Caractérisation et Méthodologie d'Analyse des Buffer Overflow

Yves KONE, Stella BITCHEBE, Alain TCHANA {Ecole Normale Supérieure de Lyon}

Compas'2018

ENS DE LYON

Common Vulnerabilities and Exposures

Le buffer overflow est la vulnérabilité la plus répandue [1] Les CVE pour Common Vulnérabilites and Exposures est une base de données publique répertoriant des vulnérabiliés de sécurité. Chaque CVE est composé de 2 parties disctinctes:

- le descriptif principal qui présente brièvement la vulnérabilité.
- les références qui sont des URLs vers des pages web de différents types.

```
Name: CVE-2021-0101
Status: Candidate
Reference: MISC:https://www.intel.com/content/www/us/en/security-center/advisory/intel-sa-00474.html
Reference: URL:https://www.intel.com/content/www/us/en/security-center/advisory/intel-sa-00474.html

Buffer overflow in the BMC firmware for Intel(R) Server BoardM10JNP2SB
before version EFI BIOS 7215, BMC 8100.01.08 may allow an
unauthenticated user to potentially enable an escalation of privilege
via adjacent access.
```

Figure 1 – Extrait du CVE de type buffer overflow (ou BOF) 2021-2182.

Problematique

Les CVEs constituent un jeu de données important pour la recherche en sécurité[2]. Neanmoins, une analyse **fine et pertinente** des CVEs peut s'avérer etre une tache très **longue et fastidieuse**.
Les principales cause sont:

- le grand nombre de CVE répertorié (12669 CVE concernant les overflow depuis 2013)
- les données sont non structurées;
- aucune méthode d'analyse connue ou d'outil automatique, il faut tout étudier soit-même à la main.

Challenge

Le principal challenge auquel nous avons du faire face est déterminer les caractéristiques d'un buffer overflow. En effet les informations étant non structurées il est assez compliqué de

- **le type de débordement;**
- **la zone mémoire;**
- **les conséquences ou effets;**
- **le contexte relatif au code;**
- **le système impacté;**
- **la compagnie, entreprise impactée.**

Méthodologie

L'algorithme est divisé en 2 parties: l'étude du descriptif principal puis de chacune des références. Pour analyser le descriptif, un texte brut en langage hulain, il faut simplement **extraire** les informations relatives á la caractérisation définie et enfin les **interpréter** correctement.

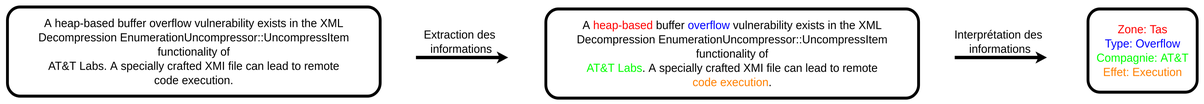

Figure 6 – Extrait du descriptif du cve 2021-2182.

Pour les références c'est un peu plus sportif. Il faut d'abord **identifier** le type de page web il s'agit, puis **identifier** les sources d'information dans la page web. Ces sources sont du texte mais peuvent être de nature différentes (titre, code, etc...). Enfin, comme pour le descriptif **extraire** les données et les **interpréter**.

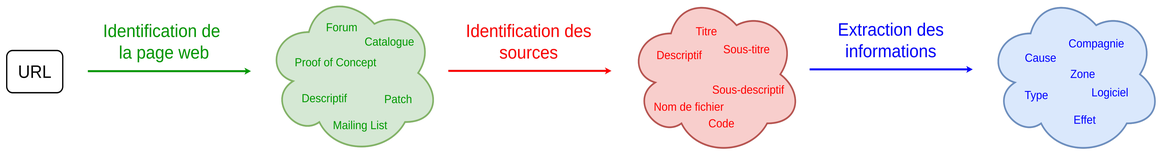

Figure 7 – Automate simplifié de l'analyse des références d'un cve.

Resultats préliminaires

Voici quelques résutlats obtenus aprés une analyse très fine de **368 CVEs** de type **buffer overflow** pour l'année 2021.

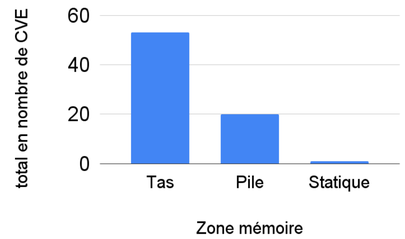

Figure 2 – Nombre de CVE par type de zone concernée.

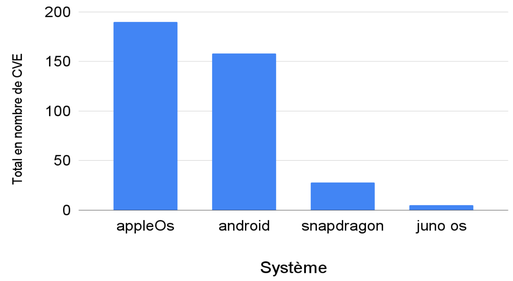

Figure 4 – Nombre de CVE par système impacté.

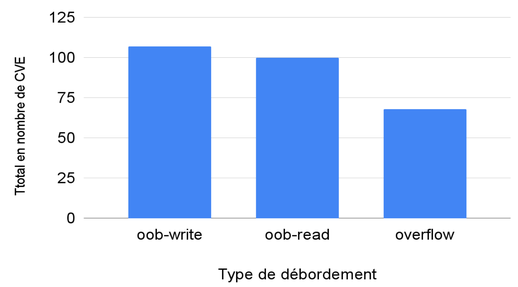

Figure 3 – Nombre de CVE par type de débordement.

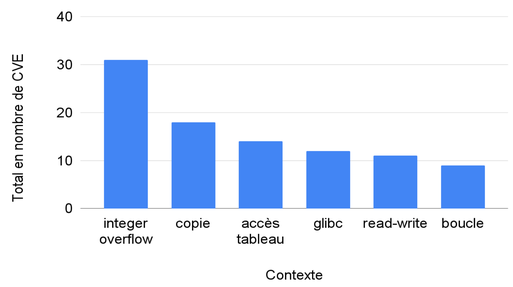

Figure 5 – Nombre de CVE par système impacté.

Travaux futurs

Dans un futur proche, nous avons pour but de mettre au point un outil capable:

- **d'extraire et interpréter du texte correctement malgré avec un tolérance aux fautes;**
- **pour chaque type de page web, identifier les sources d'informations potentielles et en extraire les données clefs;**
- **généraliser la méthode pour d'autres types de vulnérabilités.**

References

[1] Cwe/sans top 25 most dangerous software errors. https://www.sans.org/top25-software-errors, 2022.

[2] Istvan Haller and al. Dowser: A guided fuzzer for finding buffer overflow vulnerabilities. *Usenix Security*, 2013.

Contact

Email yves.kone@ens-lyon.fr
Email alain.tchana@ens-lyon.fr
Email celestine-stella.ndonga-bitchebe@ens-lyon.fr