

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:

ID	Type	Summary
1521214	🔴	Update BufferReader cursor even when it's unable to read
1518001	🔴	Assertion failure: currentEnd_ - position_ <= NurseryChunkUsableSize
1518774	🔴	ASan: intermittent heap-use-after-free in nsHTMLStreamParser::SniffStreamBytes
1516834	🔴	Intermittent AddressSanitizer: SEGV /builds/worker/workspace/build/src/media/libyuv/libyuv/source/row_gcc.cc:1945:3 in I422ToARGBRow_SSSE3
1523362	🔴	heap buffer overflow read in TabParent::RecvSetCursor
1524214	🔴	Could be use-after-free of "capturingContent" in PresShell::HandleEvent()
1521304	🔴	Assertion failure: IsAcquired() && mOwningThread == PR_GetCurrentThread(), at xpcom/threads/BlockingResourceBase.cpp:369
1524755	🔴	AddressSanitizer: Crash [@ bool InflateUTF8ToUTF16] or Assertion failure: mRangeStart <= mPtr, at dist/include/mozilla/RangedPtr.h:52
1529203	🔴	Crash when importing a data URL in the Live DOM Viewer
1506665	🔴	AddressSanitizer: heap-use-after-free @ mozilla::layers::CopyableCanvasRenderer::Initialize] with READ of size 8

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.

- **023: SECURITY FIX: May 13, 2017** *All architectures*
Heap-based buffer overflows in freetype can result in out-of-bounds writes.
A source code patch exists which remedies this problem.
- **024: SECURITY FIX: May 19, 2017** *All architectures*
 An additional mitigation is added by placing a gap of 1 MB between the stack
A source code patch exists which remedies this problem.
- **025: RELIABILITY FIX: May 22, 2017** *All architectures*
 The kernel could leak memory when processing ICMP packets with IP options.
A source code patch exists which remedies this problem.
- **026: SECURITY FIX: June 4, 2017** *All architectures*
 A race condition exists in the File::Path perl module.
A source code patch exists which remedies this problem.
- **027: SECURITY FIX: June 12, 2017** *hppa*
An integer overflow exists in two range checks of the sti(4) display driver.
A source code patch exists which remedies this problem.

Research Goal:

1. Increasing the missing vulnerabilities in the vulnerability databases.
2. Extracting 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)