



A Survey of Unix Init Schemes

Yvan Royon, Stéphane Frénot

► **To cite this version:**

Yvan Royon, Stéphane Frénot. A Survey of Unix Init Schemes. [Technical Report] RT-0338, INRIA. 2007. <inria-00155663v2>

HAL Id: inria-00155663

<https://hal.inria.fr/inria-00155663v2>

Submitted on 20 Jun 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Survey of Unix Init Schemes

Yvan Royon — Stéphane Frénot

N° 0338

June 2007

Thème COM



*R*apport
technique



A Survey of Unix Init Schemes

Yvan Royon, Stéphane Frénot

Thème COM — Systèmes communicants
Projet Ares

Rapport technique n° 0338 — June 2007 — 36 pages

Abstract:

In most modern operating systems, `init` (as in “initialization”) is the program launched by the kernel at boot time. It runs as a daemon and typically has PID 1. `init` is responsible for spawning all other processes and scavenging zombies. It is also responsible for reboot and shutdown operations.

This document describes existing solutions that implement the `init` process and/or `init` scripts in Unix-like systems. These solutions range from the legacy and still-in-use BSD and SystemV schemes, to recent and promising schemes from Ubuntu, Apple, Sun and independent developers. Our goal is to highlight their focus and compare their sets of features.

Key-words: `init` process, `init` scheme

Etat de l'art des processus init

Résumé :

Dans la plupart des systèmes d'exploitation modernes, init (comme "initialisation") est l'application lancée par le noyau au démarrage. Il s'agit d'un démon qui a le PID 1. init a pour tâches de lancer les autres processus et de nettoyer les processus zombies. Il est aussi responsable du démarrage et redémarrage du système.

Ce document décrit les implantations existantes du processus init et des scripts init pour variantes d'Unix. Ces implantations vont des solutions BSD et SystemV, pionnières et toujours utilisées, à celles plus récentes et prometteuses venant de Ubuntu, Apple, Sun, ainsi que de développeurs indépendants. Notre but ici est d'analyser les objectifs de ces implantations, et de comparer leurs fonctionnalités.

Mots-clés : processus init

The Boot Process

What happens between the moment a computer is switched on and the moment a user can log in?

Let us begin by a quick and simplified overview of the boot process in most modern operating systems.

1. The computer is switched on or rebooted;
2. The **BIOS**, stored on the motherboard, is executed from a known location (flash memory);
3. The BIOS determines the boot device (local disc, removable media, PXE from the network);
4. The boot device has a special, known sector called the Master Boot Record, which contains the **stage 1 bootloader**. The stage 1 bootloader is loaded into RAM, and points to the stage 2 bootloader. The latter can be several sectors long, and is stored on the boot device in a location unknown to the BIOS and subject to change;
5. The **stage 2 bootloader** (e.g., LILO or GRUB for Linux) uncompresses and launches the kernel;
6. The **kernel** runs a single application. For Linux, the default is `/sbin/init`, and can be modified in LILO or GRUB by passing the `init=/path/to/executable` parameter to the kernel.

Usually, `init` launches a few `getty` processes for virtual terminals, a few daemons (`cron`, `dhcp`, `acpi`), the X server, etc.

On Unix-like systems, `init` (as in “initialization”) is the program launched by the kernel at boot time. It runs as a daemon and typically has PID 1. `init` is responsible for spawning all other processes and scavenging zombies. It is also responsible for reboot and shutdown operations.

A Bit of History

The first widespread `init` is the simple solution from BSD. Its main competitor is the more featureful but more complex System V `init`.

The BSD `init` process simply launches a list of processes. It basically reads a configuration file (`/etc/rc`) where all programs to launch at startup are statically listed. If the administrator wants to modify this list, she must edit the file manually, which is clearly error-prone and has potentially disastrous effects. This is a very simple and lightweight scheme, but it is completely static.

System V, on the other hand, chose to be more flexible. It introduces the notion of *runlevel*, which can be viewed as multiple and specialized BSD *rc* files. A runlevel has an identifier and a purpose. Runlevel 1 boots into single mode, runlevel 2 boots into multi-user mode, etc. Only a fixed number of eight runlevels are available. What each runlevel does is specified in the `/etc/inittab` file.

A runlevel is a *software configuration* of the system which allows only a selected group of applications to exist. When `init` is requested to change the runlevel, it kills all applications that are not listed in the new runlevel, and starts all unstarted applications in the new runlevel.

The second great improvement in SysVinit is the notion of *rc script*. Each application contained in a runlevel comes with a wrapper script; it is an executable that accepts `start` and `stop` as parameters. This allows to launch and kill each application using the same syntax, regardless of how they are implemented and how they handle parameters internally.

Most often, the combination of *rc scripts* with runlevels makes use of symbolic links. A runlevel is a directory that contains symlinks named with a numbered prefix. When entering a runlevel, symlinks are accessed alphabetically, in the static order defined by their number. This allows to reuse the same *rc script* in multiple runlevels.

Evolution of Init Schemes

Today, most GNU/Linux distributions use a derivative of System V's `init`. Even NetBSD went from the BSD `init` to a SystemV-like `init`. However, there is room for lots of improvements, and lots of proposals exist on several topics.

The first topic is the order in which *rc scripts* are started. In 2002, the `need(8)` scheme proposed to drop statically, manually ordered lists in favour of **dependencies**. Each *rc script* lists the services it needs (e.g., the DHCP daemon needs a network connection), and `init` computes which *rc scripts* must be started, in what order. All `init` schemes except BSD, System V and three minor others have a way to express dependencies.

The second topic is the enhancement of **configuration management**. For instance, Gentoo *rc*, `initNG` and `runit` identify runlevels using a name instead of a number. This breaks the limitation on the number of runlevels, and allows to dedicate a runlevel to a context or a particular usage (e.g., "AC mode" and "battery mode" on a laptop). Similarly, `upstart`, `cinit` and Sun SMF introduce a notion of profile (or "mode" for `einit`).

Another topic is **life cycle management**. There are two main solutions when monitoring the life cycle of applications. The first one is when using *rc scripts*: `init` (or a separate *rc application*) use the `/var` directory to store which daemon has been started, with which PID. If no process with such PID is running, then the daemon has been killed. This is a reactive life cycle management, since there is a delay between the death of the daemon and `init` or *rc* noticing it. This scheme is used in System V, Gentoo *rc*, `simpleinit`, `initNG`, `minit`, `cinit`, `monit` and the LSB specifications. We must also note that *rc scripts* may differ for each distribution and `init` scheme: they handle `/var` differently. The other way to monitor the

life cycle of daemons is not to use rc scripts. A "universal father" process forks and execs daemon processes directly. The pro is that it is a proactive life cycle management: POSIX signals are used whenever a daemon dies. The con is that the daemon must handle itself portability (the environment variables and the parameters it uses), instead of delegating it to an external script. Sun SMF, launchd, upstart and einit use such a "universal father". runit and svscan use a different "father" process for each daemon.

Init schemes allow a number of operations to **instrument** the life cycle of applications. Solutions that use father processes allow mostly what POSIX signals handle: start, stop, and a hook to execute when a daemon dies unexpectedly. This hook can be used to log errors or to automatically restart ("respawn") the faulty daemon. Solutions that use rc scripts may allow a restart operation on a script (Gentoo rc, monit, NetBSD rc.d, LSB specifications), or respawn (initNG, minit, jinit, cinit), in addition to start and stop (System V and its derivatives).

Yet another topic for init enhancement is the **tools** that ease development and usage. System V uses telinit to pass commands to init. Debian introduced start-stop-daemon, that handles PID files and such. Gentoo uses rc-update to create and modify runlevels. In Sun SMF, svccfg switches profiles. Most init schemes implement their own specific tools.

Finally, the great number of existing init schemes is explained by their different focus. Some focus on startup speed (initNG parallelizes startup scripts). Some focus on implementation size (minit, cinit), in order to be usable on resource-constrained devices. Others focus on one of the topics listed above (dependencies, runlevels, life cycle management and instrumentation, tools). But the most interesting schemes are the fairly new ones backed up by Apple, Sun and Ubuntu.

Apple's launchd aims to replace all programs that start applications: xinetd, crond, anacron and init. Daemons can be started, respawned, etc, but also scheduled.

Sun's SMF provides extensive failure detection and diagnostics. It allows to run services in degraded or maintenance mode. The goal is to ease the maintenance of the system (presumably a big server) and to enhance uptime.

Ubuntu's upstart aims to replace the same tools as launchd, but also to unify devices management and power management. It transforms anything into events and triggers: a new device appearing is an event, the screen going into sleep mode is an event, etc. Actions (e.g., mount a filesystem) can be triggered when specific events are thrown. Daemons and dependencies are handled the same way: the DHCP daemon cannot start until a "network is ready" event has been thrown.

Technical Summaries

The remainder of this document are technical summaries of all init-related schemes that the authors know of. Each summary gives a short description on the init scheme, its goals and/or its origin. It then lists features in terms of configuration, life cycle, dependency and service management.

Contents

1	BSD init	7
2	System V init	8
3	The need(8) Concept	10
4	Gentoo rc	11
5	InitNG	14
6	runit	15
7	Apple launchd	17
8	Ubuntu upstart	19
9	eINIT	21
10	Pardus	22
11	LSB init Specifications	23
12	Daemontools: svscan	25
13	minit	26
14	jinit	27
15	cinit	28
16	monit	29
17	depinit	30
18	NetBSD rc.d	31
19	Solaris SMF	32
20	Other Contenders	34

1 BSD init

1.1 Home Page

Documentation for recent versions of the BSD-style init:

http://www.freebsd.org/doc/en_US.ISO8859-1/books/handbook/boot-init.html

The source code for FreeBSD and OpenBSD can be found at :

<http://www.freebsd.org/cgi/cvsweb.cgi/src/sbin/init/>

<http://www.openbsd.org/cgi-bin/cvsweb/src/sbin/init/>

1.2 About

`/sbin/init` runs a master boot script (`/etc/rc`) which orchestrates the whole boot procedure. There may be a few additional `rc.*` scripts. `init` launches them before going in multi-user mode. Because the boot process is sequential and has very few scripts, it is simple and fast.

However, if a new package needs to register itself into the boot process, it needs to edit one of the existing script files. This is very dangerous: if the package installer makes a mistake, the whole startup process will fail.

1.3 Configuration Management

BSD init only handles one or a few boot scripts when switching from single-user mode to multi-user mode, and with `halt` / `reboot` / `shutdown` operations. There are no profiles or runlevels.

1.4 Life Cycle Management

None.

1.5 Dependency Resolution

No dependencies; only an ordered, linear startup defined in `/etc/rc`.

1.6 Service Management Model

None.

1.7 Service Management Tools

None. Unix tools such as `ps` or `signal` are not part of `init`.

Communication with the outside world is achieved through signals.

2 System V init

2.1 Home Page

<http://freshmeat.net/projects/sysvinit/>

2.2 About

From the Freshmeat page:

Init's primary role is to create processes from a script stored in the `/etc/inittab` file. This package also contains well known and used utilities like `reboot`, `shutdown`, `killall`, `poweroff`, `telinit`, `sulogin`, `wall`, etc.

2.3 Configuration Management

SysVinit introduced the notion of runlevel. From the `man` page:

A runlevel is a software configuration of the system which allows only a selected group of processes to exist. The processes spawned by `init` for each of these runlevels are defined in the `/etc/inittab` file. `init` can be in one of eight runlevels: 0-6 and S. The runlevel is changed by having a privileged user run `telinit`, which sends appropriate signals to `init`, telling it which runlevel to change to.

Runlevels 0, 1, and 6 are reserved. Runlevel 0 is used to halt the system, runlevel 6 is used to reboot, and runlevel 1 is used to get the system down into single user mode. Runlevel S is used before entering runlevel 1 (single-user mode). Runlevels 2-5 are multi-user modes, and are different for each Unix / Linux distribution.

When entering a new runlevel (at system boot the runlevel is undefined, sometimes described as "N"), SysVinit executes a master script (`/etc/init.d/rc` for instance). This script executes all stop-scripts from the old runlevel, then all start-scripts from the new runlevel. Those scripts are found either in `/etc/init.d`, `/etc/rc.d` or `/sbin/init.d`. Each runlevel has a directory: `/etc/rc2.d`, `/etc/rc6.d`, etc. Those directories contain links to the real scripts. SysVinit detects whether to stop or to start a script by its name: start-scripts begin with a "S", stop-scripts with a "K" (as in "kill"). When starting the start-scripts, they are called with "start" as the first and only parameter. When starting stop-scripts, they are called with "stop" as first parameter. This is how a configuration for runlevel 2 could look like:

```
scice% ls -l /etc/rc2.d
total 0
lrwxr-xr-x 1 root root 18 Feb 20 2004 S10sysklogd -> ../init.d/sysklogd
lrwxr-xr-x 1 root root 15 Feb 20 2004 S11klogd -> ../init.d/klogd
lrwxr-xr-x 1 root root 13 Feb 20 2004 S14ppp -> ../init.d/ppp
lrwxrwxrwx 1 root root 15 Feb 24 2004 S15bind9 -> ../init.d/bind9
lrwxrwxrwx 1 root root 18 Feb 20 2004 S18quotarpc -> ../init.d/quotarpc
[...]
scice% ls -l /etc/rc6.d
```

```
lrwxrwxrwx 1 root root 15 Aug 31 16:22 /etc/rc6.d/K19aumix -> ../init.d/aumix
lrwxrwxrwx 1 root root 15 Apr 19 09:55 /etc/rc6.d/K19samba -> ../init.d/samba
lrwxrwxrwx 1 root root 13 Feb 25 2004 /etc/rc6.d/K20gpm -> ../init.d/gpm
[...]
```

(Source: nico.schotteli.us)

The number behind the S or the K is the priority. The lower the priority, the earlier the script is started.

Each service has its own init-script, which sometimes also accepts “restart” or “reload”. This way, runlevels are modular.

2.4 Life Cycle Management

The `/var/log/wtmp` binary file records all user logins and logouts. It is maintained by `init`, `login` and `getty`.

The `/var/run/utmp` binary file holds information on users and processes currently using the system. Not all applications use it, so it may be incomplete.

Also, service scripts in `/etc/init.d` can be called during runtime from the command line (not only at startup).

SysVinit can dump its state and exec a later version of itself, so no state is lost between upgrades (source: `jinit`'s page).

2.5 Dependency Resolution

None. Services are started in alphabetical order, hence the priority numbers in the script names. We just hope the administrator made no mistakes...

2.6 Service Management Model

Wrapper scripts are called with a `start` or `stop` parameter.

2.7 Service Management Tools

`telinit` switches the current runlevel.

3 The need(8) Concept

3.1 Home Page

<http://www.atnf.csiro.au/people/rgooch/linux/boot-scripts/>

3.2 About

Richard Gooch wrote this whitepaper in 2002. He argues that BSD init is not scalable (modifications are risky), and that SysVinit is too complex and ugly.

Gooch proposes to remove all master scripts and numbering schemes. `init` runs all scripts in `/etc/init.d`, in random (?) order. The scripts themselves express their dependencies with the `need` program. This ensures that a correct startup order is automatically observed. `need` also ensures that all services are run only once.

The concept of virtual service names is also introduced. For instance, `sendmail` and `qmail` both provide the `mta` service. Which implementation is used is not important, so other services simply need “`mta`”. This is done with the `provide` program.

These ideas are implemented in `simpleinit`.

3.3 Configuration Management

A runlevel is represented by a script, e.g. `/sbin/init.d/runlevel.3`, which will need an unordered set of services, as well as e.g. `runlevel.2`. This makes it easy to increment and decrement the runlevel.

However, this does not support runlevels that bypass these precedence relations.

3.4 Life Cycle Management

Same as SysVinit.

3.5 Dependency Resolution

Simplistic dependency expression: `need serviceXX`.

It is unclear how we ensure that `provide`, called in service wrapper scripts, is run before dependency resolution.

3.6 Service Management Model

Same as SysVinit.

3.7 Service Management Tools

None.

4 Gentoo rc

4.1 Home Page

<http://www.gentoo.org/doc/en/handbook/handbook-x86.xml?part=2&chap=4>

4.2 About

Gentoo uses SysVinit with modified scripts and tools. Scripts are not handled directly by `init`, but by Gentoo's `/sbin/rc`.

Runlevels have names instead of numbers; this removes the limit on their number, and allows specialized runlevels such as “unplugged” for a laptop. Additional tools allow to create, delete, modify runlevels from the command line.

A global option allows to start scripts in parallel, for a performance gain (more visible on multi-processor or multi-core machines).

Like in SysVinit, after `init` is invoked as the last step of the kernel boot sequence, it reads `/etc/inittab`. First, the `sysinit` entry is run:

```
si::sysinit:/sbin/rc sysinit
```

`rc` is called and mounts all filesystems. Second, the `bootwait` entry is run:

```
rc::bootwait:/sbin/rc boot
```

Here, `rc` switches the runlevel to `boot`. All links in the `/etc/runlevels/boot` directory correspond to services that always must be run at startup: `checkroot`, `checkfs`, `clock`, etc. Lastly, the `initdefault` entry is run:

```
id:3:initdefault:
l3:3:wait:/sbin/rc default
```

Here, `initdefault` indicates that the default runlevel is 3 (as in the old SysVinit scheme), and runlevel 3 says that it will wait until `/sbin/rc` has finished switching to runlevel `default`.

Once `rc` has finished, `init` launches virtual consoles as usual, and each console runs a `tty` with `login` (through `agetty`).

```
c1:12345:respawn:/sbin/agetty 38400 tty1 linux
```

Note that when an `agetty` dies, `init` respawns it.

Gentoo's `rc` program is currently being rewritten in C (`baselayout` package version 2, see <http://roy.marples.name/node/293>). The goal is to improve performance and not to need `bash` any more.

4.3 Configuration Management

Runlevels are no longer named `rc[0-6S]`. Instead, they have an alphanumerical name. Each runlevel has a corresponding directory in `/etc/runlevels`, which contains symbolic links to scripts in `/etc/init.d`.

The `rc-update` command is used to create, delete and modify runlevels. Each time it is used, a service tree is recomputed. This tree contains the list of services to start and their order, once all dependencies have been computed and resolved.

The `rc-status` command shows the running status for each service in a runlevel.

The `rc` command switches to a runlevel passed in argument. Switching to a runlevel means comparing the list of running services with the runlevel's service tree. If a service from the service tree is not running, it is started. If a service is running, but is neither in the service tree nor a dependency of a service in the service tree, it is stopped. This does not apply when leaving runlevels `sysinit` and `boot`.

4.4 Life Cycle Management

Each service has an init script particular to the Gentoo distribution. Such a script at least understands `start`, `stop`, `restart`, `pause`, `zap`, `status`, `inneed`, `iuse`, `needsme`, `usesme` and `broken` parameters. Life-cycle related parameters are: `start` and `stop`, `restart` (with possibly extra things to do between the stop and the re-start), `zap` (to force the status to "stopped" in case it is wrongly reported as "stopped") and `status` (to query the running status).

Once a script is successfully `started`, its status is set to "started". This status will be set to "stopped" if the script is `stopped` or `zapped`, and will be re-evaluated if the script is `status`'ed. Re-evaluation is done by checking the pid file in `/var/run` and the daemon file in `/var/lib/init.d`. This means that if a service's process is manually killed, the service's status stays "running" until a `stop`, `zap` or `status` command is issued.

4.5 Dependency Resolution

Gentoo init scripts respond to the following parameters: `inneed`, `iuse`, `needsme`, `usesme` and `broken`. The first two list the hard and soft dependencies of the scripts, the next two list the scripts that are hard- and soft-dependent on this script, and the last one lists broken dependencies (unresolved, "need"-type).

Inside the script, the following metadata tags can be used: `need`, `use`, `before`, `after`, `provide`. `need` are hard dependencies: the current script cannot run without them. `use` are soft dependencies: the current script uses them if they are present in the current runlevel or in boot, but can still run without them. `before` and `after` specify a startup order, regardless of dependencies. Finally, `provide` indicates virtual service names that the current scripts fulfills (like in Gooch's need-based init). All these tags are used when a script is manually invoked, and when a runlevel is modified.

Network dependencies are a special case. A service will need a network interface to be up (any interface other than `net.lo`). This dependency is written: `need net`.

4.6 Service Management Model

All services must have a Gentoo-specific wrapper script (an additional effort for the packager). Configurations and services can be managed through good CLI tools.

For each service, settings (preferences, configuration) are separated from the wrapper script. Scripts are in `/etc/init.d/<script>`, while settings are in `/etc/conf.d/<script>`.

The main problem is that `rc-status` does not show the exact state of services (cf. killed processes not seamlessly taken into account).

4.7 Service Management Tools

`rc-status`, `rc-config`, `rc-update`: see above.

`/sbin/rc`: switch runlevel (`sys-apps/baselayout` package).

`depscan.sh`: rebuild the dependency tree.

5 InitNG

5.1 Home Page

<http://www.initng.org/wiki>

5.2 About

InitNG's goal is to start services in parallel, for a reduced startup time.

As a side note, the tool's syntax is reversed compared to System V: instead of calling `<script> start`, InitNG is invoked via `ngc start <script>`. It makes more sense to call a single tool to manage life cycles than to call multiple scripts that duplicate similar behaviour.

The configuration files support lots of fancy features, such as respawn mode, suid, niceness, output redirect, arguments passing, environment variables passing, etc.

There is not much information on the wiki, except that it is supposed to be faster than SysVinit.

5.3 Configuration Management

ng-update manages configurations similarly to Gentoo's rc-update. The wiki does not provide more information.

5.4 Life Cycle Management

Respawn mode for daemons. Otherwise, similar to SysVinit.

5.5 Dependency Resolution

need and use dependencies, similar to Gentoo init.

Network access is a special dependency, as in Gentoo init.

5.6 Service Management Model

Every command goes through an InitNG program, instead of directly invoking rc shell scripts.

5.7 Service Management Tools

The ngc tool is used to start / stop / restart services.

The ng-update tool manages runlevels.

6 runit

6.1 Home Page

<http://smarden.org/runit/>

6.2 About

From the home page: `runit` is a cross-platform Unix init scheme with service supervision. Running and shutting down is done in three stages:

1. **Stage 1:**

`runit` starts `/etc/runit/1` and waits for it to terminate. The system's one time initialization tasks are done here. `/etc/runit/1` has full control over `/dev/console` to be able to start an emergency shell in case the one time initialization tasks fail.

2. **Stage 2:**

`runit` starts `/etc/runit/2` which should not return until the system is going to halt or reboot; if it crashes, it will be restarted. Normally, `/etc/runit/2` runs `runsvdir`.

3. **Stage 3:**

If `runit` is told to halt or reboot the system, or Stage 2 returns without errors, it terminates Stage 2 if it is running, and runs `/etc/runit/3`. The systems tasks to shutdown and halt or reboot are done here.

`runit` is "optimized for reliability and small size". If compiled and statically linked with `dietlibc`, it produces an ≈ 8.5 k executable.

6.3 Configuration Management

Runlevels are handled through the `runsvdir` and `runsvchdir` programs.

A runlevel is a directory in `/etc/runit/runsvdir/`. Symlinks indicate which are the current and previous runlevel. A runlevel is launched by `runsvdir`. Runlevel switching is performed by `runsvchdir`, which switches directories for `runsvdir`.

6.4 Life Cycle Management

Each of a runlevel's subdirectories is a service. For each service, `runsvdir` launches a `runsv`, which starts and monitors this specific service. A service contains a `./run` and an optional `./finish` script. Its state is maintained in a file by `runsv`. A named pipe allows to pass commands to `runsv`, e.g. to start, stop, run once or in respawn mode, pause (SIGSTOP), etc.

The `sv` program sends commands to one or all `runsvs`.

6.5 Dependency Resolution

Expression of dependencies is poor. If service A depends on service B, A's run script must explicitly invoke sv to ensure that B is running, e.g.:

```
#!/bin/sh
sv start B || exit 1
exec A
```

There are no optional dependencies, no virtual service names, and no tools to explore dependencies.

6.6 Service Management Model

One management process is launched for each service. A single runlevel process is launched to manage all these management processes. Commands are passed through named pipes.

6.7 Service Management Tools

sv: pass life cycle-related commands from the CLI.

runsvchdir: switch runlevel.

chpst (change process state): set limits for a service or for a user (depends on the limit, not really explicit). e.g. # open files per process, # processes per user, memory per process.

7 Apple launchd

7.1 Home Page

<http://developer.apple.com/macosx/launchd.html>

7.2 About

Starting from Mac OS X 10.4, `launchd` aims to replace `init`, the old `mach_init`, `cron` and `xinetd`. It starts daemons, shuts them down and restarts them when a request comes (`inetd` behaviour), runs programs at a specified time or later (`anacron`-like behaviour).

Each job (service) has an XML property file, replacing the different syntaxes from `crontab`, `inittab` etc. This includes program arguments, I/O redirect, `setuid`, `chroot`, etc.

Dependencies are automatically discovered (?). Services are run in parallel.

`launchd` can be controlled from a single utility, `launchctl`.

The source code has been relicensed under the ASL in august 2006.

7.3 Configuration Management

Runlevels are not supported on Mac OS (no order maintained, no selective startup of services). Directories of jobs can be used to launch a set of jobs together.

7.4 Life Cycle Management

Jobs can be started once, started in respawn mode, started in launch-on-demand mode, started based on a schedule, stopped. How a job is started is decided by its property file.

As usual, respawn mode requires that the service does not daemonize itself (`fork+exec`, `fork+exit`).

7.5 Dependency Resolution

Dependencies are a bit unusual and lack documentation. The `Uses`, `Requires` and `Provides` keywords have been dropped. Instead, services watch for changes on a file or watch for IPCs. The advantage is that we ensure dependencies are not only launched, but are also in a usable state, already initialized and ready to communicate. The drawback is that it is totally intrusive and quite obscure.

Special dependencies can be specified, also in the source code (not as metadata), but using specific APIs: disk volume mounted, network available, kernel module present (else it is launched), user Y is logged in.

7.6 Service Management Model

Everything life cycle-related goes through `launchd`, instead of using various tools (`cron`, `inet`). Every command is issued using `launchctl`, which also has a prompt mode.

Instances of `launchd` can be run for a specific user. This allows non-root users to schedule jobs.

No configuration management is provided, nor virtual service names.

7.7 Service Management Tools

`launchctl`: takes commands and passes them to `launchd` through a mutex.

8 Ubuntu upstart

8.1 Home Page

<http://upstart.ubuntu.com/>

8.2 About

upstart aims to replace `init`, `cron`, `anacron`, `at` and `inet`. It also aims to interact with `udev`, `acpi` and `module-init-tools` (`modprobe`, etc). The basic idea is that every action is triggered by an event.

Tasks (or jobs, or services) begin in the “waiting” state. Upon reception of a task-dependent set of events, the task is run and goes in the “running” state. Upon reception of another task-dependent set of events, it goes to the “dead” state (which should be “stopping”), cleans up, then returns to the “waiting” state.

An event can be triggered manually, by ACPI, by `udev`, by `upstart`, etc. Dependencies are also expressed as events: `upstart` sends an event when a task goes “running”; its dependencies simply wait for this event. Everything task-related is configured in `/etc/event.d`.

Security around events and event senders is unclear.

The project has been started before `launchd` went open source, and shares some similarities (mainly `inet` and `cron` capabilities).

Both code and documentation are quickly progressing, so some limitations presented here may already have been fixed.

8.3 Configuration Management

A notion of profile is intended, as a replacement for runlevels. It is still unspecified and unimplemented. Runlevels are still present as dependencies: “start on runlevel-2”.

`upstart` launches all tasks present in `/etc/event.d`; the order is defined by dependencies (expressed as events).

8.4 Life Cycle Management

Tasks can be waiting, running or dying. A set of events triggers state switching. Actions can be performed at each state change (start script, stop script, anything).

8.5 Dependency Resolution

Dependencies are expressed as e.g. “start on event1”. Event1 can be that a service has begun to start (“on starting s1”), has finished to stop (“on stopped s2”), etc. This unusual approach gives a lot of flexibility: it allows virtual service names, optional dependencies, and a lot more.

8.6 Service Management Model

The event system is very powerful and flexible. However, it might be a bit complex to get the whole picture of what happens on the system.

8.7 Service Management Tools

start, stop and status commands are provided to manage tasks.
initctl is used to throw events from the command line (using a mutex).

9 eINIT

9.1 Home Page

<http://einit.sourceforge.net/>

9.2 About

eINIT aims to be a dependency-based init, written entirely in C, suitable for embedded systems. It wants to get rid of rc scripts, to have flexible dependencies, to be modular (e.g. with audio output for the blind), and to support respawn mode.

9.3 Configuration Management

eINIT supports runlevels called modes. A mode is described by an XML file. It can depend on other modes (like Gentoo's boot runlevel), i.e. it comes in addition to the mode it depends on. A mode says which services must be enabled (started), disabled (stopped), and reset (restarted). A mode also defines orders of precedence: "try to start service A, and start service B if it fails". Finally, there is the notion of service groups, e.g. all network interfaces, or all graphical devices. Dependencies can be set on one, all or part of the service group.

Variables can be set globally and per-mode. (It is unclear what the use case is)

9.4 Life Cycle Management

Services are either started once, started in respawn mode, or stopped.

Dependencies are expressed inside runlevels (cf. above), not in service wrapper scripts. Services are described in the mode's XML file, with attributes such as "enable" and "disable" that indicate what to run when the service is started and stopped.

9.5 Dependency Resolution

Other attributes are "provides" and "requires" for dependencies. This allows for virtual service names. This also means that the XML description must be duplicated between modes, unless we create a mode dependency.

9.6 Service Management Model

Most management activities can be seen as switching runlevels.

eINIT supports enabling/disabling eINIT modules (e.g. TTYs, shell commands, daemons), but not services themselves. This is a much coarser granularity.

9.7 Service Management Tools

einit-control passes Commands through a Unix socket.

10 Pardus

10.1 Home Page

<http://www.pardus.org.tr/eng/projeler/comar/SpeedingUpLinuxWithPardus.html>

10.2 About

Pardus is a GNU/Linux distribution from Turkey. Two subprojects are within our scope: Mudur and Çomar.

Mudur is the init subsystem, which replaces the rc part of init. It is called by init via the `/etc/inittab` file. Mudur performs runlevel switching tasks, according to the usual System V runlevels (boot, single, reboot, etc). Mudur is written in Python.

Çomar is the "configuration management system" which manages settings and preferences, and handles the way to start applications. It needs modified rc scripts, as do most of its competitors. According to the wiki, it "handles parallel execution, access control, profile management and remote management". Çomar is written in C, and rc scripts are written in Python.

The main objective is to gain speed at startup, while being clean and hiding complexity to the user.

10.3 Configuration Management

Runlevels have the same meaning as in System V. For the sysinit runlevel, Mudur mounts pseudo-filesystems, starts udev, loads kernel modules, etc. For other runlevels, Mudur calls Çomar, which in turn handles rc scripts.

10.4 Life Cycle Management

rc scripts accept parameters such as start, stop, status, configure. The start and stop directives make use of Debian's start-stop-daemon.

10.5 Dependency Resolution

An rc script calls the start methods of the other scripts it depends on. There are no virtual service names.

10.6 Service Management Model

Command-line tools allow to call methods on rc scripts via Çomar.

10.7 Service Management Tools

Not presented.

11 LSB init Specifications

11.1 Home Page

http://www.freestandards.org/spec/refspecs/LSB_1.3.0/gLSB/gLSB/sysinit.html

11.2 About

LSB, the Linux Standards Base, defines what init scripts should conform to. This is an opinion, but the LSB seems more keen on standardizing what RedHat does than on being innovative.

11.3 Configuration Management

LSB specifies the following runlevels:

- 0: halt
- 1: single user mode
- 2: multiuser with no network services exported
- 3: normal/full multiuser
- 4: reserved for local use, default is normal/full multiuser
- 5: multiuser with xdm or equivalent
- 6: reboot

This forbids “profiles” and useful runlevel switching as seen in previous init schemes (e.g. Gentoo).

11.4 Life Cycle Management

Init files provided by LSB applications shall accept one argument, which states the action to perform:

start: start the service

stop: stop the service

restart: stop and restart the service if the service is already running, otherwise start the service

reload: cause the configuration of the service to be reloaded without actually stopping and restarting the service [optional]

force-reload: cause the configuration to be reloaded if the service supports this, otherwise restart the service

status: print the current status of the service

There is no respawn mode.

11.5 Dependency Resolution

LSB-compliant applications init scripts (but not system init scripts) must adopt this syntax for dependency declaration:

```
# Provides: boot_facility_1 [ boot_facility_2 ...]
# Required-Start: boot_facility_1 [boot_facility_2 ...]
# Required-Stop: boot_facility_1 [boot_facility_2 ...]
# Default-Start: run_level_1 [ run_level_2 ...]
# Default-Stop: run_level_1 [ run_level_2 ...]
# Short-Description: short_description
# Description: multiline_description
```

Required-stop are applications that must still run before this one stops.

The “Default-Start” and “Default-Stop” headers define which runlevels should by default run the script with a start or stop argument, respectively, to start or stop the services controlled by the init script.

“Provides” allows to have virtual service names.

11.6 Service Management Model

Out of scope.

11.7 Service Management Tools

Out of scope.

12 Daemontools: svscan

12.1 Home Page

<http://code.dogmap.org/svscan-1/>
<http://cr.yip.to/daemontools.html>

12.2 About

From the daemontools page: **daemontools** is a collection of tools for managing UNIX services. The **svscan** program can be used as an init replacement.

12.3 Configuration Management

Configurations / runlevels are supported as directories of run scripts.

svscan starts and monitors a collection of services. **svscan** starts one **supervise** process for each subdirectory of the current directory. Every five seconds, **svscan** checks for subdirectories again. If it sees a new subdirectory, it starts a new **supervise** process.

Runlevel switching is not really specified.

12.4 Life Cycle Management

supervise monitors a service. It starts the service and restarts it if it dies. To set up a new service, all **supervise** needs is a directory with a run script that runs the service.

supervise maintains status information in a binary format inside the directory `<service>/supervise`, which must be writable by **supervise**. The status information can be read using **svstat**.

There is no stop script.

12.5 Dependency Resolution

No dependencies.

12.6 Service Management Model

Put run scripts in a directory; daemontools will run the service and monitor it. It is simpler than SysVinit, but lacks features (stop command, runlevel switching).

12.7 Service Management Tools

svc is used to send signals to a service.

svstat prints the status of services monitored by **supervise**.

multilog saves error messages to one or more logs. It optionally timestamps each line and, for each log, includes or excludes lines matching specified patterns.

13 minit

13.1 Home Page

<http://www.fefe.de/minit/>

13.2 About

Minit aims to be a very small init. It can start a service, restart it, start it when another is finished, and take dependencies into account.

13.3 Configuration Management

None.

13.4 Life Cycle Management

Reads `/var/run/<service>.pid` and checks if the process has died.

13.5 Dependency Resolution

Yes, but lacks documentation.

13.6 Service Management Model

None.

13.7 Service Management Tools

`msvc`: similar to `svc` from `daemontools`.

`pidfilehack`: reads `/var/run/<service>.pid`.

14 jinit

14.1 Home Page

<http://john.fremlin.de/programs/linux/jinit/>

14.2 About

jinit is based on an extended version of Richard Gooch's need(8) scheme for init. Currently there is no provide(8) functionality.

14.3 Configuration Management

None? (or at least documentation is lacking).

14.4 Life Cycle Management

Supports respawn mode. Apart from that, same as SysVinit.

If it receives a fatal error signal like SIGSEGV it will fork a child to dump core, and exec itself (with the "panic" option so no bootscripts are run). That means it should never bring the system down unnecessarily.

14.5 Dependency Resolution

Same as the need(8) scheme.

14.6 Service Management Model

Same as SysVinit.

14.7 Service Management Tools

None. Communicates over SysV message queues.

15 cinit

15.1 Home Page

<http://linux.schottelius.org/cinit/>

15.2 About

cinit claims to be a fast, small and simple init with support for profiles. It supports parallel startup, mandatory and optional dependencies, and does seem simple to use.

Size is \approx 50KiB when statically linked against uclibc.

15.3 Configuration Management

“Profiles” are directories of services. They are put in `/etc/cinit`.

Choosing the profile at startup is supported, but profile switching is not documented.

15.4 Life Cycle Management

A service is a subfolder in a configuration (or a link to another folder). Everything is done through files or symlinks: `./on` and `./off` indicate the program to run when the service goes on/off. `./on.params` and `./on.env` contain parameters and environment variables to pass to the program.

15.5 Dependency Resolution

Dependencies are listed in `./wants` (optional) and `./needs` (mandatory).

Virtual service names are not supported. The author suggests using symlinks, e.g. `cron -> dcron`.

15.6 Service Management Model

None.

15.7 Service Management Tools

A bunch of executables:

<code>cinit.add.dependency</code>	– add a dependency to a service
<code>cinit.add.getty</code>	– add a new getty
<code>cinit.create.empty.service</code>	– create an empty service
<code>cinit.reboot</code>	– reboot in <code>/bin/sh</code>
<code>cinit.remove.getty</code>	– remove a getty service
<code>cinit.respawn.off</code>	– switch respawning off
<code>cinit.respawn.on</code>	– switch respawning on
<code>cinit.shutdown</code>	– shutdown in <code>/bin/sh</code>

16 monit

16.1 Home Page

<http://www.tildeslash.com/monit/index.php>

16.2 About

monit manages and monitors processes, files, directories and devices. It does not replace init: instead, it acts as a service manager, and can reuse rc scripts.

monit can start a process which is not running, restart a process if it does not respond, and stop a process if it uses too much resources. It can monitor files, directories and devices for changes on timestamp, checksum or size. All checks are performed at a polling interval time.

Management activities are exported through HTTP(S). Monit is configured via the `/etc/monitrc` file.

16.3 Configuration Management

Only supports groups of services which can be started / stopped together.

16.4 Life Cycle Management

A service can be (re)started and stopped. Its state is checked using pid files. This means that wrapper scripts may have to be written.

16.5 Dependency Resolution

The syntax is simplistic: `depends on serviceX` (in file `/etc/monitrc`). There are no virtual service names.

16.6 Service Management Model

Periodically check a resource (process, file...). Upon changes, actions are triggered. For example:

```
check process apache with pidfile /var/run/httpd.pid
start program = '/etc/init.d/httpd start'
stop program = '/etc/init.d/httpd stop'
if failed host www.tildeslash.com port 80 then restart
depends on apache_bin
```

16.7 Service Management Tools

HTTP interface.

17 depinit

17.1 Home Page

<http://www.nezumi.plus.com/depinit/>

17.2 About

From the home page: "Depinit is an alternative init program that can handle parallel execution, dependencies, true roll-back, pipelines, improved signaling and unmounting filesystems on shutdown. It incorporates ideas from sysvinit, simpleinit, daemontools and make. At present, it is a bit experimental, and requires good knowledge of the initialisation process to set up."

While depinit boasts more features than sysvinit, it is quite complex to setup and to use. It also has trouble working with the login program.

17.3 Configuration Management

A runlevel is a directory in `/etc/depinit`. How it operates and how it can be switched is unclear.

17.4 Life Cycle Management

Each service has its own directory in `/etc/depinit`, which will contain shell scripts named `start`, `stop`, `source`, `filter`, `dest` and/or `daemon`. Their purpose is either obvious (`start` and `stop` describe shell commands that will be called to start and stop the service), or unexplained.

17.5 Dependency Resolution

Dependencies are listed in a `depend` file within each service's directory. Each line in the file corresponds to another directory in `/etc/depinit`.

17.6 Service Management Model

All interactions with depinit are performed through the `depctl` command-line tool. Daemons are started in `respawn` mode.

17.7 Service Management Tools

`depinit`: can either replace `init` (and `rc`), or be started with a `PID > 1` and manage a subset of services.

`depctl`: `start/stop` a service, `shutdown/reboot` the system, send a signal to a daemon. These tools communicate via signals.

18 NetBSD rc.d

18.1 Home Page

<http://ezine.daemonnews.org/200108/rcsystem.html>
<http://www.netbsd.org/guide/en/chap-rc.html>
<http://cvsweb.netbsd.org/bsdweb.cgi/src/sbin/init/>

18.2 About

From the home page:

As of NetBSD 1.5, the startup of the system changed to using rc-scripts for controlling services, similar to SysVinit, but without runlevels.

18.3 Configuration Management

None.

18.4 Life Cycle Management

Service wrapper scripts understand start, stop, restart and kill. They are as limited as SysVinit's. There is no service monitoring facility.

18.5 Dependency Resolution

Dependencies are expressed in rc scripts as four keywords. REQUIRE is a mandatory dependency; PROVIDE allows virtual service names; BEFORE's explanation fell through a spatiotemporal void; KEYWORD is a tag used to include or exclude the script from ordering, e.g. with values 'nostart' or 'shutdown'.

Ordering is done via the `rcorder` command.

18.6 Service Management Model

None. Start scripts, hope they run, optionally check a pid file.

18.7 Service Management Tools

None.

19 Solaris SMF

19.1 Home Page

<http://www.sun.com/bigadmin/content/selfheal/smf-quickstart.html>

19.2 About

SMF (Service Management Facility) is a part of SUN Solaris 10 (historically a System V variant). It targets big production servers. From their page:

SMF has added the notion of the relationship between a service, its processes, and another service that is responsible for restarting the service, to the Solaris kernel. SMF understands whether a service process failed as the result of an administrator error, failure of a dependent service, software bug, or underlying hardware failure. SMF informs the appropriate restarter, which decides either to disable the service by placing it in maintenance mode because it appears to be defective, or to restart it automatically.

SMF still supports `inittab`, `inetd` and `rc` scripts, but aims at replacing them.

SMF services are started in parallel, in a dependency-based order.

Users can be given the rights to manage services and their state, without being root. Still, there are no per-user services.

Alas, the source code is closed.

19.3 Configuration Management

Runlevels are replaced by milestones. A specific set of services must be enabled (started) before a milestone is reached (incremental runlevels). Milestones are: single-user, multi-user, multi-user-server, all, none.

The notion of profile also exist: it is an XML file which lists SMF services, and whether each should be enabled or disabled.

19.4 Life Cycle Management

SMF-enabled services can be permanently enabled / disabled (persists on reboot). When enabled, they can be online (started), offline (stopped), uninitialized (config has not been read yet), degraded (runs at limited capacity), and maintenance (stopped, needs an action from the administrator).

Possible commands are: permanent or temporary enable / disable, start, stop, restart, refresh. Refresh means reloading a service's config file, which is usually triggered by `SIGHUP` on daemons (`SIGHUP` should kill non-daemon processes).

19.5 Dependency Resolution

A dependency is formed as an URI. It can be set on services, devices, network configuration, kernel configuration, milestone (runlevel), and be optional or mandatory. Shortened example from SUN's site:

```
% svcs -l network/smtp:sendmail
dependency  require_all/refresh file:///localhost/etc/mail/sendmail.cf (-)
dependency  optional_all/none  svc:/system/system-log (online)
dependency  require_all/refresh svc:/system/identity:domain (online)
dependency  require_all/refresh svc:/milestone/name-services (online)
dependency  require_all/none  svc:/network/service (online)
dependency  require_all/none  svc:/system/filesystem/local (online)
```

19.6 Service Management Model

SMF manages services, rc scripts, inet daemons, configurations in an (almost) uniform way. The goal is to obtain the best uptime possible, with as little human intervention as possible. The other big goal is to provide detailed information on failures to the administrator, in the case he must intervene.

There are lots and lots of features, with probably lots of complicated code: even the kernel is modified. This solution is ideal for servers, but not suitable for small systems.

19.7 Service Management Tools

Amongst others:

- svcs queries the system state, including old rc scripts.
- svcadm passes commands to SMF services and their restarters.
- svccfg switches profiles.

20 Other Contenders

The following projects are considered less relevant. Most are dead and/or lack documentation.

20.1 DMD

Daemons-Managing Daemon is a stalled project from the Free Software Foundation. Documentation is non-existent.

<http://directory.fsf.org/GNU/DMD.html>

20.2 twsinit

twsinit is a minimal init replacement (about 8 KB) in x86 assembler. Documentation is non-existent.

<http://www.energymech.net/users/proton/>

20.3 Busybox

BusyBox combines tiny versions of many common UNIX utilities (GNU fileutils, shellutils, textutils) into a single small executable. It contains a small init system, which is a stripped down SysVinit scheme. It can read an inittab file, but does not support runlevels.

The project is alive.

<http://www.busybox.net/>

20.4 Fedora New Init System

Fedora had planned to implement its own init replacement. Its goals were a bit confused and mixed: better startup speed, LSB compliance, respawn, logging, tools. The project is stalled / dead.

<http://fedoraproject.org/wiki/FCNewInit>

20.5 serel

Serel aims to speed up the boot sequence using parallel startup. The project is stalled.

<http://www.fastboot.org/>

	Focus	Configuration Management	Life Cycle Instrumentation	Life Cycle Monitoring	Expression of Dependencies	License
BSD init	-	No	No	No	Static ordering	BSD
SysVinit	-	Numbered runlevels	start, stop	Rc scripts (/var)	Static ordering	GPL
simpleinit	Have dependencies	No	start, stop	Rc scripts (/var)	need	GPL
Gentoo rc	Better runlevels and dependencies	Named runlevels	start, stop, restart	Rc scripts (/var)	need, use, before, after, provide	GPL
InitNG	Speed	Named runlevels	start, stop, respawn	Process (/var)	need, use	GPL
runit	Life Cycle Management	Named runlevels	POSIX signals	Father process es	manual scripting	BSD
launchd	Lots of features	No (Groups of jobs)	start, stop, inet, cron, respawn	Father process	In code: watch files & IPCs	ASL
upstart	Event-based	Profiles	Events	Father process	stopping, stopped, starting, started, any event	GPL
eINIT	Runlevels	Modes (services, dependencies, vars)	start, stop, respawn	Father process	provides, requires	BSD

	Focus	Configuration Management	Life Cycle Instrumentation	Life Cycle Monitoring	Expression of Dependencies	License
Pardus	Speed	Numbered runlevels	start, stop	Rc scripts (/var)	need	GPL
LSB	Standardize dependencies	Numbered runlevels	start, stop, restart	Rc scripts	provides, required-to-start, required-to-stop	-
svscan	Service Management	Directories of run scripts	start, respawn	Father processes	No	freedist
minit	Size	No	respawn	/var	?	GPL
jinit	need(8) deps	No	respawn	/var ?	need	GPL
cinit	Size, runlevels	Unswitchable profiles	respawn	Scripts (/var)	wants, needs	GPL
monit	Service Management	No (Groups of jobs)	start, stop, restart, quotas	Scripts (/var)	depends	GPL
NetBSD rc.d	Have rc scripts	No	start, stop, restart	Rc scripts	require, provide, before	BSD
SMF	Heavy Management	Milestones and profiles	start, stop, restart + persist + self-healing	Father Process / kernel	require, optional on URIs (svc, file...)	Closed source



Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-0803