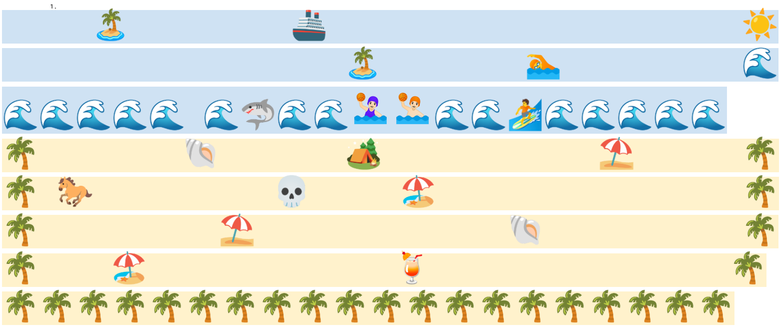
Partitioning :visited links history

[Breakout Session: Partitioning :visited links history slides]

Attendees (sign yourself in):

1. Ari Chivukula (Google Chrome)
2. Mike Taylor (Google Chrome)
3. Tom Van Goethem (Google Chrome)
4. Ian Clelland (Google Chrome)
5. Dominic Farolino (Google Chrome)
6. Austin Sullivan (Google Chrome)
7. Tab Atkins-Bittner (Google)
8. Jonathan Kingston (DuckDuckGo)
9. Adam Rice (Google Chrome)
10. Tim Nguyen (Apple)
11. Nidhi Jaju (Google Chrome)
12. Barry Pollard (Google Chrome)
13. Emilio Cobos Álvarez (Mozilla)

🅿️ Cursor beach relaxing



Minutes

Kyra: Introduction. Discuss agenda overview. Please, if you want to speak, raise your hand or put your name in the queue.

Kyra: Our goal is to improve user privacy by eliminating visited links history leaks. This is a well-known problem. 20+ years. You expose entropy about user browsing history when you have a link and style it differently based on visited state. All of these attacks rely on that. Over

the years, they've grown more sophisticated. UI interaction attacks, Timing Attacks, Pixel Color Attacks (css masking math, blend modes), Process-level attacks (e.g., Spectre).
Kyra: We're not the first people to propose a solution. In 2010, browsers implemented mitigations - the browser always returns "unvisited" when queried. Devs are limited in what styles can apply to anchors. There have been many proposed solutions. Some variations on same-origin or domain binding. Another allow-list based approach.
Kyra: The proposal: triple-key partitioning. A renderer will style a link as visited iff we have visited that link from this top-level site and frame-origin previously. The renderer takes that info, creates a partition key, asks the browser "have we seen this before", if so - style as visited.
Kyra: Why triple-key? It renders the previous attacks as useless. It only exposes info that a site also has access to (which links were click on on their site). Why top-level site? This prevents cross-site tracking. We don't want 3P iframes to share that info across top-level sites. Why frame-origin? Don't want to leak embedders/embedees.

We've had to rethink what a visited link is. What does a purple link mean?
Emilio: I've been here
Kyra: currently in Chrome, it's anything from browser history, omnibox, autofill, very different than clicking on something. We spoke with a Web Platform UX researcher and learned that this current model in Chrome doesn't align with user expectations.

To preserve privacy in this model, we're redefining what it means to be "visited". 1 - user clicking on a link. 2 - a scripted navigation. 3 - manual subframe navigation.

Dominic Farolino - Google Chrome. Are you excluding scripted navigation from subframes?

Kyra: As an "auto subframe navigation"? Yes. it's currently excluded from history backend in Chrome today. These are so frequent. It doesn't show up in history UI today. If we started showing all of those - it wouldn't match existing mental model.

This does not include autocomplete, bookmarks, subframe redirects.
Ian - Google Chrome. Does it include omnibox navigation to the current page.
Michal - I was thinking about the bookmark bar example.
Kyra: Yes, it would be considered as visited because top-level site and frame-origin match the current page. For the bookmark bar
Michal - Google Chrome. Page A, you've visited. You visit page B. It has a backlink, which is common. That should be purple. I imagine that's working, but it wasn't mentioned.
Kyra: thanks, I'll think through it.

Kyra: Should we consider other types of navigations?
Dom: what are synced visits?
Kyra: if your mobile phone syncs via profile to desktop, etc.
Tim - Apple. There are many WPT that use <a href=""> to test visited styling.
Emilio - that's a self-link, it probably should.
Ian: what about downloads?

Kyra: that’s a use case that probably makes sense.
Artur: I would assume it’s a navigation.
Ian: but does it commit?
Dom: Is committing the thing we care about?
Kyra: if we only add to history after DidFinishNavigation. That’s an implementation question as to how downloads are done.
Artur: from a security perspective, it seems fine. Reasonable to be in scope.
Kyra: would there be any case in which… if you’ve downloaded this from a different site?
Ari Chivukula (Google Chrome) - Do script navigations (or manual clicks) to #title links count as well?
Kyra: yes
Tab: everything w/ an anchor gets colored, today
Kyra: other APIs. history.pushState(). Navigation API. happy to hear thoughts.
Tab: i believe the intention is navigation API is intended to be the good way. Worth thinking thru.
Ian: pushState is used by single page apps.
Barry: What's the difference between these APIs and other scripted ones?-
Tab: there’s some navigation semantics rather than random JS sending you somewhere.
Dom: ….
Adam: Google Chrome. Domenic Denicola who worked on Navigation API is here in Seville.
Michal: why shouldn’t we consider other navigation types? Is it a question of risk? Or UX/mental model?
Kyra: there’s no semantic grouping.
Michal: As a user, I would like all of those to be purple. Unless there’s some threat, why not keep that model? Why limit reducing the scope to the minimum? Unless we have UX evidence…
Artur: I work on Google Security team. Have been trying to think through this. The problem is we’re trying to retrovit the mental model of visitedness. I don’t think it’s very well documented of what this means, by the folks who invented it. Now that we’re doing this work, we’re trying to be intentional. If you click on a link that didn’t load… have you visited it? Maybe it’s better to err on the side of overvisiting. This seems reasonable.
Michal: I’m a :visited power user.
Ari (Chrome) - It seems like any navigation the site could already be aware of (tracking clicks and/or causing them via scripts) should be treated as ‘visited’ to not give information to them they would otherwise lack.
Kyra: thanks, great point.
Ian (Chrome) - Is there any research to suggest that users consider visited-ness as a protective measure? For example, if a link to my bank shows up blue, then perhaps it is not the correct url (typo-squatting attack or homograph attack).
Kyra: there has not been a lot of user research on this topic. The last one I’m aware of is from 2004. It seems to suggest the model was intended for helping users navigate info dense environments. Great question, I don’t have an answer. It goes to Michal’s point of keeping as much state as possible.
Barry: the last 2 are definite ones we should consider (pushState, navigation API). Users would expect it.

Michal: as a follow up to UX research problem. Does it imply that we have to clear history of everyone?
Kyra: will speak to that.
Dom: I don't know if we can reliability prevent info about a navigations success or failure from getting back to site. There's a lot of ways to get this. From a privacy perspective I don't think we can plug these holes.
Emilio: when I looked in Firefox, the history DB explodes in terms of the amount of info you need to keep. E.g., you need 10 entries now instead of 1 if you clicked the same thing from 10 places. AFAIK, we don't have working attacks against us. But that's because of how we implemented things. I'm curious if you've considered, now your profile needs more data. Have you looked into this?
Kyra: we are very sensitive to that. The largest constraint we have. We haven't run the experiment yet, we're about to start it. From what we've found, you don't visit a site from that many different contexts. This is anecdotal, and some rough metrics. We don't anticipate key explosion. We believe users visit links from the same entry points, typically. Not storing omnibox navigations should save a bit as well. We are concerned.
Emilio: probably doesn't matter for 99% of users. Are you not storing info from omnibox? We do this differently in Firefox, would have to decouple.
Artur: I'm happy to chat later about security in Firefox.
Barry: there's no talk of expiring visited links?
Kyra: not yet.
Kyra: what does the experiment look like? We want to avoid "blank slate". We're going to store triple keys for 90 days in a separate database. That's experiment 1. Where we'll analyze metrics. That's rolling out soon, hopefully.

The second experiment is where we'll actually change how things appear to users.
Discussion: 1) is the platform interested? 2) should we spec it? 3) what did we miss?
Tab: from CSS, yes and yes. I've been wanted to fix this for 10 years… it's a wart on the way we do styling. It infects getComputedStyle in bizarre ways.
Emilio: that's the easy thing…
Tab: not being able to style links is really frustrating. Being able to have proper styling ability would be great.
Kyra: right now it says "browsers can do what they want to keep users safe". In your mind, how does standardization look like?
Tab: 1) we can leave as-is, but loosen restrictions on what you can do w/ visited. Under condition you're protecting users. 2) we specify triple-key. If there's agreement between browsers, then that's perfectly fine to do.

Ari (Google Chrome) - Has any consolidation gone to when this is cleared? Would the behavior match that of per-origin storage when it's cleared after storage partitioning (clearable for top-level or all below)? Will the Clear-Site-Data header be extended to support this? etc.

Kyra: the way it currently works, if you delete the visit from the backend. It deletes the corresponding entry in the visitedlinks hash table. We don't intend to change that. The way we currently are approaching it… we've altered the visited database so we have enough info to get at the partition key. I don't know about the UI end.

Ari: slightly different. Right now it's browser-level info. But you're making it site-level data. If I clear site data, I would expect to remove visited state for that site.

Artur: someone pointed out that you can use history as local state. If you can query visited state… we probably do want to clear this state to prevent this.

Ari (Google Chrome) - Is it worth exposing this to the user in some way? The same way the user can examine cookies stored by what third parties for a given top-level-frame.

Kyra: that's an interesting concept. We should ponder that.

Barry: syncing might not need to be included in the spec.

Kyra: agreed.

Adam: speaking as a user, I would like it if the visited info for sites that I don't visit frequently was deleted after some time. But I do want my wikipedia history forever.

Kyra: out of curiosity, do you feel like this is an option?

Adam: I would like it to be automatic. If we have a cookie, do we keep it? That's maybe weird. If you don't visit a site for 6 months we throw it away.

Barry: I'm surprised that's not what happens now.

Ari: would that be different… wouldn't you expect cookies to also delete? Or storage?

Adam: as a user I would like cookies to go away, but I know that's a much harder problem. But nothing relies on visited state at the moment. So we have an opportunity to introduce this now.

Artur: cookies are load bearing. Visited should not have any dependencies.

Kyra: in some future… worth pondering if it could be load bearing for sites?

Jonathan (DDG) - do you have histograms you can share?

Kyra: I can't share publicly. Based on rough calculations, we're not concerned.

Jonathan: I mean how much from a utility POV. I'm browsing bugzilla, and I'm looking at bugs… and I also have a google doc that has these links pasted into there. How often would the user be impacted?

Kyra: we don’t currently have metrics for this. Thanks for the point. Would be good information.

Artur: there’s an interested thought experiment. Some sites disable this styling now.. The question is: will anyone notice? Because it doesn’t work consistently…

Barry: if a site doesn’t do that… some links will be purple, some won’t. Current model is consistent.

Kyra: for wikipedia, colored links are helpful. Social media sites often don’t do visited links styling frequently. We’re in a unique position to train users…

Ari (Google) - Does this mean the 2010 mitigations can be removed? Is that an end goal?

Kyra: It’s not the primary goal. But this work paves the way to that being possible. If there’s appetite… we can continue on that path. But if there isn’t… we still want to do this work.

Austin (Google) - more to say on clearing, especially if 2010 mitigations are rolled back

> Should be cleared alongside cookie/storage? Treating persistent storage as cookies: https://html.spec.whatwg.org/#user-tracking

It seems like it should be cleared. We could specify this in the storage spec, maybe not when, but where. For CSD, for cookies, probably storage.

Jonathan: could be linked to max-expiry.

Michal: since you will be having a transition period. Does that also enable to do an A/B test at the same moment… would that give us an early warning that we overlooked something. Is that in the plan?

Kyra: yes, that’s in the plan. We do want to capture “what state is lost” metric.

Ian (Google) Should all same-origin history be visible? Regardless of what site the user navigated from. If you’ve clicked on wikipedia links… if you’ve entered it from 8 different sites. If 8 different origins have sent you to various wikipedia pages… later on wikipedia, can you see that you’ve visited them?

Artur: every origin knows. Wikipedia could do that. In our model

Kyra: it doesn’t differentiate by originator.

Artur: but sites could build that.

Tab: if w3c linked us to “history of seville”, later went to another page “history of spanish cities”. Would it be visited?

Artur: no, but wikipedia could build it. Probably nobody would.

Barry: repeats tab’s question.

Ian: it’s weird if it’s visited on the first link, but not the second.

Artur: can you give an example of when this would be strange?

Barry: i’m on barryblog.com, click on wikipedia link. Go to artur’s blog, click on a different wikipedia link. In both cases you’re now on wikipedia. Both cases should know that I’ve visited two pages on wikipedia.

Artur: from a security POV, if site data hasn’t been cleared, you’re not giving away any more info.

Kyra: from a user POV, i can see how the difference in jarring.

Barry: I guess it’s not clear to me “what is the proposal”.

Tab: part of the confusion is probably related to discussion on self links.

Artur: let’s say you navigate from A to B, B has a self-link. This self-link on B would not be colored as visited.

Ian: but B knows you’ve been there…

Kyra: I need to draw it out and look at implementation.

Michal: there’s a simple use case. Top 10 activities. Someone links to 4. You visit “Next”, now you’re on 5. Will the site link that as purple? It should.

Kyra: this model, not currently. But I see it. The question is Next a link? Or a scripted navigation? If latter, yes. If link, then not.

Ari: can’t you say the page you’re on is always in its own visited record.

Artur: we could duplicate all the entries… not that we want to do this.

Michal: why are scripted

Ari (Google) If you’re considering testing current vs triple key behavior, consider adding a ‘no visited info at all’ group as a baseline.

Kyra: I want to clarify, it wouldn’t be A/B. Not necessarily two arms of experiment.

Adam: for AB test. To make it a good test, you have to truncate existing history at same time you started recording history. If you compare all history in one arm, vs 90 days in other arm. It’s not a fair comparison.

Kyra: it would be imperfect, that’s a good point.

Kyra: thanks, please file issues on the explainer. Very open to hear everybody’s thoughts.

Queue - add yourself at the bottom (include topic)

-