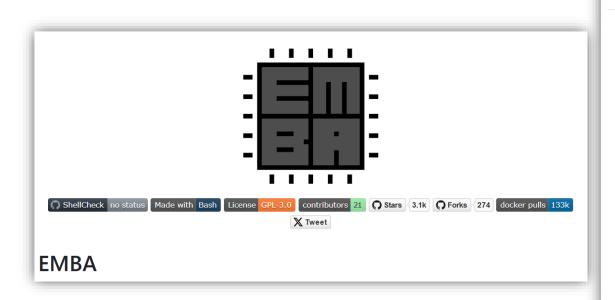
# EMBAを使ってみました

平能 耀介 ルネサスエレクトロニクス株式会社

## EMBA(Embedded Linux Analyzer)

- オープンソース(GPLv3)の組込みLinux向けの脆弱性解析ツール
- エンジニアのMichael Messnerによって開発された(Siemens Energy AG)
- 過去にBlack Hatなどのカンファレンスなどで紹介されていた
- 業務でSBOMに関わっているので、最近SBOM生成機能が追加されたことを知って調査することにした

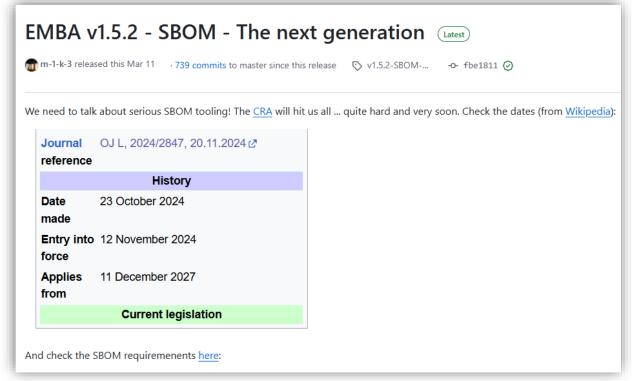


#### Publications, Talks and Live demos

- TROOPERS 25 SBOMs the right way (June 2025) Schedule/Slides [EN]
- Black Hat MEA Arsenal Riyadh 2024 (November 2024) <u>Schedule/Slides</u> [EN]
- FLOSS Weekly Episode 802: EMBA Layers Upon Layers Of Bash (September 2024)
- Leveraging Automated Firmware Analysis with the Open-Source Firmware Analyzer medium.com [EN]
- Black Hat Arsenal Asia 2024 (April 2024 Virtual) Schedule [EN]
- BSides Las Vegas USA 2023 (August 2023) Schedule/Recording [EN]
- Black Hat Arsenal USA 2023 (August 2023) Schedule [EN]
- DEF CON 31 ICS Village USA 2023 (August 2023) Schedule [EN]
- Firmware Scraper AMOS project presentation (2022/2023) Project page [EN]
- Black Hat Arsenal Europe 2022 (December 2022) Schedule/Slides/Recording [EN]

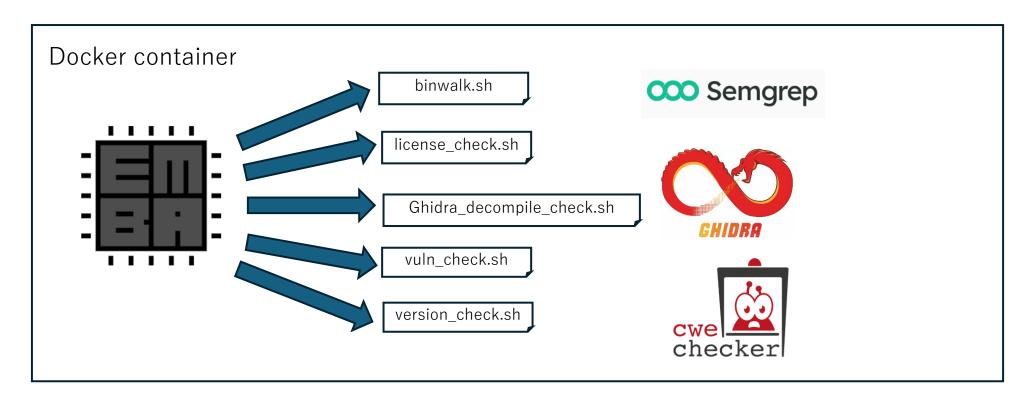
## EMBA(Embedded Linux Analyzer)

- もともとはペネトレーションテストを行う人向けのツールだが、開発者は今後はソフトウェア開発者にも使用してほしいと考えている
- 2024年10月 v1.5からSBOMの生成機能が追加された
- 2025年3月 v1.5.2 CRAに備えて今後はSBOMツールとして開発していくことになった(これまで通りfirmware analyzer機能の開発も継続)
- SBOMの精度を高めたり、関連機能の開発することに意欲的な様子



## EMBAの動作概要

- ペネトレーションテスターが手動で行っているような解析作業を集約して自動実行するもの
- docker環境に解析に必要なOSSをインストールし、それらを組み合わせて脆弱性解析を自動実行する



#### EMBAの動作概要

- binwalkなどでバイナリに含まれているデータを抽出(ファイルシステム、カーネルイメージ、ELFバイナリなど)
- 抽出したデータに対してbashスクリプトによる操作とOSSを利用して解析を行う
- 以下のような作業を自動実行してくれる
  - ✓カーネルの脆弱性検出
  - ✓ 文字列検索でライブラリのバージョン/ライセンス情報特定
  - ✔QEMU上でエミュレーションしてソフトウェアのバージョン/ライセンス情報特定
  - ✓ cwe\_checkerを利用したC/C++の危険なコードパターンのチェック
  - ✔ GhidraとSemgrepを利用したソースコードレベルの脆弱性解析

#### インストール

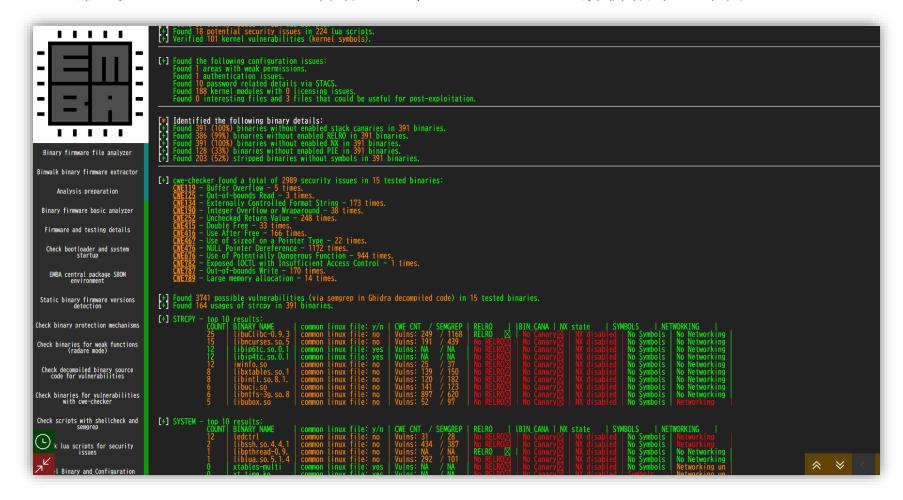
- Githubの手順に従ってリポジトリをクローンし、installer.sh を実行
- 動作環境の指定は厳しく、原則Kali LinuxとUbuntu:jammy (22.04 LTS)のみで動作する
- 動作環境を満たしていても、インストール時にエラーが発生した
- Github上にインストールができないというissueも多くあり、困っている人は多そう

### 実行

- 解析したいファイルと設定ファイルを引数に指定してCLI上で./embaを実行
- 解析対象としてRaspberry Pi OS 64bitを使用 Raspberry Pi OS downloads Raspberry Pi
- 解析にかかった時間は10時間28分
- cwe\_checkerを利用したC/C++の危険なコードパターンのチェックに最も時間がかかり、5時間程度かかった
- 動作環境
  - ✓ OS: Ubuntu 22.04.5 LTS
  - ✓ CPU: INTEL(R) XEON(R) GOLD 5512U
  - ✓ RAM: 256GB

### 結果の確認

• 解析結果はhtml形式のレポートとして保存され、ブラウザから解析結果を確認できる



### カーネルの脆弱性検出

- 解析対象バイナリから抽出したカーネルのバージョン、アーキテクチャ、kconfigを利用してカーネルをdry compile
- dry compileにより、コンパイル時に使用されたソースファイルを抽出
- カーネルのシンボル情報を抽出
- 抽出した情報とCVEの情報を照らし合わせて精度の高いCVE検出を実現
- 参考: Towards Reliable and Scalable Linux Kernel CVE Attribution in Automated Static Firmware Analyses https://arxiv.org/pdf/2209.05217
- 今回のRaspberry Pi OSの結果ではCVE-2019-3819が1つ検出された(HID系のバグ)

### バージョン/ライセンス特定

- QEMUのユーザモードエミュレーションで-V,-helpオプション指定の実行でバージョンやライセンスを特定
- 用意しておいたワードリストを利用してバイナリ中の文字列検索(strings, grep)も行う

```
-h: applet not found

[*] Emulating binary ./usr/bin/busybox with parameter -help
-help: applet not found

[*] Emulating binary ./usr/bin/busybox with parameter --help
BusyBox v1.35.0 (Debian 1:1.35.0-4+b3) multi-call binary.
BusyBox is copyrighted by many authors between 1998-2015.
Licensed under GPLv2. See source distribution for detailed copyright notices.

Usage: busybox [function [arguments]...]
    or: busybox --list[-full]
    or: busybox --show SCRIPT
    or: busybox --install [-s] [DIR]
    or: function [arguments]...
```

QEMU上で--helpコマンドを実行している

```
This module extracts version and license details from the results of the user-mode emulation module (s115).

1 Version information found catman 2.11.2 in binary /usr/bin/catman (license: GPL-3.0-or-later) (emulation).
1 Version information found Version: 1.0 in binary /usr/bin/spt/methods/rsh (license: unknown) (emulation/strict).
1 Version information found spriversion 3.9.0 in binary /usr/bin/spci (license: unknown) (emulation).
2 Version information found proces-ng 4.0.2 in binary /usr/bin/free (license: unknown) (emulation).
3 Version information found Busydox v1.35.0 (bebian 1:1.35.0-4+b3) multi-call binary in binary /usr/bin/busybox (license: GPL-4 Version information found mke2fs 1.47.0 in binary /usr/sbin/mke2fs (license: unknown) (emulation).
3 Version information found responded in binary /usr/sbin/archthool (license: unknown) (emulation).
4 Version information found version 6.1 in binary /usr/sbin/rethtool (license: unknown) (emulation).
5 Version information found (GNU coreutils) 9.1 in binary /usr/sbin/archthool (license: unknown) (emulation).
5 Version information found (GNU coreutils) 9.1 in binary /usr/sbin/archthool (license: unknown) (emulation).
6 Version information found davabi-daemon 0.8 in binary /usr/sbin/archthool (license: unknown) (emulation).
7 Version information found davabi-daemon 0.8 in binary /usr/sbin/archthool (license: unknown) (emulation).
8 Version information found davabi-daemon 0.8 in binary /usr/sbin/archthool (license: unknown) (emulation).
9 Version information found davabi-daemon 0.8 in binary /usr/sbin/archthool (license: unknown) (emulation).
9 Version information found davabi-daemon 0.8 in binary /usr/sbin/archthool (license: unknown) (emulation).
9 Version information found davabi-daemon 0.8 in binary /usr/sbin/sdrt-stop-daemon (license: unknown) (emulation).
9 Version information found schroot, version 5.2.15 in binary /usr/sbin/in/sbroot (license: unknown) (emulation).
9 Version information found schroot, version 5.2.15 in binary /usr/sbin/in/sorboot (license: u
```

検出したソフトウェアバージョンのリスト

#### cwe-checkerによる危険なコード検出

- cwe-checkerはオープンソースの静的解析ツール
- arm,x86,MIPSなどマルチアーキテクチャ対応
- ELFバイナリに対してC/C++言語のCWEに該当する危険なコードパターンを検出する

```
[+] cwe-checker found a total of 2829 and 13 different security issues in libX11.so.6.4.0 (common linux file: no):
        CWE119 - Buffer Overflow - 4 times.
        CWE125 - Out-of-bounds Read - 20 times.
        CWE134 - Externally Controlled Format String - 23 times.
        CWE190 - Integer Overflow or Wraparound - 27 times.
        CWE252 - Unchecked Return Value - 49 times.
        CWE415 - Double Free - 26 times.
        CWE416 - Use After Free - 585 times.
        CWE416 - Use of sizeof on a Pointer Type - 21 times.
        CWE477 - NULL Pointer Dereference - 1389 times.
        CWE6776 - Use of Potentially Dangerous Function - 610 times.
        CWE782 - Exposed IOCTL with Insufficient Access Control - 2 times.
        CWE787 - Out-of-bounds Write - 9 times.
        CWE788 - Large memory allocation - 64 times.
```

ライブラリに対して解析が行われた結果

## GhidraとSemgrepを利用した脆弱性解析

- バイナリをGhidraでデコンパイルして疑似コードを生成
- 疑似コードに対して静的解析ツールSemgrepを適用することでバイナリのみでソースコードレベルの 脆弱性解析を実現

• 参考:https://security.humanativaspa.it/automating-binary-vulnerability-discovery-with-ghidra-and-

semgrep/

Binary

(arm,mips,x86)



Decompile

```
int iVar2:
size_t sVar3;
   (<mark>((param_1 != (char *)0x0) && (param_2 != (char *</mark>)0x0<mark>)) &</mark>&
  (iVar2 = strcmp(param_1,param_2), iVar2 != 0)) {
 sVar3 = strlen(param_1);
 *(size_t *)(PTR_DAT_00026268 + 0x6320) = sVar3;
 sVar4 = strlen(param_2);
 *(size t *)(PTR DAT 00026268 + 0x6324) = sVar4;
 pvVar5 = malloc(sVar3 + sVar4 + 2)
  if (pvVar5 != (void *)0x0) {
   memcpy(pvVar5,param_1,sVar3 + 1)
   *(void **)(puVar1 + 0x6328) = pvVar5;
   pvVar5 = (void *)((int)pvVar5 + sVar3 + 1);
   memcpy(pvVar5,param_2,sVar4 + 1)
   *(void **)(puVar1 + 0x632c) = pvVar5;
 undefined4 *)(puVar1 + 0x6328) = 0;
  undefined4 *)(puVar1 + 0x632c) = 0;
```

```
*(voiu **)(puvari + oxoszo) - pvvars;
pvVar5 = (void *)((int)pvVar5 + sVar3 + 1);
memcpy(pvVar5, param_2, sVar4 + 1)
*(void **)(puVar1 + 0x632c) = pvVar5;
return;
```

した。 危険なmemcpy検出

疑似Cコード

## Ghidraとsemgrepを利用した脆弱性解析

• ファイルシステムに含まれているライブラリに対して脆弱性解析を行った結果

```
[+] Found 311 issues in native binary libhistory.so.8.2 (common linux file: no)
[+] Found 1088 issues in native binary libksba.so.8.14.3 (common linux file: no)
[+] Found 1547 issues in native binary aarch64-linux-gnu-objdump (common linux file: no)
[+] Found 1253 issues in native binary libreadline.so.8.2 (common linux file: no)
[+] Found 3906 issues in native binary libX11.so.6.4.0 (common linux file: no)
[+] Found 5177 issues in native binary libbfd-2.40-system.so (common linux file: no)
[+] Found 3454 issues in native binary libsystemd-core-252.so (common linux file: no)
[+] Found 1080 issues in native binary libslang.so.2.3.3 (common linux file: no)
[+] Found 2775 issues in native binary libicuuc.so.72.1 (common linux file: no)
[+] Found 28072 issues in native binary libsystemd-shared-252.so (common linux file: no)
[+] Found 28072 issues in native binary libsystemd-shared-252.so (common linux file: no)
```

どのライブラリで脆弱性が検出されたか

```
[+] Found 57943 possible vulnerabilities (via semgrep on Ghidra decompiled code) in 15 tested binaries:
    command-injection - 2 times.
    double-free - 16 times.
    format-string-bugs - 808 times.
    incorrect-unsigned-comparison - 135 times.
    incorrect-use-of-free - 54 times.
    incorrect-use-of-memset - 4 times.
    insecure-api-access-stat-lstat - 131 times.
    insecure-api-access-stat-lstat - 131 times.
    insecure-api-strcpy-stpcpy-strcat - 680 times.
    integer-truncation - 27629 times.
    integer-wraparound - 599 times.
    interesting-api-calls - 22431 times.
    off-by-one - 3 times.
    pointer-subtraction - 661 times.
    signed-unsigned-conversion - 4076 times.
    unchecked-ret-malloc-calloc-realloc - 419 times.
    unchecked-ret-setuid-seteuid - 1 times.
    unsafe-ret-snprintf-vsnprintf - 11 times.
    unterminated-string-strncpy-stpncpy - 169 times.
    use-after-free - 31 times.
```

脆弱性の種類と検出された回数

## GhidraとSemgrepを利用した脆弱性解析

• ブラウザで脆弱性が検出された部分の疑似コード確認ができる

```
long copy_history_entry(undefined8 *param_1)
   Identified source function: /logs/s16_ghidra_decompile_checks/haruspex_libhistory.so.8.2/FUN_00107d7c_00107d7c.c
Semgrep rule: external.semgrep-rules-0xdea.c.raptor-integer-truncation
Issue description:
Truncation errors occur when a primitive is cast to a primitive of a smaller size and data is lost in the conversion. The value cannot be trusted and the application will be in an undefined state.
                                                                                                                                                                                                                                                                                                                           long lVar1;
                                                                                                                                                                                                                                                                                                                         char *pcVar2;
          15 - iVar4 = _rl_get_char_len(param_4 + param_5, param_6)
                                                                                                                                                                                                                                                                                                                         size_t sVar3;
undefined8 uVar4;
   Identified source function: /logs/s16_ghidra_decompile_checks/haruspex_libhistory.so.8.2/copy_history_entry_00102ed4.c
Semgrep rule: external.semgrep-rules-0xdea.c.raptor-interesting-api-calls
Issue description:
Locate all calls to interesting and potentially insecure API functions (candidate points). The auditor can backtrace from these candidate points to find pathways allowing access from untrusted input.
                                                                                                                                                                                                                                                                                                                          15 - pcVar2 = (char *)xmalloc(sVar3 + 1)
                                                                                                                                                                                                                                                                                                                               pcVar2 = (char *)param 1[1];
if (pcVar2 != (char *)0x0) {
   Identified source function: /logs/s16_ghidra_decompile_checks/haruspex_libhistory.so.8.2/copy_history_entry_00102ed4.c
Semgrep_rule: external.semgrep-rules-0xdea.c.raptor-insecure-api-strcpy-strcat
Issue_description:
A buffer overflow condition exists when a program attempts to put more data in a buffer than it can hold, or when a program attempts to put data in a memory area outside of the boundaries of a buffer.
                                                                                                                                                                                                                                                                                                                                     sVar3 = strlen(pcVar2);
                                                                                                                                                                                                                                                                                                                                     pcVar2 = (char *)xmalloc(sVar3 + 1)
                                                                                                                                                                                                                                                                                                                                     pcVar2 = strcpy(pcVar2, (char *)param 1[1])
           16 - pcVar2 = strcpy(pcVar2, (char *)param_1[1])
                                                                                                                                                                                                                                                                                                                               uVar4 = param_1[2];
*(char **)(lVar1 + 8) = pcVar2;
*(undefined8 *)(lVar1 + 0x10) = uVar4;
[+] Identified source function: /logs/s16_ghidra_decompile_checks/haruspex_libhistory.so.8.2/copy_history_entry_00102ed4.c
Semgrep_rule: external.semgrep-rules-0xdea.c.raptor-interesting-api-calls
Issue description:
Locate all calls to interesting and potentially insecure API functions (candidate points). The auditor can backtrace from these candidate points to find pathways allowing access from untrusted input.
                                                                                                                                                                                                                                                                                                                                return lVar1;
           16 - pcVar2 = strcpy(pcVar2, (char *)param 1[1])
                                                                                                                                                                                                                                                                                                                          return 0;
```

#### SBOMについて

- Cyclone DX SBOMを出力する
- 結果を見るとCyclone DXのroot-level elementsであるmetadata, components, dependencies, vulnerabilitiesの4つがある
- NTIAが定義するSBOM最小要素は満たしていると思われる



#### SBOMについて

- 出力フォーマットもXML,JSON,PROTOBUF,CSVと多数対応
- CycloneDXからSPDXにも変換可能
- SPDXに変換した場合はauthorやコンポーネントの依存関係(Dependency Relationship)は失われる

#### NTIA SBOM最小要素

Supplier Name

Component Name

Version of the Component

Other Unique Identifiers

Dependency Relationship

Author of SBOM Data

Timestamp

```
"spdxVersion": "SPDX-2.2",
"creationInfo": {
 "comment": "This SPDX document has been converted from CycloneDX format.",
 "created": "2025-09-09T19:29:117"
                                                                                       packages
 "SPDXID": "SPDXRef-b26d5802-cf47-4830-8087-7d0f4711a3bb",
  "checksums": [ ···
  "copyrightText": "NOASSERTION",
  "description": "EMBA SBOM-group: debian pkg mgmt - name: openssl - version: 3.0.16-1~deb12u1\u002Brpt1 - descrip
  "downloadLocation": "NOASSERTION",
  "homepage": "NOASSERTION",
  "licenseConcluded": "NOASSERTION",
  "licenseDeclared": "NOASSERTION",
 "licenseInfoFromFiles": [
   "NOASSERTION"
 "name": "openssl",
 "originator": "NOASSERTION",
"supplier": "Person: Debian OpenSSL Team \u003Cpkg-openssl-devel@alioth-lists.debian.net\u003E ()",
"versionInfo": "3.0.16-1~deb12u1\u002Brpt1"
```

#### まとめと感想

- EMBAは組込みLinuxに対する脆弱性解析をメインとしたツール
- デコンパイルやエミュレーション技術、OSSを活用して解析を行っている
- SPDXの変換には課題はあるが、今後改善していく可能性はある

#### 感想

- バイナリのみで脆弱性の検出とSBOMの生成を実現しようとしているオープンソースのツールは珍しい
- ソースコードもSBOMも提供されないソフトウェアに対して独自にSBOMを生成する手段の一つとなる
- 提供されたSBOMの妥当性を検証する手段としても利用できる
- 今回の発表はEMBAを使ってみただけで、結果の正確性については確認していません

# ご清聴ありがとうございました