H3 Priorities

Sorry

Sorry

- Google argued strongly for existing H2 priorities
  - Now default disables them...

Overview of H2, H3, #2700

What are H2 priorities really?

Strict prioritization is implicit in the tree structure

Weights to share bandwidth between nodes

What are H2 priorities really?

Strict prioritization is implicit in the tree structure
Weights to share bandwidth between nodes

Conceptually, this is a very clean model, but maybe the toolbox is a bit too large?

H3 Priorities: basically keep the toolbox

Add placeholders

- Though potentially allowing 0 (or “too few”) is an issue (1,2)

Remove ‘exclusive’ prioritization

- Necessity due to avoiding intra-stream HOL blocking
- Is a solvable(but unsolved) problem (3,4)
  (assuming enough placeholders (1,2) or ~#2700)

1 https://github.com/quicwg/base-drafts/issues/2734
2 https://github.com/quicwg/base-drafts/issues/2753
3 https://github.com/quicwg/base-drafts/issues/2502#issuecomment-491246513
4 https://github.com/quicwg/base-drafts/pull/2723

What do we really need from the toolbox (browser use case)?

- Most things need to be FIFO link
  - Scripts, CSS, fonts, … need to be fully downloaded to use
  - Yet, H2 and H3 make FIFO difficult (inter-stream dependencies)

- Only some things can be Round-Robin link
  - Progressive images, video, HTML (+-)
  - Yet this is the default in H2 and H3

- So: we have several proposals that basically try to make FIFO easier

For example: A bit of sausage: PR #2700

Removes streams depending upon streams and adds a 1 byte exclusive priority

Not a new idea: Can think of it as:

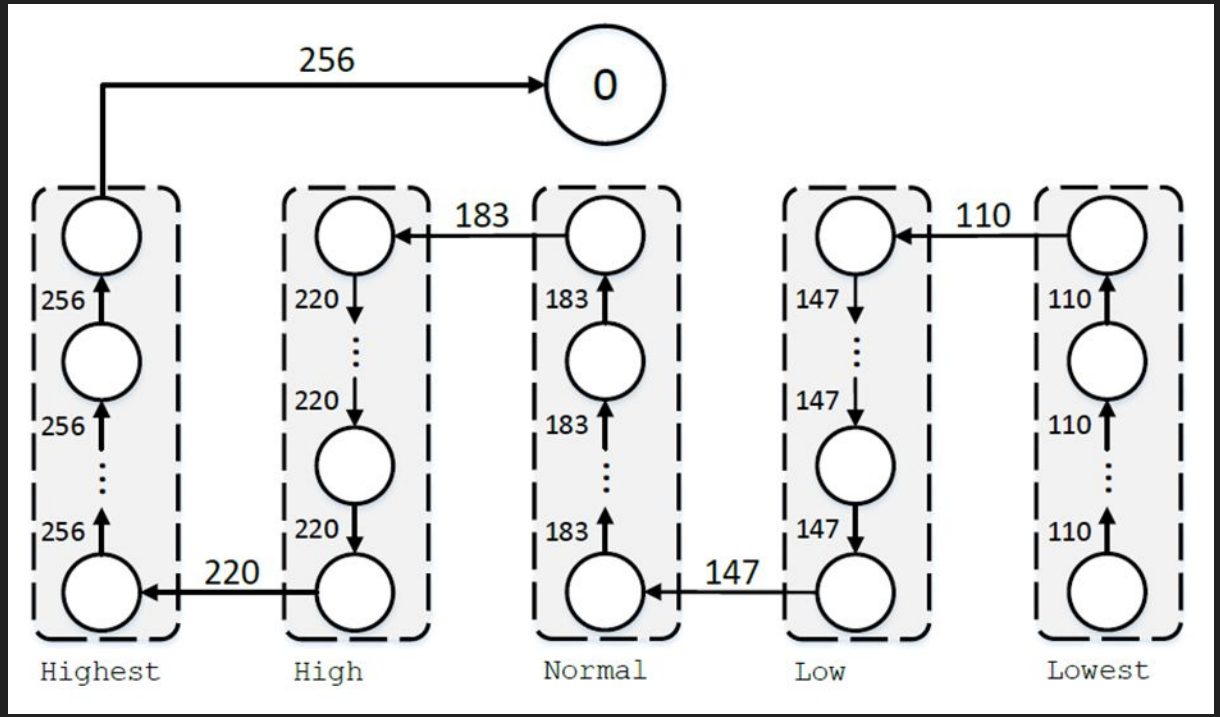
~Patrick Meenan's proposal + existing H3 placeholders

A variant of Osama Mazahir's 2014 proposal

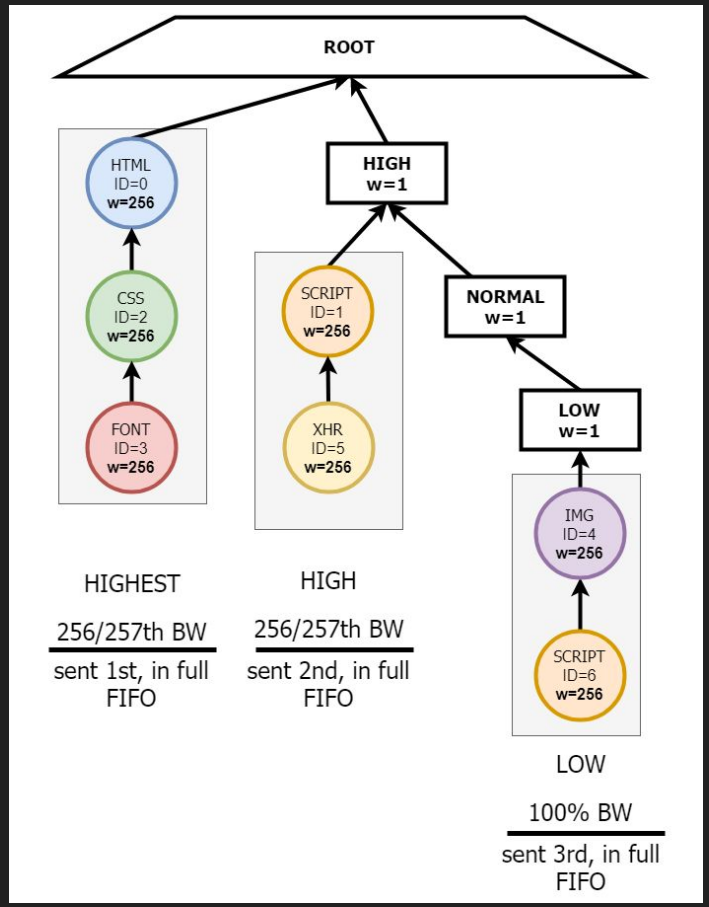
Almost everyone agrees it could be made simpler, it's a matter of what to remove

Fixes #2502 and adds back an equivalent to exclusive prioritization(#1865)

HTTP/2 original (Chrome) : First-Come-First-Served



Converting to H3: in draft-20 (Chrome): do-able but weird

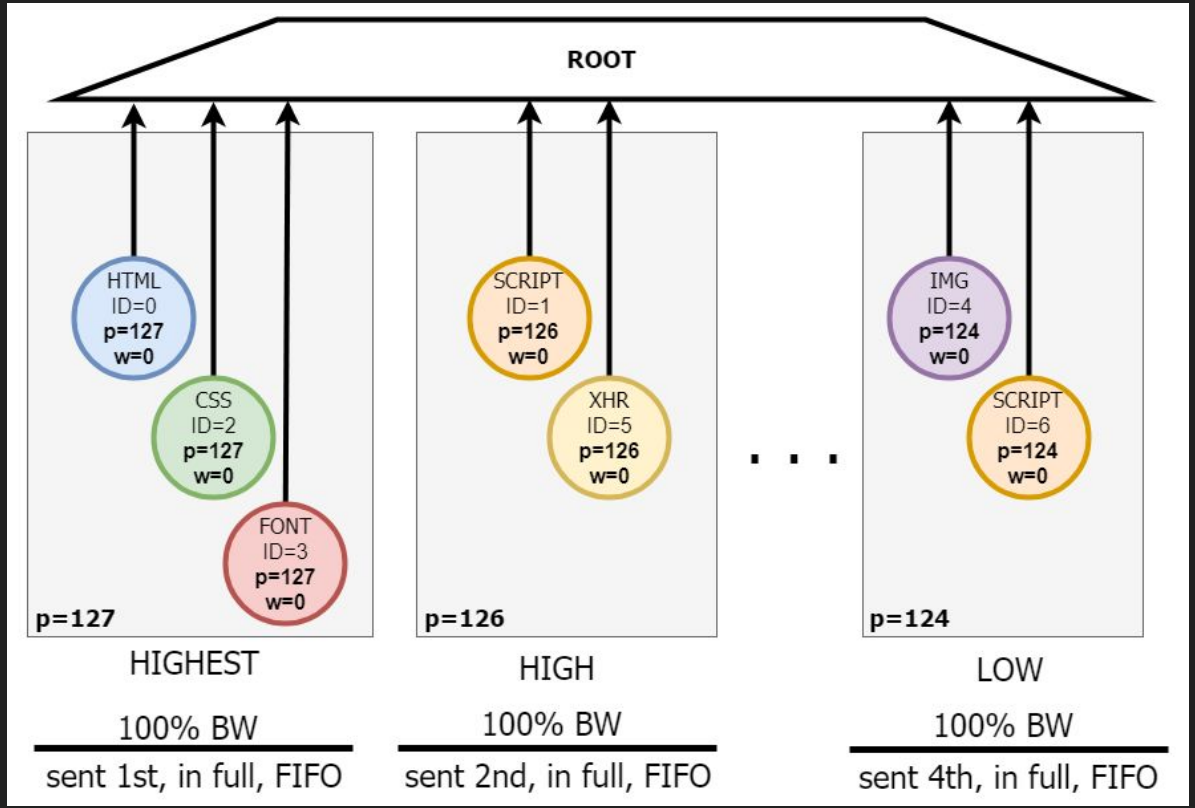


Each priority level = 1 placeholder

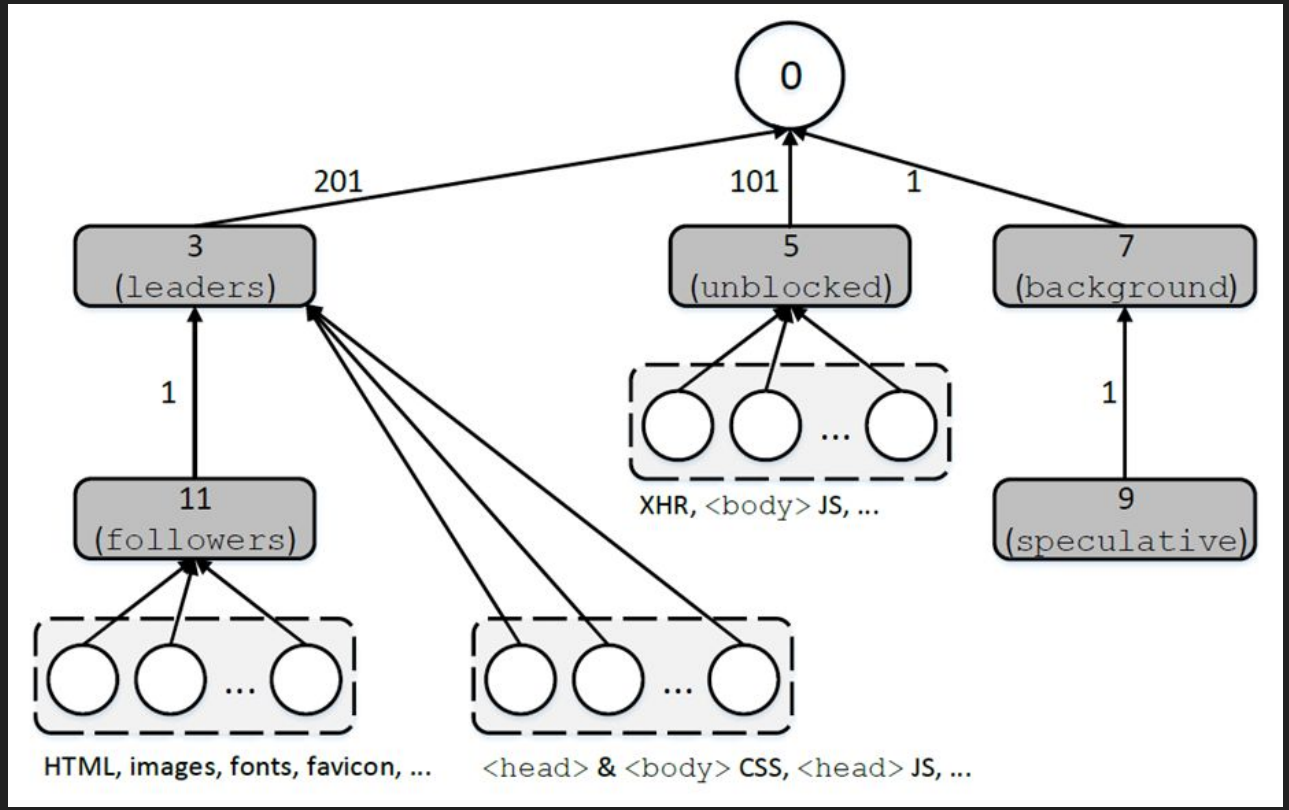
Placeholders' minimal weight = 1
- They get 1/257th of BW
- Potential solution: zero-weighting (#2723)

OR stop sending priority on request streams

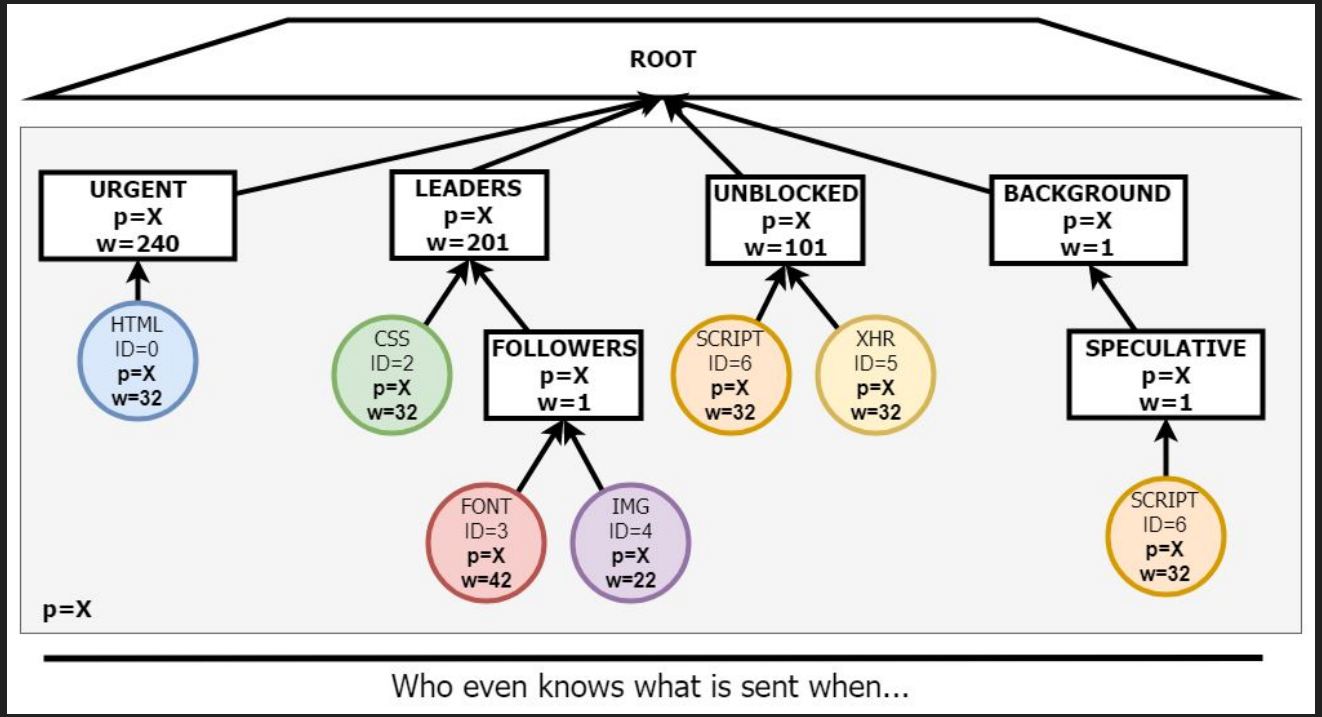
Converting to H3: in #2700 (Chrome): easy



HTTP/2 original (Firefox) : “placeholders” avant-la-lettre

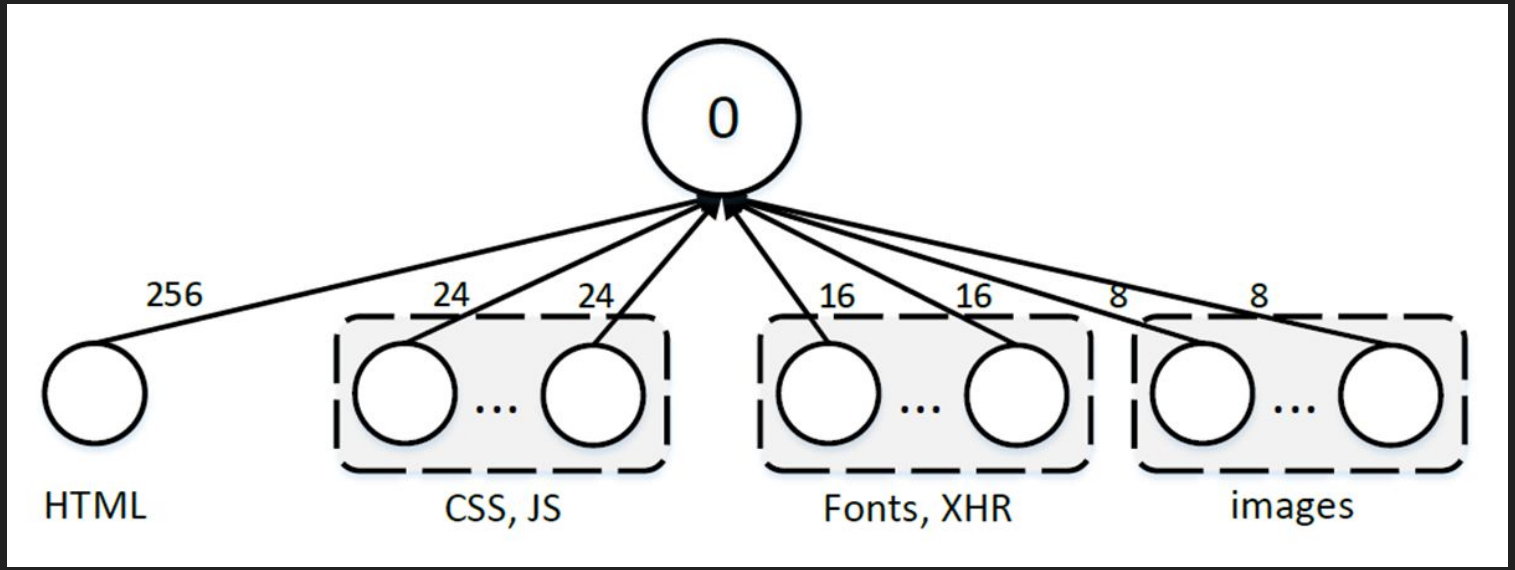


Converting to H3 (draft-20 and 2700): Easy case (Firefox) - No op*

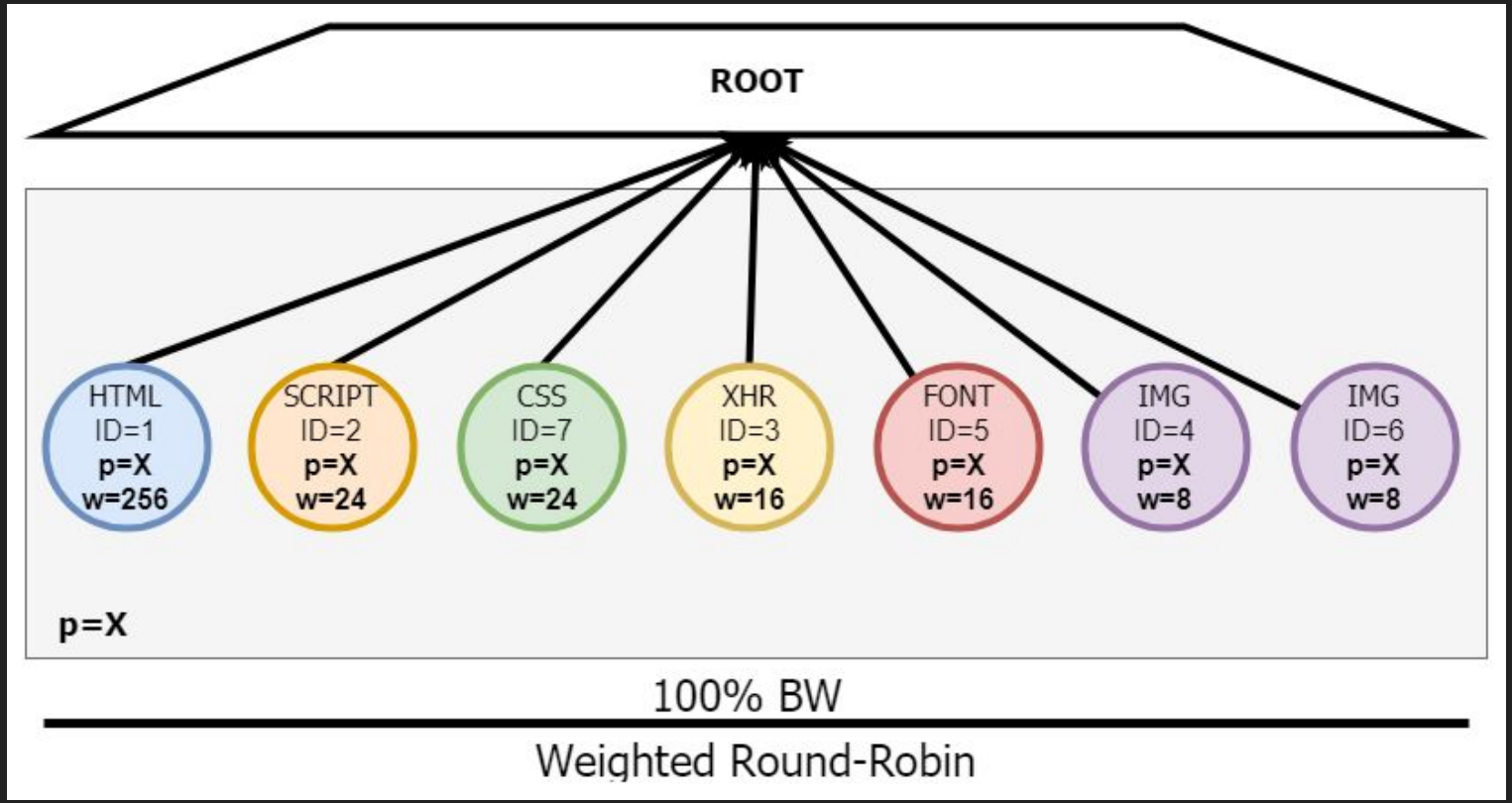


* Assuming enough placeholders, otherwise?

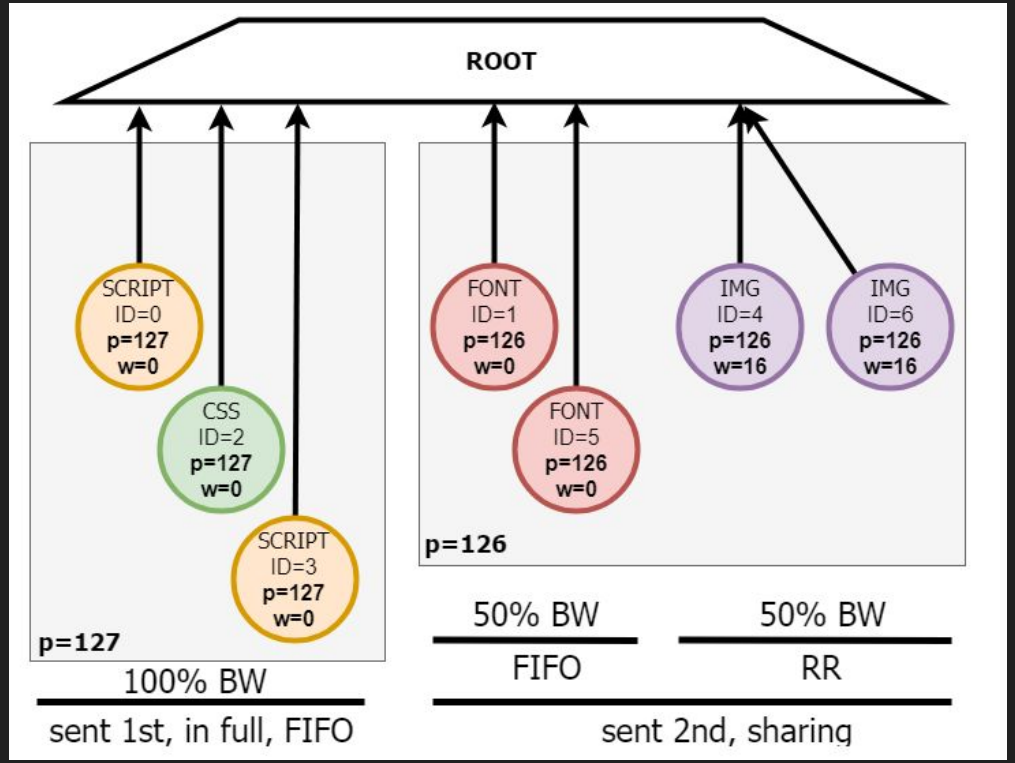
HTTP/2 original (Safari/Edge) : (Weighted) Round-Robin



Converting to H3 (draft-20 and 2700): Easy case (Safari/Edge) - No op

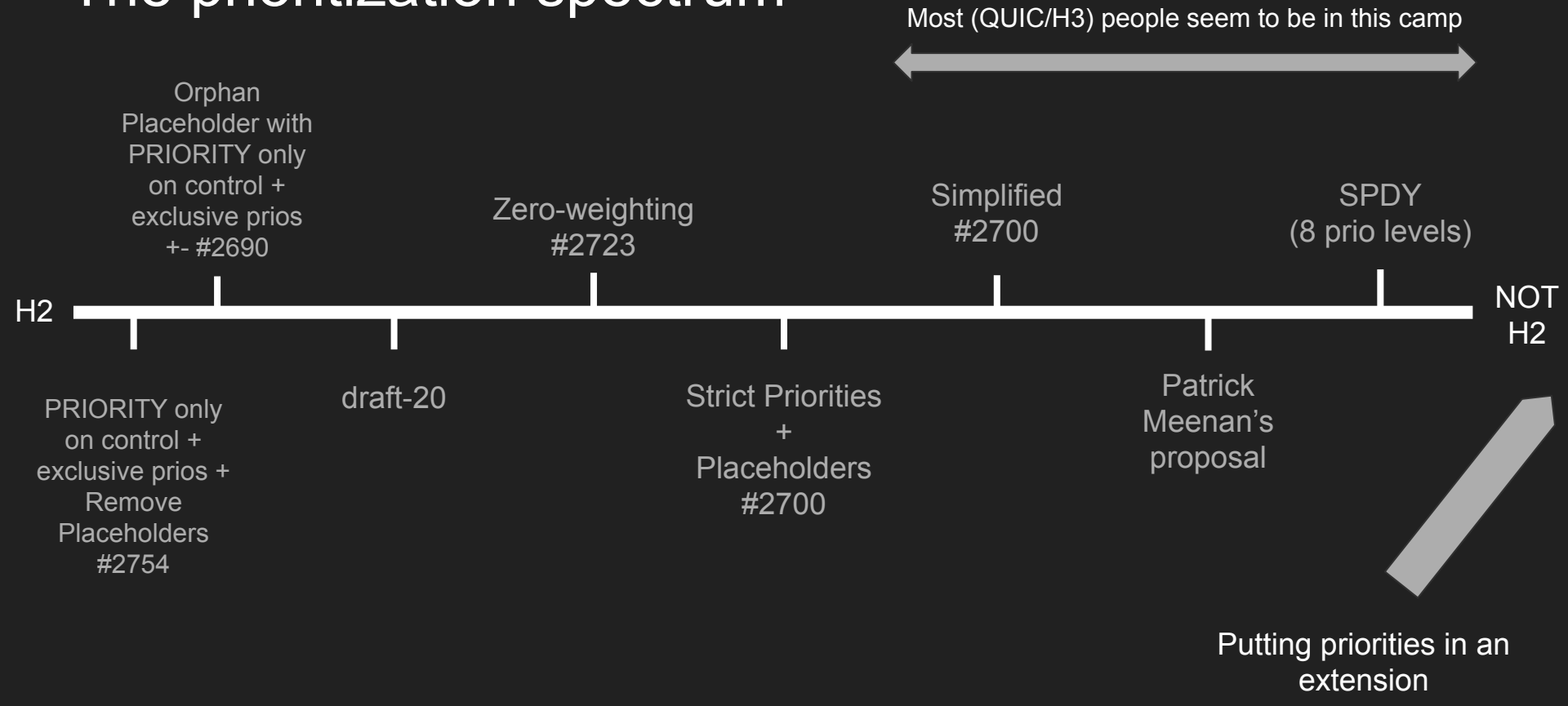


Patrick Meenan's proposal in #2700



Allows FIFO and RR with hard ordering, without needing a single placeholder

The prioritization spectrum



Issues and Goals

So what do we want to achieve?

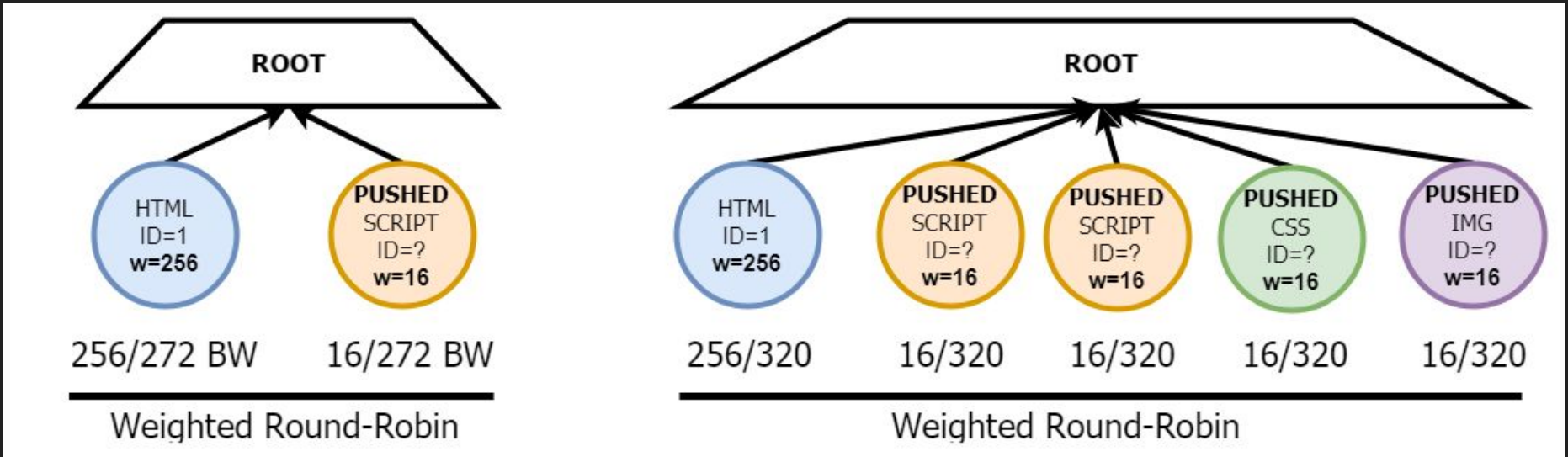
Faster page load times in browsers?

Better fair sharing in CDN to origin connections?

Preserving H2 priority functionality (H3 makes some of this optional)?

Something else?

Default Round-Robin is problematic link



"Steals" bandwidth from actually prioritized elements + delay themselves
(e.g., scripts need to be fully download before execute)

Note: not just push, many situations where this can happen!
(e.g., #2502, #2723, Edge/Safari)

What if you don’t have enough/any placeholders?

This breaks Firefox and breaks one Chrome approach

Kazuho: Could we just go back to H2 priorities? #2754

Issue #2753, #2734

Achieving wide adoption

The assumption will be the draft represents best practice

> I believe we have an obligation to either not put priorities in HTTP/3 OR Ensure HTTP/3 can implement best practices
>
> > If priorities are an extension, HTTP/3 may be slower than HTTP/2
> >
> > > Therefore, design something that will be **widely** supported

Full compliance to H2 priorities is ~25% of major CDNs link

Issue #2739

Allow Server side input

Sometimes the server knows more than the browser

- There are no JS APIs for priorities, so apps are stuck
  - How to fit an AFE priority in the tree, except max/min?

Note: Server push REALLY benefits from an initial priority

Issue #2740

Future Goal?: Priorities in the transport

Wide agreement that H2 priorities are too complex for general purpose use

Uses: WebTransport, WebRTC, HTTP/3 control streams, ...

No Issue: Seems like QUICv2?

Going Forward

1) Get feedback from this group today
2) Form a design team?
   - Get input from HTTP WG
     We decided to do only placeholders because of time…
     That was over a year ago (#441) :(
   - Get more insight in original HTTP/2 use cases
   - Implement candidate scheme in at least Chrome and Google
     - Hopefully other CDNs(ie: Cloudflare)
   - Prove it works
   - Present at Montreal+
3) ??

Important Questions

- What will people NOT implement (or NOT enable)?
  - Status Quo?
- Is functional equivalency to H2 desired/required/optional?
- Are we OK with being NOT H2 and NOT fix open issues?
- Who is willing to work on and deploy updated priorities?
- What data would be useful to the WG?
  - Browser use of priorities isn’t fully utilizing H2 features

Thanks!

To Robin Marx, who created some great illustrations (aka schema)!
- use these templates to quickly generate your own in draw.io!

To Patrick Meenan, for sending out an alternate proposal

To all the working group members who gave me feedback on ideas, slides, etc

Backup Slides

From Patrick Meenan (on browser usage)

“IMHO, an optimal ordering looks something like this:

Serialize the CSS and blocking JS in the head

Load the visible images in parallel (simple round robin)

Serialize async (and late-body) JS and load them in parallel with non-visible images. The non-visible images might be better to serialize in document order for long-scroll pages instead of loading ALL of them in parallel“

- From #2502

Theory

A single connection contains all requests for a page load

Reality (Alexa top 10 in Cr Incognito)

Google.com: 5 (+2)
YouTube.com: 10 (+6) - No video!
Facebook.com: 6 (+5) - Signup page
Baidu.com: 10 - HTTP 1.1!
Wikipedia.org: 1 - 7 requests
Qq.com: 13
Tmall.com: 22
TaoBao.com: 23
Yahoo.com: 17
Amazon.com: 15

Observation: QUIC alt-svc discovery makes this worse on Incognito

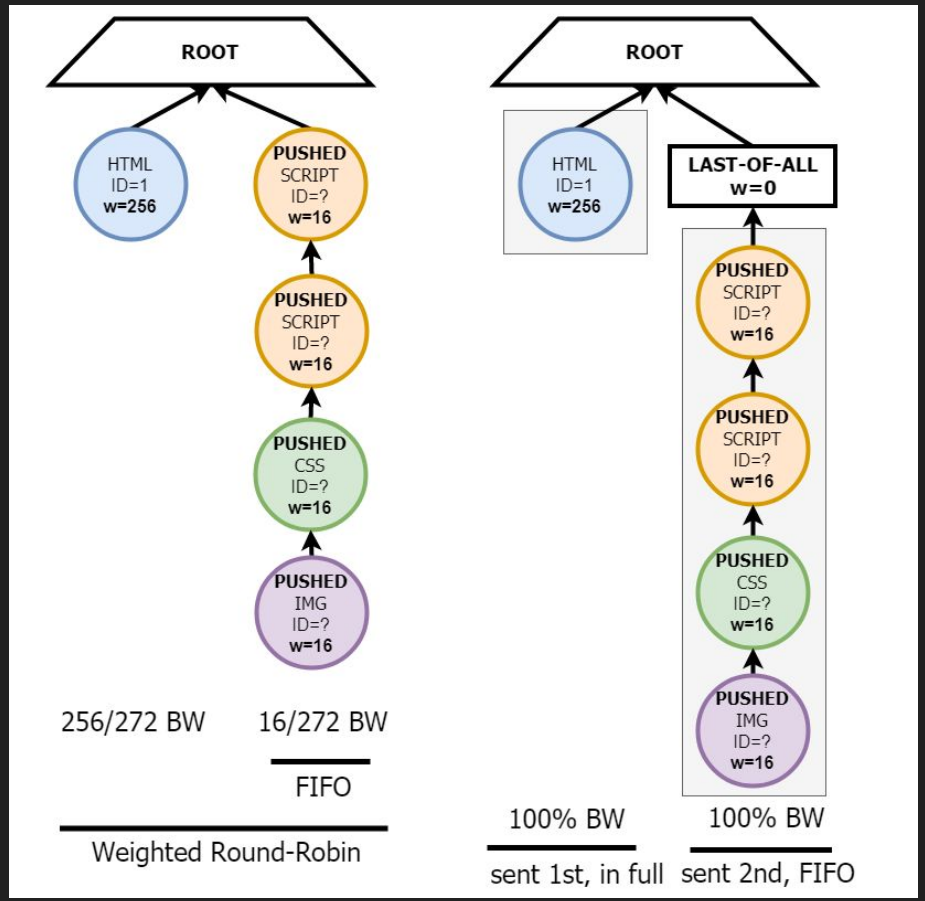
Exercising congestion control as much as priority

If priorities were absolute, a server can optimize sending behavior on one connection to minimize losses on another(ie: reduce initial burst, pacing rate, etc)

This is clearly an area of optimization and research, but it’s potentially huge

FIFO is overall better default option: PUSH link link2

Better



Best

But not 100% possible in draft-20 (but w=1 could work too…)

Proposals:
- #2690
- #2723
- #2700 comment

What can’t you do in #2700

You cannot move multiple streams to a new parent without a placeholder

- But you can in multiple operations, and I’m not aware of anyone doing this?

#2700 Framing efficiency

Priority is a byte

When >= 64 streams are open, it saves one or more bytes vs streams depending upon streams