从零打造简单的 SODUMP 工具

Author: ThomasKing

最近翻看之前的帖子,发现基于 linker init_array 加密的 SO 文件的静态分析稍微麻烦。虽然原理很清楚,但是需要 dump 之后再进行 section 修复才能放入 ida。可以看到,上述两步骤其实很机械。那么应该可以实现一个自动化工具,帮助我们解决上述问题,让我们可以精力专注于其他地方,提高效率。实现上述工具需要解决两个问题: 1> 应对各种加密算法 2> section 重建。Section 重建在 http://bbs.pediy.com/showthread.php?t=192874 已经讨论,就不再赘述,仅仅改进一些不足之处。

一、应对各种加密算法策略

在 http://bbs.pediy.com/showthread.php?t=192020&viewgoodnees=1&prefixid= 等帖子中提到的 SO 文件都基于 init_array 实现。为了静态分析,不得不分析出算法,并实现对 SO 解密,以便静态分析。而且,不同的 SO 采用不同的加密算法。这个步骤时间的花费,取决于分析人员对各种常用的加密算法 RC4、TED 等了解程度,这是比较纠结的事情。

那如何应对各种加密算法,成为工具实现的关键。直接实现各种算法,这个显然是不行的。那只能通过某种方法来绕过。想了很久,也没有很好的解决策略。不过有一点,熟悉 SO 机制的读者都知道,linker 可是应对自如!那看来只能借 linker 之手,来帮助我们实现。

Linker 除了在程序运行初加载外,还有就是通过 dl 函数会调用,即 dlopen、dlsym、dlerr和 dlclose。啰嗦下 dlopen和 dlsym 函数原型:

void * dlopen(const char * pathname, int mode);

void*dlsym(void*handle,constchar*symbol);

通常,使用 dlopen 打开加载某 SO 后,会返回一个 handle 对象,然后根据 handle 对象调用 dlsym 查找某函数符号,实现调用。仔细分析可以发现,这个过程和程序在运行初是类似的;另外,linker 只会加载某一 SO 一次,那么 linker 应该维护了一个数据表来记录。进一步分析还可以发现,这个 void *handle 对象应该指向了当前打开 SO 的数据对象。有了这个思路,翻看 dl 函数源码(位于: \bionic\linker\dlfcn.c,不是在 bionic\libdl\libdl\c),dlopen 源码:

```
void *dlopen(const char *filename, int flag)
{
    soinfo *ret;
    pthread_mutex_lock(&dl_lock);
    ret = find_library(filename);
    if (unlikely(ret == NULL)) {
        set_dlerror(DL_ERR_CANNOT_LOAD_LIBRARY);
    } else {
        ret->refcount++;
    }
    pthread_mutex_unlock(&dl_lock);
    return ret;
}
```

标红部分可以看到, ret 就是 handle。翻看 linker 源码, 其实际为: soinfo 结构体(截取部分结构)

```
struct soinfo
    const char name[SOINFO_NAME_LEN];
    Elf32_Phdr *phdr; //Elf32_Phdr 实际内存地址
    int phnum;
    unsigned entry;
    unsigned base; //SO 起始
    unsigned size; //内存对齐后占用大小
    int unused; // DO NOT USE, maintained for compatibility.
    unsigned *dynamic; //.dynamic 实际内存地址
    unsigned wrprotect_start; //mprotect 调用
    unsigned wrprotect_end;
    soinfo *next; //下一个 soinfo
    unsigned flags;
    const char *strtab; //.strtab 实际内存地址
    Elf32_Sym *symtab; //. symtab 实际内存地址
    //hash 起始位置: bucket - 2 * sizeof(int)
    unsigned nbucket; //size = nbucket * sizeof(int)
    unsigned nchain; //size = nchain * sizeof(int)
    unsigned *bucket;
    unsigned *chain;
    unsigned *plt got; //对应.dynamic: DT PLTGOT
    Elf32 Rel *plt rel; //函数重定位表
    unsigned plt_rel_count;
    Elf32_Rel *rel; //符号重定位表
    unsigned rel_count;
};
```

从这个结构中,已经可以得到各种信息,直接用于后续重建。另外,还有一点: dlopen 返回 soinfo 时,linker 已经执行了 init_array 中的函数。换句话说,已经实现了自解密,直接 dump 就是 OK。

二、改进 PLT 和 GOT 重建

{

在 http://bbs.pediy.com/showthread.php?t=192874 帖子中,PLT 和 GOT section 并不能直 接从.dynamic 中获取。重建时,通过 section 之间的排列关系,间接的进行修正。如果想对 位置变化,这种重建是无意义的。影响相对位置变化主要有:链接脚本的细微区别和 SO 经 过 DIY。比如,不同版本的 ndk 存在细微差异:

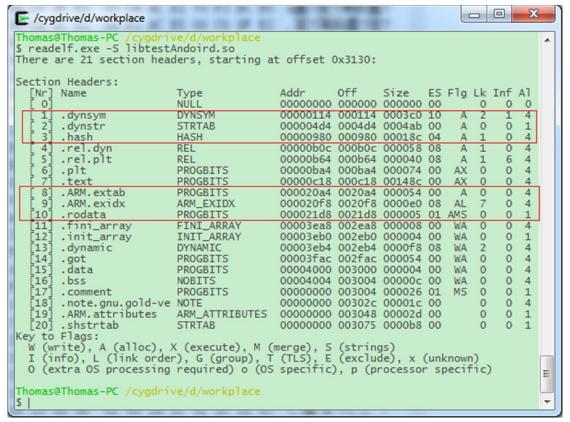


图 1

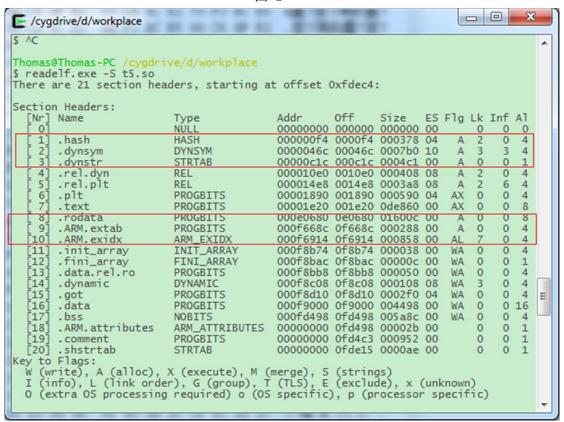


图 2

*(.igot.plt) *(.got) *(.igot) } 就是函数符号在前,其他符号在后,和之前讨论的是相反的。由于存在上述原因,对 PLT和 GOT的重建进行如下改进。

2.1 PLT section

由于优化, PLT 结构前四项是固定代码。

```
AREA .plt, CODE
; ORG 0xEB8

CODE32

STR    LR, [SP,#-4]!

LDR    LR, =(_GLOBAL_OFFSET_TABLE_ - 0xEC8)

ADD    LR, PC, LR; _GLOBAL_OFFSET_TABLE_

LDR    PC, [LR,#8]!
```

借鉴 window 内核中搜索未导出符号的思路,通过搜索前 16 个字节来确定 plt。恰好这里刚好够 16 个字节(真是无巧不成书),即:

static unsigned g_plt_start[] = {

0xe52de004, 0xe59fe004, 0xe08fe00e, 0xe5bef008 };

这样就可以得到 plt section 的起始地址。Size 前面帖子中已经说明,这里就不在赘述。

2.2 GOT section

前面已经提到 GOT 表存在两种结构。另外,.got 与.rel 的对应关系并不像.got.plt 与.rel.plt 的对应关系那么稳定,受到诸如指针间址寻址影响。但始终有一点:.got 中的项一定出现在.rel section 中。那么重建的思路就是:

Step1: 读取 DT_PLTGOT,获取__global_offset_table__地址,记为: plt_got

Step2: 读取 plt_got – 4 地址的数值, 在.rel 表中进行搜索。如果匹配, 说明 GOT 结构式: {.got, .got.plt}, 转 3.1;否则为{.got.plt, .got}转 3.2

Step3.1: 继续向前搜索,直到搜索到起始位置。

Step3.2: 调整搜索起始位置: plt_got + 4 * (3 + rel_plt_count),向后搜索,搜索到末尾。这里,存在一种情况: 若下面.data 也有重定位项且处于.data 起始连续位置,也会被搜索到。但对于.got 和.data 都对应到.rel 重定位中,重定位方式相同,统一处理也是合理的,实测不存在问题。

通过搜索,即可准确获取 GOT 起始地址和长度,完成重建。

另外,对于.data 和.rodata 的重建目前没有很好的思路,仅只通过相对位置重定位。如果有更好的思路,请告知我学习学习,衷心感谢。

三、SODUMP 出炉

经过上面讨论,SODUMP 工具基本成型,剩下的就是添砖添瓦,做成一个 apk。通过 EditText 和 Button 配合,将待 dumpSO 进行处理,记得使用 dlclose 关闭,最好使用 dlerr 把错误信息打印出来便于定位问题。废话不多说,上几张测试图:

```
LOAD:0000DC34; End of function _Unwind_GetTextRelBase
LOAD:0000DC34
 LOAD:0000DC34
                          EXPORT __gnu_armfini_31
_gnu_armfini_31 DCD 8x97A29A3E, 8x57E9111A, 8x39D872CC, 8x53427AE8, 8x84D519ED
 LOAD:0000DC38
LOAD:0000DC38
LOAD:0000DC38
                                                                                             ; DATA XREF: LOAD:off FEF41
 LOAD:0000DC38
                                                   DCD 0x27F9B5F2, 0x7093CB09, 0x93A233E3, 0x27663FB4, 0xE1844159
DCD 0x46C5B7A7, 0x91C9FC7C, 0x2EF2827F, 0xC4E6B07A, 0x54E97938
 LOAD:0000DC38
LOAD:0000DC38
                                                   DCD 0xC7C95244, 0xCa11296F, 0x7D218BA0, 0xD4BE3C30, 0x363F07B4
DCD 0x7EC9AF28, 0x392921BC, 0x25ABA7CF, 0xF01B222, 0xA806AF84
                                                   DCD 0x753E511F, 0xA4175F42, 0xEBE832B3, 0xCCCC4E9F, 0xD53495C
DCD 0xFE9382A1, 0x9762A449, 0xAD08E8F4, 0x2C335457, 0x5FDD25E2
 LOAD:0000DC38
 LOAD:0000DC38
LOAD:0000DC38
LOAD:0000DC38
                                                   DCD 0x564EA3C0, 0x6BE0D39A, 0x860B4AF3, 0x9F285CC2, 0x120D1415
DCD 0xD541F0C9, 0x4146DD1D, 0x61636776, 0x3228CCEE, 0x67DB33CC
                                                   DCD 9x714485C7, 9xD518AF41, 9x55EA2285, 9xD682A46B, 9x3E22779F
DCD 9xF8A64A7A, 9x55F8DCDE, 9x1E7981D6, 9xFF89C2EB, 9x7DD0C9C2
DCD 9xDFECB6F5, 9x8F689589, 9xE85AF8E6, 9x2B97C5DC, 9xE2B96377
DCD 9x9138443D, 9x42297A4F, 9x44C82712, 9xFC1988F3, 9x5DC2A716
 OAD:0000DC38
 LOAD:0000DC38
 LOAD:0000DC38
 OAD:0000DC38
                                                   DCD 0x6916B296, 0xC75E98E1, 0x8D05DE66, 0xBAD8A873, 0x5B979916
DCD 0xF1E6913C, 0x66A07538, 0x3C321209, 0x4794AF28, 0x433C8051
 nan-aggancas
 LOAD:0000DC38
                                                   DCD 0x3B98EAE9, 0xEBDEBDAE, 0xFEED3BEB, 0xB380F84E, 0x3C50E819
DCD 0xF9C8CE0E, 0x4C97DD2A, 0x5837CB33, 0x40903834, 0xFFFD8052
 LOAD:0000DC38
                                                   DCD 0xF682E42A, 0xABDAF27, 0x6A0B2421, 0x92A89191, 0x195914CE
DCD 0x629F215F, 0x5EE403BB
 LOAD:0000DC38
LOAD:0000DDA8 :
```

图 4 脱壳前

```
.text&.ARM.extab:0000DC38
text&.ARM.extab:0000DC38
.text&.ARM.extab:0000DC38
                                            EXPORT __gnu_armfini_31
.text&.ARM.extab:0000DC38
                             gnu armfini 31
                                                                      ; CODE XREF: gnu arm fini 05+58ip
.text&.ARM.extab:0000DC38
                                            STMFD
                                                             SP!, {R4,LR}
.text&_ARM.extab:0000DC3C
                                            MOV
                                                             R4, R0
R0, #2
.text&.ARM.extab:0000DC40
                                                             __gnu_armfini_09
R3, =(off_FEB4 - 0xDC54)
.text&.ARM.extab:0000DC44
                                            RI
                                            LDR
.text&.ARM.extab:0000DC48
.text&.ARM.extab:0000DC4C
                                                             R3, [PC,R3] ; off_FEB4
.text&.ARM.extab:0000DC50
                                            STR
                                                             R3, [R4]
text&.ARM.extab:0000DC54
                                                             SP!, {R4,PC}
                                            LDMFD
                           ; End of function __gnu_armfini_31
text&.ARM.extab:0000DC54
.text&.ARM.extab:0000DC54
.text&.ARM.extab:0000DC54
text&.ARM.extab:0000DC58
                           off_DC58
                                            DCD off_FEB4 - 0xDC54 ; DATA XREF: __gnu_armfini_31+101r
_text&_ARM_extah:00000C5C
```

图 5 脱壳后

一朋友给的某手游游戏,据说是美杜莎壳。当然,这仅仅只是该壳子的一个很小方面,这壳子很生猛的。

```
LOAD:00007BEC
LOAD:00007BEC
                                                                                  EXPORT _dhm_iWweXTSblxDrJH_LYxAJFjMaJYeNBUHdJKbHQbM_TttSKS_y
_dhm_iWweXTSblxDrJH_LYxAJFjMaJYeNBUHdJKbHQbM_TttSKS_y ; [dhm[iWweXTSblxDrJH_LYxAJFjMaJYeNBUHdJKbHQbM[TttSKS[y
USADAB R11, LR, R11, R12
 LOAD:00007BEC
LOAD:00007BEC
                                                                                                                                                                         USADA8
R11, LR, R11, R12

DCD 0x176E0333, 0x8CBA0788, 0x3E0810F, 0xD0878F3D, 0xC327872B
DCD 0x77F31DF7, 0x673FAB01, 0x6D366666, 0x337EC2B7, 0x8FC6DFA1
DCD 0x50F30BB6, 0x8B959A738, 0xEDDA6684, 0x078538BF, 0x81137F1F
DCD 0x8D7342DF, 0x1348BACF, 0x1727ED08, 0xD8BF487F, 0xD8A177B
DCD 0xED32ABF, 0x2774D33B, 0x15D71C88, 0x1793C056, 0x4227B797
DCD 0xED81423B, 0x3EC70D08, 0x3D0928B, 0x13CD8B84, 0xF3C0E7C2
DCD 0xE5187DDD, 0xCB138D04, 0x8E33D7AE, 0x8BC2EF8R, 0x46C0A727
DCD 0xE08CF704, 0x4A0736F3, 0x786C150B, 0xF573AM83, 0x81435D1BDCD
0x7B72865A, 0x3848EF6B, 0x6EF4CA9, 0x8B33986C, 0x170R09EF
DCD 0xF08A086C3, 0x5918B18D60, 0x8B330B02AE, 0x6A08FC373, 0x70718B34
DCD 0x1019726BB, 0xEF000AD4, 0x94336DEC, 0x40BC373, 0x70718B34
DCD 0x50432BF6, 0x8B8B870B, 0x7F37AB31, 0xF6CD24EC
DC 0x6055393, 0x16FB0370, 0x2B8FE6DAD2
DC 0x5055393, 0x16FB0370, 0x8B75363BS, 0xB848F74, 0x086B6B27
DCD 0x76B55393, 0x16FB037, 0x8B75363BS, 0xB848F74, 0x08B6E27
DCD 0x57BADC13, 0x13F3A5E3, 0x8D3674F, 0x58BFCC, 0x8F0C250B
DCD 0x58BC5139, 0x15F803F8, 0x8F9563BS, 0xB848F74, 0x0B36BB7
DCD 0x57BADC13, 0x13F3A5E3, 0x8D3674F, 0x5334EB57, 0xF79CB60B
DCD 0x58B6634D, 0xC238B1AB, 0x5F6F6F87, 0x58737DC6A, 0xD80B36BC
DCD 0x58A68BB0, 0x128B1AB, 0x5F6F5B7, 0x8737DC6A, 0xD80B36BC
DCD 0x58A68BB0, 0x128B1AB, 0x2F66F5B7, 0x68737DC6A, 0xD80B3BC
DCD 0x58A68BB0, 0x128B1AB, 0x2F66F5B7, 0x8737DC6A, 0xD80B3BC
DCD 0x18B9AFF, 0xC23B31AB, 0x2F66F5B7, 0x8737DC6A, 0xD80B3BC
DCD 0x18B9AFF, 0xC80B556F, 0x933ADABA, 0x6F4B6F67B, 0x575736DB
DCD 0x18176CBF, 0x20B756FB, 0x803BADAB, 0x6F4B6F67B, 0x575736DB
DCD 0x18176CBF, 0x20B756FB, 0x803BADAB, 0x6F4B6F67B, 0x575736DB
DCD 0x18176CBF, 0x20B756FB, 0x503BADAB, 0x6F4B6F67B, 0x575735DB
DCD 0x18176CBF, 0x20B756FB, 0x503BADAB, 0x6F4B6F67B, 0x575735DB
DCD 0x18176CBF, 0x20B756FB, 0x503BADAB, 0x6F4B6
 LOAD: 00007BEC
  LOAD:00007BF0
 LOAD:00007BF0
  LOAD:00007BF0
 LOAD:00007BF0
LOAD:00007BF0
LOAD:00007BF0
  LOAD:00007BF0
    LOAD: 00007BF0
  LOAD:00007BF0
LOAD:00007BF0
LOAD:00007BF0
  LOAD:00007BF0
     0 AD : 88887 BF 8
 LOAD: 00007BF0
LOAD: 00007BF0
LOAD: 00007BF0
  LOAD:00007BF0
  LOAD:00007BF0
    LOAD:00007BF0
  LOAD:00007BF0
  LOAD:00007BF0
    nan-nanazrea
                                                                                                                                                                                DCD 0xD36D29DC
```

图 6 脱壳前

```
.text&.ARM.extab:00007BEC
                                                                                         EXPORT _dHm_iWweXTSblxDrJH_LYxAJFjHaJYeNBUHdJKbHQbM_TttSKS_y
_dHm_iWweXTSblxDrJH_LYxAJFjHaJYeNBUHdJKbHQbM_TttSKS_y
; CODE XREF: sub_DB88+649lp
; sub_D888+9649lp
...
STMFD SP!, {R4,R5}; [dHm[iWweXTSblxDrJH_LYxAJFjHaJYeNBUHdJKbHQbM[TttSblxDrJH_LYxAJFjHaJYeNBUHdJKbHQbM[TttSblxDrJH_LYxAJFjHaJYeNBUHdJKbHQbM[TttSblxDrJH_LYxAJFjHaJYeNBUHdJKbHQbM]
.text&.ARM.extab:00007BEC
.text&.ARM.extab:00007BEC
.text&.ARM.extab:00007BEC
.text&.ARM.extab:00007BEC
.text&.ARM.extab:00007BEC
                                                                                                                                                                                                    ; sub_D888+
SP!, (R4,R5); [dhm
R3, [R1,#2]
R5, [R1,#1]
R4, [R1]
R12, [R1,#3]
R3, R3, L5L#16
R3, R3, R5, L5L#8
R3, R3, R4, R5, L5L#24
R3, [R0,#4]
R3, [R1,#6]
R5, [R1,#6]
R5, [R1,#6]
R1, [R1,#7]
R3, R3, L5L#16
R3, R3, R4, S, L5L#24
R3, [R0,#8]
R3, R3, R12, L5L#24
R3, [R0,#8]
R5, [R1,#9]
R4, [R1,#9]
R4, [R1,#8]
R3, R3, L5L#16
R3, R3, R12, L5L#24
R3, R3, R12, L5L#24
R3, R3, R12, L5L#8
R3, R3, R12, L5L#24
R3, R1, #8x6]
R3, R1, #8x6]
R3, R1, #8x6]
.text&_ARM.extab:00007BF0
                                                                                                                                               LDRB
LDRB
LDRB
LDRB
.text&.ARM.extab:00007BF4
.text&.ARM.extab:00007BF8
.text&.ARM.extab:00007BFC
.text&.ARM.extab:00007C00
.text&.ARM.extab:00007C04
.text&.ARM.extab:00007C08
                                                                                                                                               ORR
.text&.ARM.extab:00007C0C
.text&.ARM.extab:00007C10
                                                                                                                                               ORR
STR
.text&.ARM.extab:00007C14
                                                                                                                                               LDRB
.text&.ARM.extab:00007C18
.text&.ARM.extab:00007C1C
.text&.ARM.extab:00007C20
.text&.ARM.extab:00007C24
.text&.ARM.extab:00007C28
                                                                                                                                               LDRB
                                                                                                                                               MOV
ORR
.text&.ARM.extab:00007C2C
.text&.ARM.extab:00007C30
.text&.ARM.extab:00007C34
.text&.ARM.extab:00007C38
                                                                                                                                               nrr
                                                                                                                                               LDRB
.text&.ARM.extab:00007C3C
.text&.ARM.extab:00007C40
                                                                                                                                               LDRB
LDRB
LDRB
.text&.ARM.extab:00007C44
.text&.ARM.extab:00007C48
.text&.ARM.extab:00007C4C
                                                                                                                                               MOV
ORR
.text&.ARM.extab:00007C50
.text&.ARM.extab:00007C54
.text&.ARM.extab:00007C58
                                                                                                                                               ORR
                                                                                                                                               ORR
STR
.text&.ARM.extab:00007C5C
                                                                                                                                               LDRB
```

图 7 脱壳后

四、参考文献

Linker.c、dlfcn.c 源码

http://bbs.pediy.com/showthread.php?t=192874

http://bbs.pediy.com/showthread.php?t=190384&viewgoodnees=1&prefixid=

http://bbs.pediy.com/showthread.php?t=193720&viewgoodnees=1&prefixid=