OS の保護機構

~ *secure level* ~

Hideyuki KURASHINA

rushani@bl.mmtr.or.jp

What's "secure level"?

- 目的
  特権ユーザが「できる」ことを制限する。
- 仕組み
  - 特権ユーザプロセスのみが secure level を上げることができる (via sysctl 変数 kern.securelevel)。
  - ただし、secure level を下げることができるのは PID が 1 のプロセス (すなわち、init(8)) のみ。
- 歴史
  起源はどこかまで調べてないけれど、少なくとも 4.4BSD Lite2 (1995/06) には secure level が実装されている。

本稿では NetBSD 2.0 を対象とする。

Level -1

恒久的に安全ではないモード

- システムは level 0 のモードで動作する。
  - システムがマルチユーザモードで動作するときであっても、secure level は 0 になる。
- GENERIC カーネルを使用した時のデフォルト。
  - GENERIC カーネルコンフィギュレーションファイルには、secure level を -1 に固定する INSECURE オプションが設定されているため。

Level 0

安全ではないモード

- immutable (変更不可) フラグ、append-only (追記専用) フラグを解除できる。
- デバイスをパーミッション通りに読み書きできる。

Level 1

安全なモード

- システムの immutable フラグおよびシステムの append-only フラグを解除できなくなる。
- マウントされているファイルシステムのディスクデバイスが読み込み専用になる。
- /dev/mem および /dev/kmem が読み込み専用になる。
  - 実用上、これで困る代表的なアプリケーションが X サーバ。NetBSD では、aperture ドライバ (pkgsrc/sysutils/aperture) を使用して回避することができる。

Level 1 (Cont.)

- カーネルに格納された verified exec の指紋テーブルを変更できなくなる。
- sysctl 変数 kern.rtc_offset, net.inet.ip.forwsrcrt を変更できなくなる。
- LKM を読み込めなくなる。

Level 2

高度に安全なモード

Level 1 の制限事項が適用されることに加えて、

- マウントされているか否か関わらず、ディスクデバイスが常に読み込み専用になる。
- ディスクを新しくマウントすることができなくなる。
- ディスクの状態を読み込み専用の状態から、読み書き可能な状態に変更できなくなる。
- multi-user モードで動作している時に、ファイルシステムに対して newfs(8) を実行できなくなる。

Level 2 (Cont.)

- settimeofday(2) システムコールが時間を進めることしかできなくなる。
- IP フィルタリング機構 (ipf) のフィルタリングルールおよび NAT ルールを更新できなくなる。
- ユーザがプロセスの core ファイルの名前を変更できなくなる (変更できるのは、デフォルトのみ)。
- 高度に安全なモードから安全ではないモードに遷移する、すなわち console にアクセスするために single-user モードに移行する時には、root での認証が必須になる。これは /etc/ttys で console に “secure” マークが付いているかどうかには関係ない。
- その他、arch 固有、デバイス固有の制限が加えられる。

Misc.

- securelevel
  - FreeBSD, NetBSD, OpenBSD で secure level の内容には大きな違いはない。ただし、FreeBSD には Level 3 があり、Level 2 の IP パケットフィルタのルールの更新が分離されている。
- init(8)
  - FreeBSD ... 1 度 secure level を上げたら、2 度と下げられない。
  - NetBSD ... multi-user モードから single-user モードに遷移する際に、secure level が 0 に下がる。
  - OpenBSD ... OpenBSD と同様。
- FreeBSD 5.x 特記事項
  - ホストの環境とは独立して、jail(8) に secure level を設定できる。

Conclusion

secure level を使うと、

- 本気でシステムを破壊しようとしてきた攻撃者に対して、仮に認証機構を突破されてしまっても、一定程度の効果が期待できる。
- 深刻なオペレーションミスの予防をできる (例: うっかり fdisk(8) や disklabel(8) でディスクを破壊する心配はない)。

必要なら、

- ソースコードの中の securelevel 変数を調整して、制限したい内容をどんどん取り込んじゃえ。

References

- init(8)
- lkm(4)
- options(4)
- sysctl(3)
- rc.conf(5)
- /etc/defaults/rc.conf
- /etc/rc.d/securelevel

Appendix: Ralated Files (1/3)

```
src/sbin/init/init.c
src/sys/arch/alpha/alpha/machdep.c
src/sys/arch/amd64/amd64/sys_machdep.c
src/sys/arch/i386/i386/sys_machdep.c
src/sys/arch/news68k/news68k/disksubr.c
src/sys/arch/newsmips/newsmips/disksubr.c
src/sys/arch/xen/i386/sys_machdep.c
src/sys/dev/i2o/dpti.c
src/sys/dev/i2o/iop.c
src/sys/dev/ic/dpt.c
src/sys/dev/ic/icp_ioctl.c
src/sys/dev/ic/mlx.c
src/sys/dev/pci/mly.c
src/sys/dev/pci/twe.c
src/sys/dev/tc/stic.c
```

Appendix: Ralated Files (2/3)

```
src/sys/dev/verified_exec.c
src/sys/dev/wscons/wsdisplay_compat_usl.c
src/sys/dist/ipf/netinet/ip_fil_netbsd.c
src/sys/dist/ipf/netinet/ip_nat.c
src/sys/ipkdb/ipkdb_ipkdb.c
src/sys/kern/init_sysctl.c
src/sys/kern/kern_lkm.c
src/sys/kern/kern_sysctl.c
src/sys/kern/kern_systrace.c
src/sys/kern/kern_time.c
src/sys/kern/kern_verifiedexec.c
src/sys/kern/sys_process.c
src/sys/kern/vfs_syscalls.c
src/sys/kern/vfs_vnops.c
src/sys/miscfs/procfs/procfs_ctl.c
```

Appendix: Ralated Files (3/3)

```
src/sys/miscfs/procfs/procfs_subr.c
src/sys/miscfs/specfs/spec_vnops.c
src/sys/sys/sysctl.h
src/sys/sys/systm.h
src/sys/ufs/ext2fs/ext2fs_vnops.c
```