

DIME
Internet-Draft
Intended status: Standards Track
Expires: January 17, 2014

H. Tschofenig
Nokia Siemens Networks
July 16, 2013

Diameter Overload Piggybacking
draft-tschofenig-dime-overload-piggybacking-00.txt

Abstract

This document describes how to piggyback Diameter overload information between Diameter servers and Diameter clients.

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 17, 2014.

Copyright Notice

Copyright (c) 2013 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	2
2.	Terminology	2
3.	Information Exchanges	3
3.1.	Capability Indication	3
3.2.	Overload Information	3
4.	Security Considerations	4
5.	IANA Considerations	4
6.	Acknowledgments	4
7.	References	4
7.1.	Normative References	4
7.2.	Informative References	4
	Author's Address	4

[1.](#) Introduction

The problem statement for Diameter overload control can be found in [\[3\]](#). It describes the lack of support of conveying load information to enable load balancing of Diameter requests in case Diameter servers become overload and the inability of Diameter servers to communicate with Diameter clients to reject requests when they become severely overloaded. [\[5\]](#) goes a step further in providing an outline of architectural principles and an information model.

This document is an extension to [\[5\]](#) and defines how Diameter servers interact with Diameter clients to report about overload situations. This is accomplished by piggybacking overload information from the Diameter server to the Diameter client within existing Diameter applications, as long as extension points allow adding new AVPs.

Communication overload information to Diameter clients is the last resource when load balancing is either not available, when all available servers are already overloaded, or when a critical failure occurred since this will lead to the Diameter client rejecting requests and returning appropriate error messages to end devices via the front-end protocols (such as SIP).

[2.](#) Terminology

The key words 'MUST', 'MUST NOT', 'REQUIRED', 'SHALL', 'SHALL NOT', 'SHOULD', 'SHOULD NOT', 'RECOMMENDED', 'MAY', and 'OPTIONAL' in this specification are to be interpreted as described in [\[1\]](#).

This document re-uses terminology from the Diameter base specification [2].

[3.](#) Information Exchanges

[3.1.](#) Capability Indication

The Diameter protocol interaction starts with a Diameter client using some Diameter application. The Diameter client MUST add the Supported-Features AVP to indicate support for the functionality supported for overload control. The Supported-Features AVP MUST NOT have the M-bit set since this would require a Diameter application to be defined.

A Diameter server receiving the Supported-Features AVP from a client is therefore able to know that this client supports the Diameter overload information exchange capability. Otherwise only the Diameter Base protocol functionality [2] with DIAMETER_TOO_BUSY error message is available. This enables feature discovery and a graceful fall-back to the functionality available with the Diameter Base protocol

[3.2.](#) Overload Information

In the rare and unlikely event of an overload situation the Diameter server (or a proxy, such as a load balancer, acting on behalf of several Diameter servers) may decide to communicate to the Diameter client to reject some or even all Diameter requests. The Diameter server does so by adding the Overload-Info AVP, which contains the Overload and the Period-Of-Validity AVP. The semantic of the Overload and the Period-Of-Validity AVP is described in [5]. To inform the Diameter client to reject requests the value of the Overload AVP is set to 'INCREASING_OVERLOAD' or to 'OVERLOADED'. The Diameter server may instruct the client to gradually reduce the number of requests by using the Overload='INCREASING_OVERLOAD' marking on several subsequent messages. The Period-Of-Validity AVP allows the Diameter server to give the "rejection policy" a soft-state nature, i.e., it will automatically expire without leaving orphan state at the Diameter client in case of a Diameter server

failure or other error situations.

A Diameter client that has received information to reduce the number of Diameter requests has to evaluate the requests based on their destination and their applications.

In case the Diameter server recovers and is able to accept more Diameter requests it can signal this changed state to the client with the 'DECREASING_OVERLOAD' or the 'NO_OVERLOAD' directive. Waiting for the expiry of the state is another option at the disposal of the Diameter server.

Tschofenig

Expires January 17, 2014

[Page 3]

Internet-Draft

Diameter Overload Piggybacking

July 2013

[4.](#) Security Considerations

This document utilizes the AVP defined in [\[5\]](#). The ability to use end-to-end signaling allows the Diameter AVP level security mechanisms, described in [\[4\]](#), to be re-used. Since the communicated rejection policies are bound to the application and the realm from an earlier request the ability to inject fake message is less likely.

[5.](#) IANA Considerations

This document does not require actions by IANA.

[6.](#) Acknowledgments

Add your name here.

[7.](#) References

[7.1.](#) Normative References

- [1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [2] Fajardo, V., Arkko, J., Loughney, J., and G. Zorn, "Diameter Base Protocol", [RFC 6733](#), October 2012.

[7.2.](#) Informative References

- [3] McMurry, E. and B. Campbell, "Diameter Overload Control

Requirements", [draft-ietf-dime-overload-reqs-07](#) (work in progress), June 2013.

- [4] Korhonen, J. and H. Tschofenig, "Diameter AVP Level Security: Keyed Message Digests, Digital Signatures, and Encryption", [draft-korhonen-dime-e2e-security-02](#) (work in progress), July 2013.
- [5] Tschofenig, Hannes., "Diameter Overload Architecture and Information Model", July 2013.

Author's Address

Tschofenig	Expires January 17, 2014	[Page 4]
------------	--------------------------	----------

Internet-Draft	Diameter Overload Piggybacking	July 2013
----------------	--------------------------------	-----------

Hannes Tschofenig
Nokia Siemens Networks
Linnoitustie 6
Espoo 02600
Finland

Phone: +358 (50) 4871445
Email: Hannes.Tschofenig@gmx.net
URI: <http://www.tschofenig.priv.at>

