

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 10, 2011

A. Niemi
K. Kiss
Nokia
S. Loreto
Ericsson
July 09, 2010

Session Initiation Protocol (SIP) Event Notification Extension for
Notification Rate Control
draft-ietf-sipcore-event-rate-control-04

Abstract

This document specifies mechanisms for adjusting the rate of Session Initiation Protocol (SIP) event notifications. These mechanisms can be applied in subscriptions to all SIP event packages.

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 10, 2011.

Copyright Notice

Copyright (c) 2010 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

Internet-Draft

Event Rate Control

July 2010

described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1.	Introduction	4
2.	Definitions and Document Conventions	5
3.	Overview	5
3.1.	Use Case for limiting the maximum rate of notifications .	5
3.2.	Use Case for setting a minimum rate for notifications . .	6
3.3.	Use Case for specifying the average rate of notifications	7
3.4.	Requirements	7
3.5.	The maximum rate mechanism for Resource List Server . . .	8
3.6.	Basic Operation	10
4.	Operation of the maximum rate mechanism	11
4.1.	Subscriber Behavior	11
4.2.	Notifier Behavior	11
4.3.	Selecting the maximum rate	12
4.4.	Buffer Policy Description	13
4.4.1.	Partial State Notifications	13
4.4.2.	Full State Notifications	14
4.5.	Estimated Bandwidth Savings	14
5.	Operation of the minimum rate mechanism	14
5.1.	Subscriber Behavior	15
5.2.	Notifier Behavior	15
6.	Operation of the average rate mechanism	16
6.1.	Subscriber Behavior	16
6.2.	Notifier Behavior	17
6.3.	Calculating the timeout	18
7.	Usage of "min-interval", "max-interval" and "average-interval" in a combination	19
8.	Syntax	20
8.1.	"min-interval", "max-interval" and "average-interval" Header Field Parameters	20
8.2.	Augmented BNF Definitions	20
8.3.	Event header field usage in responses to the NOTIFY request	21
9.	IANA Considerations	21

10.	Security Considerations	22
11.	Acknowledgments	22
12.	References	22
12.1.	Normative References	22
12.2.	Informative References	23
	Authors' Addresses	24

[1.](#) Introduction

The SIP events framework [[RFC3265](#)] defines a generic framework for subscriptions to and notifications of events related to SIP systems. This framework defines the methods SUBSCRIBE and NOTIFY, and introduces the concept of an event package, which is a concrete application of the SIP events framework to a particular class of events.

One of the things the SIP events framework mandates is that each event package specification defines an absolute maximum on the rate at which notifications are allowed to be generated by a single notifier. Such a limit is provided in order to reduce network load.

All of the existing event package specifications include a maximum notification rate recommendation, ranging from once in every five seconds [[RFC3856](#)], [[RFC3680](#)], [[RFC3857](#)] to once per second [[RFC3842](#)].

Per the SIP events framework, each event package specification is also allowed to define additional throttle mechanisms which allow the subscriber to further limit the rate of event notification. So far none of the event package specifications have defined such a mechanism.

The resource list extension [[RFC4662](#)] to the SIP events framework also deals with rate limiting of event notifications. The extension allows a subscriber to subscribe to a heterogeneous list of resources with a single SUBSCRIBE request, rather than having to install a subscription for each resource separately. The event list

subscription also allows rate limiting, or throttling of notifications, by means of the Resource List Server (RLS) buffering notifications of resource state changes, and sending them in batches. However, the event list mechanism provides no means for the subscriber to set the interval for the throttling; it is strictly an implementation decision whether batching of notifications is supported, and by what means.

This document defines an extension to the SIP events framework defining the following three "Event" header field parameters that allow a subscriber to set a Minimum, a Maximum and an Average rate of event notifications generated by the notifier:

min-interval: specifies a minimum notification time period between two notifications, in seconds.

max-interval: specifies a maximum notification time period between two notifications, in seconds.

average-interval: specifies an average cadence at which notifications are desired, in seconds.

The requirements and model are further discussed in [Section 3](#). All these mechanisms are simply timer values that indicate the minimum, maximum and average time period allowed between two notifications. As a result of these mechanisms, a compliant notifier will adjust the rate at which it generates notifications.

These mechanisms are applicable to any event subscription, both single event subscription and event list subscription.

[2](#). Definitions and Document Conventions

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](#)] and indicate requirement levels for compliant implementations.

Indented passages such as this one are used in this document to provide additional information and clarifying text. They do not contain normative protocol behavior.

[3.](#) Overview

[3.1.](#) Use Case for limiting the maximum rate of notifications

A presence client in a mobile device contains a list of 100 buddies or presentities. In order to decrease the processing and network load of watching 100 presentities, the presence client has employed a Resource List Server (RLS) with the list of buddies, and therefore only needs a single subscription to the RLS in order to receive notification of the presence state of the resource list.

In order to control the buffer policy of the RLS, the presence client sets a maximum rate ("min-interval" parameter), i.e. a minimum time interval between two notifications. Alternatively, the presence client could set the maximum rate for the resource list via a list manipulation interface, e.g., using the XML Configuration Access Protocol (XCAP) [[RFC4825](#)].

The RLS will buffer notifications that do not comply with the maximum rate and batch all of the buffered state changes together in a single

notification only when allowed by the maximum rate. The maximum rate applies to the overall resource list, which means that there is a hard cap imposed by the maximum rate to the number of notifications the presence client can expect to receive.

For example, with a "min-interval" of 20 seconds, the presence application can expect to receive a notification at a minimum of every 20 seconds.

The presence client can also modify the "min-interval" parameter during the lifetime of the subscription. For example, if the User Interface (UI) of the application shows inactivity for a period of time, it can simply pause the notifications by setting the "min-interval" parameter to the subscription expiration time, while still keeping the subscription alive. When the user becomes active again,

the presence client can resume the stream of notifications by re-subscribing with a "min-interval" parameter set to the earlier used value. Application of the mechanism defined by [RFC 5839](#) [[RFC5839](#)] can also eliminate for the presence client to receive a (full-state) notification carrying the latest resource state after the subscription refresh.

[3.2.](#) Use Case for setting a minimum rate for notifications

A location application is monitoring the movement of a target.

In order to decrease the processing and network load, the location application has made a subscription with a set of location filters [[I-D.ietf-geopriv-loc-filters](#)] that specify trigger criterias, for example, to send an update only when the target has moved at least n meters. However, the application is also interested in receiving the current state periodically even if the state of the target has not changed enough to satisfy any of the trigger criteria, i.e. has not moved at least n meters within the period.

In order to control the actual state, the location application sets a minimum rate ("max-interval" parameter), i.e. a maximum time interval between two notifications.

The location application can also modify the "max-interval" parameter during the lifetime of the subscription. When the subscription to the movement of a target is made, the notifier may not have the location information available. Thus, the first notification might be empty, or certain values might be absent. An important use case is placing constraints on when complete state should be provided after creating the subscription. The "max-interval" parameter indicates the time to the notifier when a complete state information should be notified. Once state is acquired and the second

notification is sent, the subscriber updates or changes the "max-interval" parameter to a more sensible value. This update can be performed in the 200 OK response to the NOTIFY request that contains the complete state information.

[3.3.](#) Use Case for specifying the average rate of notifications

The previous mechanisms introduce a static and instantaneous rate

control. However there are some applications that would work better with an adaptive rate control. This section illustrates the tracking scenario.

A tracking application is monitoring a target.

In order to decrease the processing and network load, the tracking application wants to make a subscription that dynamically increases the interval between notifications if the target has sent out several notifications recently.

In order to set an adaptive rate control, the application defines a average cadence ("average-interval" parameter) at which notifications are desired. The "average-interval" parameter value is used by the notifier to dynamically calculate the maximum time allowed between two notifications. In order to dynamically calculate the maximum, the Notifier takes into consideration the frequency at which notifications have been sent recently.

This type of rate control allows the notifier to dynamically increase or decrease the Notification frequency.

The tracking application can also modify the "average-interval" parameter during the lifetime of the subscription.

[3.4.](#) Requirements

- REQ1: The subscriber must be able to set the minimum time period ("min-interval" parameter) between two notifications in a specific subscription.
- REQ2: The subscriber must be able to set the maximum time period ("max-interval" parameter) between two notifications in a specific subscription.
- REQ3: The subscriber must be able to set an average cadence ("average-interval" parameter) at which notifications are desired in a specific subscription.

- REQ4: It must be possible to apply all together, or in any

combination, the "min-interval", "max-interval" and "average-interval" mechanisms in a specific subscription.

- REQ5: It must be possible to use of the different rate control mechanisms in subscriptions to any events.
- REQ6: It must be possible to use the different rate control mechanisms together with any other event filtering mechanisms.
- REQ7: The notifier must be allowed to use a policy in which the minimum time period between two notifications is adjusted from the value given by the subscriber.

For example, due to congestion reasons, local policy at the notifier could temporarily dictate a policy that in effect increases the subscriber-configured minimum time period between two notifications. In another example, the notifier can decrease the proposed minimum time by the subscriber to match it with the remaining expiry time left for the subscription.

- REQ8: The different rate control mechanisms must discuss corner cases for setting the time periods between two notifications. At a minimum, the mechanisms must include discussion of the situation resulting from a minimum, maximum or average time period which exceeds the subscription duration, and should provide mechanisms for avoiding this situation.
- REQ9: The different rate control mechanisms must be possible to be installed, modified, or removed in the course of an active subscription.
- REQ10: The different rate control mechanisms must allow for the application of authentication and integrity protection mechanisms to subscriptions invoking that mechanism.

3.5. The maximum rate mechanism for Resource List Server

When applied to a list subscription [[RFC4662](#)], the maximum rate mechanism has some additional considerations. Specifically, the maximum rate applies to the aggregate notification stream resulting from the list subscription, rather than explicitly controlling the notification of each of the implied constituent events. Moreover, the RLS can use the maximum rate mechanism on its own to control the rate of the back-end subscriptions to avoid overflowing its buffer.

The notifier is responsible for sending out event notifications upon state changes of the subscribed resource. We can model the notifier as consisting of three components: the event state resource(s), the Resource List Server (RLS) (or any other notifier), a notification buffer, and finally the subscriber, or watcher of the event state, as shown in Figure 1.

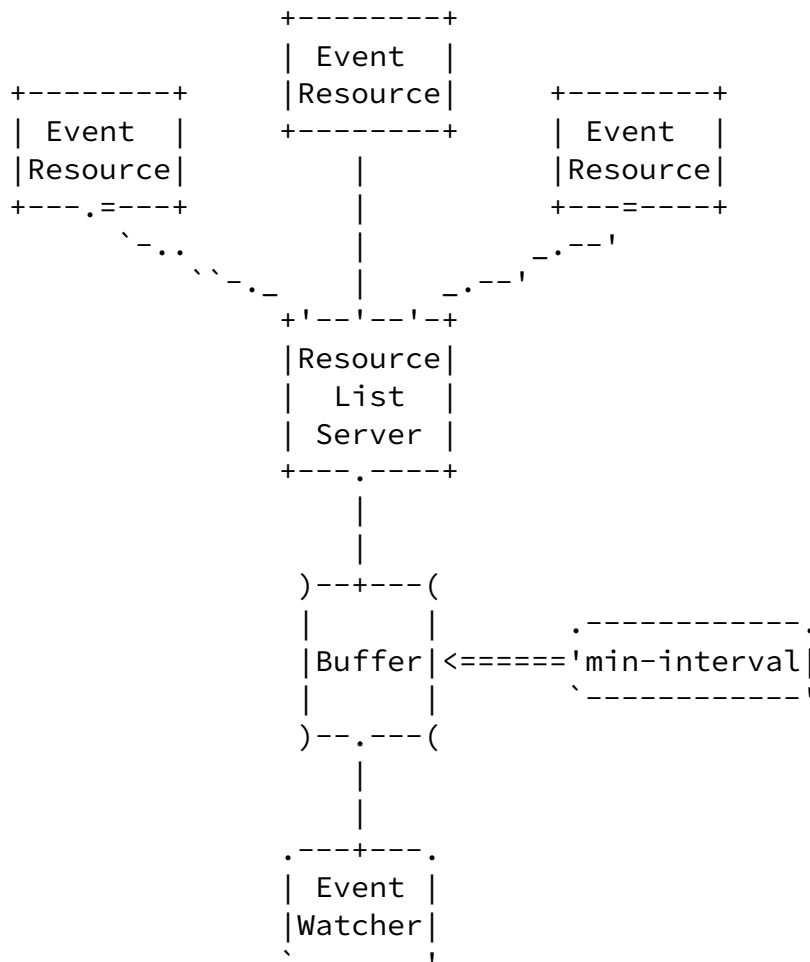


Figure 1: Model for the Resource List Server (RLS) Supporting Throttling

In short, the RLS reads event state changes from the event state resource, either by creating a back end subscription, or by other means; it packages them into event notifications, and submits them into the output buffer. The rate at which this output buffer drains is controlled by the subscriber via the maximum rate mechanism. When a set of notifications are batched together, the way in which overlapping resource state is handled depends on the type of the resource state:

In theory, there are many buffer policies that the notifier could implement. However, we only concentrate on two practical buffer policies in this specification, leaving additional ones for further study and out of the scope of this work. These two buffer policies depend on the mode in which the notifier is operating.

Full-state: Last (most recent) full state notification of each resource is sent out, and all others in the buffer are discarded. This policy applies to those event packages that carry full-state notifications.

Partial-state: The state deltas of each buffered partial notification per resource are merged, and the resulting notification is sent out. This policy applies to those event packages that carry partial-state notifications.

[3.6.](#) Basic Operation

A subscriber that wants to limit the rate of event notification in a specific event subscription does so by including a "min-interval" Event header parameter as part of the SUBSCRIBE request. The "min-interval" value indicates the minimum time allowed between transmission of two consecutive notifications in a subscription.

Note that the witnessed time between two consecutive received notifications may not conform to the "min-interval" value for a number of reasons. For example, network jitter and retransmissions may result in the subscriber receiving the notifications with smaller intervals than the "min-interval" value recommends.

A subscriber that wants to have a maximum notification time period in a specific event subscription does so by including a "max-interval" Event header parameter as part of the SUBSCRIBE request. The "max-interval" value indicates the maximum time allowed between transmission of two consecutive notifications in a subscription.

A subscriber that wants to have an average cadence for the notifications in a specific event subscription does so by including a

"average-interval" Event header parameter as part of the SUBSCRIBE request.

A subscriber that wants to update a previously agreed event rate control parameter does so by including the updated "min-interval", "max-interval" or "average-interval" Event header parameter as part of a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request.

A notifier that supports the different rate control mechanisms will comply with the value given in "min-interval", "max-interval" and "average-interval" parameters and adjust its rate of notification accordingly. However, if the notifier needs to lower the subscription expiration value or a local policy or other implementation-determined constraint at the notifier can not satisfy the rate control request, then the notifier can adjust (i.e. increase or decrease) opportunely the subscriber requested rate control.

[4.](#) Operation of the maximum rate mechanism

[4.1.](#) Subscriber Behavior

In general, the way in which a subscriber generates SUBSCRIBE requests and processes NOTIFY requests is according to [RFC 3265](#) [[RFC3265](#)].

A subscriber that wishes to apply a maximum rate to notifications in a subscription MUST construct a SUBSCRIBE request that includes a minimum time interval between two consecutive notifications included in the "min-interval" Event header field parameter. The value of this parameter is an integral number of seconds in decimal.

A subscriber that wishes to update the previously agreed maximum rate of notifications MUST include the updated "min-interval" Event header field parameter in a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request. If the Event header field of the SUBSCRIBE request did not include the "min-interval" parameter, the subscriber MUST NOT include an initial value of the "min-interval" Event header field parameter in a 200-class response to the NOTIFY request.

A subscriber that wishes to remove the maximum rate control from notifications MUST indicate so by not including a "min-interval" Event header field parameter in a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request.

There are two main consequences for the subscriber when applying the maximum rate mechanism: state transitions may be lost, and event notifications may be delayed. If either of these side effects constitute a problem to the application that is to utilize event notifications, developers are instructed not to use the mechanism.

[4.2.](#) Notifier Behavior

In general, the way in which a notifier processes SUBSCRIBE requests and generates NOTIFY requests is according to [RFC 3265](#) [[RFC3265](#)].

A notifier that supports the maximum rate mechanism MUST extract the value of the "min-interval" Event header parameter from a SUBSCRIBE request or a 200-class response to the NOTIFY request and use it as the suggested time allowed between two notifications. This value can be adjusted by the notifier, as defined in [Section 4.3](#). If the Event header field of the SUBSCRIBE request did not include the "min-interval" parameter, the notifier MUST ignore an initial value of the "min-interval" Event header field parameter in a 200-class response to the NOTIFY request, if present.

A compliant notifier MUST reflect back the possibly adjusted minimum time interval in a "min-interval" Subscription-State header field parameter of the subsequent NOTIFY requests. The indicated "min-interval" value is adopted by the notifier, and the notification rate is adjusted accordingly.

A notifier that does not understand this extension will not reflect the "min-interval" Subscription-State header field parameter in the NOTIFY requests; the absence of this parameter indicates to the subscriber that no rate control is supported by the notifier.

A compliant notifier MUST NOT generate notifications more frequently than the maximum rate allows for, except when generating the notification either upon receipt of a SUBSCRIBE request (the first notification), when the subscription state is changing from "pending"

to "active" state or upon termination of the subscription (the last notification). Such notifications reset the timer for the next notification, even though they do not need to abide by it.

When a local policy dictates a maximum rate for notifications, a notifier will not generate notifications more frequently than the local policy maximum rate, even if the subscriber is not asking for maximum rate control. The notifier MAY inform the subscriber about such local policy maximum rate using the "min-interval" Subscription-State header field parameter included in the subsequent NOTIFY requests.

Retransmissions of NOTIFY requests are not affected by the maximum rate mechanism, i.e., the maximum rate mechanism only applies to the generation of new transactions. In other words, the maximum rate mechanism does not in any way break or modify the normal retransmission mechanism specified in [RFC 3261](#) [[RFC3261](#)].

[4.3.](#) Selecting the maximum rate

Special care needs to be taken when selecting the "min-interval" value. Using the "min-interval" syntax it is possible to insist both very short and very long intervals between notifications.

For example, the maximum rate could potentially set a minimum time value between notifications that exceeds the subscription expiration value. Such a configuration would effectively quench the notifier, resulting in exactly two notifications to be generated. If the subscriber requests a "min-interval" value greater than the subscription expiration, the notifier MUST lower the "min-interval" value and set it to the expiration time left. According to [RFC 3265](#) [[RFC3265](#)] the notifier may also shorten the subscription expiry anytime during an active subscription. If the subscription expiry is shortened during an active subscription, the notifier MUST also lower the "min-interval" value and set it to the reduced expiration time.

In some cases it makes sense to pause the notification stream on an existing subscription dialog on a temporary basis without terminating the subscription, e.g. due to inactivity on the application UI. Whenever a subscriber discovers the need to perform the notification pause operation, it SHOULD set the "min-interval" value to the remaining subscription expiration value. This results in receiving

no further notifications until the subscription expires or the subscriber sends a SUBSCRIBE request resuming notifications.

The notifier MAY decide to increase or decrease the proposed maximum rate value by the subscriber based on its local policy, static configuration or other implementation-determined constraints. The notifier MUST include the adjusted "min-interval" value in the Subscription-State header field's "min-interval" parameter in each of the NOTIFY requests. In addition, different event packages MAY define additional constraints to the allowed "min-interval" intervals. Such constraints are out of the scope of this specification.

[4.4.](#) Buffer Policy Description

[4.4.1.](#) Partial State Notifications

With partial notifications, the notifier will always need to keep both a copy of the current full state of the resource F, as well as the last successfully communicated full state view F' of the resource in a specific subscription. The construction of a partial notification then involves creating a difference of the two states, and generating a notification that contains that difference.

When the maximum rate mechanism is applied to the subscription, it is important that F' is replaced with F only when the difference of F and F' was already included in a partial state notification to the subscriber allowed by the maximum rate mechanism. Additionally, the notifier implementation SHOULD check to see that the size of an accumulated partial state notification is smaller than the full

state, and if not, the notifier SHOULD send the full state notification instead.

[4.4.2.](#) Full State Notifications

With full state notifications, the notifier only needs to keep the full state of the resource, and when that changes, send the resulting notification over to the subscriber.

When the maximum rate mechanism is applied to the subscription, the notifier receives the state changes of the resource, and generates a

notification. If there is a pending notification, the notifier simply replaces that notification with the new notification, discarding the older state.

[4.5.](#) Estimated Bandwidth Savings

It is difficult to estimate the total bandwidth savings accrued by using the maximum rate mechanism over a subscription, since such estimates will vary depending on the usage scenarios. However, it is easy to see that given a subscription where several full state notifications would have normally been sent in any given interval set by the "min-interval" parameter, only a single notification is sent during the same interval when using the maximum rate mechanism, yielding bandwidth savings of several times the notification size.

With partial-state notifications, drawing estimates is further complicated by the fact that the states of consecutive updates may or may not overlap. However, even in the worst case scenario, where each partial update is to a different part of the full state, a rate controlled notification merging all of these n partial states together should at a maximum be the size of a full-state update. In this case, the bandwidth savings are approximately n times the size of the header fields of the NOTIFY request.

It is also true that there are several compression schemes available that have been designed to save bandwidth in SIP, e.g., SigComp [[RFC3320](#)] and TLS compression [[RFC3943](#)]. However, such compression schemes are complementary rather than competing mechanisms to the maximum rate mechanism. After all, they can both be applied simultaneously, and in such a way that the compound savings are as good as the sum of applying each one alone.

[5.](#) Operation of the minimum rate mechanism

[5.1.](#) Subscriber Behavior

In general, the way in which a subscriber generates SUBSCRIBE requests and processes NOTIFY requests is according to [RFC 3265](#)

[\[RFC3265\]](#).

A subscriber that wishes to apply a minimum rate to notifications in a subscription MUST construct a SUBSCRIBE request that includes a maximum time interval between two consecutive notifications included in the "max-interval" Event header field parameter. The value of this parameter is an integral number of seconds in decimal.

A subscriber that wishes to update the previously agreed minimum rate of notifications MUST include the updated "max-interval" Event header field parameter in a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request. If the Event header field of the SUBSCRIBE request did not include the "max-interval" parameter, the subscriber MUST NOT include an initial value of the "max-interval" Event header field parameter in a 200-class response to the NOTIFY request.

A subscriber that wishes to remove the minimum rate control from notifications MUST indicate so by not including a "max-interval" Event header field parameter in a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request.

The main consequence for the subscriber when applying the minimum rate mechanism is that it can receive a notification even if nothing has changed in the current state of the notifier. However, [RFC 5839](#) [RFC5839] defines a mechanism that allows sending only an etag instead of the full resource state in a notification if the state has not changed.

[5.2.](#) Notifier Behavior

In general, the way in which a notifier processes SUBSCRIBE requests and generates NOTIFY requests is according to [RFC 3265](#) [RFC3265].

A notifier that supports the minimum rate mechanism MUST extract the value of the "max-interval" Event header parameter from a SUBSCRIBE request or a 200-class response to the NOTIFY request and use it as the suggested maximum time allowed between two notifications. If the Event header field of the SUBSCRIBE request did not include the "max-interval" parameter, the notifier MUST ignore an initial value of the "max-interval" Event header field parameter in a 200-class response to the NOTIFY request, if present.

The notifier MAY decide to increase or decrease the proposed minimum

rate value based on its local policy, static configuration or other implementation-determined constraints. A compliant notifier **MUST** reflect back the possibly adjusted maximum time interval in a "max-interval" Subscription-State header field parameter of the subsequent NOTIFY requests. The indicated "max-interval" value is adopted by the notifier, and the notification rate is adjusted accordingly.

A notifier that does not understand this extension, will not reflect the "max-interval" Subscription-State header field parameter in the NOTIFY requests; the absence of this parameter indicates to the subscriber that no rate control is supported by the notifier.

A compliant notifier **MUST** generate notifications whenever the time since the most recent notification exceeds the value in the "max-interval" parameter. Depending on the event package and subscriber preferences indicated in the SUBSCRIBE request, the NOTIFY request sent as a result of a max-interval expiration **MUST** contain either the current full state or the partial state showing the difference between the current state and the last successfully communicated state.

Retransmissions of NOTIFY requests are not affected by the minimum rate mechanism, i.e., the minimum rate mechanism only applies to the generation of new transactions. In other words, the minimum rate mechanism does not in any way break or modify the normal retransmission mechanism.

[6.](#) Operation of the average rate mechanism

[6.1.](#) Subscriber Behavior

In general, the way in which a subscriber generates SUBSCRIBE requests and processes NOTIFY requests is according to [RFC 3265](#) [[RFC3265](#)].

A subscriber that wishes to apply an average rate to notifications in a subscription **MUST** construct a SUBSCRIBE request that includes a proposed average time interval between two consecutive notifications included in a "average-interval" Event header field parameter. The value of this parameter is an integral number of seconds in decimal.

A subscriber that wishes to update the previously agreed average rate of notifications **MUST** include the updated "average-interval" Event header field parameter in a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request. If the Event header field of the SUBSCRIBE request did not include the "average-interval"

parameter, the subscriber MUST NOT include an initial value of the

"average-interval" Event header field parameter in a 200-class response to the NOTIFY request.

A subscriber that wishes to remove the average rate control from notifications MUST indicate so by not including a "average-interval" Event header field parameter in a subsequent SUBSCRIBE request or a 200-class response to the NOTIFY request.

The main consequence for the subscriber when applying the average rate mechanism is that it can receive a notification even if nothing has changed in the current state of the notifier. However, [RFC 5839](#) [RFC5839] defines a mechanism that allows sending only an etag instead of the full resource state in a notification if the state has not changed.

[6.2.](#) Notifier Behavior

In general, the way in which a notifier processes SUBSCRIBE requests and generates NOTIFY requests is according to [RFC 3265](#) [RFC3265].

A notifier that supports the average rate mechanism MUST extract the value of the "average-interval" Event header parameter from a SUBSCRIBE request or a 200-class response to the NOTIFY request and use it to calculate the maximum time allowed between two transactions as defined in [Section 6.3](#). If the Event header field of the SUBSCRIBE request did not include the "average-interval" parameter, the notifier MUST ignore an initial value of the "average-interval" Event header field parameter in a 200-class response to the NOTIFY request, if present.

The notifier MAY decide to increase or decrease the proposed average time interval based on its local policy, static configuration or other implementation-determined constraints. A compliant notifier MUST reflect back the possibly adjusted average time interval in an "average-interval" Subscription-State header field parameter of the subsequent NOTIFY requests. The indicated "average-interval" value is adopted by the notifier, and the notification rate is adjusted accordingly.

A notifier that does not understand this extension will not reflect

the "average-interval" Subscription-State header parameter in the NOTIFY requests; the absence of this parameter indicates to the subscriber that no rate control is supported by the notifier.

A compliant notifier SHOULD generate notifications whenever the time since the most recent notification exceeds the value calculated using the formula defined in [Section 6.3](#).

The average rate mechanism is implemented as follows:

- 1) When a subscription is first created, the notifier creates a record that keeps track of the number of notifications that have been sent in the "period". This record is initialized to contain a history of having sent one message every "average-interval" seconds for the "period".
- 2) The "timeout" value is calculated according to the equation given in [Section 6.3](#).
- 3) If the timeout period passes without a NOTIFY request being sent in the subscription, then the current resource state is sent (subject to any filtering associated with the subscription).
- 4) Whenever a NOTIFY request is sent (regardless of whether due to a timeout or a state change), the notifier updates the notification history record, recalculates the value of "timeout," and returns to step 3.

Retransmissions of NOTIFY requests are not affected by the timeout, i.e., the timeout only applies to the generation of new transactions. In other words, the timeout does not in any way break or modify the normal retransmission mechanism specified in [RFC 3261](#) [[RFC3261](#)].

[6.3](#). Calculating the timeout

The formula used to vary the absolute pacing in a way that will meet the average rate requested over the period is given in equation (1):

$$\text{timeout} = (\text{average-interval} \wedge 2) * \text{count} / \text{period} \quad (1)$$

The output of the formula, "timeout", is the time to the next notification, expressed in seconds. The formula has three inputs:

average-interval: The value of the "average-interval" parameter conveyed in the Subscription-State header field, in seconds.

period: The rolling average period, in seconds. The value of the "period" parameter MUST be greater than the value of the "average-interval" parameter.

count: The number of notifications that have been sent during the last "period" of seconds not including any retransmissions of requests.

In case both the maximum rate and the average rate mechanisms are

used in the same subscription the formula used to dynamically calculate the timeout is given in equation (2):

$$\text{timeout} = \text{MAX}[\text{min-interval}, (\text{average-interval} \wedge 2) * \text{count} / \text{period}] \quad (2)$$

min-interval: The value of the "min-interval" parameter conveyed in the Event header field, in seconds.

The formula in (2) makes sure that for all the possible values of the "min-interval" and "average-interval" parameters, with "average-interval" > "min-interval", the timeout never results in a lower value than the value of the "min-interval" parameter.

In some situation it may be beneficial for the Notifier to achieve an average rate in a different way than the algorithm detailed in this document allows. However, the Notifier MUST comply with any "min-interval" or "max-interval" parameters that have been negotiated.

[7.](#) Usage of "min-interval", "max-interval" and "average-interval" in a combination

Applications can subscribe to an event package using all the rate control mechanisms individually, or in combination; in fact there is

no technical incompatibility among them. However there are some combinations of the different rate control mechanisms that make little sense to be used together. This section lists all the combinations that are possible to insert in a subscription; the utility to use each combination in a subscription is also analyzed.

min-interval and max-interval: this combination allows to reduce the notification frequency rate, but at the same time assures the reception of a notification every time the most recent notification exceeds a specified interval.

A subscriber SHOULD choose a "max-interval" value higher than the "min-interval" value, otherwise the notifier MUST adjust the subscriber provided "max-interval" value to a value equivalent or higher than the "min-interval" value.

min-interval and average-interval: it works in a similar way as the combination above, but with the difference that the interval at which notifications are assured changes dynamically.

A subscriber SHOULD choose a "average-interval" value higher than the "min-interval" value, otherwise the notifier MUST adjust the subscriber provided "average-interval" value to a value equivalent or higher than the "min-interval" value.

max-interval and average-interval: as both the parameters are designed as minimum rate mechanisms, this combination makes sense only in some corner cases.

A subscriber SHOULD choose a "max-interval" value higher than the "average-interval" value, otherwise the notifier MUST NOT consider the "max-interval" value.

min-interval, max-interval and average-interval: this combination makes little sense to be used although not forbidden.

A subscriber SHOULD choose a "max-interval" and "average-interval" values higher than the "min-interval" value, otherwise the notifier MUST adjust the subscriber provided "max-interval" and

"average-interval" values to a value equivalent or higher than the "min-interval" value.

A subscriber SHOULD choose a "max-interval" value higher than the "average-interval" value, otherwise the notifier MUST NOT consider the "max-interval" value.

[8.](#) Syntax

This section describes the syntax extensions required for the different rate control mechanisms.

[8.1.](#) "min-interval", "max-interval" and "average-interval" Header Field Parameters

The "min-interval", "max-interval" and "average-interval" parameters are added to the rule definitions of the Event header field and the Subscription-State header field in [RFC 3265](#) [[RFC3265](#)] grammar. Usage of this parameter is described in [Section 4](#), [Section 5](#) and [Section 6](#).

[8.2.](#) Augmented BNF Definitions

This section describes the Augmented BNF [[RFC5234](#)] definitions for the new syntax elements. Note that we derive here from the ruleset present in [RFC 3265](#) [[RFC3265](#)], adding additional alternatives to the alternative sets of "event-param" and "subexp-params" defined therein.

```
event-param      =/ min-interval-param
subexp-params    =/ min-interval-param
min-interval-param = "min-interval" EQUAL delta-seconds
```

```
event-param      =/ max-interval-param
subexp-params    =/ max-interval-param
max-interval-param = "max-interval" EQUAL delta-seconds
```

```
event-param      =/ average-interval-param
subexp-params    =/ average-interval-param
average-interval-param = "average-interval" EQUAL delta-seconds
```

8.3. Event header field usage in responses to the NOTIFY request

This table expands the table described in [Section 7.2 of RFC 3265 \[RFC3265\]](#) allowing the Event header field to appear in a 200-class response to a NOTIFY request. A UA that wishes to update the value for minimum, maximum or average rate of notifications can do so by including all desired values for the rate control parameters in an Event header field of the 200-class response to a NOTIFY request.

Header field	where proxy	ACK	BYE	CAN	INV	OPT	REG	PRA	SUB	NOT
Event	2xx	-	-	-	-	-	-	-	-	0

9. IANA Considerations

This specification registers three new SIP header field parameters, defined by the following information which is to be added to the Header Field Parameters and Parameter Values sub-registry under <http://www.iana.org/assignments/sip-parameters>.

Header Field	Parameter Name	Predefined Values	Reference
Event	min-interval	No	[RFCxxxx]
Subscription-State	min-interval	No	[RFCxxxx]
Event	max-interval	No	[RFCxxxx]
Subscription-State	max-interval	No	[RFCxxxx]
Event	average-interval	No	[RFCxxxx]
Subscription-State	average-interval	No	[RFCxxxx]

(Note to the RFC Editor: please replace "xxxx" with the RFC number of this specification, when assigned.)

10. Security Considerations

Naturally, the security considerations listed in [RFC 3265 \[RFC3265\]](#), which the rate control mechanisms described in this document extends, apply in entirety. In particular, authentication and message

integrity SHOULD be applied to subscriptions with this extension.

[RFC 3265](#) [[RFC3265](#)] recommends the integrity protection of the Event header field of SUBSCRIBE requests. Implementations of this extension SHOULD also provide integrity protection for the Event header field included in the 200-class response to the NOTIFY request.

When the maximum rate mechanism involves partial state notifications, the security considerations listed in [RFC 5263](#) [[RFC5263](#)] apply in entirety.

[11.](#) Acknowledgments

Thanks to Pekka Pessi, Dean Willis, Eric Burger, Alex Audu, Alexander Milinski, Jonathan Rosenberg, Cullen Jennings, Adam Roach, Hisham Khartabil, Dale Worley, Martin Thomson, Byron Campen, Alan Johnston, Michael Procter and Janet Gunn for support and/or review of this work.

Thanks to Brian Rosen for the idea of the minimum and average rate mechanisms, and Adam Roach for the work on the averaging algorithm and other feedback.

[12.](#) References

[12.1.](#) Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", [RFC 3261](#), June 2002.
- [RFC3265] Roach, A., "Session Initiation Protocol (SIP)-Specific Event Notification", [RFC 3265](#), June 2002.
- [RFC4662] Roach, A., Campbell, B., and J. Rosenberg, "A Session Initiation Protocol (SIP) Event Notification Extension for

Resource Lists", [RFC 4662](#), August 2006.

[RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, [RFC 5234](#), January 2008.

[RFC5263] Lonnfors, M., Costa-Requena, J., Leppanen, E., and H. Khartabil, "Session Initiation Protocol (SIP) Extension for Partial Notification of Presence Information", [RFC 5263](#), September 2008.

12.2. Informative References

- [I-D.ietf-geopriv-loc-filters]
Mahy, R., Rosen, B., and H. Tschofenig, "Filtering Location Notifications in the Session Initiation Protocol (SIP)", [draft-ietf-geopriv-loc-filters-11](#) (work in progress), March 2010.
- [RFC3320] Price, R., Bormann, C., Christoffersson, J., Hannu, H., Liu, Z., and J. Rosenberg, "Signaling Compression (SigComp)", [RFC 3320](#), January 2003.
- [RFC3680] Rosenberg, J., "A Session Initiation Protocol (SIP) Event Package for Registrations", [RFC 3680](#), March 2004.
- [RFC3842] Mahy, R., "A Message Summary and Message Waiting Indication Event Package for the Session Initiation Protocol (SIP)", [RFC 3842](#), August 2004.
- [RFC3856] Rosenberg, J., "A Presence Event Package for the Session Initiation Protocol (SIP)", [RFC 3856](#), August 2004.
- [RFC3857] Rosenberg, J., "A Watcher Information Event Template-Package for the Session Initiation Protocol (SIP)", [RFC 3857](#), August 2004.
- [RFC3943] Friend, R., "Transport Layer Security (TLS) Protocol Compression Using Lempel-Ziv-Stac (LZS)", [RFC 3943](#), November 2004.
- [RFC4825] Rosenberg, J., "The Extensible Markup Language (XML) Configuration Access Protocol (XCAP)", [RFC 4825](#), May 2007.
- [RFC5839] Niemi, A. and D. Willis, "An Extension to Session Initiation Protocol (SIP) Events for Conditional Event Notification", [RFC 5839](#), May 2010.

Internet-Draft

Event Rate Control

July 2010

Authors' Addresses

Aki Niemi
Nokia
P.O. Box 407
NOKIA GROUP, FIN 00045
Finland

Phone: +358 50 389 1644
Email: aki.niemi@nokia.com

Krisztian Kiss
Nokia
323 Fairchild Dr
Mountain View, CA 94043
US

Phone: +1 650 391 5969
Email: krisztian.kiss@nokia.com

Salvatore Loreto
Ericsson
Hirsalantie 11
Jorvas 02420
Finland

Email: salvatore.loreto@ericsson.com

