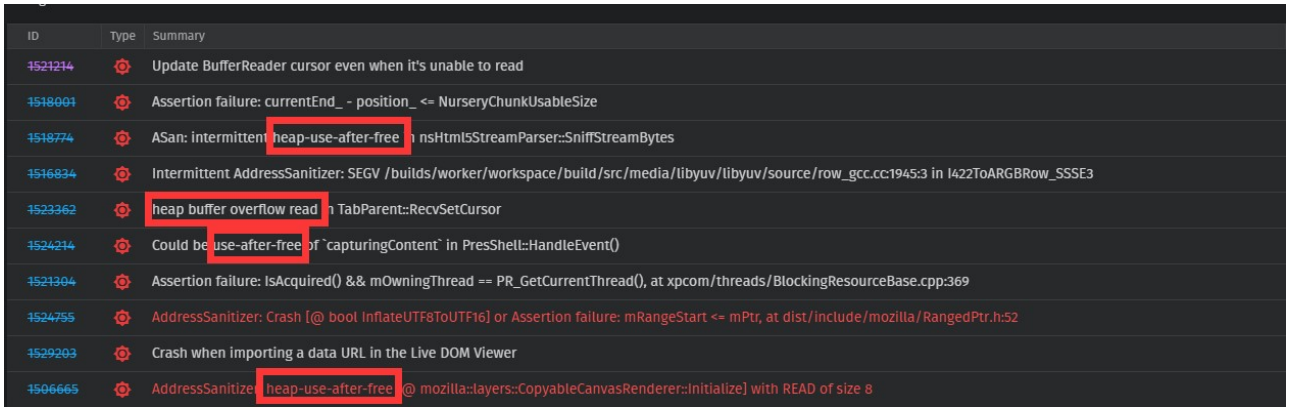
Improving the quality of vulnerability data

Motivation:

1. The vulnerability information in vulnerability databases is usually incomplete. Many vulnerabilities entries lack the vulnerability type. For example, in CVE-2000-0998, the vulnerability type (CWE) is missing.

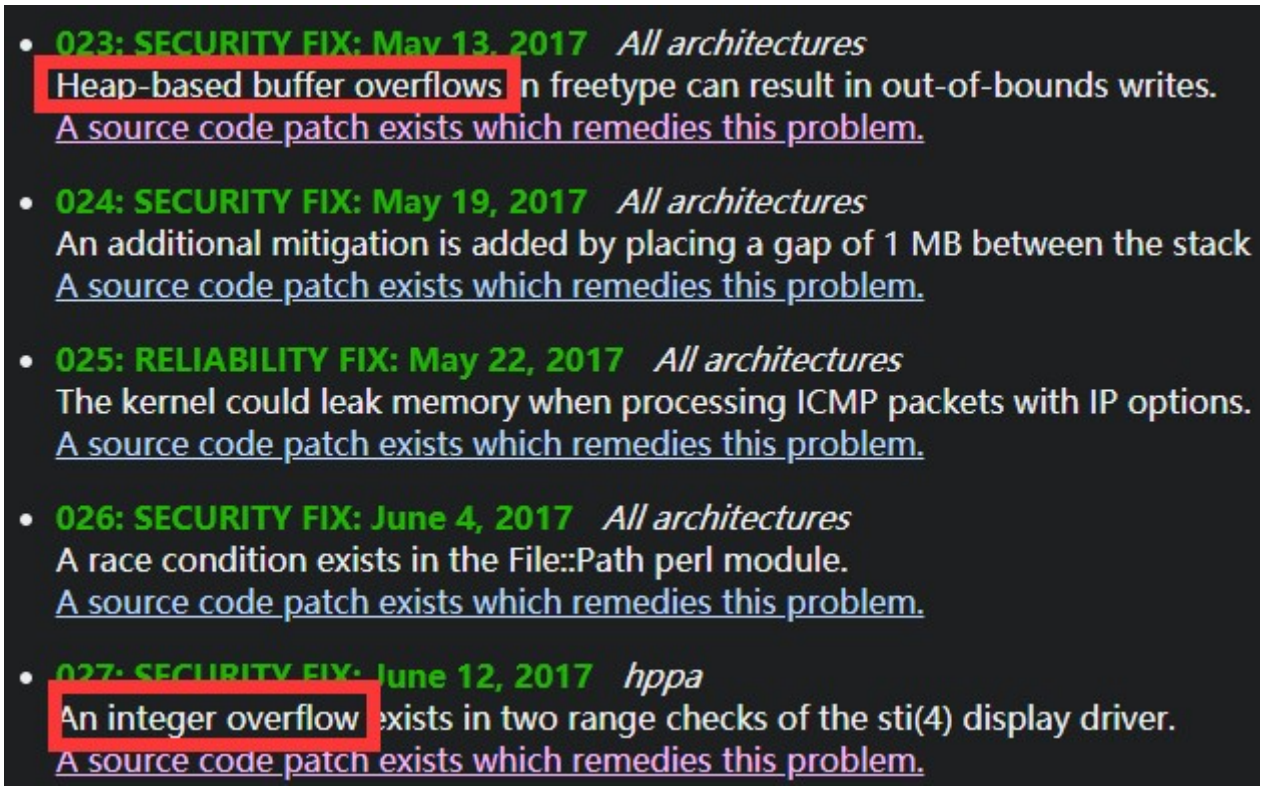
2. A CVE may contains multiple vulnerabilities. Such as CVE-2019-9788: "Mozilla developers and community members reported memory safety bugs present in Firefox 65, Firefox ESR 60.5, and Thunderbird 60.5. Some of these bugs showed evidence of memory corruption and we presume that with enough effort that some of these could be exploited to run arbitrary code. This vulnerability affects Thunderbird < 60.6, Firefox ESR < 60.6, and Firefox < 66."

   The types of these vulnerabilities can be learned from bug description in Bugzilla, such as the buglist corresponding to this CVE:



3. In vulnerability databases， the impact of a vulnerability is usually reflected by the impact score in CVSS, but the specific impact of the vulnerability is included in the description of the vulnerability. Such as the description of CVE-2000-0998: "Format string vulnerability in top program allows local attackers to **gain root privileges** via the "kill" or "renice" function."

4. Some vulnerabilities are not included in the vulnerability databases. For example, in the vulnerability list of OpenBSD 6.0, Vuln-023, Vuln-027 are not included in the vulnerability databases.



Research Goal:

1. Increasing the missing vulnerabilities in the vulnerability databases.
2. Extrcting the attributes of vulnerability automatically. The attributes of vulnerability are as follow:
    - Vulnerability type
    - Vulnerability impact
3. A Named Entity Recognition model in cybersecurity

Vulnerability Data Source

1. NVD
2. SecurityFocus
3. Vendor security advisory(e.g. http://www.openbsd.org/security.html, www.mozilla.org/en-US/security/advisories/)
4. Software bug tracking system(e.g. bugzilla.mozilla.org)