INTERNET-DRAFT
Intended Status: Standards Track
Expires: April 21, 2016
Fabio Chiussi

October 19, 2015

Subscription-Less Web Push Framework
draft-chiussi-webpush-subscription-less-framework-01

Abstract

Subscription is a integral part of the current Web Push service. However, the current explicit subscription requirement makes it difficult to use web push in a number of use cases where there may be a need to reach User Agents that are not subscribed to the Web Push service. In addition, the current Web Push subscription model does not provide a mechanism to achieve coordination and control among subscriptions associated to different applications. This document describes a framework for making subscription more flexible and solve these issues.

Table of Contents

1. Introduction

The notion of subscription is one of the pillars of the current Web Push service [I-D. draft-thomson-webpush-protocol][W3CAPI]. There are good reasons to require explicit subscription from the User Agent (UA) [RFC6973]. In fact, by its very nature, Web Push is an invasive service, which requires some form of explicit acceptance by the UA and some form of regulation by the Push Service to make sure that the push capabilities of different applications do not proliferate in such a way that the volume and invasiveness of push traffic becomes uncontrollable.

The Web Push explicit subscription model as it is currently defined, however, has two limitations.

First, the requirement for explicit subscription prevents, or at least makes it awkward and potentially ineffective, the use of Web Push in an increasing number of relevant use cases in which there is a need for a mechanism to reach User Agents in a timely fashion, but it is unknown whether all the User Agents that need to be reached in a certain location have subscribed to any Web Push service. For similar reasons, the explicit subscription requirement limits the usefulness of Web Push in broadcast use cases.

Second, the Web Push explicit subscription model as it is currently defined, does not provide a mechanism to achieve or impose coordination among subscription traffic generated by different applications. As a consequence, the Web Push mechanism is not scalable to a large number of subscriptions to different applications. Although this is intentional in the design at this time, it is recognized as rather crude constraint that does not quite address the control of the total volume of Web Push traffic across applications directed to the same UA, which is ultimately what matters from a UA perspective, and as such it is the most important factor that defines a satisfactory User Experience with Web Push. Some provisions to ameliorate this problem are included in [I-D. draft-thomson-webpush-protocol], but more comprehensive mechanisms for a global management of Web Push traffic across applications still need to be defined to make Web Push as useful and usable as possible.

Both these current limitations point towards the need to define a trusted Web Push Subscription Authority (WPSA), in charge of managing the Web Push subscription traffic "globally," i.e., across all applications, directed to each UA at any given location.

If such a WPSA is trusted, it can also be used to provide a "subscription-less" Web Push to accommodate the additional use cases mentioned above. It should be noted that such a subscription-less Web Push does not mean uncontrolled Web Push. Quite the contrary, the WPSA is in charge of controlling the global Web Push traffic to each UE, and thus can block or throttle offending subscription-less Web Push traffic.

In both cases, the WPSA can apply policies that can be rather elaborate and globally defined, either across applications, or depending on the urgency and criticality of reaching the UA with a Web Push with a certain service in a given situation.

This document defines a Web Push subscription framework using a trusted WPSA to provide both a subscription-less form of Web Push and a mechanism for subscription management that achieves coordination and control among applications.

The two main challenges in the design of WPSA are the definition of trusted authority and the architecture of the WPSA to make it scalable and applicable on a local basis.

1.1. Terminology

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].

This document uses the terminology defined in [I-D. draft-thomson-webpush-protocol].

In addition, this document uses the following terms.

- Web Push Subscription Authority (WPSA): A trusted entity in charge of managing subscriptions across applications and enabling subscription-less Web Push, based on globally defined policies.

2. Subscription-Less Web Push Use Cases

Several emerging popular use cases need a mechanism to push information to a set of UAs. Because of the characteristics of these use cases, it is in general unknown or irrelevant whether the UA has subscribed to any Web Push services.

These use cases require a form of Web Push that does not depend on explicit subscription, which we refer to as "subscription less" Web Push.

At least four such uses cases come to mind.

1. Local Emergency, Alert, and Urgent Notification Services. These may include hyper local services in smart building or smart venues such as amber alert, local emergencies, crowd management, missing person, and other similar services.

   A common characteristic of these service is that the information to be pushed has a very clear "value" to the user and needs to be delivered in a timely fashion. In other words, the information is so valuable and timely for the user that it is clear that the user is willing to receive unsolicited traffic and trade some privacy for the benefit of receiving the information in real time, regardless of subscriptions. These services are rapidly proliferating and user are starting to expect them.

   Clearly, requiring subscription for such services would not be practical. An unsubscribed user may even be unaware that the services exist in a certain area. On the other hand, the very nature of these services demands that when the emergency, alert, or urgent notification arises, all UAs in the area are notified, not just those that have explicitly subscribed.

2. Dormant Mobile Device Presence Determination. With the proliferation of hyper-local location based services that need to be triggered when a user enters or exits a certain geographical area (especially exiting constitutes a challenge), there is a strong need for a lightweight mechanism capable of pinging and waking up a dormant mobile device.

   Ideally, such a mechanism would only involve the mobile browser, and thus be usable without requiring the installation of native clients on the device, be air-interface agnostic and capable to operate over technologies used for indoor positioning, and thus be usable to sense presence with small cells, WiFi, Bluetooth, etc., and operate without requiring paging.

   Web Push potentially fits all these requirements. However, also in this case, in order for Web Push to serve the purpose, a subscription-less flavor of Web Push would be required, since the goal is to sense presence of all UAs, not just those that have subscribed to a Web Push service.

3. Venues with Relaxed User Privacy Expectations. The use of Web Push would be beneficial in Smart Buildings and other environments where the user has a manifested expectation of tapping into available services in an unsolicited fashion. Examples include enterprise smart applications and venues where a provider clearly "owns" the connectivity (e.g., in a store associated to a captive portal) or there is an established expectation by the user that a

provider is making available certain services to all the UAs in that space.

In such scenarios, the user may have an established tolerance or even an expectation for a controlled volume of information being pushed to the device. Also in this case, the notion of subscription is implicit because it is clearly specific to time and location, since the subscription is associated with the duration of the user presence in a certain venue.

These use cases, which are rapidly growing in popularity are an indication that some notion of "subscription-less Web Push," or more precisely Web Push with some form of relaxed subscription requirements, is actually desirable and useful to increase the applicability of Web Push.

4. Broadcast Local Web Push. In general, for services that need to broadcast information using Web Push to an audience in a location, the requirement for explicit subscription often reduces the effectiveness of the service. A controlled Web Push service that does not rely on explicit subscription would be very useful to enable more effective services of this kind.

3. Subscription Management Across Applications

The second limitation of the current Web Push subscription model is the fact that it does not directly provide a mechanism to control the global volume of traffic destined to a given UA. Instead, the current control relies on judiciously distributing subscriptions to applications and potentially placing a limit on the number of applications that can be allowed to use Web Push.

This approach, in addition to curtailing the scalability of Web Push across different applications, may not provide effective control on the quality of the User Experience when Web Push is used.

The quality of experience is largely determined by the global volume and timing of Web Push traffic that a UA receives, rather than what is specifically generated by an individual application. In addition, limiting the number of applications that can use Web Push is recognized as a rather crude way to control the global traffic, since it is not the behavior of a single application that defines the quality of experience, but rather the combined behavior of all the relevant applications in a certain location.

The metric of interest is not simply the volume of traffic, but the timing of the traffic. Since applications are scarcely aware of one another, it is certainly conceivable or even likely that the combination of traffic may be perceived by the user as significantly more intrusive than the traffic from each individual applications.

All these reasons point to the need for an entity in charge of managing Web Push across applications.

4. Web Push Subscription Authority

Both the subscription-less Web Push service and the capability of managing Web Push subscriptions globally across applications are achieved by introducing a trusted WPSA in charge of controlling the traffic generated by Web Push services and destined to each UA.

The WPSA imposes policies that control the combined volume and profile of Web Push traffic destined to each UA and provide coordination among traffic generated by different applications.

By its nature, the WPSA is a "policer" of Web Push traffic across applications. As such, it must provide mechanisms to throttle traffic generated by each application, and when necessary, prioritize traffic from one application versus another.

A first challenge in defining the WPSA is the definition of trust underlying this entity. Trust must be established in two ways. In order to enable subscription-less Web Push, the UAs must trust the WPSA to control the nature and volume of allowed subscription-less traffic. In order to enable global Web Push traffic management, the applications must recognize the WPSA with the authority of prioritizing different applications based on globally-defined policies.

A second challenge in defining the WPSA is the distributed nature of the WPSA function. For example, most the use cases advocating subscription-less Web Push are highly-local in nature. Accordingly, the WPSA needs to be capable to apply policies that are global across applications, but may be location specific. Since local instances of the WPSA may be required for scalability of such a model, a simple mechanism to discover them and associate them with each location is also required.

5. Security Considerations

TBD.

6. IANA Considerations

TBD.

7. References

7.1 Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[I-D. draft-thomson-webpush-protocol] M. Thomson et al., "Generic Event Delivery Using HTTP Push", draft-thomson-webpush-protocol-00 (work in progress), April 2015.

7.2 Informative References

[W3CAPI] Sullivan, B., Fullea, E., and M. van Ouwerkerk, "Web Push API", ED push-api, February 2015, <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.

Authors' Addresses

Fabio Chiussi
Seattle, WA 98116
Email: fabiochiussi@gmail.com

WEBPUSH
Internet-Draft
Intended status: Standards Track
Expires: April 17, 2016

M. Thomson
Mozilla
E. Damaggio
B. Raymor, Ed.
Microsoft

October 15, 2015

Generic Event Delivery Using HTTP Push
draft-ietf-webpush-protocol-01

Abstract

A simple protocol for the delivery of realtime events to user agents is described. This scheme uses HTTP/2 server push.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Many applications on mobile and embedded devices require continuous access to network communications so that real-time events - such as incoming calls or messages - can be delivered (or "pushed") in a timely fashion. These devices typically have limited power reserves, so finding more efficient ways to serve application requirements greatly benefits the application ecosystem.

One significant contributor to power usage is the radio. Radio communications consume a significant portion of the energy budget on a wireless device.

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 create new push message delivery subscriptions, and
- o monitor for new push messages.

Requesting the delivery of events is particularly important for the Web Push API. The subscription, management and monitoring functions are currently fulfilled by proprietary protocols; these are adequate, but do not offer any of the advantages that standardization affords.

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.

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

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

push message receipt: A message delivery confirmation sent from the push service to the application server.

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

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 [RFC7540] is used.

2. Overview

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

```
 +-------+           +--------------+       +-------------+
 |  UA   |           | Push Service |       | Application |
 +-------+           +--------------+       +-------------+
     |                      |                      |
     |      Subscribe       |                      |
     |--------------------->|                      |
     |       Monitor        |                      |
     |<====================>|                      |
     |                      |                      |
     |          Distribute Push Resource           |
     |-------------------------------------------->|
     |                      |                      |
     :                      :                      :
     |                      |     Push Message     |
     |    Push Message      |<---------------------|
     |<---------------------|                      |
     |                      |                      |
```

At the very beginning of the process, a new message subscription is created by the user agent and then distributed to its application server. This subscription is the basis of all future interactions between the actors.

To offer more control for authorization, a message subscription is modeled as two resources with different capabilities:

- A subscription resource is used to receive messages from a subscription and to delete a subscription. It is private to the user agent.
- A push resource is used to send messages to a subscription. It is public and shared by the user agent with its application server.

It is expected that a unique subscription will be distributed 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 has security and privacy implications.

Subscriptions have a limited lifetime. They can also be terminated by either the push service or user agent at any time. User agents and application servers must be prepared to manage changes in subscription state.

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 to create push message subscriptions (see Section 3). A URL for the push service is configured into user agents.

push message subscription: This resource provides read and delete access for a message subscription. A user agent receives push messages (Section 6) using a push message subscription. Every push message subscription has exactly one push resource associated with it.

push: A push resource is used by the application server to request the delivery of a push message (see Section 5). A link relation of type "urn:ietf:params:push" is used to identify a push resource.

push message: A push message resource is created to identify push messages that have been accepted by the push service. The push message resource is also used to acknowledge receipt of a push message.

receipt subscribe: A receipt subscribe resource is used by an application server to create a receipt subscription (see Section 4). A link relation of type "urn:ietf:params:push:receipt" is used to identity a receipt subscribe resource.

receipt subscription: An application server receives delivery confirmations (Section 5.1) for push messages using a receipt subscription.

3. Subscribing for Push Messages

A user agent sends a POST request to its configured push service resource to create a new subscription.

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

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

The push service MUST provide a URI for the push resource corresponding to the push message subscription using a link relation of type "urn:ietf:params:push".

The push service MUST provide a URI for a receipt subscribe resource in a link relation of type "urn:ietf:params:push:receipt".

An application-specific method is used to distribute the push and receipt subscribe URIs to the application server. Confidentiality protection and application server authentication MUST be used to ensure that these URIs are not disclosed to unauthorized recipients (see Section 8.3).

```
HTTP/1.1 201 Created
Date: Thu, 11 Dec 2014 23:56:52 GMT
Link: </p/JzLQ3raZJfFBR0aqvOMsLrt54w4rJUsV>;
        rel="urn:ietf:params:push"
Link: </receipts/xjTG79I3VuptNWS0DsFu4ihT97aE6UQJ>;
        rel="urn:ietf:params:push:receipt"
Location: https://push.example.net/s/LBhhw0OohO-Wl4Oi971UGsB7sdQGUibx
```

4. Subscribing for Push Message Receipts

An application server requests the creation of a receipt subscription by sending a HTTP POST request to the receipt subscribe resource distributed to the application server by a user agent.

```
POST /receipts/xjTG79I3VuptNWS0DsFu4ihT97aE6UQJ HTTP/1.1
Host: push.example.net
```

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

```
HTTP/1.1 201 Created
Date: Thu, 11 Dec 2014 23:56:52 GMT
Location: https://push.example.net/r/3ZtI4YVNBnUUZhuoChl6omUvG4ZM9mpN
```

An application server that sends push messages to a large population of user agents incurs a significant load if it has to monitor a receipt subscription for each user agent. Reuse of receipt subscriptions is critical in reducing load on application servers. A receipt subscription can be used for all resources that have the same receipt subscribe URI.

A push service SHOULD provide the same receipt subscribe URI to all user agents. Application servers SHOULD reuse receipt subscription URIs if the receipt subscribe URI provided with the push resource is identical to the one used to create the receipt subscription. Checking that the receipt subscribe URI is identical allows the application server to avoid creating unnecessary receipt subscriptions.

5. Requesting Push Message Delivery

An application server requests the delivery of a push message by sending a HTTP request to a push resource distributed to the application server by a user agent. The push message is included in the body of the request.

```
POST /p/JzLQ3raZJfFBR0aqvOMsLrt54w4rJUsV HTTP/1.1
Host: push.example.net
Push-Receipt: https://push.example.net/r/3ZtI4YVNBnUUZhuoChl6omUvG4ZM9mpN
Content-Type: text/plain;charset=utf8
Content-Length: 36

iChYuI3jMzt3ir20P8r_jgRR-dSuN182x7iB
```

A 201 (Created) response indicates that the push message was accepted. A URI for the push message resource that was created in response to the request is included in the Location header field. This does not indicate that the message was delivered to the user agent.

```
HTTP/1.1 201 Created
Date: Thu, 11 Dec 2014 23:56:55 GMT
Location: https://push.example.net/d/qDIYHNcfAIPP_5ITvURr-d6BGtYnTRnk
```

A push service MAY generate a 413 (Payload Too Large) status code in response to 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.

5.1. Requesting Push Message Receipts

An application server can use the Push-Receipt header field to request a confirmation from the push service when a push message is delivered and acknowledged by the user agent. The Push-Receipt header field is a URI-Reference as defined in Section 2.7 of [RFC7230].

```
Push-Receipt = URI-reference
```

The application sets the Push-Receipt header field value to a receipt subscription URI. This receipt subscription resource MUST be created from the same receipt subscribe resource which was returned with the push message subscription response (see Section 3).

5.2. Push Message Time-To-Live

A push service can improve the reliability of push message delivery considerably by storing push messages for a period. User agents are often only intermittently connected, and so benefit from having short term message storage at the push service.

Delaying delivery might also be used to batch communication with the user agent, thereby conserving radio resources.

Some push messages are not useful once a certain period of time elapses. Delivery of messages after they have ceased to be relevant is wasteful. For example, if the push message contains a call notification, receiving a message after the caller has abandoned the call is of no value; the application at the user agent is forced to suppress the message so that it does not generate a useless alert.

An application server can use the TTL header field to limit the time that a push message is retained by a push service. The TTL header field contains a value in seconds that describes how long a push message is retained by the push service.

```
TTL = 1*DIGIT
```

Once the Time-To-Live (TTL) period elapses, the push service MUST NOT attempt to deliver the push message to the user agent. A push service might adjust the TTL value to account for time accounting errors in processing. For instance, distributing a push message within a server cluster might accrue errors due to clock skew or propagation delays.

A push service is not obligated to account for time spent by the application server in sending a push message to the push service, or delays incurred while sending a push message to the user agent. An application server needs to account for transit delays in selecting a TTL header field value.

Absence of the TTL header field is interpreted as equivalent to a zero value. A Push message with a zero TTL is immediately delivered if the user agent is available to receive the message. After delivery, the push service is permitted to immediately remove a push message with a zero TTL. This might occur before the user agent acknowledges receipt of the message by performing a HTTP DELETE on the push message resource. Consequently, an application server cannot rely on receiving acknowledgement receipts for zero TTL push messages.

If the user agent is unavailable, a push message with a zero TTL expires and is never delivered.

A push service MAY choose to retain a push message for a shorter duration than that requested. It indicates this by including a TTL header field in the response that includes the actual TTL. This TTL value MUST be less than or equal to the value provided by the application server.

6. Receiving Push Messages

A user agent requests the delivery of new push messages by making a GET request to a push message subscription resource. The push service does not respond to this request, it instead uses HTTP/2 server push [RFC7540] 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 push message resource that was created by the push service when the application server requested message delivery. The response body is the entity body from the most recent request sent to the push resource.

The following example request is made over HTTP/2.

```
HEADERS      [stream 7] +END_STREAM +END_HEADERS
  :method        = GET
  :path          = /s/LBhhw0OohO-Wl4Oi971UGsB7sdQGUibx
  :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          = /d/qDIYHNcfAIPP_5ITvURr-d6BGtYnTRnk
  :authority     = push.example.net

HEADERS      [stream 4] +END_HEADERS
  :status        = 200
  date           = Thu, 11 Dec 2014 23:56:56 GMT
  last-modified  = Thu, 11 Dec 2014 23:56:55 GMT
  cache-control  = private
  content-type   = text/plain;charset=utf8
  content-length = 36

DATA         [stream 4] +END_STREAM
  iChYuI3jMzt3ir20P8r_jgRR-dSuN182x7iB
```

In response to this request, the push service MUST generate a server push for all push messages that have not yet been delivered. In addition, the push service SHOULD return link references to the push and receipt subscribe resources.

A user agent can request the contents of the push message subscription resource immediately by including a Prefer header field [RFC7240] with a "wait" parameter set to "0".

A 204 (No Content) status code with no associated server pushes indicates that no messages are presently available. This could be because push messages have expired.

6.1. Acknowledging Push Messages

To ensure that a push message is properly delivered to the user agent at least once, the user agent MUST acknowledge receipt of the message by performing a HTTP DELETE on the push message resource.

```
DELETE /d/qDIYHNcfAIPP_5ITvURr-d6BGtYnTRnk HTTP/1.1
Host: push.example.net
```

If the push service receives the acknowledgement and the application has requested a delivery receipt, the push service MUST deliver a success response to the application server monitoring the receipt subscription resource.

If the push service does not receive the acknowledgement within a reasonable amount of time, then the message is considered to be not yet delivered. The push service SHOULD continue to retry delivery of the message until its advertised expiration.

The push service MAY cease to retry delivery of the message prior to its advertised expiration due to scenarios such as an unresponsive user agent or operational constraints. If the application has requested a delivery receipt, then the push service MUST deliver a failure response to the application server monitoring the receipt subscription resource.

6.2. Receiving Push Message Receipts

The application server requests the delivery of receipts from the push service by making a HTTP GET request to the receipt subscription resource. The push service does not respond to this request, it instead uses HTTP/2 server push [RFC7540] to send push receipts when messages are acknowledged (Section 6.1) by the user agent.

Each receipt is pushed in response to a synthesized GET request. The GET request is made to the same push message resource that was created by the push service when the application server requested message delivery. A successful response includes a 410 (GONE) status code with no data.

The following example request is made over HTTP/2.

```
HEADERS      [stream 13] +END_STREAM +END_HEADERS
  :method        = GET
  :path          = /r/3ZtI4YVNBnUUZhuoChl6omUvG4ZM9mpN
  :authority     = push.example.net
```

The push service permits the request to remain outstanding. When the user agent acknowledges the message, the push service pushes a delivery receipt to the application server. A 410 (Gone) status code confirms that the message was delivered and acknowledged.

```
PUSH_PROMISE [stream 13; promised stream 82] +END_HEADERS
  :method        = GET
  :path          = /d/qDIYHNcfAIPP_5ITvURr-d6BGtYnTRnk
  :authority     = push.example.net

HEADERS      [stream 4] +END_STREAM
                        +END_HEADERS
  :status        = 410
  date           = Thu, 11 Dec 2014 23:56:56 GMT
```

If the user agent fails to acknowledge the receipt of the push message and the push service ceases to retry delivery of the message prior to its advertised expiration, then the push service MUST push a failure response with a status code of 5XX (TBD).

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 subscription is requested.

A server that wishes to redistribute load can do so using alternative services [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

Storage of push messages based on the TTL header field comprises a potentially significant amount of storage for a push service. A push service is not obligated to store messages indefinitely. A push service is able to indicate how long it intends to retain a message to an application server using the TTL header field (see Section 5.2).

A user agent that does not actively monitor for push messages will not receive messages that expire during that interval.

Push messages that are stored and not delivered to a user agent are delivered when the user agent 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 push message subscription resource that has only expired messages results in response as though no push message were ever sent.

Push services might need to limit the size and number of stored push messages to avoid overloading. To limit the size of messages, the push service MAY return the 413 (Payload Too Large) status code for messages that are too large. To limit the number of stored push messages, the push service MAY either expire messages prior to their advertised Time-To-Live or reduce their advertised Time-To-Live.

7.3. Subscription Expiration

In some cases, it may be necessary to terminate subscriptions so that they can be refreshed. This applies to both push message subscriptions and receipt subscriptions.

A push service can remove a subscription at any time. If a user agent or application server has an outstanding request to a subscription resource (see Section 6), this can be signaled by returning a 400-series status code, such as 410 (Gone).

A user agent or application server can request that a subscription be removed by sending a DELETE request to the push message subscription or receipt subscription URI.

A push service MUST return a 400-series status code, such as 404 (Not Found) or 410 (Gone) if an application server attempts to send a push message to a removed or expired push message subscription.

7.4. Implications for Application Reliability

A push service that does not support reliable delivery over intermittent network connections or failing applications on devices, forces the device to acknowledge receipt directly to the application server, incurring additional power drain in order to establish (usually secure) connections to the individual application servers.

Push message reliability can be important if messages contain information critical to the state of an application. Repairing state can be costly, particularly for devices with limited communications capacity. Knowing that a push message has been correctly received avoids costly retransmissions, polling and state resynchronization.

The availability of push message delivery receipts ensures that the application developer is not tempted 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.

However, reliability might not be necessary for messages that are transient (e.g. an incoming call) or messages that are quickly superceded (e.g. the current number of unread emails).

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. All URIs therefore use the "https" scheme. This provides confidentiality and integrity protection for subscriptions 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 service 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.

The URIs provided for push resources MUST NOT provide any basis to correlate communications for a given user agent. It MUST NOT be possible to correlate any two push resource URIs based solely on their contents. This allows a user agent to control correlation across different applications, or over time.

Similarly, the URIs provided by the push service to identify a push message MUST NOT provide any information that allows for correlation across subscriptions. Push message URIs for the same subscription MAY contain information that would allow correlation with the associated subscription or other push messages for that subscription.

User and device information MUST NOT be exposed through a push or push message URI.

In addition, push URIs established by the same user agent or push message URIs for the same subscription MUST NOT include any information that allows them to be correlated with the user agent.

Note: This need not be perfect as long as the resulting anonymity set (see [RFC6973], Section 6.1.1) is sufficiently large. A push 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 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 subscription, or whether push messages can be delivered to the user agent. A push service MAY choose to authorize requests 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 is managed using capability URLs for the push message subscription, push, and receipt subscription resources (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.

Capability URLs act as bearer tokens. Knowledge of a push message subscription URI implies authorization to either receive push messages or delete the subscription. Knowledge of a push URI implies authorization to send push messages. Knowledge of a push message URI allows for reading and acknowledging that specific message. Knowledge of a receipt subscription URI implies authorization to receive push receipts. Knowledge of a receipt subscribe URI implies authorization to create subscriptions for receipts.

Note that the same receipt subscribe URI could be returned for multiple push message subscriptions. Using the same value for a large number of subscriptions allows application servers to reuse receipt subscriptions, which can provide a significant efficiency advantage. A push service that uses a common receipt subscribe URI loses control over the creation of receipt subscriptions. This can result in a potential exposure to denial of service; stateless resource creation can be used to mitigate the effects of this exposure.

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 capability URL.

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.

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

A malicious application with a valid push URI could use the greater resources of a push service to mount a denial of service attack on a user agent. Push services 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 receive too many push messages.

End-to-end confidentiality mechanisms, such as those in [API], prevent an entity with a valid push message subscription URI from learning the contents of push messages. 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 user agent availability, or genuine service outages.

8.5. Logging Risks

Server request logs can reveal subscription-related URIs. Acquiring a push message subscription URI enables the receipt of messages or deletion of the subscription. Acquiring a push URI permits the sending of push messages. Logging could also reveal relationships between different subscription-related URIs for the same user agent. Encrypted message contents are not revealed to the push service.

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

9. IANA Considerations

This protocol defines new HTTP header fields in Section 9.1. New link relation types are identified using the URNs defined in Section 9.2.

9.1. Header Field Registrations

HTTP header fields are registered within the "Message Headers" registry maintained at <https://www.iana.org/assignments/message-headers/>.

This document defines the following HTTP header fields, so their associated registry entries shall be added according to the permanent registrations below (see [RFC3864]):

Header Field Name	Protocol	Status	Reference
TTL	http	standard	Section 5.2
Push-Receipt	http	standard	Section 5.1

The change controller is: "IETF (iesg@ietf.org) - Internet Engineering Task Force".

9.2. Link Relation URNs

This document registers 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.

These 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 resource for sending push messages.

Specification: (this document)

Contact: The Web Push WG (webpush@ietf.org)

URN: urn:ietf:params:push:receipt

Description: This link relation type is used to identify a resource for creating new push message receipt subscriptions.

Specification: (this document)

Contact: 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-07 (work in progress), May 2015.

[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.

[RFC3864] Klyne, G., Nottingham, M., and J. Mogul, "Registration Procedures for Message Header Fields", BCP 90, RFC 3864, September 2004.

[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.

[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.

[RFC7240] Snell, J., "Prefer Header for HTTP", RFC 7240, June 2014.

[RFC7540] Belshe, M., Peon, R., and M. Thomson, "Hypertext Transfer Protocol Version 2", RFC 7540, May 2015.

11.2. Informative References

[API] Sullivan, B., Fullea, E., and M. van Ouwerkerk, "Web Push API", ED push-api, February 2015, <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.

Appendix A. Change Log

[[The RFC Editor is requested to remove this section at publication.]]

A.1. Since draft-ietf-webpush-protocol-00

Editorial changes for Push Message Time-To-Live

Editorial changes for Push Acknowledgements

Removed subscription expiration based on HTTP cache headers

Authors' Addresses

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

Email: martin.thomson@gmail.com

Elio Damaggio
Microsoft
One Microsoft Way
Redmond, WA 98052
US

Email: elioda@microsoft.com

Brian Raymor (editor)
Microsoft
One Microsoft Way
Redmond, WA 98052
US

Email: brian.raymor@microsoft.com