracy in certain circumstances, without breaking the encryption. Other work, referenced from [spot-me], has shown that the language spoken in encrypted conversations can also be recognised. This is potentially a security risk for some applications. How significant these results are and how they generalise to other codecs is still an open question. This memo discusses ways in which this traffic analysis risk may be mitigated.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Scenario-Dependent Risk

Whether the information leak analysed in [spot-me] is significant highly depends on the application. In the worst case, using the rate information to recognize a pre-recorded message knowing the set of all possible messages would lead to near-perfect accuracy. Even when the audio is not pre-recorded, there is a real possibility of being able to recognize contents from encypted audio when the dialog is highly structured (e.g. when the evesdropper knows that only a handful of possible sentences are possible) and thus contain only little information. On the other end, recognizing unconstrained conversational speech from the rate information alone appears to be highly unlikely at best. In fact, such a task is already considered a hard problem even when one has access to the unencrypted audio.

In practical SRTP scenarios, it must also be considered how significant the information leak is when compared to other SRTP-related information, such as the fact that the source and destination

IP addresses are available.

3. Guidelines for use of VBR Audio with SRTP

It is the responsibility of the application designer to determine the appropriate trade-off between security and bandwidth overhead. As a general rule, VBR codecs should be considered safe in the context of encrypted unstructured calls. However, applications that make use of pre-recorded messages where the contents of such pre-recorded messages may be of any value to an evesdropper (i.e., messages beyond standard greeting messages) SHOULD NOT use codecs in VBR mode. Interactive voice response (IVR) applications would be particularly vulnerable since an evesdropper could easily use the rate information to easily recognize the prompts being played out.

It is safe to use variable rate coding to adapt the output of a voice codec to match characteristics of a network channel, for example for congestion control purposes, provided this adaptation done in a way that does not expose any information on the speech signal. That is, if the variation is driven by the available network bandwidth, not by the input speech (i.e., if the packet sizes and spacing are constant unless the network conditions change). VBR speech codecs can safely be used in this fashion with SRTP while avoiding leaking information on the contents of the speech signal that might be useful for traffic analysis.

4. Guidelines for use of Voice Activity Detection with SRTP

Many speech codecs employ some form of voice activity detection (VAD) to either suppress output frames, or generate some form of lower-rate comfort noise frames, during periods when the speaker is not active. If VAD is used on an encrypted speech signal, then some information about the characteristics of that speech signal can be determined by watching the patterns of voice activity. This information leakage is less than with VBR coding since there are only two rates possible.

The information leakage due to VAD in SRTP audio sessions can be much reduced if the sender adds an unpredictable "overhang" period to the end of active speech intervals, so obscuring their actual length. An RTP sender using VAD with encrypted SRTP audio SHOULD insert such an overhang period at the end of each talkspurt, delaying the start of the silence/comfort noise by a random interval. The length of the overhang applied to each talkspurt must be randomly chosen in such a way that it is computationally infeasible for an attacker to reliably estimate the length of that talkspurt. This may be more important for short talk spurts, since is seems easier to distinguish between

different single word reponses based on the exact word length, than to glean meaning from the duration of a longer phrase. The audio data comprising the overhang period must be packetised and transmitted in RTP packets in a manner that is indistinguishable from the other data in the talkspurt.

The overhang period SHOULD have an exponentially-decreasing probability distribution function. This ensures a long tail, while being easy to compute. It is RECOMMENDED to use an overhang with a "half life" of a few hundred milliseconds (this should be sufficient to obscure the presence of inter-word pauses and the lengths of single words spoken in isolation, for example the digits of a credit card number clearly enunciated for an automated system, but not so long as to significantly reduce the effectiveness of VAD for detecting listening pauses). Despite the overhang (and no matter what the duration is), there is still a small amount of information leaked about the start time of the talkspurt due to the fact that we cannot apply an overhang to the start of a talkspurt without unacceptably affecting intelligibility. For that reason, VAD SHOULD NOT be used in encrypted IVR applications where the content of pre-recorded messages may be of any value to an eavesdropper.

The application of a random overhang period to each talkspurt will reduce the effectiveness of VAD in SRTP sessions when compared to non-SRTP sessions. It is, however, still expected that the use of VAD will provide a significant bandwidth saving for many encrypted sessions.

5. Padding the output of VBR codecs

For scenarios where VBR is considered unsafe, the codec SHOULD be operated in constant bit rate (CBR) mode. However, if the codec does not support CBR, RTP padding SHOULD be used to reduce the information leak to an insignificant level. Packets may be padded to a constant size ([spot-me] achieves good results by padding to the next multiple of 16 octets, but the amount of padding needed to hide the variation in packet size will depend on the codec), or may be padded to a size that varies with time. In the case where the size of the padded packets varies in time, the same concerns as for VAD apply. That is, the padding SHOULD NOT be reduced without waiting for a certain (random) time. The RECOMMENDED "hold time" is the same as the one for VAD.

Note that SRTP encrypts the count of the number of octets of padding added to a packet, but not the bit in the RTP header that indicates that the packet has been padded. For this reason, it is RECOMMENDED to add at least one octet of padding to all packets in a media

stream, so an attacker cannot tell which packets needed padding.

6. Security Considerations

The security considerations of [RFC3711] apply.

7. IANA Considerations

No IANA actions are required.

8. Acknowledgements

This memo is based on the discussion in [spot-me]. ZRTP [RFC6189] contain a similar recommendation; the purpose of this memo is to highlight these issues to a wider audience, since they are not specific to ZRTP. Thanks are due to Phil Zimmermann, Stefan Doehla, Mats Naslund, Gregory Maxwell, David McGrew, Mark Baugher, Koen Vos, and Ingemar Johansson for their comments and feedback on this memo.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

9.2. Informative References

[RFC6189] Zimmermann, P., Johnston, A., and J. Callas, "ZRTP: Media Path Key Agreement for Unicast Secure RTP", RFC 6189, April 2011.

[spot-me] Wright, C., Ballard, L., Coull, S., Monrose, F., and G. Masson, "Spot me if you can: Uncovering spoken phrases in encrypted VoIP conversation", Proceedings of the IEEE Symposium on Security and Privacy 2008, May 2008.

Authors' Addresses

Colin Perkins
University of Glasgow
School of Computing Science
Glasgow G12 8QQ
UK

Email: csp@csperkins.org

Jean-Marc Valin
Octasic Inc.
4101 Molson Street, Suite 300
Montreal, Quebec H1Y 3L1
Canada

Email: Jean-Marc.Valin@octasic.com