ipsecme
Internet-Draft
Intended status: Standards Track
Expires: December 13, 2012

R. Zhang, Ed.
S. Zhou
ZTE Corporation

June 11, 2012

Updates to the IKEv2 Extension for IKEv2/IPsec High Availablity draft-zhang-ipsecme-ipsecha-00

Abstract

This document updates RFC 6311, Protocol Support for High Availability of IKEv2/IPsec. This document analyzes RFC 6311, and proposes some updates.

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 December 13, 2012.

Table of Contents

1. Introduction

RFC 6311 [RFC6311] defines an extension to the IKEv2 protocol to solve the main issues of "IPsec Cluster Problem Statement" in the commonly deployed hot standby cluster, and provides implementation advice for other issues. The main issues solved are the synchronization of IKEv2 Message ID counters, and of IPsec replay counters. This document analyzes RFC 6311, and proposes some updates.

1.1. Terminology

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. Security Analysis of RFC 6311

Note: this section is just for informative purpose, and should be removed or moved to Appendix in the subsequent versions.

2.1. Observation 1: State Data of the New Active Member Being Stale and Unreliable

Since normally synchronization is carried out periodically between the active cluster member and standby members, the state data of the new active member is stale. The state data of the new active member just reflects the state data of the old active member at the last state synchronization time point T1 before the failover events occurs, rather than at the time point T2 when the failover event happens. So the state data of the new active member can NOT reflect the actual state data of the old active member at T2. This implies that the state data of the new active member is NOT reliable. Specifically, the new active member does NOT know what messages the old active member has sent and received from T1 and T2, and does NOT know the next sender's Message ID value of the old active member, and does NOT know the last Message ID value recevied from the peer of the old active member, and does NOT know whether requested messages from T1 to T2 by the old active member have been acknowledged, and does NOT know whether the old active member has acknowledged requested messages from T1 to T2 from the peer. Further, this implies that the new active member does NOT know whether some new IPsec SAs have been created, and whether some old IPsec SAs have been deleted and some old IPsec SAs have been updated, and whether the old IKE SA has been deleted and a new IKE SA has been created from T1 to T2.

2.2. Observation 2: A Parameter Flaw in Section 5.1: Processing Rules for IKE Message ID Synchronization

In RFC 6311 [RFC6311], M1, P1, M2 and P2 are set as follows.

1. M1 is the next sender's Message ID to be used by the member. M1 MUST be chosen so that it is larger than any value known to have been used. It is RECOMMENDED to increment the known value at least by the size of the IKE sender window.

2. P1 SHOULD be 1 more than the last Message ID value received from the peer, but may be any higher value.

3. M2 MUST be at least the higher of the received M1, and one more than the highest sender value received from the cluster. This includes any previous received synchronization messages.

4. P2 MUST be the higher of the received P1 value, and one more than the highest sender value used by the peer.

The setting of M2 is wrong, the reason is as follows. In the following disccussion, suppose the peer is a IPsec client or VPN gateway, not a IPsec cluster. First, based on observation 1, M1 chosen by the new active member is NOT accurate or reliable. Under some circumstances, M1 may be much lower than the highest sender value received from the cluster, if a lot of IKEv2 messages are exchanged from T1 to T2. Second, one more than the highest sender value received from the cluster just covers the received messages, does NOT cover the sent but unreached messages from the old active member. Suppose the highest sender value received from the cluster is RN0, and the sender window of the cluster is SW. Then the old active member may have sent SW messages with message ID RNO+1, ... RNO+SW messages, and these SW messages may have NOT reached the peer. If M2 is set as RN0+1, then the Message ID of the new active number will start from RNO+1. Since RNO+1 is lower than RNO+SW, this will cause some problems.

1. Case A: since RNO+1 is lower than RNO+SW, the Message ID of the first sent message from the new active member is lower than the Messsage ID of the last message sent by the old active member. This means that the message ID are NOT monotonically incremental.

2. Case B: the message IDs RNO+1,RNO+2,..., RNO+SW will be used twice in the two groups of messages sent by the old active member and new active member. This means that the peer may receive two different request messages with the same Message ID.

3. Case C: the new active member may receive an acceptable but mismatched response message for a request message with message ID within RNO+1,RNO+2,..., RNO+SW. For example, the peer may generate a response message to a request from the old active member with a Message ID RNO+2, and the new active member may receive this message as the response for its request message with the same Message ID. Under some circumstances, this may cause some security risks.

When the peer can act as a reliable archor point, the flaw can be fixed.

2.3. Observation 3: Mishandling of Simultaneous Failover

When simultaneous failover happens, the state data of new active members of the two clusters are NOT reliable, and can NOT act as reliable archor points to perform the IKEv2 message ID synchronization. Suppose there exists two clusters: X and Y. And the old and new active member of X is x1 and x2, respectively. The old and new active member of Y is y1 and y2, respectively. And suppose x2 initiates the synchronization process. First, based on observation 1, M1 chosen by x2 is NOT accurate or reliable. Additionally, because y2's state data is stale, y2's highest sender value received from the cluster X is not exactly identical to that of y1's. Under some circumstances, y2's highest sender value received from the cluster X may be much lower than y1's, if a lot of IKEv2 messages are exchanged between X1 and Y1 from T1 to T2. As a consequence, M2 chosen by y2 may be NOT reliable. Second, based on the same reason, P1 chosen by x2 is NOT accurate or reliable. The reason is that x2's last Message ID value received from cluster Y may be much lower than x1's. Similarily, the highest sender value used by y2 may be much lower than the highest sender value used by y1. As a consequence, P2 chosen by y2 may be NOT reliable, and may be even lower than some message IDs used by y1. Overall, the mishanling of simultaneous failover may cause some problems.

1. Case A: the Message ID of the first sent message from x2 are lower than the Messsage ID of the last message sent by x1. This means that the message ID are NOT monotonically incremental.

2. Case B: Some message IDs will be used twice in the two groups of messages sent by x1 and x2. This means that Y2 may receive two different request messages with the same Message ID.

3. Case C: x2 may receive an acceptable but mismatched response message for a request message.

4. Case D: x2 may accept a request message sent by y1. When y2's P2 is lower than the message IDs used by y1, the request message of y1 with a Message ID greater than P2 will be accepted by x2. This indicates that a request message from y1 may be accepted twice, one by x1, and one by x2.

5. Case E: y2 may accept a response message sent by x1. When y2's P2 is lower than the message IDs used by y1, the response message of x1 with a Message ID greater than P2 may be accepted by y2. Under some circumstances, this may cause some security risks.

6. Case F: from T1 to T2, x1 and y1 may have deleted the old IKE SA, and created a new IKE SA. However, since x2 and y2 have NOT updated their state data, after the simulateous failover event occurs, x2 and y2 may still use the old IKE SA. This may cause some security risks

Unlike observation 2, in the case of simulatenous failover, the flaw can NOT be easily fixed.

3. Update to RFC 6311, Avoiding Simultaneous Failover

When a failover event occurs, the synchronization state of the new active member should be set to UNSYNED. Meanwhile, the new active member should generate a synchronization request message and send it to the peer. When a cluster member in UNSYNED state receives a synchronization request, it should reply a Simultaneous Failover Failure response message to the peer to avoid the simultaneous failover failure, and terminates the synchronization process. Correspondingly, when the new active member of the peer cluster in UNSYNED state receives the Simultaneous Failover Failure message, it will terminate the synchronization process as well.

Note: the details of message type and protocol will be done in the subsequent versions. A new parameter called synchronization state is integrated into the IKE SA state data. After the failover event occurs, the synchronization state is set to UNSYNED. And after the synchronization process is finished, the synchronization state is set to SYNED. A simple state machine is introduced.The synchronization state of a IPsec VPN and IPsec gateway is always SYNED. The initial state of a new active member is UNSYNED. After the success of synchronization, the synchronization of the new active member will be changed to SYNED. An IPsec party in SYNED state is capable of performing synchronization for the opposite party in UNSYNED state.

4. Update to RFC 6311, Section 5.1: Processing Rules for IKE Message ID Synchronization

1. Parameter modification: M2 MUST be at least the higher of the received M1 and M1', where M1' is the peer cluster's sender window size plus the highest sender value received from the peer cluster. This includes any previous received synchronization messages. This modification ensures that M2 is larger than the largest Message ID sent by the old active member.

5. Security Considerations

Security implications will be discussed in the subsequent versions.

6. IANA Considerations

This document introduces a new IKEv2 Notification Message type: Simultaneous Failover Failure. This new type needs to be assigned a value.

Name	Value
Simultaneous Failover Failure	TBA

7. References

7.1. Normative References

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

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

[RFC4302] Kent, S., "IP Authentication Header", RFC 4302, December 2005.

[RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, December 2005.

[RFC5996] Kaufman, C., Hoffman, P., Nir, Y., and P. Eronen, "Internet Key Exchange Protocol Version 2 (IKEv2)",

RFC 5996, September 2010.

[RFC6311] Singh, R., Kalyani, G., Nir, Y., Sheffer, Y., and D. Zhang, "Protocol Support for High Availability of IKEv2/ IPsec", RFC 6311, July 2011.

7.2. Informative References

[RFC6027] Nir, Y., "IPsec Cluster Problem Statement", RFC 6027, October 2010.

Authors' Addresses

Ruishan Zhang (editor)
ZTE Corporation
889 Bibo Rd, Zhangjiang Hi-Tech Park
Shanghai 201203
R.R.China

Email: zhang.ruishan@zte.com.cn

Sujing Zhou
ZTE Corporation
No.68 Zijinghua Rd. Yuhuatai District
Nanjing, Jiang Su 210012
R.R.China

Email: zhou.sujing@zte.com.cn