

WEBPUSH  
Internet-Draft  
Intended status: Standards Track  
Expires: November 13, 2014

M. Thomson  
Mozilla  
May 12, 2014

**Generic Event Delivery Using HTTP Push**  
**draft-thomson-webpush-http2-00**

Abstract

A simple scheme for the delivery of realtime events to clients is described. This scheme uses HTTP/2 push.

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 November 13, 2014.

Copyright Notice

Copyright (c) 2014 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

<a href="#">1.</a>	<a href="#">Introduction . . . . .</a>	<a href="#">2</a>
<a href="#">1.1.</a>	<a href="#">Conventions and Terminology . . . . .</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Overview . . . . .</a>	<a href="#">4</a>
<a href="#">3.</a>	<a href="#">Delivering Push Messages . . . . .</a>	<a href="#">5</a>
<a href="#">4.</a>	<a href="#">Registering . . . . .</a>	<a href="#">5</a>
<a href="#">5.</a>	<a href="#">Channels . . . . .</a>	<a href="#">6</a>
<a href="#">6.</a>	<a href="#">Monitoring and Receiving Push Messages . . . . .</a>	<a href="#">6</a>
<a href="#">7.</a>	<a href="#">Store and Forward Operation . . . . .</a>	<a href="#">6</a>
<a href="#">8.</a>	<a href="#">IANA Considerations . . . . .</a>	<a href="#">7</a>
<a href="#">9.</a>	<a href="#">Security Considerations . . . . .</a>	<a href="#">7</a>
<a href="#">10.</a>	<a href="#">References . . . . .</a>	<a href="#">8</a>
<a href="#">10.1.</a>	<a href="#">Normative References . . . . .</a>	<a href="#">8</a>
<a href="#">10.2.</a>	<a href="#">Informative References . . . . .</a>	<a href="#">8</a>
	<a href="#">Author's Address . . . . .</a>	<a href="#">9</a>

## [1.](#) Introduction

Mobile computing devices are increasingly relied upon for a great many applications. Mobile devices typically have limited power reserves, so finding more efficient ways to serve application requirements is an important part of any mobile platform.

One significant contributor to power usage mobile devices is the radio. Cellular - and to a lesser extent, IEEE 802 radios - consume a significant portion of the energy budget on a wirelessly connected mobile device.

Many applications require continuous access to network communications so that real-time events - such as incoming calls or messages - can be conveyed (or "pushed") to the user in a timely fashion. Uncoordinated use of persistent connections or sessions from multiple applications can contribute to unnecessary use of the device radio, since each independent session independently incurs overheads. In particular, keep alive traffic used to ensure that middleboxes do not prematurely time out sessions, can result in significant waste. Maintenance traffic tends to dominate over the long term, since events are relatively rare.

Consolidating all real-time events into a single session ensures more efficient use of network and radio resources. A single service consolidates all events, distributing those events to applications as they arrive. This requires just one session, avoiding duplicated overhead costs.



The Web Push API [[API](#)] describes an API that enables the use of a consolidated push service from web applications. This expands on that work by describing a protocol that can be used to:

- o request the delivery of an event to an attached device
- o register a new device, create new event delivery channels, and monitor those channels for new events

This is intentionally split into these two categories because requesting the delivery of events is required for immediate use by the Web Push API. The registration, management and monitoring functions are currently fulfilled by proprietary protocols; these are adequate, but do not offer any advantages that standardization affords.

The monitoring function described in this document is intended to be replaceable, enabling the use of monitoring schemes that are better optimized for the network environment and the device. For instance, using notification systems like the GSM Short Message Service (SMS) can take advantage of the native paging capabilities of a cellular network, avoiding the ongoing maintenance cost of a persistent TCP connection.

This document intentionally does not describe how a push server is discovered. Discovery of push servers is left for future efforts, if it turns out to be necessary at all. Devices are expected to be configured with a push server URL.

Similarly, discovery of support for and negotiation of use of alternative monitoring schemes is left to documents that extend this basic protocol.

### **1.1. Conventions and Terminology**

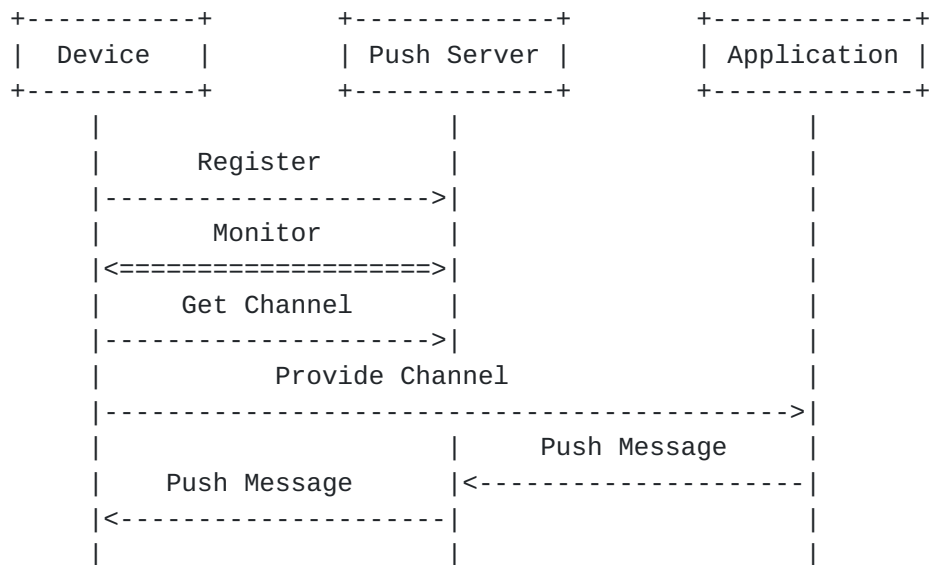
In cases where normative language needs to be emphasized, this document falls back on established shorthands for expressing interoperability requirements on implementations: the capitalized words "MUST", "MUST NOT", "SHOULD" and "MAY". The meaning of these is described in [[RFC2119](#)].

This document will use the terminology from [[API](#)], though "application" will be used in preference to "webapp", since the described protocols are not restricted to web use. This document introduces the term "device", which refers to the consumer of push messages.



## 2. Overview

A general model for push services includes three basic actors: a device, a push server, and an application.



At the very beginning of the process, the device registers with the push server. This establishes a shared session between the device and push server that will be used to aggregate push messages from all applications that the device interacts with.

The registration response includes details on how the device is expected to monitor for incoming push messages. This document describes one such mechanism, though more efficient means of monitoring could be optionally defined (and this is expressly permitted).

A registration after creation has no channels associated with it. New channels can be requested by the device and then distributed to applications. It is expected that devices will distribute a different channel to each application, with the potential for multiple channels being provided to the same application.

Applications use channels to deliver push messages to devices, via the push server.

Both registrations and channels have a limited lifetime. These will need to be refreshed over time.



### **3. Delivering Push Messages**

A push channel is identified with an HTTP URI [[I-D.ietf-httpbis-p1-messaging](#)]. An application can request the delivery of a push message by sending an HTTP PUT request to this URI, including the push message in the body of the request.

A push server can acknowledge the end-to-end delivery of a push message by responding with a 200 (OK) status code. A push server that stores the message for later delivery (see [Section 7](#)) could respond with a 202 (Accepted) status code to indicate that the message was stored, but not delivered.

### **4. Registering**

A device that wishes to establish a new or replacement registration sends an HTTP POST request to its configured push server URL. The request contains no entity body.

The push server creates a new registration in response to this request, creating two new resources and allocating an HTTP URI for each. These URIs are included in link relations that are included in Link header fields in the response.

**monitor:** A link relation of type "...:push:monitor" includes the URL of a resource that the device can monitor for events. Monitoring is described in [Section 6](#).

**channel:** A link relation of type "...:push:channel" includes a URL of a resource where the device can create new channels. Creating channels is described in [Section 5](#).

The push server includes the "monitor" link relation in a Location header field.

The push server MUST include expiration information in the response to this request in either the Expires header field, or by setting a "max-age" parameter on a Cache-Control header field. The Cache-Control header field MUST include the "private" directive [[I-D.ietf-httpbis-p6-cache](#)].

The push server SHOULD also provide the "channel" link and expiration information in response to requests to the "monitor" resource.





## 5. Channels

A client sends a POST request to the "channel" resource to create a new channel.

A response with a 201 status code includes the channel URI in the Location header field.

A channel can expire. Servers indicate this using the Expires header field, or by setting a "max-age" parameter on a Cache-Control header field.

A client can explicitly delete a channel by sending a DELETE request to channel URI.

## 6. Monitoring and Receiving Push Messages

A device monitors for new events by making a GET request to the monitor resource. The server does not respond to these request, it instead uses server push [[I-D.ietf-httpbis-http2](#)] to send the contents of push messages as applications send them.

Each push message consists of a synthesized GET request to the channel URI that was the target of the push. The response body is the entity body from the PUT request.

A device can request the monitor resource immediately by including a Prefer header field [[I-D.snell-http-prefer](#)] with a "wait" parameter set to "0".

A server that wishes to redistribute load can do so using The alternative services mechanisms that are part of HTTP/2 [[I-D.ietf-httpbis-alt-svc](#)]. The ALTSVC frame type allows for redistribution of load whilst retaining the same monitor resource. Once a device has established a replacement connection, it can notify the server of imminent shutdown using a GOAWAY frame, which allows the server to respond to the long-standing GET request and gracefully shut down the connection. This allows for seamless migration between servers.

## 7. Store and Forward Operation

Push servers are not obligated to store messages for any time. If a client is not actively monitoring for push messages, messages can be lost.

Push servers are permitted to store messages for some time to allow for limited recovery from transient faults. If a message is stored,



but not delivered, the push server can indicate the probable duration of storage by including expiration information in the response to the push request.

Messages that were stored and not delivered to a client MAY be delivered when a client commences monitoring. These messages should include a Last-Modified header field. If a server stores push messages, a GET request to a channel URI returns the last message sent by an application to that channel.

Push servers that store push messages might need to limit the size of push messages to avoid being subject to overloading. Push servers that don't store can stream the payload of push messages to devices. This can use HTTP/2 flow control to limit the state commitment this requires. However, push servers MAY place an upper limit on the size of push messages that they permit.

## **8. IANA Considerations**

TODO: register link relation types, as necessary.

## **9. Security Considerations**

This protocol does not specify a single authorization framework for managing access to push servers, either by devices or applications. Thus, there is a very real possibility that this could be exploited to mount denial of service attacks on the push service.

Push services MAY choose to authorize requests based on any HTTP-compatible means available, of which there are numerous options.

A malicious application can use the greater resources of a push server to mount a denial of service attack on devices. Push servers MAY limit the rate at which push messages are sent to devices.

One basic protection against misuse of push channels is to ensure that only authorized applications are given channel URIs and to make it extremely difficult to successfully guess a valid channel URI. Encoding a large amount of random entropy in the URI path is one technique for ensuring that channel URIs are able to act as bearer tokens.

A push server is able to see (and modify) the content of push messages. Applications that depend on end-to-end guarantees MUST use object security mechanisms to protect the confidentiality and integrity of push messages. Since the application frequently resides both in the server that originates the push messages and in the device, and there is a requirement for direct communications between



these two instances to exchange channel information, this provides a perfect medium for key agreement.

A push server is also able to deny service. In addition to malicious reasons, delivery of push messages can fail due to transient faults in networks, the device being temporarily unavailable, or other similar faults. Therefore, applications that rely on the content of push messages being delivered MUST ensure that they provide other means of delivering messages to devices.

## **10. References**

### **10.1. Normative References**

- [I-D.ietf-httpbis-alt-svc]  
mnot, m., McManus, P., and J. Reschke, "HTTP Alternative Services", [draft-ietf-httpbis-alt-svc-01](#) (work in progress), April 2014.
- [I-D.ietf-httpbis-http2]  
Belshe, M., Peon, R., and M. Thomson, "Hypertext Transfer Protocol version 2", [draft-ietf-httpbis-http2-12](#) (work in progress), April 2014.
- [I-D.ietf-httpbis-p1-messaging]  
Fielding, R. and J. Reschke, "Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing", [draft-ietf-httpbis-p1-messaging-26](#) (work in progress), February 2014.
- [I-D.ietf-httpbis-p6-cache]  
Fielding, R., mnot, m., and J. Reschke, "Hypertext Transfer Protocol (HTTP/1.1): Caching", [draft-ietf-httpbis-p6-cache-26](#) (work in progress), February 2014.
- [I-D.snell-http-prefer]  
Snell, J., "Prefer Header for HTTP", [draft-snell-http-prefer-18](#) (work in progress), January 2013.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

### **10.2. Informative References**

- [API] Sullivan, B. and E. Fulllea, "Web Push API", Editor's Draft push-api, May 2014, <<https://w3c.github.io/push-api/index.html>>.



Author's Address

Martin Thomson  
Mozilla  
331 E Evelyn Street  
Mountain View, CA 94041  
US

Email: [martin.thomson@gmail.com](mailto:martin.thomson@gmail.com)