

WEBPUSH
Internet-Draft
Intended status: Standards Track
Expires: June 15, 2015

M. Thomson
Mozilla
December 12, 2014

Generic Event Delivery Using HTTP Push
draft-thomson-webpush-http2-02

Abstract

A simple protocol for the delivery of realtime events to clients is described. This scheme uses HTTP/2 server 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 June 15, 2015.

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.

Internet-Draft

HTTP Web Push

December 2014

Table of Contents

1.	Introduction	2
1.1.	Conventions and Terminology	4
2.	Overview	4
2.1.	HTTP Resources	6
3.	Registration	6
4.	Subscribing	7
5.	Requesting Push Message Delivery	7
6.	Receiving Push Messages	8
7.	Operational Considerations	9
7.1.	Load Management	9
7.2.	Push Message Expiration	10
7.3.	Registration and Subscription Expiration	10
7.4.	Implications for Application Reliability	11
8.	Security Considerations	11
8.1.	Confidentiality from Push Service Access	11
8.2.	Privacy Considerations	12
8.3.	Authorization	13
8.4.	Denial of Service Considerations	13
8.5.	Logging Risks	14
9.	IANA Considerations	14
10.	Acknowledgements	16
11.	References	16
11.1.	Normative References	16
11.2.	Informative References	17
	Author's Address	17

[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. Radio communications consumes 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 W3C 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 a push message to a user agent,
- o register a new user agent,
- o create new push message delivery subscriptions, and
- o monitor for new push messages.

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 of the 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 user agent. For instance, using notification systems like the 3GPP Short Message Service (SMS) [[TS.3GPP.23.040](#)] 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 service is discovered. Discovery of push services is left for future efforts, if it turns out to be necessary at all. User agents are expected to be configured with a URL for a push service.

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 defines the following terms:

application: Both the sender and ultimate consumer of push messages. Many applications have components that are run on a user agent and other components that run on servers.

application server: The component of an application that runs on a server and requests the delivery of a push message.

push message: A message, sent from an application server to a user agent via a push service.

push service: A service that delivers push messages to user agents.

registration: A session that is established between the user agent and the push service. A registration has any number of associated subscriptions.

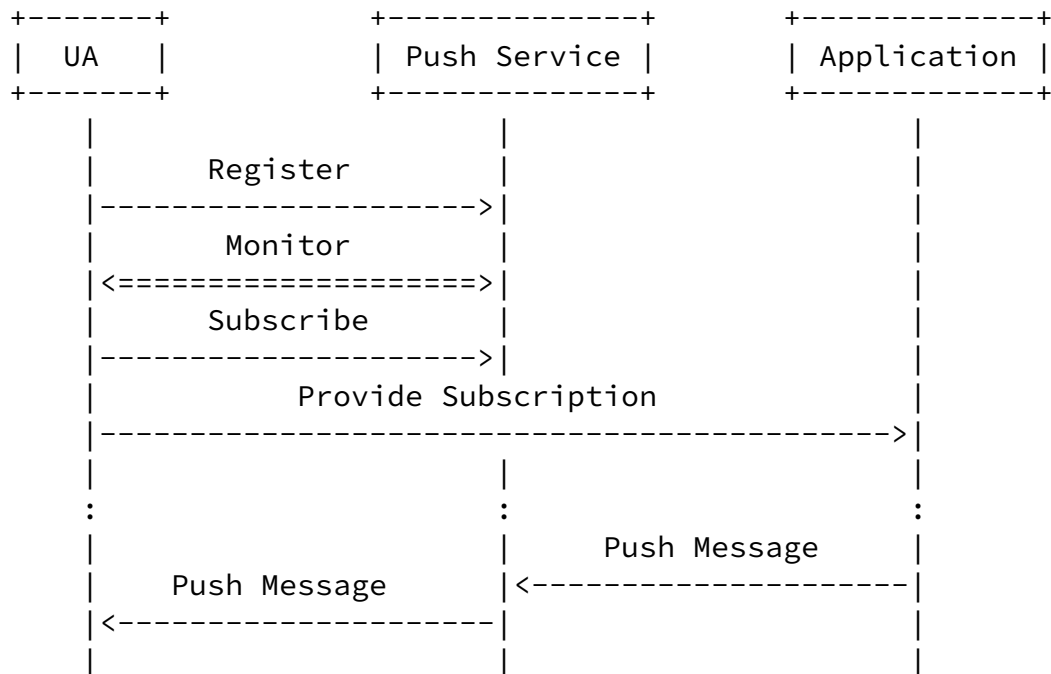
subscription: A message delivery context that is established between the user agent and the push service and shared with applications. All push messages are associated with a subscription.

user agent: A device and software that is the recipient of push messages.

Examples in this document use the HTTP/1.1 message format [RFC7230]. Many of the exchanges can be completed using HTTP/1.1, where HTTP/2 is necessary, the more verbose frame format from [I-D.ietf-httpbis-http2] is used.

2. Overview

A general model for push services includes three basic actors: a user agent, a push service, and an application (server).



At the very beginning of the process, the device creates a registration resource at the push service. The registration is the basis of all future interactions between the user agent and push

service. The registration aggregates push messages that the user agent receives from all subscriptions.

The registration resource describes how the user agent is expected to monitor for incoming push messages. This document describes one such mechanism, though more efficient means of monitoring could be defined.

The registration resource also describes how the user agent might create a subscription. A new subscription is created by the user agent and then distributed by to an application server.

It is expected that a different subscription will be provided to each application; however, there are no inherent cardinality constraints in the protocol. Multiple subscriptions might be created for the same application, or multiple applications could use the same subscription. Note however that sharing subscriptions can have security and privacy implications.

Application servers use subscriptions to deliver push messages to devices, via the push service.

Both registrations and subscriptions have a limited lifetime. They can also be terminated by either push service or user agent at any time. User agents and application servers need to be prepared to handle changes in registrations and subscriptions; the protocol

described here supports any number of either resource concurrently with minimal overhead.

[2.1](#). HTTP Resources

This protocol uses HTTP resources [[RFC7230](#)] and link relations [[RFC5988](#)]. The following resources are defined:

push service: This resource is used in Registration ([Section 3](#)). It is configured into user agents.

registration: A link relation of type "urn:ietf:params:push:reg" refers to a registration resource. This resource is used by a user agent in requesting the delivery of push messages. A process for monitoring for messages using this resource is described in

[Section 6.](#)

subscribe: A link relation of type "urn:ietf:params:push:sub" refers to a resource where a user agent can create new subscriptions. Creating a subscription is described in [Section 4](#).

subscription: A link relation of type "urn:ietf:params:push" refers to a subscription resource. Subscription resources are used to deliver push messages. An application server sends push messages ([Section 5](#)) and a user agent receives push messages ([Section 6](#)) using this resource.

[3.](#) Registration

A user agent that wishes to establish a new or replacement registration sends an HTTP POST request to its configured push service resource.

A request to create a registration contains no entity body. A future specification MAY define a format and semantics for the entity body on this request.

The push service creates a new registration and subscribe resource in response to this request. URIs for these resources are included in Link header fields in the response. The push server includes the "registration" link relation in a Location header field.

For example, the following request requests the creation of a new registration:

```
POST /register HTTP/1.1
Host: push.example.net
```

The following response might be generated in response to this request:

```
HTTP/1.1 201 Created
Date: Thu, 11 Dec 2014 23:56:47 GMT
Link: </monitor/1G_GIPMorg_n-IrQvqZr6g>;
      rel="urn:ietf:params:push:reg"
Link: </subscribe/1G_GIPMorg_n-IrQvqZr6g>;
```

```
rel="urn:ietf:params:push:sub"
Location: https://push.example.net/reg/1G_GIPMorg_n-IrQvqZr6g
Cache-Control: max-age=8640000, private
```

The push server SHOULD provide a URI for the associated "subscribe" resource and any known expiration information in response to requests to the "registration" resource.

[4.](#) Subscribing

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

```
POST /subscribe/1G_GIPMorg_n-IrQvqZr6g HTTP/1.1
Host: push.example.net
```

A response with a 201 (Created) status code includes a URI for the subscription in the Location header field.

```
HTTP/1.1 201 Created
Date: Thu, 11 Dec 2014 23:56:52 GMT
Link: </p/LBhhw0Ooh0-Wl40i971UGsB7sdQGUibx>;
      rel="urn:ietf:params:push"
Location: https://push.example.net/p/LBhhw0Ooh0-Wl40i971UGsB7sdQGUibx
Cache-Control: max-age:864000, private
```

[5.](#) Requesting Push Message Delivery

A push subscription is an HTTP resource [[RFC7230](#)].

An application server requests the delivery of a push message by sending an HTTP PUT request to this resource, including the push message in the body of the request.

```
PUT /p/LBhhw0Ooh0-Wl40i971UGsB7sdQGUibx HTTP/1.1
```


Host: push.example.net
Content-Type: text/plain;charset=utf8
Content-Length: 15

Hello, World!

A simple 200 response is sufficient to indicate that the push message was accepted. This does not indicate that the message was delivered to the user agent.

HTTP/1.1 200 OK
Date: Thu, 11 Dec 2014 23:56:55 GMT
Cache-Control: max-age=600

A push service MAY generate a 413 (Payload Too Large) status code in response to PUT requests that include an entity body that is too large. Push services MUST NOT generate a 413 status code in responses to an entity body that is 4k (4096 bytes) or less in size.

A push service that does not store messages can stream the payload of push messages to the user agent. Flow control SHOULD be used to limit the state commitment for delivery of large messages.

6. Receiving Push Messages

A user agent requests the delivery of new push messages by making a GET request to the "registration" resource. The push service does not respond to this request, it instead uses HTTP/2 server push [[I-D.ietf-httpbis-http2](#)] to send the contents of push messages as they are sent by application servers.

Each push message is pushed in response to a synthesized GET request. The GET request is made to the same "subscription" URI that is used by the application server to send the message. The response body is the entity body from the most recent PUT request sent to the subscription resource.

The following example request is made over HTTP/2.

```
HEADERS      [stream 7] +END_STREAM +END_HEADERS
  :method      = GET
  :path        = /monitor/1G_GIPMorg_n-IrQvqZr6g
  :authority   = push.example.net
```

The push service permits the request to remain outstanding. When a push message is sent by an application server, a server push is associated with the initial request. The response includes the push message.

```
PUSH_PROMISE [stream 7; promised stream 4] +END_HEADERS
  :method      = GET
  :path        = /p/LBhhw0Ooh0-Wl40i971UGsB7sdQGUibx
  :authority   = push.example.net
```

```
HEADERS      [stream 4] +END_HEADERS
  :status     = 200
  date        = Thu, 11 Dec 2014 23:56:55 GMT
  last-modified = Thu, 11 Dec 2014 23:56:55 GMT
  cache-control = private
  content-type = text/plain;charset=utf8
  content-length = 15
```

```
DATA          [stream 4] +END_STREAM
  Hello World!\r\n
```

A user agent can request the contents of the "registration" resource immediately by including a Prefer header field [[RFC7240](#)] with a "wait" parameter set to "0". The push server SHOULD return a link reference to the "registration" resource and expiration information in response to this request. This request also triggers the delivery of all push messages that the push service has stored but not yet delivered.

A user agent can request the last push message for a "subscription" resource by sending GET requests to its URI. A 200 (OK) status response contains the last push message sent to the subscription. A 204 (No Content) status code indicates that no messages are presently available.

[7.](#) Operational Considerations

A push service is likely to have to maintain a very large number of open TCP connections. Effective management of those connections can depend on being able to move connections between server instances.

[7.1.](#) Load Management

A user agent MUST support the 307 (Temporary Redirect) status code [[RFC7231](#)], which can be used by a push service to redistribute load at the time that a new registration is requested.

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](#)]. Alternative services allows for redistribution of load whilst maintaining the same URIs for various resources. User agents can ensure a graceful transition by using the GOAWAY frame once it has established a replacement connection.

[7.2.](#) Push Message Expiration

Push services typically store messages for some time to allow for limited recovery from transient faults. If a push message is stored, but not delivered, the push service can indicate the probable duration of storage by including expiration information in the response to the push request.

A push service is not obligated to store messages indefinitely. If a user agent is not actively monitoring for push messages, those messages can be lost or overridden by newer messages on the same subscription.

Push messages that were stored and not delivered to a client are delivered when a client recommences monitoring. Stored push messages SHOULD include a Last-Modified header field (see [Section 2.2 of RFC7232](#)) indicating when delivery was requested by an application server.

A GET request to a "subscription" resource that has expired messages results in a 204 (No Content) status response, equivalent to if no push message were ever sent.

Push services might need to limit the size and number of stored push messages to avoid overloading. In addition to using the 413 (Payload Too Large) status code for too large push messages, a push service MAY expire push messages prior to any advertised expiration time.

[7.3.](#) Registration and Subscription Expiration

In some cases, it may be necessary to terminate registrations or subscriptions so that they can be refreshed.

A push service might choose to set a fixed expiration time. If a resource has a known expiration time, expiration information is included in responses to requests that create the resource, or in requests that retrieve a representation of the resource.

Expiration is indicated using either the Expires header field, or by setting a "max-age" parameter on a Cache-Control header field (see

[RFC7234]). The Cache-Control header field MUST also include the "private" directive.

A push service can invalidate a registration or subscription at any time. If a user agent has an outstanding request to the "registration" resource (see [Section 6](#)), this can be signaled by returning a 400-series response code, such as 410 (Gone). A push service uses server push to indicate that a subscription has expired; a pushed 400-series status code for the subscription resource signals the termination of a subscription.

A user agent can request that a registration or subscription be removed by sending a DELETE request to the corresponding URI.

[7.4.](#) Implications for Application Reliability

An application developer might find it tempting to create alternative mechanisms for message delivery in case the push service fails to deliver a critical message. Setting up a polling mechanism or a backup messaging channel in order to compensate for these shortcomings negates almost all of the advantages a push service provides.

Applications are encouraged to instead provide a means to detect situations where push messages were not delivered and recover gracefully. For instance, an application server might include a sequence number in push messages; a gap in the sequence can then be used to trigger some form of state resynchronization. For instance, the missing messages might be requested from the application server directly. Push service failures are expected to be rare, therefore performance optimization of any recovery mechanism might be unnecessary.

8. Security Considerations

This protocol MUST use HTTP over TLS [[RFC2818](#)]; this includes any communications between user agent and push service, plus communications between the application and the push service. Thus, all URIs use the "https" scheme. This provides confidentiality and integrity protection for registrations and push messages from external parties.

8.1. Confidentiality from Push Service Access

The protection afforded by TLS does not protect content from the push service. Without additional safeguards, a push service is able to see and modify the content of the messages.

Applications are able to provide additional confidentiality, integrity or authentication mechanisms within the push message itself. The application server sending the push message and the application on the user agent that receives it are frequently just different instances of the same application, so no standardized protocol is needed to establish a proper security context. The process of providing the application server with subscription information provides a convenient medium for key agreement.

The Web Push API codifies this practice by requiring that each push subscription created by the browser be bound to a browser generated encryption key. Pushed messages are authenticated and decrypted by the browser before delivery to applications. This scheme ensures that the push service is unable to examine the contents of push messages.

The public key for a subscription ensures that applications using that subscription can identify messages from unknown sources and discard them. This depends on the public key only being disclosed to entities that are authorized to send messages on the channel. The push server does not require access to this public key.

8.2. Privacy Considerations

Push message confidentiality does not ensure that the identity of who

is communicating and when they are communicating is protected. However, the amount of information that is exposed can be limited.

Subscription URIs MUST NOT provide any basis to correlate communications for a given user agent. It MUST NOT be possible to correlate any two subscription URIs based solely on the content of the subscription URIs. This allows a user agent to control correlation across different applications, or over time.

In particular, user and device information MUST NOT be exposed through the subscription URI.

In addition, subscription URIs established by the same user agent MUST NOT include any information that allows them to be correlated with the associated registration or the user agent. The push service is the only entity that needs to be able to correlate subscriptions with a registration.

Note: This need not be perfect as long as the resulting anonymity set (see [\[RFC6973\], Section 6.1.1](#)) is sufficiently large. A subscription URI necessarily identifies a push service or a single server instance. It is also possible that traffic analysis could be used to correlate subscriptions.

A user agent MUST be able to create new registrations and subscriptions with new identifiers at any time.

[8.3.](#) Authorization

This protocol does not define how a push service establishes whether a user agent is permitted to create a registration or subscription, or whether push messages can be delivered to the user agent. A push service MAY choose to authorize request based on any HTTP-compatible authorization method available, of which there are numerous options. The authorization process and any associated credentials are expected to be configured in the user agent along with the URI for the "push service".

Authorization for sending push messages is managed using capability URLs (see [\[CAP-URI\]](#)). A capability URL grants access to a resource based solely on knowledge of the URL. Capability URLs are used for their "easy onward sharing" and "easy client API" properties. These

make it possible to avoid relying on relationships between push services and application servers, with the protocols necessary to build and support those relationships.

A subscription URI therefore acts as a bearer token: knowledge of the URI implies authorization to send push messages. Subscription URIs MUST be extremely difficult to guess. Encoding a large amount of random entropy (at least 120 bits) in the path component ensures that it is difficult to successfully guess a valid subscription URI.

8.4. Denial of Service Considerations

Discarding unwanted messages at the user agent based on message authentication doesn't protect against a denial of service attack on the user agent. Even a relatively small volume of push messages can cause battery-powered devices to exhaust power reserves. Limiting the number of entities with access to push channels limits the number of entities that can generate value push requests of the push server.

An application can limit where push messages can originate by limiting the distribution of subscription URIs to authorized entities. Ensuring that subscription URIs are hard to guess ensures that only applications servers that have been given a subscription URI can use it.

A malicious application with a valid subscription use the greater resources of a push service to mount a denial of service attack on a user agent. Push service SHOULD limit the rate at which push messages are sent to individual user agents. A push service or user

agent MAY terminate subscriptions ([Section 7.3](#)) that receives too many push messages.

End-to-end confidentiality mechanisms, such as those in [\[API\]](#), prevent an entity with a registration URI from learning the contents of push messages. In both cases, push messages that are not successfully authenticated will not be delivered by the API, but this can present a denial of service risk.

Conversely, a push service is also able to deny service to user agents. Intentional failure to deliver messages is difficult to

distinguish from faults, which might occur due to transient network errors, interruptions in device availability, or genuine service outages.

[8.5.](#) Logging Risks

Server request logs can reveal registration and subscription URIs. Acquiring a registration URI permits the creation of new subscriptions, as well as potentially enabling the receipt of messages. Acquiring a subscription URI permits the sending of push messages. Logging could also reveal relationships between different subscription URIs for the same registration, or between different registrations for the same device.

Limitations on log retention and strong access control mechanisms can ensure that URIs are not learned by unauthorized entities.

[9.](#) IANA Considerations

This document registers three URNs for use in identifying link relation types. These are added to a new "Web Push Identifiers" registry according to the procedures in [Section 4 of \[RFC3553\]](#); the corresponding "push" sub-namespace is entered in the "IETF URN Sub-namespace for Registered Protocol Parameter Identifiers" registry.

The "Web Push Identifiers" registry operates under the IETF Review policy [[RFC5226](#)].

Registry name: Web Push Identifiers

URN Prefix: urn:ietf:params:push

Specification: (this document)

Repository: [Editor/IANA note: please include a link to the final registry location.]

Index value: Values in this registry are URNs or URN prefixes that start with the prefix "urn:ietf:params:push". Each is registered independently.

New registrations in the "Web Push Identifiers" are encouraged to include the following information:

URN: A complete URN or URN prefix.

Description: A summary description.

Specification: A reference to a specification describing the semantics of the URN or URN prefix.

Contact: Email for the person or group making the registration.

Index value: As described in [[RFC3553](#)], URN prefixes that are registered include a description of how the URN is constructed. This is not applicable for specific URNs.

Three values are entered as the initial content of the "Web Push Identifiers" registry.

URN: urn:ietf:params:push

Description: This link relation type is used to identify a web push subscription.

Specification: (this document)

Contact: Martin Thomson (martin.thomson@gmail.com) or the Web Push WG (webpush@ietf.org)

URN: urn:ietf:params:push:reg

Description: This link relation type is used to identify a web push registration.

Specification: (this document)

Contact: Martin Thomson (martin.thomson@gmail.com) or the Web Push WG (webpush@ietf.org)

URN: urn:ietf:params:push:sub

Description: This link relation type is used to identify a resource that can be used to create new web push subscriptions.

Specification: (this document)

Contact: Martin Thomson (martin.thomson@gmail) or the Web Push WG (webpush@ietf.org)

10. Acknowledgements

Significant technical input to this document has been provided by Costin Manolache, Robert Sparks, Mark Nottingham, Matthew Kaufman and many others.

11. References

11.1. Normative References

- [CAP-URI] Tennison, J., "Good Practices for Capability URLs", FPWD capability-urls, February 2014, <<http://www.w3.org/TR/capability-urls/>>.
- [I-D.ietf-httpbis-alt-svc] Nottingham, 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.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2818] Rescorla, E., "HTTP Over TLS", [RFC 2818](#), May 2000.
- [RFC3553] Mealling, M., Masinter, L., Hardie, T., and G. Klyne, "An IETF URN Sub-namespace for Registered Protocol Parameters", [BCP 73](#), [RFC 3553](#), June 2003.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", [BCP 26](#), [RFC 5226](#), May 2008.
- [RFC5988] Nottingham, M., "Web Linking", [RFC 5988](#), October 2010.
- [RFC7230] Fielding, R. and J. Reschke, "Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing", [RFC 7230](#), June 2014.

Internet-Draft

HTTP Web Push

December 2014

- [RFC7231] Fielding, R. and J. Reschke, "Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content", [RFC 7231](#), June 2014.
- [RFC7232] Fielding, R. and J. Reschke, "Hypertext Transfer Protocol (HTTP/1.1): Conditional Requests", [RFC 7232](#), June 2014.
- [RFC7234] Fielding, R., Nottingham, M., and J. Reschke, "Hypertext Transfer Protocol (HTTP/1.1): Caching", [RFC 7234](#), June 2014.
- [RFC7240] Snell, J., "Prefer Header for HTTP", [RFC 7240](#), June 2014.

11.2. Informative References

- [API] Sullivan, B. and E. Fulla, "Web Push API", ED push-api, December 2014, <<https://w3c.github.io/push-api/>>.
- [RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", [RFC 6973](#), July 2013.
- [TS.3GPP.23.040] 3GPP, "Technical realization of the Short Message Service (SMS)", 3GPP TS 23.040 12.2.0, October 2014.

Author's Address

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

Email: martin.thomson@gmail.com

Thomson

Expires June 15, 2015

[Page 17]