Caractérisation et Méthodologie d'Analyse des Buffer Overflow

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

Compas'2018

ENS DE LYON

CVE : Common Vulnerabilities and Exposures

L'ensemble des vulnérabilités de sécurité est répertorié dans une base de données publique sous le nom de CVE. Les CVE constituent un jeu de données important pour la recherche en sécurité[1]. Chaque entrée dans la base de données est composée 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.

Problématique

Les buffers overflow consitituent la vulnérabilité la plus répandue [2]. Neanmoins, une analyse **fine et pertinente** des CVE pour les extraire peut s'avérer une tâche très **longue et fastidieuse**.
Les principales causes 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 est de déterminer les caractéristiques d'un buffer overflow. En effet les informations étant non structurées il est assez compliqué de XXX:

- **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 d'extraction de CVE est divisé en 2 parties: l'étude du descriptif principal puis de chacune des références. Pour l'étude du descriptif, qui est constitué de texte brut en langage humain, il faut simplement ***extraire*** les informations relatives à la caractérisation définie et les **interpréter** correctement.

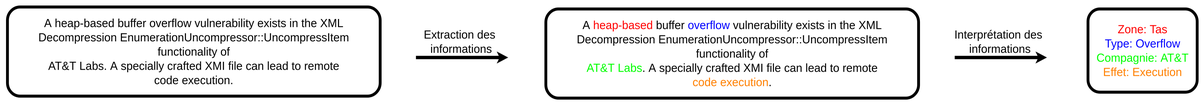

Figure 6 – Extrait du descriptif du cve 2021-2182.

Pour les références, il faut tout d'abord **identifier** le type de page web ça veut dire quoi?? dont il s'agit et ensuite les sources d'information dans la page web. Ces sources sont en général constituées de texte et peuvent être de nature différentes (titre, code, etc...). Enfin, comme pour le descriptif **extraire** les données et les **interpréter**.

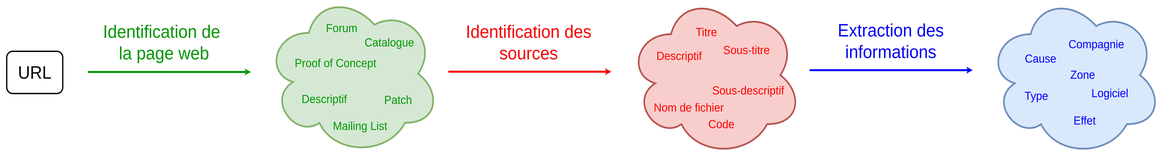

Figure 7 – Automate simplifié de l'analyse des références d'un cve. –» elle est hors du cadre!!

Résultats Préliminaires

Les figures 2-5 présentent quelques résutlats obtenus après une analyse rigoureuse de **368 CVEs** de type **buffer overflow** pour l'année 2021.

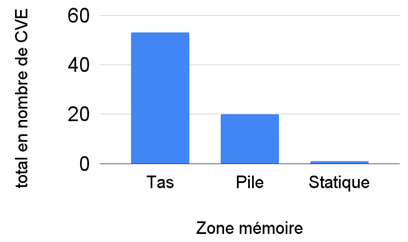

Figure 2 – Nombre de CVE par type de zone concernée.

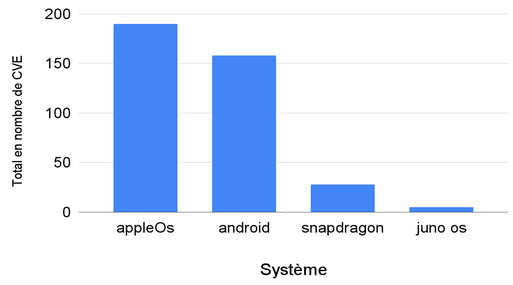

Figure 4 – Nombre de CVE par système impacté.

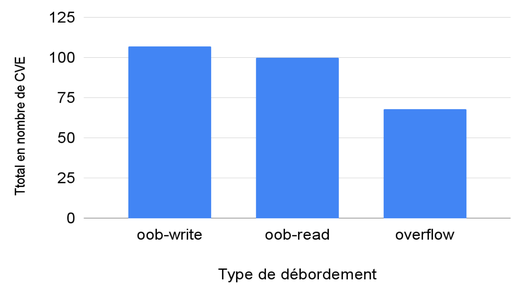

Figure 3 – Nombre de CVE par type de débordement.

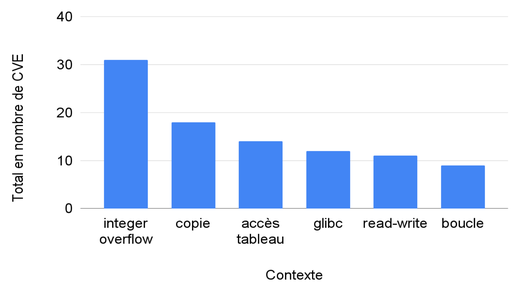

Figure 5 – Nombre de CVE par système impacté.

Travaux Futurs

La conception d'un outil pour :

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

Références

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

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

Contact

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