wget && segfault - Résumé

Salut à tous, **winw** m'a montré récemment un truc assez sympa. Dans un terminal, tapez la commande

Vous obtiendrez un segfault. C'est assez sympa, d'autant plus que wget est quand même un binaire largement utilisé. Les bugs comme celui-ci se font rares ! On s'est alors demandé ce qu'on pourrait bien en faire. Nous ne nous sommes donc pas arrêtés là, et on a cherché la cause du problème. Pour cela, nous nous sommes armés de ce bon vieux gdb, ainsi que des sources de la dernière version de wget en date (1.16.3) disponible ici :

http://ftp.gnu.org/gnu/wget/wget-1.16.3.tar.gz

Dans un premier temps, nous avons recompilé le binaire afin d'en avoir une version non strippée et donc avoir accès aux symboles. Dans le dossiers des sources de wget :

Ensuite nous avons provoqué le segfault dans gdb puis affiché la backtrace pour trouver où se situe le problème

Le segfault se produit dans la fonction *getproxy* se trouvant dans **retr.c**

Après quelques petites recherches, on remarque que le pointeur u sur une structure *url* est un pointeur null, et du coup à la ligne

la tentative d'accès au champ *host* de la structure provoque le segfault.

Très bien, nous avons isolé la cause du segfault. Cependant, comment se fait-il que le pointeur u passé à *getproxy* soit nul ? Nous remontons alors un peu la backtrace.
Dans *retrieve_url*, toujours dans le même fichier

On voit l'appel à *getproxy*

Et on voit plus haut que *u* est défini comme ceci :

En mettant un breakpoint à l'entrée de la fonction *retrieve_url,* on se rend compte que le paramètre *orig_parsed* est déjà un pointeur nul. On continue et on remonte la backtrace d'un cran, en allant voir la fonction *retrieve_tree* situé dans le fichier **recur.c**

On voit l'appel à la fonction *retrieve_url* ici

Nous avons dit que le paramètre *url_parsed* était nul. Ce pointeur est défini une ligne au dessus :

Cette fois-ci, aucun des paramètres passés à *url_parse* ne sont nuls. Cette fonction renvoie donc un pointeur nul. En mettant un breakpoint juste après l'appel à cette fonction, on peut voir ce qu'il y a dans *url_err* : Le numéro 8.
Le code d'erreur 8 est défini dans le fichier **url.c** (dans lequel il y a la fonction *url_parse*)

Effectivement, dans la fonction *url_parse*, nous avons la vérification suivante :

Je vous rappelle que l'argument que nous avons passé à wget était **-r %3a** or **%3a** est le code ASCII de **:** . En amont, wget a détecté notre **:** et a donc considéré que c'était une adresse IPv6. Celle-ci étant invalide, *i*) renvoie *false*, et nous avons le code d'erreur. Tout est bien et se passe comme prévu par les développeurs à ce moment là.

L'erreur, c'est dans le fichier **recur.c** avec ces lignes :

Il n'y a aucune vérification de faite sur le retour de la fonction *url_parse*, et le pointeur *url_parsed* est utilisé sans vérifier s'il est nul, ou non.
Nous avons donc, logiquement, un segfault. De notre point de vue, cet oubli ne permet aucune exploitation, mais c'était une analyse intéressante. Un fix est de vérifier que la fonction *url_parse* a renvoyé un pointeur non nul, de la manière suivante :

Nous avons d'ailleurs proposé un fix à GNU. Nous verrons s'il sera accepté !

Ce problème n'existe pas si le paramètre **-r** est omis, puisque cet oubli de vérification se situe seulement dans le fichier **recur.c**, et nulle part ailleurs.