Caractérisation et Méthodologie d'Analyse des Buffer Overflow

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

Compas'2018

Common Vulnerabilities and Exposures

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 – Exemple d'un CVE concernant les buffer overflow en 2021.

Problematique

Les CVEs constituent un Dataset important pour la recherche en sécurité (12669 CVE concernant les overflow depuis 2013). Neanmoins, une analyse fine et pertinente des CVEs peut s'avérer etre une tache très longue et fastidieuse.
A titre d'exemple nous avons passé plus de 6 mois (et plusieurs boîtes de vitamines) pour obtenir nos premiers résultats. Les principales cause sont:

- **le grand nombre de CVE répertorié;**
- **les données sont non structurées, parser les CVE ne suffit pas;**
- **pas de méthode d'analyse ou d'outil automatique, il faut tout faire a la main.**

Catégorisation

Catétgorisation et extraction et interprétation des informations Notre objectif est de proposer une méthodologie d'analyse des CVEs de type BOF et concevoir un outil d'analyse automatique. Nous avions donc un besoin d'identifier les caractéristiques des BOF: Après plusieurs passages, nous avons catégories suivantes:

- **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 consiste à analyser le descriptif principal puis les références. Pour analyser le descriptif il faut extraire les informations pertinente et les interpréter.

Pour les références c'est un peu plus subtil. Il faut d'abord, en fonction du type de page web, identifier les sources d'information puis extraire celles qui nous intéressent et ensuite les interpréter

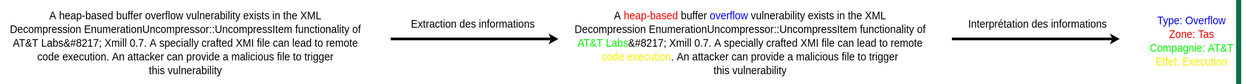

Figure 2 – Automate simplifié de l'analyse d'un CVE.

L'algorithme consiste à analyser le descriptif principal puis les références. Pour analyser le descriptif il faut extraire les informations pertinente et les interpréter.

Figure 3 – Automate simplifié de l'analyse d'un CVE.

References

Resultats préliminaires

Nous avons obtenus quelques résultats

Contact

Linkedin
Email yves.kone@ens-lyon.fr