

Internet Architecture Board (IAB)
Request for Comments: 8752
Category: Informational
ISSN: 2070-1721

M. Thomson
M. Nottingham
March 2020

Report from the IAB Workshop on Exploring Synergy between Content Aggregation and the Publisher Ecosystem (ESCAPE)

Abstract

The Exploring Synergy between Content Aggregation and the Publisher Ecosystem (ESCAPE) Workshop was convened by the Internet Architecture Board (IAB) in July 2019. This report summarizes its significant points of discussion and identifies topics that may warrant further consideration.

Note that this document is a report on the proceedings of the workshop. The views and positions documented in this report are those of the workshop participants and do not necessarily reflect IAB views and positions.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Architecture Board (IAB) and represents information that the IAB has deemed valuable to provide for permanent record. It represents the consensus of the Internet Architecture Board (IAB). Documents approved for publication by the IAB are not candidates for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <https://www.rfc-editor.org/info/rfc8752>.

Copyright Notice

Copyright (c) 2020 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 (<https://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.

Table of Contents

1. Introduction
 - 1.1. Mention of Specific Entities
2. Use Cases
 - 2.1. Instant Navigation
 - 2.2. Offline Content Sharing
 - 2.3. Other Use Cases
 - 2.3.1. Book Publishing
 - 2.3.2. Web Archiving
3. Interactions between Web Publishers and Aggregators
 - 3.1. Incentives for Web Packages

- 3.2. Operational Costs
- 3.3. Content Regulation
- 3.4. Web Performance
- 4. Systemic Effects
 - 4.1. Consolidation
 - 4.1.1. Consolidation of Power in Linking Sites
 - 4.1.2. Consolidation of Power in Publishers
 - 4.1.3. Consolidation of User Preferences
 - 4.2. Effect on Web Security
 - 4.3. Privacy of Content
- 5. AMP Issues Unrelated to Web Packaging
 - 5.1. AMP Governance
 - 5.2. Constraints on the AMP Format
 - 5.3. Performance
 - 5.4. Implementation of Paywalls
- 6. Venues for Future Discussion
- 7. Security Considerations
- 8. Informative References
- Appendix A. About the Workshop
 - A.1. Agenda
 - A.1.1. Thursday 2019-07-18
 - A.1.2. Friday 2019-07-19
 - A.2. Workshop Attendees
- Appendix B. Web Packaging Overview
 - B.1. Authority in HTTPS
 - B.2. Authority in Web Packaging
 - B.3. Applicability
 - B.4. The AMP Format, Google Search Results, and Web Packaging
- IAB Members at the Time of Approval
- Authors' Addresses

1. Introduction

The Internet Architecture Board (IAB) holds occasional workshops designed to consider long-term issues and strategies for the Internet, and to suggest future directions for the Internet architecture. This long-term planning function of the IAB is complementary to the ongoing engineering efforts performed by working groups of the Internet Engineering Task Force (IETF).

The IAB convened the ESCAPE Workshop to examine some proposed changes to the Internet and the Web, and their potential effects on the Internet publishing landscape. Of particular interest was the Web Packaging proposal from Google, under consideration in the IETF, the W3C's Web Incubator Community Group (WICG), and the Web Hypertext Application Technology Working Group (WHATWG).

In considering these proposals, we heard about both positive effects of Web Packaging and concerns that it could have significant effects on the relationship between publishers (e.g., news web sites) and content aggregators (e.g., search engines and social networks). As such, our focus was primarily on this relationship, rather than technical discussion.

Online publishers do not regularly participate in standards activities directly. A workshop format was used to solicit input from them. The workshop had 27 participants from a diverse set of backgrounds, including a small number of attendees from publishers, one aggregator (Google), plus representatives from browsers, the Accelerated Mobile Pages (AMP) community, Content Distribution Networks (CDNs), network operators, academia, and standards bodies. See the workshop call for papers [CFP] for more information and a complete listing of submissions.

As intended, the workshop was primarily a forum for discussion, so it did not reach definite conclusions. Instead, this report is the

primary output of the workshop, as a record of that discussion.

This report documents the use cases discussed in Section 2 and explains the interactions between publishers and aggregators that might be affected by it in Section 3. Appendix A includes more details about the workshop itself. For those unfamiliar with Web Packaging, Appendix B provides a summary as background material.

1.1. Mention of Specific Entities

Participants agreed to conduct the workshop under the Chatham House Rule [CHATHAM-HOUSE], so this report does not attribute statements to individuals or organizations without express permission. Submissions to the workshop were public and thus attributable; they are used here to provide substance and context.

2. Use Cases

Much of the workshop concentrated on discussion of the validity and relative merits of the use cases that might be enabled by Web Packaging. See Appendix B for an overview of Web Packaging.

2.1. Instant Navigation

The largest use of Web Packaging so far is in Google Search, where packages are intended to improve the perceived performance of navigation to pages that are linked from search results when "clicked".

To enable this, when a linking (or referring) web page includes links to pages on another site, it also provides the browser with a packaged copy of the target content, signed by the origin of the target content. In effect, the referring page provides a cache for the target page's content. If navigation to one of those links occurs, having the Web Package gives a browser the assurance that the cache didn't change the content, so it can treat that content as if it were acquired directly from the server for the target page -- even though it came from a different server. In many cases, this results in significantly lower perceived delay in displaying the target page.

A vital characteristic of this technique is that the browser does not contact the target site before navigation. The browser does not make any requests to sites until after navigation occurs, and only then if the site requires additional content or makes a request directly.

Similar improvements could also be realized by downloading content (packaged or otherwise) directly from the target site through a technique called "prefetching". However, doing so would reveal information about the user's activity on the linking page to those sites -- even when the user never actually navigates to it.

| Note: This technique that uses Web Packaging is also referred
| to as "privacy-preserving prefetch". This document avoids that
| term as there was some contention at the workshop about which
| aspects of privacy might be preserved by the technique.

Sites bundled with Web Packaging can additionally be constructed in a way that ensures that they render without needing any additional network access. This makes it possible to provide near-instantaneous navigation. The proposed changes to web navigation in support of loading Web Packages is designed to support this use case.

Workshop participants recognized the value of web performance for usability, as well as for business metrics like retention and bounce rates. Such improvements were seen as a valuable goal, but publishers raised questions about whether they justified the cost of

supporting an additional format, while others raised concerns about different aspects of the Web Packaging proposal.

2.2. Offline Content Sharing

Another primary use case discussed was the ability to share web content between devices where neither has an active connection to the Internet. One of the stated goals of Web Packaging is to enable sharing of content offline.

Several participants reported that in areas where Internet access is expensive, slow, or intermittent, the use of direct peer-to-peer file exchange (e.g., "saving a website and sharing it on a USB stick") is commonplace. Most web browsers already have some affordances for this, but these are recognized as in need of improvements.

In the discussion, several rejected an assumed requirement of this use case -- that there be no difference between the treatment of a "normal" web page and that of one loaded from an offline Web Package.

The ability for a Web Package to provide clear attribution for content was seen as valuable by some participants for a range of reasons. However, reservations were expressed about the subtleties of the properties that signatures provide and the effect of this on web security; see also Sections 4.2 and 2.3.2.

Many participants pointed out that using "unsigned bundles" -- that is, Web Packages without signed exchanges -- could be adequate for this use case, since most users don't need cryptographic proof of the site's identity. However, some expressed concerns that this might worsen the propagation of falsehood.

Some suggested that the value of signed exchanges was not realized in small-scale interpersonal exchange of information but in the building of systems for content delivery that might include capabilities like discovery and automated distribution. The contention here was that effective use of digital signatures in offline distribution of content implied considerably more infrastructure than was described in current proposals.

No definite conclusions about offline sharing were reached during the workshop.

2.3. Other Use Cases

A session on the second morning concentrated on two other significant potential use cases for Web Packages: book publishing and Web archiving. These were not seen as "primary" by the proponents of Web Packaging; the original intent was not to spend significant time on these subjects, but there was considerable interest from attendees.

2.3.1. Book Publishing

The potential application of a packaging format to book publishing was discussed, with particular reference to ways that books differ from web content. Specialists from that industry pointed out that book delivery can vary greatly from typical web content delivery.

Workshop participants briefly explored existing solutions. PDF was seen as particularly challenging for this use case, due to its limitations, and EPUB has constraints that also make it challenging for publishers.

Although Web Packaging might help to address this use case, the question of how to identify book content was not resolved. The use of signed exchanges in this context might offer means of tying

content in books to a website, but several limitations inherent in doing that were identified.

In particular, book publication specialists represented that books don't have the same requirements for timeliness or currency as web pages. For instance, Dave Cramer's submission [CRAMER] observed that Moby Dick was published over 61,000 days ago, which is considerably longer than the proposed limit of 7 days for signed exchanges. The limited length of time that a Web Package can be considered valid was discussed at some length.

Additionally, the risk of a publisher going out of business during the lifetime of a book is significant, because books -- at least successful ones -- often span generations in their applicability. To that end, having a means of attributing content to a publisher was considered less practical and potentially undesirable (much like the discussion above regarding "unsigned bundles").

There were other aspects of book publication that participants saw as challenging for packaging. For example, it is currently not understood what it means to refer to distinct parts of a book. Participants saw this as an area where providing stable references for bundles of content might offer possibilities, but nothing concrete came from that discussion.

The potential for active content in a bundle to use web APIs to enrich content or enable new features was considered valuable. Models for enabling paywalls were discussed at some length (see Section 5.4).

2.3.2. Web Archiving

Web archiving is a complicated discipline that is made more difficult by the complex nature of the Web itself.

From an archival standpoint, the potential for web content to be provided in a self-contained form was viewed positively. Several improvements to the structure of Web Packaging were considered, such as providing complete sets of content and the use of Memento [MEMENTO].

Though there were potential applications of a packaging scheme, many challenges were recognized as requiring additional work on the part of content producers to be fully effective. For example, JavaScript is needed to render some archived content faithfully, but attributing that content to an origin in all scenarios is challenging.

If packaging were to be widely deployed, it might improve the situation for archival replay. In particular, the speculation is that there would be less "live leakage" as packaged content might be less likely to refer to live resources that currently tend to "leak" into views of archives. It was also noted that subresources might also be more likely to be packaged, especially those that are needed for deferred representations (i.e., after JavaScript execution on the page or some user interactions). Other potential applications and enhancements are discussed in [ALAM].

Participants discussed the use of a signature for non-repudiation at some length. In one case related to the Internet Archive, a public figure disputed the accuracy of archived content, asserting that the original content was modified either at the source or in the archive.

Some participants initially saw digital signatures as a way to address such issues of provenance. As similar problems exist in other areas, such as in book publication, medical research, and news, a solution to this problem was considered to have broad

applicability.

However, the discussion ultimately concluded that providing non-repudiation in retrospect is challenging. Signing keys are not expected to remain secure for long periods. If keys are leaked afterwards, an attacker could retroactively generate fraudulent signatures. Alternative solutions were discussed, such as providing independent archives for the same data, using consensus protocols, or using an append-only construct like a Haber-Stornetta log [AOLOG], all of which can be used to increase the difficulty of altering or misrepresenting established archives.

3. Interactions between Web Publishers and Aggregators

A significant motivation for holding the workshop was to provide a forum where publishers could discuss the impact of Web Packaging on the online publishing ecosystem. Of primary interest was whether Web Packages might effectively enable a transfer of power from publishers to aggregators.

Both publishers and aggregators at the workshop expressed the importance of maintaining a positive relationship. Publishers in particular expressed the need to be able to trust that aggregators won't misrepresent their work or de-emphasize it for reasons unrelated to quality and perceived value to the user.

One key question from [BERJON] was discussed:

| Web Packaging has other uses, but it is primarily seen by a large
| proportion of its stakeholders as a solution to problems that AMP
| created. Before we agree to solve those issues, should we not ask
| if AMP was a useful approach in the first place -- and useful to
| whom?

In examining this issue, discussion focused on the current incentive model offered by aggregators. The costs that publishers incur for participation in that system were considered. Considerable time was spent on AMP; a summary of that discussion can be found in Section 5.

We also considered the question of whether standardizing Web Packaging confers credibility to aggregators exercising unwelcome control over publisher content or whether the technical safeguards Web Packaging provides could allow aggregators to relax their restrictions on the kinds of content they're willing to cache and serve. No conclusions were drawn.

3.1. Incentives for Web Packages

Submissions to the workshop indicated that the use of inducements involving better placement and formatting of links to publisher content had a significant effect on the uptake of related technology. For example, in [DEPUYDT-NELSON]:

| [...] The Washington Post has always placed a great deal of trust
| in Google to represent its content--and their reward for doing so
| is more traffic, which positively impacts the business.

During the workshop, several online publishers indicated that if it weren't for the privileged position in the Google Search carousel given to AMP content, they would not publish in that format.

Publishers that do produce AMP said they see a non-trivial increase in traffic as a result of deploying AMP content. For example, Yahoo Japan reported a 60% increase in traffic as a result of deploying AMP on Yahoo Travel [OTSU]. There was no data presented as to whether this increase was due to better placement in Google Search results,

the inherent benefits of the AMP Cache, or the use of the AMP format.

Anecdotal evidence was offered by another large publisher that saw a 10% drop in traffic as a result of accidentally disabling AMP content. However, increases in traffic might not result in similarly proportioned increases in revenue, as observed in [BREWSTER].

3.2. Operational Costs

Several participants pointed out that introducing a new, parallel format for Web content incurs operational costs. In particular, supporting any new format -- such as Web Packaging, Apple News, or Facebook Instant Articles -- requires not only initial development of tooling (some generic and some specific to a site's requirements) but also an ongoing investment in maintaining its operability. Some participants expressed concern about the impact upon small publishers with limited technical and financial resources, especially in the current publishing climate.

Increased exposure from new formats might not always justify the added expense of providing articles in that format [BREWSTER]. However, a standardized format might help publishers reduce the cost of maintaining multiple formats.

3.3. Content Regulation

The use of Web Packaging as a tool for avoiding censorship was not a significant topic of discussion, except to note that publishers often have regulatory requirements regarding removal or correction of content.

Reference was made to the desire to remove videos of a recent shooting [CHRISTCHURCH] and the potential difficulty in doing so if content were available as Web Packages. Legal requirements to remove content come from multiple angles: copyright violations, illegal content, editorial corrections or errors, and right to erasure provisions in the European Union General Data Protection Regulation [GDPR] were mentioned. One participant speculated that making it more difficult to remove material in this way might discourage regulators from censoring content.

In this context, participants observed that it would be difficult to create mechanisms to track and control content served as a Web Package without compromising the stated goal of censorship resistance.

3.4. Web Performance

Understanding the effect that Web Packaging might have on web performance was a matter of some contention.

Some informal analysis from the Google Search deployment was presented (later published in [AMP-PERF]) that showed significant performance improvements in metrics related to navigation time resulting from the combination of prefetch, prerendering, and the AMP format. These results are suggestive of a possibility that Web Packaging could provide some of that improvement on its own, but no data was presented that apportioned the improvement among the three components.

Though data was presented to demonstrate potential rather than be a definitive result, discussions raised a number of questions that suggest the need for further study. Attendees suggested that future measurements consider the effect of signed bundles distinct from the enhancements derived from the AMP format. Future research in this area might also consider the effectiveness of different strategies on

devices with varying capabilities, bandwidth, power consumption requirements, or network conditions.

Of particular interest is the additional work required to fetch and render multiple web pages in preparation for navigation. This might ultimately use fewer connections but comes with an increased network and CPU cost for clients. Some participants pointed out that different clients or applications might require different tuning -- for example, when users have limited (or expensive) bandwidth or for sites with less clear knowledge about the use of outbound links.

Workshop participants also expressed interest in learning about the effect of Web Packages on subsequent navigations within the target site.

In discussion, some participants suggested that their experience supported a theory that operating a cache at the linking site was most effective and the additional work done prior to navigation in terms of fetching and preparing content was what provided the most gains; others suggested that the benefits inherent in the AMP format was a dominant factor.

Understanding the complete effect of Web Packaging on web performance will require further work.

4. Systemic Effects

It is not straightforward to estimate how a proposed technology change might affect all of the parts of a system -- including not only other components, but also things like end-user rights and the balance of power between parties -- ahead of time. To date, when evaluating proposals, the IETF has generally focused on more immediate concerns, such as interoperability and security.

Moreover, people often find new uses for successful standards [SUCCESS] after they are deployed. It is rarely possible to accurately predict all applications of a protocol or format, whether they are harmful or beneficial. Refusing standardization only impedes both outcomes.

With the understanding that predictions are difficult to make, there was considerable speculation at the workshop about the possible effect of Web Packaging on the Web. Some of that speculation is informed by experience, but that experience is necessarily limited in scope. This section attempts to capture that discussion.

4.1. Consolidation

Concerns about the consolidation of power on the Internet have significantly increased lately, as a result of several factors. While the IAB, the Internet Society, and others are examining this phenomenon to understand it better, it is nevertheless prudent to consider whether proposals for changes to how the Internet works favors or counters consolidation. Favoring entities with existing advantages -- like resources, size, or market share -- is not necessarily a factor that disqualifies a new proposal, but it needs to be considered as a cost of enabling that technology.

Although the outcomes of adopting Web Packaging are unclear, the workshop revealed several concerns for consolidation risks for all involved parties: users, publisher sites, linking sites, and services they each rely on.

4.1.1. Consolidation of Power in Linking Sites

Several participants noted that Web Packaging's enabling of instant

navigation (Section 2.1) might advantage larger linking sites -- such as social networks or search engines -- over smaller ones in the same industry because doing so requires careful selections of which links to optimize, so as not to create unneeded traffic.

For example, a news article often has many links, but not all of them are equally likely to be followed. Deciding which ones to prefetch requires considerable data collection and engineering, so this technique might not be feasible for smaller entities. Additionally, some participants noted that this technique favors sites that have a linear set of ranked links, like search results; it is more difficult to apply to a page of news (for example) because predicting what link a user will follow is less obvious.

This technique also requires access to a cache with terms of use compatible with the requirements of the site. It was pointed out that the Google AMP Cache has policies that might be acceptable to many, and there are other caches. Sites operated by entities other than Google already use this cache, though it was observed that a site that does not host its own cache suffers a minor performance degradation.

4.1.2. Consolidation of Power in Publishers

Participants seemed to agree that if performance is a strong enough differentiator, the effective use of Web Packaging might turn out to be a condition for success for online publishers. Google Search's choice to privilege content that is served using HTTPS was pointed out as showing that this sort of influence can be effective. Equally, it is not necessarily the case that standardization of new capabilities will affect such policies materially, as noted in [YASSKIN]:

| It seems unlikely that any decisions we make in a packaging or
| distribution system will affect the considerations aggregators use
| when deciding how to rank recommendations or the power this gives
| them over publishers.

The most common concern raised in the discussion was the effect of this technology on smaller publishers who might be less able to optimize the packages they produce, where their primary differentiation in the market has previously been the quality of their content.

4.1.3. Consolidation of User Preferences

In typical operation of the Web, servers have an opportunity to tailor content to the needs of their users. In contrast, a static Web Package has few options for individualization, as the content is generated once and used by many.

As a result, publishers noted that AMP provides less opportunity to customize content for their customers. Their concerns included not only personalizing content based on what they know about the user but also optimizing the package for specific browsers. Other participants observed in relation to this that Web Packaging might also have a consolidating effect in the browser market.

Some participants brought up the possibility of customization by providing multiple packages, including multiple variants of resources in a single package, or performing customization after the package was loaded. However, other participants pointed out that all of these options have negative side effects, either in complexity or reduced performance arising from larger bundles or delayed customization.

4.2. Effect on Web Security

One session explored the impact of introducing a new security model for the Web. Currently, sites rely on connection-oriented security (provided by TLS [TLS]), but Web Packaging adds a limited form of object security. That is, the package protects the integrity of a message, rather than providing integrity and confidentiality for its delivery. Object security is not a new concept in the context of the Web; designs like SHTTP [SHTTP] are as old as HTTPS. Though the intent is for Web Packaging to have a far more narrow applicability, it provides fewer security guarantees than HTTPS, since it provides only authentication, no confidentiality with respect to the cache, and no assurance of liveness.

Object-based security -- such as proposed in Web Packaging -- allows the use of content regardless of how it is obtained; some participants noted that third parties gain greater control over the distribution of content, reducing the ability of publishers to retract or alter content over the validity period of signed content.

Another topic of discussion was composition attacks. In its proposed form, Web Packaging only provides authentication of independent resources, not a web page as a single unit, allowing an attacker to control the composition of resources. This weakness was acknowledged as a known shortcoming of the current proposal that would be addressed.

The issue of managing the trade-off between control and performance in caches arose. While participants recognized that problems with resource composition already occur by accident -- for example, when a cache stores different versions of resources -- Web Packaging allows an attacker more direct control over what resources are available to clients.

For example, an attacker might be able to cause content with a security flaw to be used up to a week past the time that the defect was fixed.

As an example of how Web Packaging might change the risk profile for sites, participants discussed recovery from cross-site scripting attacks. It is already the case that a brief exposure to this class of attack can result in an attacker gaining persistent access, but mechanisms exist that can be used to avoid or correct issues, like cache validation and Clear Site Data [CLEAR-DATA]. These measures are not available to clients unless they connect to the site.

The discussion pointed out that these concerns are not new or uniquely enabled by Web Packaging. However, it was pointed out that new features are routinely subject to higher security and privacy expectations. In an example unrelated to Web Packaging but with similar trade-offs, shared compression of multiple resources has significant performance benefits. The risk with shared compression is the potential for exposing encrypted information through side channels. Though sites can use shared compression without this exposure, shared compression will likely only be enabled once it is clear that measures to prevent accidental information exposure are understood to be effective in a broad set of deployments.

The discussion also addressed the question of whether concerns might equally apply to the typical use of a CDN as a third-party provider of the content. Some participants concluded that CDNs are typically in a contractual relationship with the sites they serve and so are more likely to have their interests aligned.

4.3. Privacy of Content

Discussion and submissions raised concerns regarding how serving content using Web Packages might adversely affect privacy of individuals. There are challenges here, but the very narrow applicability of Web Packaging to what is effectively static content limits the privacy risk. The conclusion was that, provided sufficient care is taken in implementation, the use of Web Packages does not substantially increase the information that an aggregator gains about what content is consumed.

Concretely, an aggregator knows what content it serves in anticipation of navigation. This is -- at least in theory -- substantially the same as the content that the aggregator might receive if it performed the navigation itself. Assuming that content is stripped of personalization, the aggregator gains no new information.

5. AMP Issues Unrelated to Web Packaging

On multiple occasions, discussion at the workshop concentrated on problems that arise as a result of constraints on the AMP format or details of its inclusion in Google Search. For instance, the requirement to make pages expose their metadata is unlikely to be affected by any standardization of a packaging format as that requirement is independent of the process of delivering content.

This section provides some detail on aspects of the discussion that touched on AMP more generally in this way. Some treatment of these points is considered relevant as some of the discussion at the workshop, even under the remit of discussing Web Packaging, concentrated on the effect of AMP on the ecosystem.

Note: Of the four formats mentioned in the workshop call for papers [CFP], only AMP sent representatives to the workshop. The discussion was therefore concentrated around AMP; this section should not be read to imply anything about other formats.

Discussion and submissions referred to a commitment [AMP-LESSONS] to allow publishers to use content that met specific criteria to access privileged positions in search results, regardless of their adoption of AMP. Participants felt that this approach might address some of these concerns if it were adopted and durable. For instance, the use of Web Packaging might be sufficient to remove some constraints on active content on the basis that the active content would be attributed to the publisher and not the AMP Cache.

5.1. AMP Governance

There was interest from workshop participants in the governance model used for AMP. In particular, the question of how independent the AMP project would be of Google and Google Search arose.

Three of the seven members of the AMP Technical Steering Committee, the body that governs AMP, are Google employees, which gives Google considerable influence over the project. It was asserted that the governance structure was intended to be more independent of Google over time. The understanding was that any consumer of the format, such as Google Search, would make an independent assessment about whether to use or require different aspects of the AMP project products.

5.2. Constraints on the AMP Format

Sites often implement AMP by creating a separate set of content in parallel to their regular HTML content. Publishers noted this as a high cost, particularly for smaller sites. It was pointed out that

websites can serve AMP-compliant content exclusively. However, several publishers referred to limitations in the format that made it unsuitable for their needs.

Many cited reasons for this duplication were related to the necessity of running arbitrary active content (typically, JavaScript). For example:

- * AMP provides a framework for supporting user authentication, but publishers asserted that using this framework was not considered practical.
- * AMP content does not support rendering of certain content, which can affect the ability of publishers to innovate content production.
- * The AMP model for the implementation of paywalls (Section 5.4) was claimed to be inimical to some publisher business models.

More broadly, they considered AMP's constraints on the use of active content as problematic, since they prevent the use of capabilities that are provided on equivalent non-AMP pages. Reference was made to a proposed `<amp-script>` element -- which has since been made fully available -- that seeks to provide limited access to some dynamic content.

5.3. Performance

Publishers observed that using the AMP format does not provide any guarantee of performance gains and, in some cases, could contribute to performance degradation. It was suggested that this was most problematic for sites that are already well-tuned for performance.

5.4. Implementation of Paywalls

The use of paywalls by web publishers to control access to content in return for payment is increasingly common. One popular approach is to offer a limited number of articles without payment while insisting on a paid subscription to access further articles.

On several occasions, participants expressed dissatisfaction with the difficulty of integrating paywall authorization when using AMP. In particular, they said AMP encourages publishers to include an article's full content, hidden by default but easily accessible to motivated users. The discussion extended to workarounds like cookie syncing [COOKIE-SYNC], which is used as part of authorization and is a consequence of having cached content hosted on the linking site rather than the target site.

The same topic came up concerning book publication, where publishers indicated that having a means of enabling different methods of distribution without also facilitating unconstrained copying of book content was necessary.

This conflation of AMP issues with those addressed by Web Packaging was recurrent in the discussion. As observed in [DAS], these concerns might be addressed by linking to a signed bundle.

6. Venues for Future Discussion

Web Packaging work continues in multiple forums. Questions about the core format and signatures are being discussed on the `wpack@ietf.org` mailing list (<https://www.ietf.org/mailman/listinfo/wpack>). Changes to web browsers as proposed in [LOADING] will be discussed on the Fetch specification repository (<https://github.com/whatwg/fetch/issues/784>).

7. Security Considerations

Proposals discussed at the workshop might have a significant security impact, and these topics were discussed in some depth; see Section 4.2.

8. Informative References

[ALAM] Alam, S., Weigle, M., Nelson, M., Klein, M., and H. Van de Sompel, "Supporting Web Archiving via Web Packaging", 6 June 2019, <<https://www.iab.org/wp-content/IAB-uploads/2019/06/sawood-alam-2.pdf>>.

[AMP-LESSONS] Ubl, M., "Standardizing lessons learned from AMP", 8 March 2018, <<https://blog.amp.dev/2018/03/08/standardizing-lessons-learned-from-amp/>>.

[AMP-PERF] Steinlauf, E., "The Speed Benefit of AMP Prerendering", 14 August 2019, <<https://developers.googleblog.com/2019/08/the-speed-benefit-of-amp-prerendering.html>>.

[AOLOG] Haber, S. and W. Stornetta, "How to time-stamp a digital document", Journal of Cryptology, Vol. 3, Issue 2, pp. 99-111, DOI 10.1007/bf00196791, 1991, <<https://doi.org/10.1007/bf00196791>>.

[BERJON] Berjon, R., "ESCAPE: The New York Times Position", 9 July 2019, <<https://www.iab.org/wp-content/IAB-uploads/2019/07/NYT-ESCAPE.pdf>>.

[BREWSTER] Brewster, A., "ESCAPE Position / Patch.com", 6 June 2019, <<https://www.iab.org/wp-content/IAB-uploads/2019/06/patch.pdf>>.

[BUNDLE] Yasskin, J., "Bundled HTTP Exchanges", Work in Progress, Internet-Draft, draft-yasskin-wpack-bundled-exchanges-02, 26 September 2019, <<https://tools.ietf.org/html/draft-yasskin-wpack-bundled-exchanges-02>>.

[CFP] Internet Architecture Board, "Exploring Synergy between Content Aggregation and the Publisher Ecosystem Workshop 2019", 3 May 2019, <<https://www.iab.org/activities/workshops/escape-workshop/>>.

[CHATHAM-HOUSE] Chatham House, "Chatham House Rule", <<https://www.chathamhouse.org/chatham-house-rule>>.

[CHRISTCHURCH] Stevenson, R. and J. Anthony, "'Thousands' of Christchurch shootings videos removed from YouTube, Google says", 16 March 2019, <<https://www.stuff.co.nz/business/111330323/facebook-working-around-the-clock-to-block-christchurch-shootings-video>>.

[CLEAR-DATA] West, M., "Clear Site Data", W3C Working Draft, 30 November 2017, <<https://www.w3.org/TR/clear-site-data/>>.

[COOKIE-SYNC] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., and C. Diaz, "The Web Never Forgets", CSS '14: Proceedings of the 2014 ACM SIGSAC Conference on

- Computer and Communications Security, pp. 674-689,
DOI 10.1145/2660267.2660347, 2014,
<<https://doi.org/10.1145/2660267.2660347>>.
- [CRAMER] Cramer, D., "Packaging Books", 2 June 2019,
<<https://www.iab.org/wp-content/IAB-uploads/2019/06/cramer-position-paper.pdf>>.
- [DAS] Das, S., "The Implication of Signed Exchanges on
E-Commerce", 7 June 2019, <https://www.iab.org/wp-content/IAB-uploads/2019/06/IAB-Position-Paper_-Signed-Exchanges.pdf>.
- [DEPUYDT-NELSON] DePuydt, M. and M. Nelson, "Signed Exchanges and The
Importance of Trust in Aggregator/Publisher
relationships", 4 June 2019, <<https://www.iab.org/wp-content/IAB-uploads/2019/06/washpost.pdf>>.
- [GDPR] European Union, "General Data Protection Regulation", EU
Regulation 2016/679, 27 April 2016, <<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e2606-1-1>>.
- [HTTP] Fielding, R., Ed. and J. Reschke, Ed., "Hypertext Transfer
Protocol (HTTP/1.1): Message Syntax and Routing",
RFC 7230, DOI 10.17487/RFC7230, June 2014,
<<https://www.rfc-editor.org/info/rfc7230>>.
- [LOADING] Yasskin, J., "Loading Signed Exchanges", 4 September 2019,
<<https://wicg.github.io/webpackage/loading.html>>.
- [MEMENTO] Van de Sompel, H., Nelson, M., and R. Sanderson, "HTTP
Framework for Time-Based Access to Resource States --
Memento", RFC 7089, DOI 10.17487/RFC7089, December 2013,
<<https://www.rfc-editor.org/info/rfc7089>>.
- [ORIGIN] Barth, A., "The Web Origin Concept", RFC 6454,
DOI 10.17487/RFC6454, December 2011,
<<https://www.rfc-editor.org/info/rfc6454>>.
- [OTSU] Ohtsu, S., "Deployment Experience of Signed HTTP Exchanges
with AMP as a Publisher", 4 June 2019,
<<https://www.iab.org/wp-content/IAB-uploads/2019/06/shigeki-ohtsu.pdf>>.
- [SHTTP] Rescorla, E. and A. Schiffman, "The Secure HyperText
Transfer Protocol", RFC 2660, DOI 10.17487/RFC2660, August
1999, <<https://www.rfc-editor.org/info/rfc2660>>.
- [SUCCESS] Thaler, D. and B. Aboba, "What Makes for a Successful
Protocol?", RFC 5218, DOI 10.17487/RFC5218, July 2008,
<<https://www.rfc-editor.org/info/rfc5218>>.
- [SXG] Yasskin, J., "Signed HTTP Exchanges", Work in Progress,
Internet-Draft, draft-yasskin-http-origin-signed-
responses-08, 4 November 2019,
<<https://tools.ietf.org/html/draft-yasskin-http-origin-signed-responses-08>>.
- [TAG-DC] Betts, A., Ed., "Distributed and syndicated content", W3C
TAG Finding, 27 July 2017,
<<https://www.w3.org/2001/tag/doc/distributed-content/>>.
- [TLS] Rescorla, E., "The Transport Layer Security (TLS) Protocol
Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018,

<<https://www.rfc-editor.org/info/rfc8446>>.

[YASSKIN] Yasskin, J., "Chrome's position on the ESCAPE workshop", 6 June 2019, <<https://www.iab.org/wp-content/IAB-uploads/2019/06/chrome.html>>.

Appendix A. About the Workshop

The ESCAPE Workshop was held on 2019-07-18 and the morning of 2019-07-19 at Cisco's facility in Herndon, Virginia, USA.

Workshop attendees were asked to submit position papers. These papers are published on the IAB website [CFP].

The workshop was conducted under the Chatham House Rule [CHATHAM-HOUSE], meaning that statements cannot be attributed to individuals or organizations without explicit authorization.

A.1. Agenda

This section outlines the broad areas of discussion on each day.

A.1.1. Thursday 2019-07-18

Web Packaging Overview: A technical summary of Web Packaging was provided, plus a longer discussion of a range of use cases.

Web Packaging and Aggregators: The use of Web Packaging from the perspective of a content aggregator was given.

Web Packaging and Publishers: After a break, presentations from web publishers talked about the benefits and costs of Web Packaging. This included some discussion of the effect of developing AMP-conformant versions of content from a publisher perspective.

Web Packaging and Security: This session concentrated on how the Web Packaging proposal might affect the web security model.

Alternatives to Web Packaging: This session looked at alternative technologies, including those that were attempted in the past and some more recent ideas for addressing the use case of making web navigations more performant.

A.1.2. Friday 2019-07-19

Web Archival: This session talked about the potential application of a technology like Web Packaging in addressing some of the myriad problems faced by web archival systems.

Book Publishing: The effect of technologies for bundling and distribution of books was discussed.

Conclusions: A wrap-up session attempted to capture key takeaways from the workshop.

A.2. Workshop Attendees

Attendees of the workshop are listed with their primary affiliation as it appeared in submissions. Attendees from the program committee (PC), the Internet Architecture Board (IAB), and the Internet Engineering Steering Group (IESG) are also marked.

- * Sawood Alam, Old Dominion University
- * Jari Arkko, Ericsson (IAB)
- * Richard Barnes, Cisco
- * Robin Berjon, New York Times (PC)

- * Zack Bloom, Cloudflare
- * Abraham Brewster, Patch.com
- * Alissa Cooper, Cisco (IESG, IAB)
- * Dave Cramer, Hachette Book Group
- * Melissa DePuydt, Washington Post
- * Levi Durfee, AMP Advisory Committee
- * Rudy Galfi, Google
- * Joseph Lorenzo Hall, Center for Democracy & Technology (PC)
- * Matthew Nelson, Washington Post
- * Michael Nelson, Old Dominion University
- * Mark Nottingham, Fastly (IAB, PC)
- * Shigeki Ohtsu, Yahoo
- * Eric Rescorla, Mozilla
- * Adam Roach, Mozilla (IESG)
- * Rich Salz, Akamai Technologies
- * Wendy Seltzer, W3C
- * David Strauss, Pantheon (PC)
- * Chi-Jiun Su, Hughes
- * Ralph Swick, W3C
- * Martin Thomson, Mozilla (IAB, PC)
- * Jeffrey Yasskin, Google
- * Dan York, Internet Society
- * Benjamin Young, John Wiley & Sons

Appendix B. Web Packaging Overview

Web Packaging is comprised of two separate technologies: resource bundling [BUNDLE] and signed exchanges [SXG].

In both the submissions and workshop discussion, the most controversial aspect of the technology is the use of signed exchanges as an alternative means of providing authority over a particular resource, for a few different reasons.

This appendix explains how authority works on the Web and how Web Packaging proposes to change that.

B.1. Authority in HTTPS

The Web currently uses HTTPS [HTTP] to establish a server's authority -- that is, to give an assurance that the content came from where the URL implies. The combination of URI scheme (https), domain name (or host), and port number are formed into a single identifier, the origin [ORIGIN] to which content is attributed.

Web browsers use the certificate offered as part of a TLS connection [TLS] to servers in determining whether a server is authoritative for that origin; see [ORIGIN] and Section 9.1 of [HTTP]. Content is attributed to a given URL only if it is received from a connection to a server that is authoritative for the associated origin.

As an example, a web browser seeking to load "https://example.com/index.html" makes a TLS connection to a server. As part of the TLS connection establishment, the server offers a certificate for the name "example.com". If the browser accepts the certificate, it will then make requests for URLs on the "https://example.com" origin on that connection and consider any answers from the server to be authoritative.

This notion of authority is a crucial property of web security: only content that is attributed to the same web origin can access all information in that origin, including the content of most resources as well as state associated with the origin, such as cookies. This separation ensures that sites can keep secrets from each other, even when they are both loaded in the same browser.

B.2. Authority in Web Packaging

Web Packaging, through the use of signed exchanges, aims to provide an alternative means of establishing authority. A signed exchange is an expression of an HTTP request and response (an exchange) with certain information stripped and a digital signature applied.

The signature is made with a similar certificate to the one a server might offer in HTTPS -- that certificate can also be used for HTTPS -- but it includes a special attribute that denotes its suitability for signed exchanges.

A web browser that has been provided with a signed exchange can verify the signature and, if the signature is valid and the certificate is acceptable, use the content from the signed exchange. Critically, the web browser does not make an HTTPS connection to a server to get the content or to verify the signature.

In effect, Web Packaging moves from a model where authority is derived from the delivery method (i.e., TLS) to an object security model, where authority is derived from a signature on objects. In doing so, it aims to render the means of delivery irrelevant to determinations of security.

B.3. Applicability

Web Packaging does not claim to supplant the authority model of the Web completely, but it does provide an alternative that might be used under certain narrow conditions. In particular, Web Packaging is intended for use with content that is not secret from an entity that is aware of the existence of that content.

In aid of this goal, Web Packaging does not include information from exchanges that is related to the process of acquiring content nor does it include any information that is related to individual requests. For instance, use of the Set-Cookie header field is expressly forbidden, as it often contains information that is related to a particular user.

B.4. The AMP Format, Google Search Results, and Web Packaging

The relationship between the AMP Project <<https://amp.dev/>> and Web Packaging is complicated. The AMP Project, sponsored by Google, establishes a profile of HTML with a stated goal of providing support for the best practices for the format, with a strong emphasis on performance. The format tightly constrains the use of HTML features but also offers a library of components that provide sanitized implementations of many commonly used capabilities.

The connection to Web Packaging is bound up in the way that Google Search treats AMP content specially. AMP content provides two properties that Google Search exploits: metadata exposure and static analysis of active content.

AMP content provides metadata in a form that can be reliably extracted, using the microformats defined by the Schema.org project <<https://schema.org/>>. This aspect of AMP has no effect on the discussion, except to the extent that this relates to Google Search and their use of this metadata in populating the carousel.

Constrained use of active content -- such as JavaScript -- in AMP makes it possible to analyze content to verify that actions taken are narrowly limited. This static analysis assures that AMP content can be served without affecting other content on the same site. For Google Search, this is what enables the loading of AMP content alongside search content and other AMP resources.

To provide preloading, Google operates the Google AMP Cache <<https://developers.google.com/amp/cache/>>, from which AMP content is served. As a consequence, browsers attribute the content to the origin [ORIGIN] of the AMP Cache and not the publisher, creating some confusion about how content is attributed, as discussed in the W3C finding on distributed content [TAG-DC].

An important goal of Web Packaging is to attribute content loaded from a cache, such as the Google AMP Cache, to the publisher that created that content. For more on this, see Section 2.1.

IAB Members at the Time of Approval

Internet Architecture Board members at the time this document was approved for publication were:

Jari Arkko
Alissa Cooper
Stephen Farrell
Wes Hardaker
Ted Hardie
Christian Huitema
Zhenbin Li
Erik Nordmark
Mark Nottingham
Melinda Shore
Jeff Tantsura
Martin Thomson
Brian Trammell

Authors' Addresses

Martin Thomson

Email: mt@lowentropy.net

Mark Nottingham

Email: mnot@mnot.net