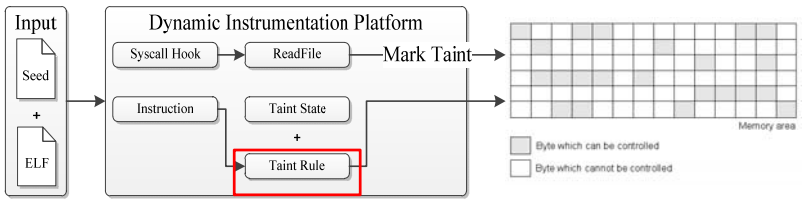
# One Engine To Serveém All: Inferring Taint Rules Without Architectural Semantics

Zheng Leong Chua*, Yanhao Wang*, Teodora Baluta, Prateek Saxena,
Zhenkai Liang, **Purui Su** (purui@iscas.ac.cn)

*NDSS 2019 : Network and Distributed System Security Symposium*

## 污点分析




- 污点分析分析是**追踪**目标数据沿程序执行路径流动过程的技术
- 污点分析主要被用作漏洞机理分析，模糊测试和恶意代码分析等程序分析环节
- 污点传播规则是决定污点分析技术**准确性**的关键

## 污点规则设计缺陷

- 目前的制定和实现方法是通过人工读取指令架构的开发者手册、理解语义后再转换为污点规则，这种方法的缺陷明显
  - 实现效率低：对人工专业技术要求高，设计实现工作繁琐
  - 准确性极低：校验困难，修复困难
  - 平台扩展性差：无法应对多平台，多指令集
  - 无法处理特殊指令：and eax，0x0

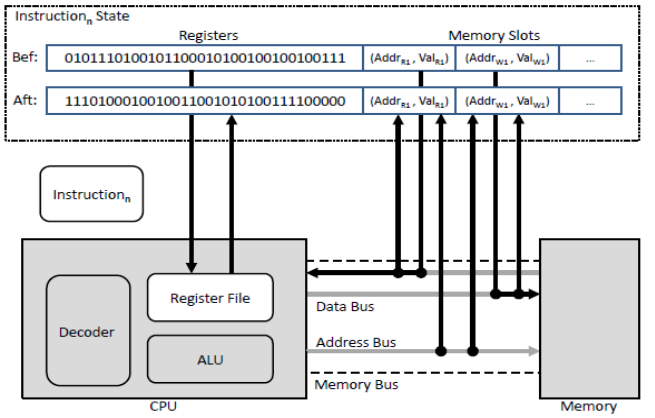
```
if (True) { T[eax][0:8]   = T[eax][0:8];
            T[eax][8:16]  = T[eax][8:16];
            T[eax][16:24] = T[eax][16:24];
            T[eax][24:32] = T[eax][24:32];}
```

过污染错误

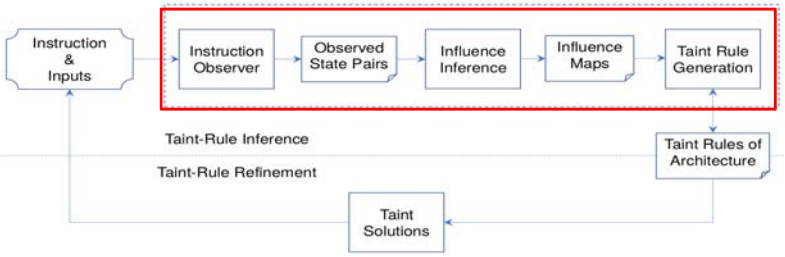
Code 4. libdft : and eax, 0xff

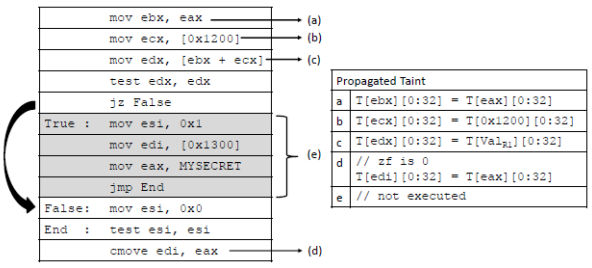
## 方案概述

- 污点传播规则
  - 污点规则本质是描述指令行为对系统状态（CPU及内存）造成的影响
  - 污点规则可等价为指令源和目的操作数之间的**数据流映射关系**
- 污点规则自动生成方案设计思路
  - 通过指令级模糊测试技术尽可能覆盖指令的所有行为，探测指令的源和目的操作数，结合算法分析指令源操作数对目的操作数各bit位的数据流影响关系。
  - 尽可能少的依赖指令架构的先验知识 （提供跨平台支持）
  - 不依赖于任何人工对指令语义的理解 （保证自动化，准确率及扩展）
- 关键技术问题
  - 直接数据流依赖关系提取：简单指令规则提取
  - 控制依赖关系提取：条件指令规则提取
  - 污点传播规则完整性优化




## 系统结构




## 真实漏洞分析结果

TABLE I SUMMARY OF CVES. NUM IS NUMBER OF INSTRUCTIONS. RCE IS REMOTE CODE EXECUTION, S-OF IS STACK OVERFLOW, I-D IS INTEGER DIVISION-BY-ZERO, I-U IS INTEGER UNDERFLOW, FP-D IS FLOATING-POINT DIVISION-BY-ZERO, HC IS HEAP CORRUPTION. * REPRESENTS CVES WHICH HAVE INDIRECT DATA PROPAGATION.

| CVE | Prog | Type | Num | OS |
|---|---|---|---|---|
| CA-1999-14 | bind | RCE | 857915 | Linux |
| CA-1999-14 | bind | I-U | 866934 | Linux |
| CVE-1999-0009 | bind | RCE | 230825 | Linux |
| CVE-2001-0013 | bind | RCE | 216774 | Linux |
| CA-2003-07 | sendmail | RCE | 82999 | Linux |
| CVE-1999-0131 | sendmail | S-OF | 920086 | Linux |
| CVE-1999-0206 | sendmail | RCE | 90918 | Linux |
| CVE-1999-0047* | sendmail | RCE | 192953 | Linux |
| CA-2003-12* | sendmail | RCE | 200018 | Linux |
| CVE-2001-0653 | sendmail | I-U | 76049 | Linux |
| CVE-2002-0906 | sendmail | RCE | 106421 | Linux |
| CVE-1999-0878 | wu-ftpd | RCE | 168604 | Linux |
| CAN-2003-0466 | wu-ftpd | RCE | 98976 | Linux |
| CVE-1999-0368 | wu-ftpd | RCE | 185949 | Linux |
| CVE-2003-0352 | rpcss | RCE | 45328 | WinXP |
| CVE-2002-0649 | mssql | RCE | 213584 | WinXP |
| CVE-2002-0649 | mssql | RCE | 551212 | Win2k |
| CVE-2002-1816 | atphttpd | RCE | 168119 | Linux |
| CVE-2001-0414 | ntpd | RCE | 26100 | Linux |
| CVE-2003-0201 | smbd | RCE | 623815 | Linux |
| CVE-2002-1816 | ghttpd | RCE | 48398 | Linux |
| CVE-2015-6031 | miniupnp | S-OF | 358896 | Linux |
| CVE-2016-9112 | openjpeg2 | I-DIV | 614908 | Linux |
| CVE-2013-4788 | glibc | S-OF | 9725 | Linux |
| CVE-2017-14245 | libsndfile | FP-DIV | 121700 | Linux |
| CVE-2017-7476 | gnulib | HC | 367930 | Linux |

TaintInduce能够准确定位漏洞位置

## 分析精度对比

| Binary | libdft Trace Total | libdft Unique | libdft Tainted | libdft Mismatch T | libdft Mismatch U | libdft Reason for Mismatch: Impl Rules T | Impl Rules U | Ins Supp T | Ins Supp U | Gen Rules T | Gen Rules U | Triton Trace Total | Triton Unique | Triton Tainted | Triton Mismatch T | Triton Mismatch U | Triton Reason for Mismatch: Impl Rules T | Impl Rules U | Ins Gr T | Ins Gr U | TI Ins Samples T | TI Ins Samples U | Gen Rules T | Gen Rules U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| base64 | 283132 | 6244 | 42071 | 10660 | 33 | 6110 | 16 | 4482 | 9 | 68 | 8 | 281159 | 6182 | 42194 | 11267 | 150 | 149 | 5 | 10688 | 37 | 281 | 93 | 149 | 15 |
| who | 2031615 | 19175 | 549350 | 201757 | 94 | 114592 | 79 | 85842 | 12 | 1323 | 3 | 320765 | 6699 | 45284 | 13908 | 158 | 212 | 8 | 13380 | 82 | 213 | 61 | 103 | 7 |
| uniq | 2097151 | 6242 | 932802 | 69476 | 56 | 63889 | 36 | 4270 | 17 | 1317 | 3 | 326395 | 5527 | 9655 | 388 | 6 | 0 | 0 | 388 | 6 | 0 | 0 | 0 | 0 |
| md5sum | 2670592 | 7712 | 11886 | 482 | 69 | 295 | 40 | 135 | 17 | 52 | 12 | 310780 | 6042 | 375 | 94 | 7 | 45 | 2 | 25 | 3 | 24 | 2 | 0 | 0 |
| tiffsplit | 655359 | 6434 | 339112 | 749 | 51 | 710 | 35 | 18 | 8 | 21 | 8 | 515325 | 5888 | 110503 | 859 | 125 | 416 | 15 | 280 | 70 | 147 | 35 | 16 | 5 |
| tiff2pdf | 1048575 | 8854 | 648397 | 2771 | 65 | 2632 | 46 | 106 | 12 | 33 | 7 | 550396 | 5377 | 120306 | 867 | 87 | 545 | 9 | 182 | 42 | 139 | 35 | 1 | 1 |
| tiff2rgba | 2684353 | 6697 | 2329387 | 96191 | 88 | 80687 | 67 | 15300 | 13 | 204 | 8 | 523517 | 6223 | 78515 | 1688 | 147 | 300 | 16 | 767 | 86 | 593 | 38 | 28 | 7 |
| bmp2tiff | 2162687 | 6061 | 1774856 | 103814 | 29 | 103762 | 15 | 17 | 7 | 35 | 7 | 527613 | 5259 | 79874 | 4790 | 50 | 52 | 11 | 4709 | 26 | 15 | 6 | 14 | 7 |
| objdump | 2682463 | 10568 | 677219 | 12255 | 116 | 8593 | 79 | 3325 | 25 | 337 | 12 | 433370 | 5539 | 18614 | 426 | 40 | 120 | 8 | 222 | 23 | 84 | 9 | 0 | 0 |
| readelf | 1106687 | 8413 | 233400 | 9060 | 60 | 6480 | 38 | 2506 | 15 | 74 | 7 | 324622 | 6249 | 1257 | 2 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 |
| **Total** | **17422614** | **31050** | **7538480** | **507215** | **661** | **387750** | **451** | **116001** | **135** | **3464** | **75** | **4113942** | **12987** | **506577** | **34289** | **771** | **1839** | **74** | **30643** | **376** | **1496** | **279** | **311** | **42** |

TaintInduece与对比污点工具93.27%的污点分析结果一致，不一致的结果中仅有0.28%是由于TaintInduce造成，且这部分可以通过完整性检查解决，剩下的不一致都是由于对比工具的规则错误及指令支持异常造成的

## 跨平台性

| Architecture | Type of Instructions: Arith S | Arith P | Arith T | Comp S | Comp P | Comp T | Jump S | Jump P | Jump T | Mov S | Mov P | Mov T | Cond S | Cond P | Cond T | FPU S | FPU P | FPU T | SIMD S | SIMD P | SIMD T | MISC S | MISC P | MISC T | Instruction set: Sound | Support | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **x86** | 28 | 43 | 43 | 0 | 8 | 8 | 33 | 33 | 33 | 32 | 32 | 32 | 60 | 60 | 60 | 59 | 83 | 86 | 176 | 251 | 259 | 23 | 26 | 28 | 411 | 536 | 549 |
| **x64** | 23 | 38 | 38 | 0 | 10 | 10 | 35 | 35 | 35 | 39 | 39 | 39 | 60 | 60 | 60 | 59 | 83 | 86 | 176 | 251 | 259 | 23 | 26 | 29 | 415 | 542 | 556 |
| **Aarch64** | 39 | 64 | 64 | 0 | 3 | 3 | 3 | 3 | 3 | 43 | 43 | 43 | 11 | 11 | 11 | 18 | 37 | 37 | 61 | 188 | 196 | 13 | 13 | 13 | 188 | 362 | 370 |
| **MIPS-I** | 22 | 26 | 26 | 4 | 4 | 4 | 7 | 7 | 7 | 14 | 14 | 14 | - | | | - | | | - | | | 1 | 1 | 1 | 48 | 52 | 52 |