Éviter les mises en cache de durée imprévisible des fichiers statiques

Problème qu'on a avec la conf apache/tomcat par défaut :

- apache/nginx/tomcat mettent "Last-Modified" pour les fichiers statiques
- les navigateurs s'en servent pour mettre en cache : https://publicobject.com/2015/03/26/how-do-http-caching-heuristics-work/

Le résultat est fragile, notamment si on veut modifier un fichier n'ayant pas été modifié depuis longtemps.

Exemple :

- en mars 2021, une personne accède à https://pod.univ-reunion.fr/static/css/pod.css
- la date du fichier est 2017-07. Il est donc vieux de 3 ans 6 mois, ou 1340 jours
- ce fichier est servi par nginx qui renvoie "Last-Modified: ... Jul 2017 ...", mais aucun Expires/Cache-Control
- n'ayant pas d'instructions précises, le navigateur met en cache le fichier pour 134 jours (10% de 1340)
- donc si l'utilisateur accède au fichier 4 mois plus tard, le navigateur peut fournir la version dans son cache, sans faire aucune requête vers le serveur
- si le fichier est modifié, la seule solution possible côté serveur pour s'assurer que l'utilisateur a la dernière version : modifier la page HTML pour faire un lien vers pod.css?v=2021-05 (la valeur importe peu du moment qu'elle est différente)

Solution : **forcer une durée fixe de mise en cache**

Vincent Bonamy à Rouen a forcé un Cache-Control précis sur ses apaches, du genre : https://www.esup-portail.org/wiki/display/ESUPMU/Optimisation+HTTP+Apache#OptimisationHTTPApache-Cachenavigateur

Pour la majorité des applications cette solution est sans problème. Le seul comportement imprévu concerne les pages qui ne forcent pas de "Cache-Control/Expires" mais qui sont néanmoins potentiellement dynamiques :

- en PHP par défaut, aucun header cache n'est mis tant qu'une session n'est pas utilisée
- en tomcat, une jsp simple ne force pas de headers non plus.

Solution proposée sur un reverse proxy

(solution en production à l'université Paris 1 depuis 2018)

Ajouter Cache-Control quand Last-Modified est mis et pas de Expires/Cache-Control

- pour apache >= 2.4.10 :

```
Header setifempty Cache-Control "max-age=86400" "expr=-n resp('Last-Modified') && -z resp('Expires')"
```

- pour nginx (nécessite lua) :

```
header_filter_by_lua_block {
  if (not ngx.header["Cache-Control"] and not ngx.header["Expires"] and ngx.header["Last-Modified"]) then
     ngx.header["Cache-Control"] = "max-age=86400"
  end
}
```

NB : 86400 = 24h

Notes

Squid

squid suit globalement les même règles que les navigateur (cf http://www.squid-cache.org/Doc/config/refresh_pattern/) :

- si le serveur envoie uniquement Last-Modified, squid va mettre le fichier en cache
- si le navigateur demande le fichier sans faire reload, il aura le fichier du cache squid.
- par contre si le navigateur fait un reload ou un force-reload, la requête en reload est propagée sur le serveur

Conclusion : même sur un navigateur vierge, on peut avoir une ancienne version à cause du cache squid.

Durée max de mise en cache de squid pour la conf suggérée "refresh_pattern . 0 20% 4320"

- conservation 20% => 2 fois plus longue que les 10% du navigateur
- cache max pour les fichiers n'ayant pas de Cache-Control/Expires : 4320 = 3 jours