CoRE
Internet Draft
Intended status: Standards track
Expires: March 2014

A. Bhattacharyya
S. Bandyopadhyay
A. Pal
Tata Consultancy Services Ltd.

September 28, 2013

CoAP option for no server-response
draft-tcs-coap-no-response-option-02

Abstract

There can be typical M2M scenarios where responses from the data sink to the data source against request/ notification from the source might be considered redundant. This kind of open-loop exchange (with no reverse path from the sink to the source) may be desired while updating resources or notifying about the updated status of a resource in constrained systems looking for maximized throughput with minimized resource consumption. CoAP already provides a non-confirmable (NON) mode of exchange where The receiving end-point does not respond with ACK. However, the receiving end-point responds the sender with a status code indicating "the result of the attempt to understand and satisfy the request".

This draft introduces a header option: 'No-Resp' to suppress responses from the receiver and discusses exemplary use cases which motivated this proposition based on real experience. This option also provides granularity by allowing suppression of a particular class or a combination of classes of responses. This option is applicable for both request/ response as well as resource-observe mode and may be effective for both unicast and multicast scenarios.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on March 28, 2014.

Table of Contents

1. Introduction

This draft proposes a new header option 'No-Resp' for Constrained Application Protocol (CoAP). This option enables the sender end-point to explicitly express its disinterest in getting responses back from the receiving end-point. By default this option expresses disinterest in any kind of response. This option should be applicable along with non-confirmable (NON) updates/ notifications. At present this option will have no effect if used with confirmable (CON) mode.

Along with the technical details this draft presents some practical application scenarios which should bring out the utility of this option.

1.1. Granular suppression of responses

This option enables granularity by allowing the sender to choose the typical class or combination of classes of responses which it is disinterested in. For example, a sender may explicitly tell the receiver that no response is required unless something 'bad' happens and a response of class 4.xx or 5.xx is to be fed back to the sender. No response is required in case of 2.xx classes. A similar scheme is described in Section 3.7 of [I-D.ietf-core-groupcomm] on the server side. Here the server may perform granular suppression for group communication. But in this case the server itself decides whether to suppress responses or not. This option enables the

clients to explicitly inform the server about the disinterest in responses.

1.2. Terminology

The terms used in this draft are in conformance with those defined in [I-D.ietf-core-coap].

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.

2. Potential benefits

If this option is opportunistically used with fitting M2M applications then the concerned systems may benefit in the following aspects:

- Reduction of network clogging
- Reduction in server-side loading
- Reduction in battery consumption at the constrained end-point
- Reduction in communication cost at the constrained end-point
- May help to satisfy hard real-time requirements (since, waiting due to closed loop latency is completely avoided)

3. Exemplary application scenarios

The described scenarios are confined within a communication pattern where there is a direct communication channel between a constrained device (the device may well be a constrained gateway) and an unconstrained backend. Also, we consider only the scenario of data updates which may happen in the following 2 forms:

1) Through a push to the server by the client using PUT or POST (request/ response)
2) Through notifications by the server to client in response to an 'observe' request by the client

The application scenarios are classified into 2 categories as below:

Category 1) Data-source=constrained device; Data-sink=backend.

Category 2) Data-source=backend; Data-sink=constrained device.

Next sub-section describes the user stories and the potential benefits in each of the cases through the use of No-Resp option. An Intelligent Traffic System (ITS) is considered as the base

application. The application scenarios are formed out of its different aspects.

3.1. Frequent update of geo-location from vehicles to backend (Category 1)

Each vehicle in ITS is equipped with a sensor-gateway comprising sensors like GPS and Accelerometer. The sensor-gateway connects to the Internet using a low-bandwidth cellular (e.g. GPRS) connection. The GPS co-ordinates are periodically updated to the backend server by the gateway. In case of ITS the update rate is adaptive to the motional-state of the vehicle. If the vehicle moves fast the update rate is high as the position of the vehicle changes rapidly. If the vehicle is static or moves slowly then the update rate is low. This ensures that bandwidth and energy is not consumed unnecessarily. The motional-state of the vehicle is inferred by a local analytics, running on the sensor-gateway, using the accelerometer data and the rate of change in GPS co-ordinates. The back-end server hosts applications which use the updates for each vehicle and produce necessary information for remote users.

The application may act in request/ response mode where the sensor-gateways push data to the backend. Or it can act in resource-observe mode where the backend initiates the exchange by sending observe requests to the sensor-gateways and receive updates in the form of notifications.

Retransmitting a location co-ordinate which is already passed by a vehicle is not efficient as it adds redundant traffic to the network. So, the updates are done in NON mode. However, given the thousands of vehicles updating frequently, the NON exchange will also trigger huge number of status responses from the backend. Each response in the air is of 4bytes of application layer plus several bytes originating from the lower layers. Thus the cumulative load on the network will be quite significant.

On the contrary, if the edge devices explicitly declare that they do not need any status response then significant load will be reduced from the network and the server as well.

3.1.1. Benefits using No-Resp

Thus mapping the above scenario to the benefits mentioned in section 2 reveals that use of 'No-Resp' will help in:

-> Reduction of network clogging
-> Reduction in server-side loading

-> Help in achieving real-time requirements as the application is not bound by any delay due to closed loop latency

3.2. A fleet-tracking application running on a PDA or smart-phone (Category 2)

Now, let us consider the other side of the ITS backend. Say, the security personnel in a city are on high alert and want to track a high-speed train carrying some important statesmen. The application on the hand-held device subscribes to the ITS backend server and receives continuous real-time updates.

If the handheld has to send a status response for each notification it receives then that will cost the device both in terms of communication cost and battery life. This can be avoided if, similar to the above scenario, the backend explicitly specifies that no response from the handheld is required. So, the backend uses 'No-Resp' for each of the notifications.

3.2.1. Benefits using No-Resp

Thus mapping the above scenario to the benefits mentioned in section 2 reveals that use of 'No-Resp' will help in:

-> Reduction in battery consumption at the constrained end-point
-> Reduction in communication cost at the constrained end-point
-> Help in achieving real-time requirements as the application is not bound by any delay due to closed loop latency.

3.3. Multicasting traffic congestion information to PDAs/ smart-phones using resource-observe (Category 2 with pseudo multicast)

The ITS might have an application which runs some analytics at the backend and determines the instantaneous traffic congestion spots in a city. The analytics is done based on the real-time geo-location updates received from the vehicles within the system. The backend application multicasts the results of the analytics to the constrained handheld devices which subscribed for real-time updates on congestion points. So, in stricter terms, it is a pseudo multicast using resource observe. In this case the backend may use No-Resp option along with NON notifications to reduce the traffic generated due to simultaneous status responses from hundreds of subscribed handheld devices.

3.3.1. Using granular response suppression

However, an intelligent application may use the granularity feature of this option such that the responses are fed-back to the backend when notification to particular devices causes errors. So the notifications may contain No-Resp saying that a response is to be suppressed only in success conditions and all responses in case of errors should be fed back. The server might eventually stop sending notification to the subscribed clients which responded with consecutive 'bad' responses. This will indirectly help saving network bandwidth.

3.3.2. Benefits using No-Resp

Thus mapping the above scenario to the benefits mentioned in section 2 reveals that use of 'No-Resp' will help in:

-> Reduction of network clogging
-> Reduction in battery consumption at the constrained end-point
-> Reduction in communication cost at the constrained end-point

4. Option Definition

The properties of this option are as in Table 1.

Number	C/E	Name	Data Format	Length	Default
TBD	E	No-Resp	uint	0-1	(none)

Table 1: Option Properties

This option has a maximum length of 1 byte. When present with an empty value this option would express the sender's disinterest in all kinds of responses by default.

This option may contain values to indicate disinterest in a particular class or combination of classes of responses as described in the next sub-section.

This option is presently defined for update requests (e.g., PUT) in NON mode or for non-confirmable update notifications against an observe request. At present this option should have no effect if used with a CON request.

4.1. Achieving granular suppression

The option length is 1 byte when this option is to be used for granular response suppression. This option is defined as a bit-map to achieve granularity as shown in Table 2.

Value	Binary Representation	Description
0	00000000	No suppression
1	00000001	Suppress 2.xx success responses
2	00000010	Suppress 4.xx client errors
4	00000100	Suppress 5.xx server errors

Table 2: Option values

XOR of the values defined for suppressing particular classes will result in suppression of a combination of classes of response. So, a value of 5 (binary: 00000101) will result in suppression of all 2.xx and 5.xx classes of responses.

It is to be noted that a value of 7 will indicate that all types of responses to be suppressed which is same as No-Resp with empty value.

5. Example

This section illustrates two examples of exchanges based on the scenario narrated in section 3.1. Examples for other scenarios can be easily conceived based on these illustrations.

5.1. Request/response Scenario

Figure 1 shows a typical request with this option. The depicted scenario occurs when the vehicle#n moves very fast and update rate is high. The vehicle is assigned a dedicated resource: vehicle-stat-<n>, where <n> can be any string uniquely identifying the vehicle. The update requests are in NON mode. The No-Resp option causes the server not to reply with any status code.

```
Client Server
|      |
|      |
+----->| Header: PUT (T=NON, Code=3, MID=0x7d38)
| PUT  | Token: 0x53
|      | Uri-Path: "vehicle-stat-00"
|      | Content Type: text/plain
|      | No-Resp: (empty)
|      | Payload:
|      | "VehID=00&RouteID=DN47&Lat=22.5658745&Long=88.4107966667&
|      | Time=2013-01-13T11:24:31"
|      |
[No response from the server. Next update in 20 secs.]
|      |
+----->| Header: PUT (T=NON, Code=3, MID=0x7d39)
| PUT  | Token: 0x54
|      | Uri-Path: "vehicle-stat-00"
|      | Content Type: text/plain
|      | No-Resp: (empty)
|      | Payload:
|      | "VehID=00&&RouteID=DN47Lat=22.5649015&Long=88.4103511667&
|      | Time=2013-01-13T11:24:51"
```

Figure 1: Exemplary unreliable update with no-resp option.

5.2. Resource-observe Scenario

The resource-observe variant of the scenario of section 3.1 is depicted in Figure 2.

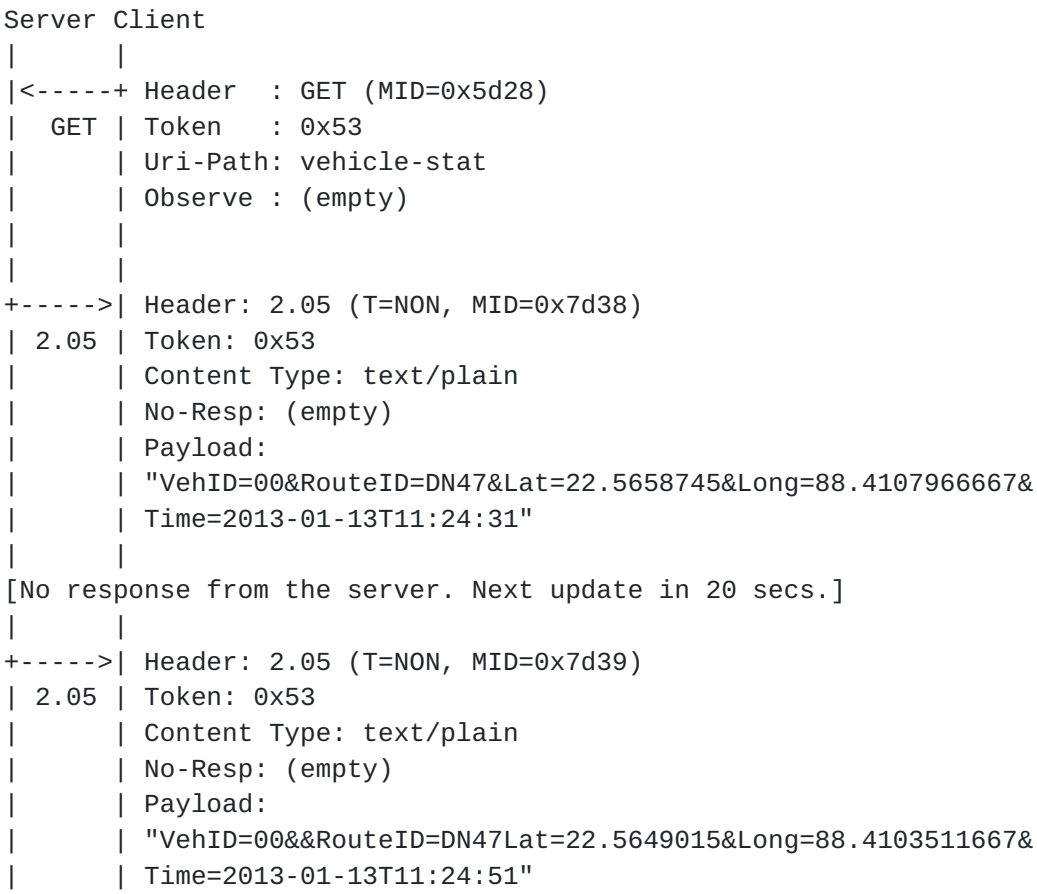

Figure 2: Exemplary unreliable update in resource-observe mode with no-resp option.

6. IANA Considerations

The IANA is requested to add the following option number entries:

Number	Name	Reference
TBD	No-Resp	Section 4 of this document

7. Security Considerations

The No-Resp option defined in this document presents no security considerations beyond those in Section 11 of the base CoAP specification [I-D.ietf-core-coap].

8. Acknowledgments

Thanks to Carsten Bormann, Esko Dijk and Bert Greevenbosch for their valuable inputs.

9. References

9.1. Normative References

[I-D.ietf-core-coap]

Shelby, Z., Hartke, K. and Bormann, C.,"Constrained Application Protocol (CoAP)", draft-ietf-core-coap-18, June 28, 2013

[I-D.ietf-core-observe]

Hartke, K.,"Observing Resources in CoAP", draft-ietf-core-observe-09, July 15, 2013

[I-D.ietf-core-groupcomm]

Rahman, A. and Dijk, E.,"Group Communication for CoAP", draft-ietf core-groupcomm-12, July 30, 2013

Authors' Addresses

Abhijan Bhattacharyya
Tata Consultancy Services Ltd.
Kolkata, India

Email: abhijan.bhattacharyya@tcs.com

Soma Bandyopadhyay
Tata Consultancy Services Ltd.
Kolkata, India

Email: soma.bandyopadhyay@tcs.com

Arpan Pal
Tata Consultancy Services Ltd.
Kolkata, India

Email: arpan.pal@tcs.com