RMT
Internet Draft
<draft-ietf-rmt-sec-discussion-06>
Intended status: Informational
Expires: September 15, 2011

B. Adamson
Naval Research Laboratory
V. Roca
INRIA
H. Asaeda
Keio University
March 14, 2011

Security and Reliable Multicast Transport Protocols: Discussions and Guidelines

draft-ietf-rmt-sec-discussion-06

Abstract

This document describes general security considerations for the Reliable Multicast Transport (RMT) Working Group set of building blocks and protocols. An emphasis is placed on risks that might be resolved in the scope of transport protocol design. However, relevant security issues related to IP Multicast control-plane and other concerns not strictly within the scope of reliable transport protocol design are also discussed. The document also begins an exploration of approaches that could be embraced to mitigate these risks. The purpose of this document is to provide a consolidated security discussion and provide a basis for further discussions and potential resolution of any significant security issues that may exist in the current set of RMT standards.

Table of Contents

1. Introduction

The Reliable Multicast Transport (RMT) Working Group has produced a set of building block (BB) and protocol instantiation (PI) specifications for reliable multicast data transport. Some present PIs defined within the scope of RMT include Asynchronous Layered Coding (ALC) [RFC5775], NACK-Oriented Reliable Multicast (NORM) [RFC5740], and the File Delivery over Unidirectional Transport (FLUTE) [I-D.ietf-rmt-flute-revised] application that is built on top of ALC. These can be considered "Content Delivery Protocols" (CDP) as described in [Neumann05]. In this document, the term CDP will refer indifferently to either ALC or NORM, with their associated BBs.

The use of these BBs and PIs raises some new security risks. For instance, these protocols share a novel set of Forward Error Correction (FEC) and congestion control building blocks that present some new capabilities for Internet transport, but may also pose some new security risks. Yet some security risks are not related to the particular BBs used by the PIs, but are more general. Reliable multicast transport sessions are expected to involve at least one sender and multiple receivers. Thus, the risk of and avenues to attack are implicitly greater than that of point-to-point (unicast) transport sessions. Also the nature of IP multicast can expose other coexistent network flows and services to risk if malicious users exploit it. The classic Any-Source Multicast (ASM) [RFC1112] model of multicast routing allows any host to join an IP multicast group and send traffic to that group. This poses many potential security challenges. And, while the emerging Source-Specific Multicast (SSM) [RFC3569], [RFC4607] model that enables users to receive multicast data sent only from specified sender(s) simplifies some challenges, there are still specific issues. For instance, possible areas of attack include those against the control plane where malicious hosts join IP multicast groups to cause multicast traffic to be directed to parts of the network where it is not needed or desired. This may indirectly cause denial-of-service (DoS) to other network flows. Also, attackers may transmit erroneous or corrupt messages to the group or employ strategies such as replay attack within the "data plane" of protocol operation.

The goals of this document are therefore to:

1. Define the possible general security goals: protecting the network infrastructure, and/or the protocol, and/or the content, and/or the user (e.g., its privacy);
2. List the possible elementary security services that will make it possible to fulfill the general security goals. Some of these services are generic (e.g., object and/or packet integrity), while others are specific to RMT protocols (e.g., congestion control specific security schemes);
3. List some technological building blocks and solutions that can provide the desired security services;
4. Highlight the CDP and the use-case specificities that will impact security. Indeed, the set of solutions proposed to fulfill the security goals will greatly be impacted by these considerations;

In some cases, the existing RMT documents already discuss the risks and outline approaches to solve them, at least partially. The purpose of this document is to consolidate this content and provide a basis for further discussion and potential resolution of any significant security issues that may exist.

1.1 Conventions Used in this Document

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 [RFC2119].

2. Quick Introduction to RMT Protocols and their Use

2.1 The Two Families of CDP

The ALC and NORM classes of CDP are designed to reliably deliver content to a group of multicast receivers. However, ALC and NORM have a different set of features and limitations. ALC supports a unidirectional delivery model where there is no feedback from the receivers to senders. Reliability is achieved by the joint use of carousel-based transmission techniques associated to FEC encoding to recover from missing (erased) packets.

On the opposite, NORM achieves reliability by means of FEC encoding (as with ALC) and feedback from the receivers. More specifically, NORM leverages Negative Acknowledgement techniques to control the senders' transmission of content. The advantage is that the sender need not transmit any more information than necessary to satisfy the receivers' need for fully reliable transfers. However, while NORM specifies feedback control techniques to allow it to scale to large group sizes, it is not as massively scalable as ALC. Additionally, the NORM feedback control mechanisms add some header content and protocol implementation complexity.

The appropriate choice of a CDP depends upon application needs, deployment constraints, and network connectivity considerations. And while there are many common security considerations for these two classes of CDP, there are also some unique considerations for each.

2.2 RMT Protocol Characteristics

This section focuses on the RMT protocol characteristics that impact the choice of the technological building blocks, and the way they can be applied. Both ALC and NORM have been designed with receiver group size scalability. While ALC targets massively scalable sessions (e.g., millions of receivers), NORM is less ambitious, essentially because of the use of feedback messages.

The ALC and NORM protocols also differ in the communication paths:

- sender to receivers: ALC and NORM, for bulk data transfer and signaling messages;
- receivers to sender: NORM only, for feedback messages;
- receivers to receivers: NORM only for control messages;

Note that the fact ALC is capable of working on top of purely unidirectional networks does not mean that no back-channel is available (Section 2.3).

The NORM and ALC protocols support a variety of content delivery models where transport may be carefully coordinated among the sender and receivers or with looser coordination and interaction. This leads to a number of different use cases for these protocols.

2.3 Target Use Case Characteristics

This section focuses on the target use cases and their special characteristics. These details will impact both the choice of the technological building blocks and the way they can be applied. One can distinguish the following use case features:

- Purely unidirectional transport versus symmetric bidirectional transport versus asymmetric bidirectional transport. Most of the time, the amount of traffic flowing to the source is limited, and one can overlook whether the transport channel is symmetric or not. The nature of the underlying transport channel is of paramount importance, since many security building blocks will require a bidirectional communication;
- Massively scalable versus moderately scalable session. Here we do not define precisely what the terms "massively scalable" and "moderately scalable" mean.
- Known set of receivers versus unknown set of receivers: I.e., does the source know at any point of time the set of receivers or not? Of course, knowing the set of receivers is usually not compatible with massively scalable sessions;
- Dynamic set of receivers versus fixed set of receivers: I.e., does the source know at some point of time the maximum set of receivers or will it evolve dynamically?

- High rate data flow versus small rate data flow: Some security building blocks are CPU-intensive and are therefore incompatible with high data rate sessions (e.g., solutions that digitally sign all packets sent).
- Protocol stack available at both ends: A solution that requires some unusual features within the protocol stack will not always be usable. Some target environments (e.g., embedded systems) provide a minimum set of features and extending them (e.g., to add IPsec) is not necessarily realistic;
- Multicast routing and other layer-3 protocols in use: E.g., SSM routing is often seen as one of the key service to improve the security within multicast sessions, and some security building blocks require specialized versions of layer-3 protocols (e.g., IGMP/MLD with security extensions). In some cases these assumptions might not be realistic.

Depending on the target goal and the associated security building block used, other features might be of importance. For instance TESLA requires a loose time synchronization between the source and the receivers. Several possible techniques are available to provide this, but some of them may be feasible only if the target use case has the appropriate characteristics.

3. Some Security Threats

The IP architecture provides common access to notional control and data planes to both end and intermediate systems. For the purposes of discussion here, the "control plane" mechanisms are considered those with message exchanges between end systems (typically computers) and intermediate systems (typically routers) (or among intermediate systems) while the "data plane" encompasses messages exchanged among end systems, usually pertaining to the transfer of application data. The security threats described here are introduced within the taxonomy of control plane and data plane IP mechanisms.

3.1 Control-Plane Attacks

In this discussion, "control-plane" in the context of Internet Protocol systems refers to signaling among end systems and intermediate systems to facilitate routing and forwarding of packets. For IP multicast, this notably includes Internet Group Management Protocol (IGMP), Multicast Listener Discovery protocol (MLD), and multicast routing protocol messaging. While control-plane attacks may be considered outside of the scope of the transport protocol specifications discussed here, it is important to understand the potential impact of such attacks with respect to the deployment and operation of these protocols. For example, awareness of possible IP Multicast control-plane manipulation that can lead to unauthorized (or unexpected) monitoring of data plane traffic by malicious users may lead a transport application or protocol implementation to support encryption to ensure data confidentiality and/or privacy. Also, these types of attack also have bearing on assessing the real risks of potentially more complex attacks against the transport mechanisms themselves. In some cases, the solutions to these control-plane risk areas may reduce the impact or possibility of some data-plane attacks that are discussed in this document.

The presence of these types of attack may necessitate that policy-based controls be embedded in routers to limit the distribution (including transmission and reception) of multicast traffic (on a group-wise and/or traffic volume basis) to different parts of the network. Such policy-based controls are beyond the scope of the RMT protocol specifications. However, such network protection mechanisms may reduce the opportunities for or effectiveness of some of the data-plane attacks discussed later. For example, reverse-path checks can significantly limit opportunities for attackers to conduct replay attacks when hosts actually do use IPsec. Also, future IP Multicast control protocols may wish to consider providing security mechanism to prevent unauthorized monitoring or manipulation of messages related to group membership, routing, and activity. The sections below describe some variants of control-plane attacks.

3.1.1 Control Plane Monitoring

While this may not be a direct attack on the transport system, it may be possible for an attacker to gain useful information in advancing attack goals by monitoring IP Multicast control plane traffic including group membership and multicast routing information. Identification of hosts and/or routers participating in specific multicast groups may readily identify systems vulnerable to protocol-specific exploitation. And, with regards to user privacy concerns, such "side information" may be relevant to this emerging aspect of network security as described in Section 4.4.

3.1.2 Unauthorized (or Malicious) Group Membership

One of the simplest attacks is that where a malicious host joins an IP multicast group so that potentially unwanted traffic is routed to the host's network interface. This type of attack can turn a legitimate source of IP traffic into a "attacker" without requiring any access privileges to the source host or routers involved. This type of attack can be used for denial-of-service purposes or for the real attacker (the malicious joiner) to gain access to the information content being sent. Similarly, some routing protocols may permit any sender (whether joined to the specific group or not) to transmit messages to a multicast group.

It is possible that malicious hosts could also spoof IGMP/MLD messages, joining groups posing as legitimate hosts (or spoof source traffic from legitimate hosts). This may be done at intermediate locations in the network or by hosts co-resident with the authorized hosts on local area networks. Such spoofing could be done by raw packet generation or with replay of previously-recorded control messages.

For the sake of completeness, it should be noted that multicast routing protocol control messaging may be subject to similar threats if sufficient protocol security mechanisms are not enabled in the routing infrastructure. [RFC4609] describes security threats to the PIM-SM multicast routing infrastructures.

3.2 Data-Plane Attacks

This section discusses some types of active attacks that might be conducted "in-band" with respect to the reliable multicast transport protocol operating within the data plane of network data transfer. I.e., the "data-plane" here refers to IP packets containing end-to-end transport content to support the reliable multicast transfer. The passive attack of unauthorized data-plan monitoring is discussed above since such activity might be made possible by the vulnerabilities of the IP Multicast control plane. To cover the two classes of RMT protocols, the active data-plane attacks are categorized as 1) those where the attacker generates messages posing as a data sender, and 2) those where the attacker generates messages posing as a receiver providing feedback to the sender(s) or group. Additionally, a common threat to protocol operation is that of brute-force, rogue packet generation. This is discussed briefly below, but the more subtle attacks that might be conducted are given more attention as those fall within the scope of the RMT transport protocol design. Additionally, special consideration is given to that of the "replay attack" [see Section 3.2.4], as it can be applied across these different categories.

3.2.1 Rogue Traffic Generation

If an attacker is able to successfully inject packets into the multicast distribution tree, one obvious denial-of-service attack is for the attacker to generate a large volume of apparently authenticate traffic (and if authentication mechanisms are used, a "replay" attack strategy might be used). The impact of this type of attack can be significant since the potential for routers to relay the traffic to multiple portions of a networks (as compared to a single unicast routing path). However, other than the amplified negative impact to the network, this type of attack is no different than what is possible with rogue unicast packet generation and similar measures used to protect the network from such attacks could be used to contain this type of brute-force attack. Of course, the pragmatic question of whether current implementations of such protection mechanisms support IP Multicast SHOULD be considered.

3.2.2 Sender Message Spoofing

Sender message spoofing attacks are applicable to both CDP: ALC (sender-only transmission) and NORM (sender-receiver exchanges). Without an authentication mechanism, an attacker can easily generate sender messages that could disrupt a reliable multicast transfer session. And with FEC-based transport mechanisms, a single packet with an apparently-correct FEC payload identifier[RFC5052] but a corrupted FEC payload could potentially render an entire block of transported data invalid. Thus, a modest injection rate of corrupt traffic could cause severe impairment of data transport. Additionally, such invalid sender packets could convey out-of-bound indices (e.g., bad symbol or block identifiers) that can lead to buffer overflow exploits or similar issues in implementations that insufficiently check for invalid data.

An indirect use of sender message spoofing would be to generate messages that would cause receivers to take inappropriate congestion-control action. In the case of the layered congestion control mechanisms proposed for ALC use, this could lead to the receivers erroneously leaving groups associated with higher bandwidth transport layers and suffering unnecessarily low transport rates. Similarly, receivers may be misled to join inappropriate groups directing unwanted traffic to their part of the network. Attacks with similar effect could be conducted against the TCP-Friendly Multicast Congestion Control (TFMCC) [RFC4654] approach proposed for NORM operation with spoofing of sender messages carrying congestion control state to receivers.

3.2.3 Receiver Message Spoofing

These attacks are limited to CDP that use feedback from receivers in the group to influence sender and other receiver operation. In the NORM protocol, this includes negative-acknowledgement (NACK) messages fed back to the sender to achieve reliable transfer, congestion control feedback content, and the optional positive acknowledgement features of the specification. It is also important to note that for ASM operation, NORM receivers pay attention to the messages of other receivers for the purpose of suppression to avoid feedback implosion as group size grows large.

An attacker that can generate false feedback can manipulate the NORM sender to unnecessarily transmit repair information and reduce the goodput of the reliable transfer regardless of the sender's transmit rate. Contrived congestion control feedback could also cause the sender to transmit at an unfairly low rate.

As mentioned, spoofed receiver messaging may not be directed only at senders, but also at receivers participating in the session. For example, an attacker may direct phony receiver feedback messages to selected receivers in the group causing those receivers to suppress feedback that might have otherwise been transmitted. This attack could compromise the ability of those receivers to achieve reliable transfer. Also, suppressed congestion control feedback could cause the sender to transmit at a rate unfair to those attacked receivers if their fair congestion control rate were lower.

3.2.4 Replay Attacks

The infamous "replay attack" (injection of a previously transmitted packet to one or more participants) is given special attention here because of the special consequences it can have on RMT protocol operation. Without specific protection mechanisms against replay (e.g., duplicate message detection), it is possible for these attacks to be successful even when security mechanisms such as packet authentication and/or encryption are employed.

3.2.4.1 Replay of Sender Messages

Generally, replay of recent protocol messages from the sender will not harm transport, and could potentially assist it, unless it is of sufficient volume to result in the same type of impact as the "rogue traffic generation" described above. However, it is possible that replay of sufficiently old messages may cause receivers to think they are "out of sync" with the sender and reset state, compromising the transfer. Also, if sender transport data identifiers are reused (object identifiers, FEC payload identifiers, etc), it is possible that replay of old messages could corrupt data of a current transfer.

3.2.4.2 Replay of Receiver Messages

Replay of receiver messages are problematic for the NORM protocol, because replay of NACK messages could cause the sender to unnecessarily transmit repair information for an FEC coding block. Similarly, the sender transmission rate might be manipulated by replay of congestion control feedback messages from receivers in the group. And the way that NORM senders estimate group round-trip timing (GRTT) could allow a replay attack to manipulate the senders' GRTT estimate to an unnecessarily large value, adding latency to the reliable transport process.

4. General Security Goals

The term "security" is extremely vast and encompasses many different meanings. The goal of this section is to clarify what "security" means when considering the CDP defined in the IETF RMT working group. However, the scope can also encompass additional applications, like streaming applications. This section only focuses on the desired general goals. The following sections will then discuss the possible elementary services that will be required to fulfill these general goals, as well as the underlying technological building blocks.

The possible final goals include, in decreasing order of importance:

- network protection: the goal is to protect the network from attacks, no matter whether these attacks are voluntary (i.e., launched by one or several attackers) or non voluntary (i.e., caused by a misbehaving system, where "system" can designate a building block, a protocol, an application, or a user);
- protocol protection: the goal is to protect the RMT protocol itself, e.g., to avoid that a misbehaving receiver prevents other receivers to get the content, no matter whether this is done intentionally or not;
- content protection: to goal is to protect the content itself, for instance to guaranty the integrity of the content, or to make sure that only authorized clients can access the content;
- and user protection: the goal is often to protect the user privacy.

4.1 Network Protection

Protecting the network is of course of primary importance. An attacker should not be able to damage the whole infrastructure by exploiting some features of the RMT protocol. Unfortunately, recent past has shown that the multicast routing infrastructure is relatively fragile, as well as the applications built on top of it. Since the RMT protocols may use congestion control mechanisms to regulate sender transmission rate, the protocol security features should ensure that the sender may not be manipulated to transmit at incorrect rates (most importantly not at an excessive rate) to any parts of the receiver group. In the case of NORM, the security mechanisms should ensure that the feedback suppression mechanisms are protected to prevent badly-behaving network nodes from purposefully causing feedback implosion. In the case of ALC, where layered congestion control may be used via dynamic grou/layer membership, this extends to considerations of excessive manipulation of the multicast router control plane.

4.2 Protocol Protection

Protecting the protocols is also of importance, since the higher the number of clients, the more serious the consequences of an attack. This is all the more true as scalability is often one of the desired goals of CDP. Ideally, receivers should be sufficiently isolated from one another, so that a single misbehaving receiver does not affect others. Similarly, an external attacker should not be able to break the system, i.e., resulting in unreliable operation or delivery of incorrect content.

4.3 Content Protection

The content itself should be protected when meaningful. This level of security is often the concern of the content provider (and its responsibility). For instance, in case of confidential (or non-free) content, the typical solution consists in encrypting the content. It can be done within the upper application, i.e., above the RMT protocol, or within the transport system.

But other requirements may exist, like verifying the integrity of a received object, or authenticating the sender of the received packets. To that goal, one can rely on the use of building blocks integrated within, or above, or beneath the RMT protocol.

One may also consider that offering the packet sender authentication and content integrity services are basic requirements that should fulfill any RMT system that operates within an open network, where any attacker can easily inject spurious traffic in an ongoing NORM or ALC session. In that case this goal is not the responsibility of the content provider but the responsibility of the administrator who deploys the RMT system itself.

4.4 Privacy

Finally the user should be protected, and more specifically its privacy. In general, there is no privacy issue for data sender: the data sender's address is announced to all prospective receivers prior to their joins. Moreover receivers need to specify the source address(es) as well as the IP multicast address in SSM communication upon their subscription. The situation is different if we consider receivers since their address should not be disclosed publicly.

Data receivers use IGMP or MLD protocols to notify their upstream routers to join or leave IP multicast session. The recent IGMPv3 [RFC3376] and MLDv2 [RFC3810] do not adopt the "report suppression mechanism". Report suppression makes the receiver host withdraw its own report when the host hears a report scheduled to be sent from other host joining the same group. Eliminating the report suppression mechanism does not contribute to minimizing the number of responses, but enables the router to keep track of host membership status on a link. Due to this specification, operators who maintain upstream routers that attach multicast data receiver can recognize data receivers' addresses by tracing IGMP/MLD report messages. Although such traced data may be useful for capacity planning or accounting from operator's perspective, the detail information including receivers' IP addresses should be carefully treated.

As described in Section 3.1.2, unauthorized users may spoof IGMP/MLD query messages and trace receivers' addresses on the same LAN. Currently, IGMP/MLD protocols do not protect this attack. It is desired for these protocols to ignore invalid query messages and provide receiver's privacy by some means.

5. Elementary Security Techniques

The goals defined in Section 4 will be fulfilled by means of underlying security techniques, provided by one or several technological building blocks. This section only focuses on these elementary security techniques. Some general techniques traditionally available are:

Technique	Goal
packet integrity	Enable session participants to verify that a packet has not been inappropriately modified in transit.
packet source authentication	Enable a receiver to verify the source of a packet.
packet group authentication	Enable a receiver to verify that a packet originated or was modified only within the group and has not been modified by nonmembers in transit; Additionally, if attribution of any modifications by the group is required, certain group authentication mechanisms may provide this capability.
packet non-repudiation	Enable any third party to verify the source of a packet such that the source cannot repudiate having sent the packet.
packet anti-replay	Enable a receiver to detect that a packet is the same as a previously-received packet
object integrity	Enable a receiver to verify the integrity of a whole object. Such object integrity verification should be possible for any singular object or any composition of sub-objects which together constitute a larger object structure.
object source authentication	Enable a receiver to verify the source of an object.
object confidentiality	Enable a source to guarantee that only authorized receivers can access the object data.

Table 1: General Security Techniques

Some additional techniques are specific to the RMT protocols:

Technique	Goal
congestion control security	Prevent an attacker from modifying the congestion control protocol normal behavior (e.g., by reducing the transmission (NORM) or reception (ALC) rate, or on the opposite increasing this rate up to a point where congestion occurs)
group management	Ensure that only authorized receivers (as defined by a certain group management policy) join the RMT session and possibly inform the source
backward group secrecy	Prevent a new group member to access the information in clear sent to the group before he joined the group
forward group secrecy	Prevent a former group member to access the information in clear sent to the group after he left the group

Table 2: RMT-Specific Security Techniques

These techniques are usually achieved by means of one or several technological building blocks. The target use case where the RMT system will be deployed will greatly impact the choice of the technological building block(s) used to provide these services, as explained in Section 2.3.

6. Technological Building Blocks

Here is a list of techniques and building blocks that are likely to fulfill one or several of the goals listed above:

- IPsec;
- Group MAC;
- Digital signatures;
- TESLA;
- SSM communication model;

Each of them is now quickly discussed. In particular we identify what service it can offer, its limitations, and its field of application (adequacy with respect to the CDP and the target use case).

6.1 IPsec

6.1.1 Benefits

One direct approach using existing standards is to apply IPsec [RFC4301] to achieve the following properties:

- source authentication and packet integrity (IPsec AH or ESP)
- confidentiality by means of encryption (IPsec ESP)

6.1.2 Requirements

It is expected that the approach to apply IPsec for reliable multicast transport sessions is similar to that described for OSPFv3 security[RFC4552]. The following list proposes the IPsec capabilities needed to support a similar approach to RMT protocol security:

- Mode - Transport mode IPsec security is required;
- Selectors - source and destination addresses and ports, protocol.
- For some uses, preplaced, manual key support may be required to support application deployment and operation. For automated key management for group communication the Group Secure Association Key Management Protocol (GSAKMP) described in [RFC4535] may be used to emplace the keys for IPsec operation.

Note that a periodic rekeying procedure similar to that described in RFC 4552 can also be applied with the additional benefit that the reliable transport aspects of the CDP provide robustness to any message loss that might occur due to ANY timing discrepancies among the participants in the reliable multicast session.

6.1.3 Limitations

It should be noted that current IPsec implementations may not provide the capability for anti-replay protection for multicast operation. In the case of the NORM protocol, a sequence number is provided for packet loss measurement to support congestion control operation. This sequence number can also be used within a NORM implementation for detecting duplicate (replayed) messages from sources (senders or receivers) within the transport session group. In this way, protection against replay attack can be achieved in conjunction with the authentication and possibly confidentiality properties provided by an IPsec encapsulation of NORM messages. NORM receivers generate a very low volume of feedback traffic and it is expected that the 16-bit sequence space provided by NORM will be sufficient for replay attack protection. When a NORM session is long-lived, the limits of the sender repair window are expected to provide protection from replayed NACKs as they would typically be outside of the sender's current repair window. It is suggested that IPsec implementations that can provide anti-replay protection for IP Multicast traffic, even when there are multiple senders within a group, be adopted. The GSAKMP document has some discussion in this area.

6.2 Group MAC

6.2.1 Benefits

The use of Group MAC (Message Authentication Codes) within the CDP Simple Authentication Schemes for the ALC and NORM Protocols [I-D.ietf-rmt-simple-auth-for-alc-norm] is a simple solution to provide a loss tolerant group authentication/integrity service for all the packets exchanged within a session (i.e., the packets generated by the session's sender and the session's receivers). This scheme is easy to deploy since it only requires that all the group members share a common secret key. Because Group MAC heavily relies on fast symmetric cryptographic building blocks, CPU processing remains limited both at the sender and receiver sides, which makes it suitable for high data rate transmissions, and/or lightweight terminals. Finally, the transmission overhead remains limited.

6.2.2 Requirements

This scheme only requires that all the group members share a common secret key, possibly associated to a re-keying mechanism (e.g., each time the group membership changes, or on a periodic basis).

6.2.3 Limitations

This scheme cannot protect against attacks coming from inside the group, where a group member impersonates the sender and sends forged messages to other receivers. It only provides a group-level authentication/integrity service, unlike the TESLA and Digital Signature schemes. Note that the Group MAC and Digital Signature schemes can be advantageously used together, as explained in Simple Authentication Schemes for the ALC and NORM Protocols [I-D.ietf-rmt-simple-auth-for-alc-norm].

6.3 Digital Signatures

6.3.1 Benefits

The use of Digital Signatures within the CDP Simple Authentication Schemes for the ALC and NORM Protocols [I-D.ietf-rmt-simple-auth-for-alc-norm] is a simple solution to provide a loss-tolerant authentication/integrity service for all the packets exchanged within a session (i.e., the packets generated by the session's sender and the session's receivers). This scheme is easy to deploy since it only requires that the participants know the packet sender's public key, which can be achieved with either Public Key Infrastructure (PKI) or by preplacement of these keys.

6.3.2 Requirements

This scheme is easy to deploy since it requires only that the participants know the packet sender's public key, which can be achieved with either PKI or by preplacement of these keys.

6.3.3 Limitations

When RSA [RsaPaper] asymmetric cryptography is used, the digital signatures approach has two major shortcomings:

- it is limited by high computational costs, especially at the sender, and
- it is limited by high transmission overheads.

This scheme is well suited to low data rate flows, when transmission overheads are not a major issue. For instance it can be used as a complement to TESLA for the feedback traffic coming from the session's receivers. The use of ECC ("Elliptic Curve Cryptography") significantly relaxes these constraints, especially when seeking for higher security levels. For instance, the following key size provide equivalent security:

Symmetric Key Size	RSA Key Size	ECC Key Size
80 bits	1024 bits	160 bits
112 bits	2048 bits	224 bits

However in some cases, the Intellectual Property Rights (IPR) considerations for ECC may limit its use, so the other techniques are presented here as well. Note that the Group MAC and Digital Signature schemes can be advantageously used together, as explained in Simple Authentication Schemes for the ALC and NORM Protocols [I-D.ietf-rmt-simple-auth-for-alc-norm].

6.4 TESLA

6.4.1 Benefits

The use of TESLA [RFC5776] within the CDP offers a loss tolerant, lightweight, authentication/integrity service for the packets generated by the session's sender. Depending on the time synchronization and bootstrap methods used, TESLA can be compatible with massively scalable sessions. Because TESLA heavily relies on fast symmetric cryptographic building blocks, CPU processing remains limited both at the sender and receiver sides, which makes it suitable for high data rate transmissions, and/or lightweight terminals. Finally, the transmission overhead remains limited.

6.4.2 Requirements

The security offered by TESLA relies heavily on time. Therefore the session's sender and each receiver need to be loosely synchronized in a secure way. To that purpose, several methods exist, depending on the use case: direct time synchronization (which requires a bidirectional transport channel), using a secure Network Time Protocol (NTP) [RFC5905] infrastructure (which also requires a bidirectional transport channel), or a Global Positioning System (GPS) device, or a clock with a time-drift that is negligible in front of the TESLA time accuracy requirements.

The various bootstrap parameters must also be communicated to the receivers, using either an in-band or out-of-band mechanism, sometimes requiring bidirectional communications. So, depending on the time synchronization scheme and the bootstrap mechanism method, TESLA can be used with either bidirectional or unidirectional transport channels.

6.4.3 Limitations

One limitation is that TESLA does not protect the packets that are generated by receivers, for instance the feedback packets of NORM. These packets must be protected by other means.

Another limitation is that TESLA requires some buffering capabilities at the receivers in order to enable the delayed authentication feature. This is not considered though as a major issue in the general case (e.g., FEC decoding of objects within an ALC session already requires some buffering capabilities, that often exceed that of TESLA), but it might be one in case of embedded environments.

6.5 Source-Specific Multicast

Source-Specific Multicast (SSM) [RFC3569], [RFC4607] amends the classical Any-Source Multicast (ASM) model by creating logical IP multicast "channels" that are defined by the multicast destination address *and* the specific source address(es). Thus for a given "channel", only the specific source(s) can inject packets that are distributed to the receivers. This form of multicast has group management benefits since a source can independently control the "channels" it creates.

6.5.1 Requirements

Use of SSM requires that the network intermediate systems explicitly support it. Additionally, hosts operating systems are required to support the IGMPv3/MLDv2 extensions for SSM, and the CDP implementations need to support the IGMPv3/MLDv2 API, including management of the <srcAddr; dstMcastAddr> "channel" identifiers.

6.5.2 Limitations

CDP such as NORM that use signaling from receivers to multicast senders will need to use unicast addressing for feedback messages. In the case of NORM, its timer-based feedback suppression requires support of the sender

NORM_CMD(REPAIR_ADV) message to control receiver feedback. In some topologies, use of unicast feedback may require some additional latency (increased backoff factor) for safe operation. The security of the unicast feedback from the receivers to sender will need to be addressed separately since the IP multicast model, including SSM, does not provide the sender knowledge of authorized group members.

6.5.3 Source-Based and Receiver-Based Attacks

The security threats are categorized into "source-based" and "receiver-based" attacks [RFC4609]. In short, the former is a DoS attack against the multicast networks, in which data is sent to numerous and random group addresses, and the latter is a DoS attack against multicast routers, in which innumerable IGMP/MLD joins are sent from a client.

Regarding source-based attack, there are some security benefits in SSM. Since data-plane traffic for an SSM "channel" is limited to that of a single, specific source address, it is possible that network intermediate systems may impose mechanism that prevent injection of traffic to the group from inappropriate (perhaps malicious) nodes. This can reduce the risk for denial-of-service and some of the other attacks described in this document. While SSM alone is not a complete security solution, it can simplify secure RMT operation.

On the contrary, SSM is not robust against receiver-based attack. An SSM capable router constructs a Shortest-Path Tree (SPT) with no shared tree coordination. Therefore, even if a host triggers invalid or unavailable channel subscriptions, the upstream router starts establishing all SPTs with no intellectual decision. What is worse is that these multicast routers cannot recognize the original router that is attacked and cannot stop the attack itself.

6.6 Summary

The following table summarizes the pros/cons of each authentication/integrity scheme used at application/transport level (where "-" means bad, "0" means neutral, and "+" means good):

	RSA Digital Signature	**ECC Digital Signature**	**Group MAC**	**TESLA**
True authentication and integrity	Yes	Yes	No (group security)	Yes
Immediate authentication	Yes	Yes	Yes	No
Processing load	-	0	+	+
Transmission overhead	-	0	+	+
Complexity	+	+	+	-

7. Security Infrastructure

Deploying the elementary technological building blocks often requires that a security infrastructure exists. Such security infrastructure can provide:

- Public Key Infrastructure (PKI) for trusted third party vetting of, and vouching for, user identities. PKI also allows the binding of public keys to users, usually by means of certificates.
- Group Key Management with rekeying schemes that are either periodic or triggered by some higher level event. It is required in particular when the group is dynamic and forward/backward secrecy are important. This is also required to improve the scalability of the CDP (since key management is done automatically, using a key server topology), or the security provided by the CDP (since the underlying cryptographic keys will be changed frequently)

It is expected that some CDP deployments may use existing client-server security infrastructure models so that receivers may acquire any necessary security material and be authenticated or validated as needed for group participation. Then, the reliable delivery of session data content will be provided via the applicable RMT protocols. Note that in this case the security infrastructure itself may limit the scalability of the group size or other aspects of reliable multicast transfer. The IETF Multicast Security (MSEC) Working Group has developed some protocols that can be applied to achieve more scalable and effective group communication security infrastructure[RFC4046]. It is encouraged that these mechanisms be considered in the development of security for CDP.

8. New Threats Introduced by the Security Scheme Itself

Introducing a security scheme, as a side effect, can sometimes introduce new security threats. For instance, signing all packets with asymmetric cryptographic schemes (to provide a source authentication/content integrity/anti-replay service) opens the door to DoS attacks. Indeed, verifying asymmetric-based cryptographic signatures is a CPU intensive task. Therefore an attacker can easily overload a receiver (or a sender in case of NORM) by injecting a significant number of faked packets.

9. Consequences for the RMT and MSEC Working Group

To meet the goals outlined in this document, it is expected that the RMT and MSEC Working Groups may need to develop some supporting protocol security mechanisms. It is also possible to cooperate with the Multicast Backbone (MBONE) Deployment (MBONED) Working Group for defining operational considerations.

9.1 RMT Transport Message Security Encapsulation Header

An alternative approach to using IPsec to provide the necessary properties to protect RMT protocol operation from the application attacks described earlier, is to extend the RMT protocol message set to include a message encapsulation option. This encapsulation header could be used to provide authentication, confidentiality, and anti-replay protection as needed. Since this would be independent of the IP layer, the header might need to provide a source identifier to be used as a "selector" for recalling security state (including authentication certificate(s), sequence state, etc) for a given message. In the case of the NORM protocol, a `NormNodeId` field exists that could be used for this purpose. In the case of ALC, the security encapsulation mechanism would need to add this function. The security encapsulation mechanism, although resident "above" the IP layer, could use GSAKMP [RFC4535] or a similar approach for automated key management.

10. IANA Considerations

This document has no actions for IANA.

11. Security Considerations

This document is a general discussion of security for the RMT protocol family. But specific security considerations are not applicable as this document does not introduce any new techniques.

12. Acknowledgments

The authors would like acknowledge Magnus Westerlund for stimulating the working group activity in this area. Additionally George Gross and Ran Atkinson contributed many ideas to the discussion here.

13. References

13.1 Normative References

[I-D.ietf-rmt-flute-revised] Paila, T., Walsh, R., Luby, M., Roca, V., and R. Lehtonen, "FLUTE - File Delivery over Unidirectional Transport", Internet-Draft draft-ietf-rmt-flute-revised-12 (work in progress), February 2011.

[I-D.ietf-rmt-simple-auth-for-alc-norm] Roca, V., "Simple Authentication Schemes for the ALC and NORM Protocols", Internet-Draft draft-ietf-rmt-simple-auth-for-alc-norm-03 (work in progress), July 2010.

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, August 1989.

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

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[RFC3569] Bhattacharyya, S., "An Overview of Source-Specific Multicast (SSM)", RFC 3569, July 2003.

[RFC3810] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.

[RFC4046] Baugher, M., Canetti, R., Dondeti, L., and F. Lindholm, "Multicast Security (MSEC) Group Key Management Architecture", RFC 4046, April 2005.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

[RFC4535] Harney, H., Meth, U., Colegrove, A., and G. Gross, "GSAKMP: Group Secure Association Key Management Protocol", RFC 4535, June 2006.

[RFC4552] Gupta, M. and N. Melam, "Authentication/Confidentiality for OSPFv3", RFC 4552, June 2006.

[RFC4607] Holbrook, H. and B. Cain, "Source-Specific Multicast for IP", RFC 4607, August 2006.

[RFC4609] Savola, P., Lehtonen, R., and D. Meyer, "Protocol Independent Multicast - Sparse Mode (PIM-SM) Multicast Routing Security

Issues and Enhancements", RFC 4609, October 2006.

[RFC4654] Widmer, J. and M. Handley, "TCP-Friendly Multicast Congestion Control (TFMCC): Protocol Specification", RFC 4654, August 2006.

[RFC5052] Watson, M., Luby, M., and L. Vicisano, "Forward Error Correction (FEC) Building Block", RFC 5052, August 2007.

[RFC5740] Adamson, B., Bormann, C., Handley, M., and J. Macker, "NACK-Oriented Reliable Multicast (NORM) Transport Protocol", RFC 5740, November 2009.

[RFC5775] Luby, M., Watson, M., and L. Vicisano, "Asynchronous Layered Coding (ALC) Protocol Instantiation", RFC 5775, April 2010.

[RFC5776] Roca, V., Francillon, A., and S. Faurite, "Use of Timed Efficient Stream Loss-Tolerant Authentication (TESLA) in the Asynchronous Layered Coding (ALC) and NACK-Oriented Reliable Multicast (NORM) Protocols", RFC 5776, April 2010.

13.2 Informative References

[Neumann05] Neumann, C., Roca, V., and R. Walsh, "Large Scale Content Distribution Protocols", ACM Computer Communications Review (CCR) Vol. 35 No. 5, October 2005.

[RFC5905] Mills, D., Martin, J., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

[RsaPaper] Rivest, R.L., Shamir, A., and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", Communications of the ACM 21, pp. 120-126, 1978.

Authors' Addresses

Brian Adamson
Naval Research Laboratory
Washington, DC, 20375
USA
EMail: adamson@itd.nrl.navy.mil
URI: http://cs.itd.nrl.navy.mil

Vincent Roca
INRIA
Montbonnot, 38334
France
EMail: vincent.roca@inria.fr
URI: http://planete.inrialpes.fr/~roca/

Hitoshi Asaeda
Keio University
5322 Endo
Fujisawa, Kanagawa 252-8520
Japan
EMail: asaeda@wide.ad.jp
URI: http://www.sfc.wide.ad.jp/~asaeda/