The Unix history in a repository
44 years of Unix history in one Git repo

Diomidis Spinellis
Department of Management Science and Technology
Athens University of Economics and Business

www.spinellis.gr
dds@aueb.gr
@CoolSWEng
LIVE FREE OR DIE

UNIX®

TRADEMARK OF BELL LABS®
January 23, 2002

Dear UNIX® enthusiasts,

Caldera International, Inc. hereby grants a fee free license that includes the rights use, modify and distribute this named source code, including creating derived binary products created from the source code. The source code for which Caldera International, Inc. grants rights are limited to the following UNIX Operating Systems that operate on the 16-Bit PDP-11 CPU and early versions of the 32-Bit UNIX Operating System, with specific exclusion of UNIX System III and UNIX System V and successor operating systems:

32-bit 32V UNIX
16 bit UNIX Versions 1, 2, 3, 4, 5, 6, 7

Caldera International, Inc. makes no guarantees or commitments that any source code is available from Caldera International, Inc.

The following copyright notice applies to the source code files for which this license is granted.

Copyright(C) Caldera International Inc. 2001-2002. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code and documentation must retain the above copyright notice, this list of conditions and the following disclaimer. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

All advertising materials mentioning features or use of this software must display the following acknowledgement:

This product includes software developed or owned by Caldera International, Inc.

Neither the name of Caldera International, Inc nor the names of other contributors may be used to endorse or promote products derived from this software without specific prior written permission.

USE OF THE SOFTWARE PROVIDED FOR UNDER THIS LICENSE BY CALDERA INTERNATIONAL, INC. AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL CALDERA INTERNATIONAL, INC. BE LIABLE FOR ANY DIRECT, INDIRECT INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Very truly yours,

/signed/ Bill Broderick

Bill Broderick
Director, Licensing Services

* UNIX is a registered trademark of The Open Group in the US and other countries.
Motivation

• Explore evolution of programming style
• Consolidate digital artifacts of historical importance
• Collect and record history that is fading away
• Provide a data set of digital archeology and repository mining
## In Numbers ...

<table>
<thead>
<tr>
<th>Metric</th>
<th>Unix history</th>
<th>Linux history</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start date</td>
<td>20/06/1972</td>
<td>17/09/1991</td>
</tr>
<tr>
<td>Start files</td>
<td>13</td>
<td>92</td>
</tr>
<tr>
<td>Start lines</td>
<td>4768</td>
<td>917,812</td>
</tr>
<tr>
<td>End files</td>
<td>63,049</td>
<td>51,396</td>
</tr>
<tr>
<td>End lines</td>
<td>27,388,943</td>
<td>21,525,436</td>
</tr>
<tr>
<td>Number of commits</td>
<td>677,197</td>
<td>611,735</td>
</tr>
<tr>
<td>Number of merges</td>
<td>2,048</td>
<td>48,821</td>
</tr>
<tr>
<td>Number of authors</td>
<td>921</td>
<td>18,465</td>
</tr>
<tr>
<td>Days with activity</td>
<td>12,827</td>
<td>5,126</td>
</tr>
</tbody>
</table>
Repository Contents

- Research Edition Unix: V1, V3–V7
- Unix 32V
- BSD 1, 2, 3, 4, 4.1, 4.2, 4.3 *, 4.4 *
- 386BSD 0.0, 0.1
- FreeBSD 1.0–10.0

- Tags
- Contributors
- Branches and merges
subject: Study of UNIX

date: September 14, 1972
from: T. R. Bashkow

Messrs. W. S. Bartlett
D. P. Clayton
D. H. Copp
Mmes. G. J. Hansen
J. Hintz
Mr. L. J. Kelly
Miss R. L. Klein

Messrs. J. J. Ludwig
J. F. Maranzano
Mrs. G. Pettit
Messrs. J. E. Ritacco
B. A. Tague
D. W. Vogel
Mrs. L. S. Wright

On Tuesday, September 19, at 9:30 a.m. in Room 2A-418 at Murray Hill, I will give a talk on my study of the UNIX operating system. The emphasis will be on the structure, functional components, and internal operation of the system.

MH-8234-TRB-mbh

T. R. Bashkow

Copy to
Mr. G. L. Baldwin
/ initialize inodes for special files (inodes 1 to 40.)

mov $40,,r1 / set r1=i-node-number 40.

1:
jsr r0,i,get / read i-node 'r1' from disk into inode area of
core and write modified inode out (if any)
mov $100017,i.flgs / set flags in core image of inode to indi-
cate allocated, read (owner, non-owner),
write (owner, non-owner)

movb $1,i.nlks / set no. of links = 1
movb $1,i.uid / set user id of owner = 1
jsr r0,setimod / set imod=1 to indicate i-node modified, also
stuff time of modification into i-node
dec r1 / next i-node no. = present i-node no.-1
bgt 1b / has i-node 1 been initialized; no, branch

// initialize i-nodes r1,...,47. and write the root device, binary, etc.,
directories onto fixed head disk. user temporary, initialization prog.
UNIX IMPLEMENTATION

/ uo -- unix

cold = 0
orig = 0 . /orig = 0. relocatable

rkda = 177412 / disk address reg
rkds = 177400 / driv status reg
rkcs = 177404 / control status reg
crsr = 174000 / receiver status reg
rcbr = 174002 / receiver buffer reg
tcsr = 174004 / xmt status reg
tcbr = 174006 / xmt buffer reg
tcst = 177340 / dec tape control status
tccm = 177342 / dec tape command reg
tccw = 177344 / word count
tcba = 177346 / bus addr
tcdt = 177350 / data reg
dcs = 177460 / drum control status
dae = 177470 / drum address extension
dks = 177546 / clock status reg
prs = 177550 / paper tape reader status
pro = 177552 / buffer
ppr = 177554 / punch status
ppb = 177556 / punch buffer
/lpb = 177514 / line printer status
/lpdb = 177516 / line printer buffer
cks = 177560 / console read status
tkb = 177560 / console read
tps = 177564 / punch status
tpb = 177564 / punch buffer
ps = 177776 / processor status

halt = 0
wait = 1
rti = 2

nproc = 16 . /Number of processes
nmfs = 50.
nnty = 841

if cold / ignored if cold = 0

/core = orig+40000 /specifies beginning of user's core
tcore = core+20000 /specifies end of user's core (4096 wo.

/core 4:14
init by copy
4:14 unkni:0 / bus error
1:1 fsym:0 / illg in tr
14:1 unkni:0 / trace and trap (see Sec. B.1 page )
10:1 unkni:0 / trap
14:1 panic:0 / pwr
3:1 rtsym:0 / emt
3:1 syse:0 / sys

mov @data, r0 / r0 = base addr. of assembled directories.
mov @u.off, u.fofp / pointer to u.off in u.fopf (holds file
offset)

1:
mov (r0)+, r1 /r1 = 41, ..., 47: "0" in the assembled directory
header signals last
beq 1f / assembled directory has been written onto drum
jir r0, imap / locate the inode map bit for i-node 'r1'
bis3 mq, (r2) / set the bit to indicate the i-node is not
available
jir r0, iget / read inode 'r1' from disk into inode area of
/core and write modified i-node on drum (if any)
mov (r0)+, i.fofs / set flags in core image of inode from
assembled directories header
movb (r0)+, i.nlks / set no. of links from header
movb (r0)+, i.uid / set user id of owner from header
jir r0, setmod / set inmod to indicate inode modified: also,
stuff time of modification into i-node
mov (r0)+, i.count / set byte count for write call equal to
size of directory
mov r0, u.base / set buffer address for write to top of directory
cir u.off / clear file offset used in "seek" and 'tell'
add u.count, r0 / r0 points to the header of the next directory
jir r0, writei / write the directory and i-node onto drum
br 1b / do next directory

.endif

/next 2 instructions not executed during cold boot.
bis $2000, $b0 / $b0 I/O queue entry for superblock on drum;
jir r0, poke / read drum superblock

1:
tstb $b0+1 / has I/O request been honored (for drum)?
bne 1b / no, continue to idle.

1:
decb sysflg / normally sysflag=0, indicates executing in system
sys exec; 2f; 1f / generates trap interrupt; trap vector =
sysent; 0
br panic / execute file/etc/init

1:
2f:0 this is not the init.proc

2:
</etc/init> / UNIX looks for strings term, noted by null

panic:
cir ps

1:
dec $0
bne 1b
dec $5
bne 1b
jmp *$173700 / rom loader address

Issue D Date 3/17/72 ID IMO.1-1 Section E.0
FUNCTION: systork creates a new process. This process is referred to as the child process.
This new process's core image is a copy of that of the caller of "systork".
The only distinction is the return location and the fact that (u.ro) in the old process
(parent) contains the process id (p.pid) of the new process (child). This id is
used by "syswait". "systork" works in the following manner:
1) The process status table (p.stat) is searched to find a process number
   that is unused. If none are found an error occurs.
2) When one is found, it becomes the child process number and its status
   (p.stat) is set to active.
3) If the parent had a control tty, the interrupt character in that
   tty buffer is cleared.
4) The child process is put on the lowest priority run queue via "put".
5) A new process name is gotten from mp.id (actually gets a unique
   number) and is put in the child's unique id table; the process
   id (p.pid)
6) The process name of the parent is then obtained and placed in the
   unique identifier of the parent process of the child (p.p.pid). The parent
   process name is then put in (u.ro)
7) The child process is then written out to disk by "pswap" i.e.,
   the parent process is copied onto disk and the child is born
8) The parent process number is then restored to (u.ro)
9) The child process name is put in (u.ro)
10) The pc on the stack, sp + 18 is incremented by 2 to create the
    return address for the parent process.
11) The uro list is then searched to see what files the parent had
    opened. For each file the parent had opened, the corresponding tsp
    entry must be updated to indicate that the child process also has
    opened the file. A branch to sysset is then made.

CALLING SEQUENCE
ARGUMENTS
INPUTS
from shell:
pu.stat status of a process active, dead, unused
parent process number
u.typ pointer to parents process control tty buffer
mp.id process name generated
w.tp list of index into the tsp table
tsp - table of open files
UNIX PROGRAMMER’S MANUAL
Second Edition

K. Thompson
D. M. Ritchie

June 12, 1972

Copyright © 1972
Bell Telephone Laboratories, Inc.

No part of this document may be reproduced, or distributed outside the Laboratories, without the written permission of Bell Telephone Laboratories.
The owner section gives the name of the person or persons to be consulted in case of difficulty. The rule has been that the last one to modify something owns it, so the owner is not necessarily the author. The owner's nicknames stand for:

ken      K. Thompson
dmr      D. M. Ritchie
jfo      J. F. Ossanna
rhm      R. Morris
doug     M. D. McIlroy
lem      L. E. McMahon
llc      L. L. Cherry
csr      C. S. Roberts

These nicknames also happen to be UNIX user ID's, so messages may be transmitted by the mail command or, if the addressee is logged in, by write.

At the beginning of this document is a table of contents, organized by section and alphabetically within each section. There is also a permuted index derived from the table of contents. Within each index entry, the title of the writeup to which it refers is followed by the appropriate section number in parentheses. This fact is important because there is considerable name duplication among the sections, arising principally from commands which exist only to exercise a particular system call.

This manual was prepared using the UNIX text editor ed and the formatting program roff.
I. SYSTEM CALLS

break ........................................ set program break
cent ........................................... catch ENT traps
chdir ......................................... change working directory
close .......................................... change mode of file
close .......................................... change owner of file
close .......................................... close open file
close .......................................... create file
creat .......................................... execute program file
cexec .......................................... terminate execution
cexit .......................................... create new process
cfstat ......................................... status of open file
cfstat .......................................... get user ID
cgetuid ........................................ get typewriter mode
cgtty ........................................... set low-priority status
chung ........................................... catch illegal instruction trap
cintr ........................................... catch or inhibit interrupts
clink ........................................... link to file
cmake .......................................... destroy process
ccreate ........................................ create directory
cset ........................................... set date modified of file
cqueue .......................................... mount file system
cmount .......................................... open file
copen .......................................... open file
ccatch or inhibit quits read file
cread ........................................... release processor
crelease ....................................... move read or write pointer
crelease ....................................... set user ID
cset ........................................... delay execution
csleep .......................................... get file status
cstatus ......................................... set system time
cstate .......................................... set mode of typewriter

time .......................................... sync

II. SUBROUTINES

getc ........................................... assure synchronization
getc ........................................... find read or write pointer
getc ........................................... get time of year
getc ........................................... dismount file system
getc ........................................... remove (delete) file
getc ........................................... wait for process
getc ........................................... write file

III. SPECIAL FILES

dn0 ........................................... arctangent
d0 ........................................... convert ASCII to floating
dp ........................................... convert ASCII to integer
dp ........................................... convert floating-point constants
dp ........................................... convert time to ASCII
dp ........................................... exponential function
dp ........................................... floating-point simulator
dp ........................................... convert floating to ASCII
dp ........................................... communicate with gccs
dp ........................................... gerts
dp ........................................... get character
dp ........................................... compute hypotenuse
dp ........................................... convert integer to ASCII
dp ........................................... logarithm base e
dp ........................................... print string on typewriter
dp ........................................... read name list
dp ........................................... print time
dp ........................................... write character or word
dp ........................................... quicker sort
dp ........................................... storage allocator
dp ........................................... sine, cosine
dp ........................................... square root
dp ........................................... transfer depending on value

dp ........................................... 801 ACU

dp ........................................... 201 Dataphone

dt ........................................... line printer

dt ........................................... core memory

dt ........................................... magtape

dt ........................................... punched paper tape

dt ........................................... RP disk

dt ........................................... RK disk

dt ........................................... RP disk

dt ........................................... DECTape

dt ........................................... console typewriter

dt ........................................... remote typewriter

V. FILE FORMATS

dp ........................................... assembler and loader output

dp ........................................... archive file

dp ........................................... archive file

dp ........................................... core image file
directory
file system
ident
passwd
tap
uids
utmp
wtmp
directory format
file system format
GCOS ident cards
password file
DECTape format
map names to user ID's
logged-in user information
accounting files

VI. USER MAINTAINED PROGRAMS

basic
bc
bj
cal
chas
cref
das
dli
dpt
moo
ptx
tmg
tti
DEC supplied BASIC
compile B program
the game of black jack
print calendar
prepare symbol table
cross-reference table
disassembler
load DEC binary paper tapes
read DEC ASCII paper tapes
the game of MOO
permut ed index
compile tmg1 program
the game of tic-tac-toe

VII. MISCELLANEOUS

ascii
bproc
getty
glob
init
kbd
login
msh
tabs
map of ASCII
boot procedure
adapt to typewriter
argument expander
initializer process
map of TTY 37 keyboard
how to log onto system
mini Shell
set tab stops on typewriter

- ix -
Research Editions

- **1st**: (Nov 1971) Printed PDP-11 kernel
- **2nd**: (Jun 1972) Dump DECTape fragments of programs
- **3rd**: (Feb 1973) 90% C kernel
- **4th**: (Nov 1973) only troff manual
- **5th**: (June 1974): No manual source
- **6th**: (May 1975): Complete, widely distributed
1978
By John Raiser and Tom London
Bell Labs Holmdel
VAX as a large PDP-11
   — swapping, not paging
15 Berkeley Snapshots

- BSD (1978): ex, Pascal, tools
- 2BSD: vi, termcap, csh, ...
- 3BSD (1979): VM
- 4BSD (1980): CSRG/DARPA (email, ^Z, signals)
- 4.1c2BSD (1982): TCP/IP, ftp, rsh, rlogin, ...
- ...
- 4.3BSD (1988) performance, BIND
- 4.3BSD Net/1 (1988) no AT&T licensing
- ...
- 4.4BSD-Lite Release/2 (1995) last enhancements
Metadata

• Date
• Author
• Commit parents
\RESEARCH \V1 DEVELOPMENT
\WORK ON FILE US5.S

[...]

\RESEARCH \V5 DEVELOPMENT
\WORK ON FILE USR/SYS/KEN/SLP.C

[...]

\RESEARCH \V5 DEVELOPMENT
\WORK ON FILE USR/SYS/DMR/KL.C

[...]

\RESEARCH \V6 DEVELOPMENT
\WORK ON FILE RAT/R.G

[...]

\RESEARCH \V7 DEVELOPMENT
\WORK ON FILE USR/SRC/CMD/SH/BLOK.C
commit a20c66bfa7223fbaf148f13b47356ad4326e9265
Author: Bill Joy <wnj@ucbvax.Berkeley.EDU>
Date: Sun Oct 24 22:38:00 1982 -0800

create netif directory

SCCS-vsn: 4.5

[...]

commit e449c1d11a5de86f09ea0c5b80461d7d6cb01792
Author: Eric Schmidt <schmidt@ucbvax.Berkeley.EDU>
Date: Wed May 9 18:47:25 1979 -0800

BSD 2 development

Work on file src/net/netcp.c
add debugging information to server smtp code: _debug to set
Debug, _verbose to set verbose, and _showq to show the send queue

SCCS-vsn: 3.18

index 1ec3c41..1eadb47 100644
--- a/usr/src/usr.sbin/sendmail/src/srvrsmtp.c
+++ b/usr/src/usr.sbin/sendmail/src/srvrsmtp.c
@@ -55,7 +57,9 @@ static struct cmd CmdTab[] =
  "helo",
  CMDHELO,
 # ifdef DEBUG
-  "showq",
+  "_showq",
+  CMDDBGSHOWQ,
+  "_debug",
+  CMDDBGDEBUG,
+  "_verbose",
+  CMDDBGVERBOSE,
 # endif DEBUG
      NULL,
      CMDERROR,
3,3;
@@ -245,6 +252,16 @@ smtp()

break;
+
+  case CMDDBGDEBUG: /* set debug mode */
+                  Debug = atoi(p);
+                  message("200", "Debug = \%d", Debug);
+                  break;
# endif DEBUG
register struct inode *ip;

ip = fp->f_inode;
c = u.u_count;

loop:

/*
 * If error or all done, return.
 */
if (u.u_error)
    return;
plock(ip);
if(c == 0) {
    prele(ip);
    u.u_count = 0;
    return;
}

usr/src/libc/gen/timezone.c
(Dennis Ritchie 1979-01-10 14:58:45 -0500 76) static struct zone {
    int offset;
    char *stdzone;
    char *dzone;
} zonetab[] = {
    \{-1*60, "MET", "MET DST"},
    [...]
}
...
lib/libc/gen/timezone.c
(Jordan K. Hubbard 1996-07-12 18:57:58 +0000 81) \{-1\}
usr/src/libc/gen/timezone.c
(Bill Joy 1980-12-22 00:40:25 -0800 97) \};
usr/src/libc/gen/timezone.c
(Bill Joy 1980-12-22 00:40:25 -0800 98)
usr/src/libc/gen/timezone.c
(Keith Bostic 1987-03-28 19:27:07 -0800 106) char *
lib/libc/gen/timezone.c
(Ed Schouten 2009-12-05 19:31:38 +0000 107) _tztab(int zone, int dst)
lib/libc/gen/timezone.c
(Rodney Grimes 1994-05-27 05:00:24 +0000 108) {
    struct zone *zp;
    char sign;
    for (zp = zonetab; zp->offset != -1; ++zp) /* static tables */
    {
        if (zp->offset == zone) {
            if (dst & zp->dzone) return(zp->dzone);
            if (!dst & zp->stdzone) return(zp->stdzone);
        }
    }
    if (zone < 0) {
        zone = -zone;
        sign = '+';
    } else
    {
        sign = '-';
    }
    \(\text{void})\text{snprintf}(czone, sizeof(czone),
lib/libc/gen/timezone.c
(Warner Losh 1998-01-21 21:46:36 +0000 126) "GMT%c%d:%02d",sign,zone /60,zone % 60);
lib/libc/gen/timezone.c
(Rodney Grimes 1994-05-27 05:00:24 +0000 128) return(czone);
lib/libc/gen/timezone.c
(Rodney Grimes 1994-05-27 05:00:24 +0000 129) \}
Creation process

- Gather primary material (11GB)
- Populate author maps, author details
- Import command
  - Release snapshots
  - SCCS
  - (CVS), Git
- Build script
- Lookaside reference files
## Data Sources

<table>
<thead>
<tr>
<th>Tag</th>
<th>Data source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research-V1</td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v1/svntree-20081216.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v1/svntree-20081216.tar.gz</a></td>
</tr>
<tr>
<td>Research-V5</td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v5/v5root.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v5/v5root.tar.gz</a></td>
</tr>
<tr>
<td>Research-V6</td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v6/v6root.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v6/v6root.tar.gz</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v6/v6src.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v6/v6src.tar.gz</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v6/v6doc.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v6/v6doc.tar.gz</a></td>
</tr>
<tr>
<td>BSD-1</td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/ucb/1bsd.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/ucb/1bsd.tar.gz</a></td>
</tr>
<tr>
<td>BSD-2</td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/ucb/2bsd.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/ucb/2bsd.tar.gz</a></td>
</tr>
<tr>
<td>Research-V7</td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Henry_Spencer_v7/v7.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Henry_Spencer_v7/v7.tar.gz</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Henry_Spencer_v7/v7.patches.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Henry_Spencer_v7/v7.patches.tar.gz</a></td>
</tr>
<tr>
<td>Bell-32V</td>
<td><a href="http://www.tuhs.org/Archive/VAX/Distributions/32V/32v_usr.tar.gz">http://www.tuhs.org/Archive/VAX/Distributions/32V/32v_usr.tar.gz</a></td>
</tr>
<tr>
<td>BSD-3</td>
<td><a href="http://www.tuhs.org/Archive/4BSD/Distributions/3bsd.tar.gz">http://www.tuhs.org/Archive/4BSD/Distributions/3bsd.tar.gz</a></td>
</tr>
<tr>
<td>BSD-4</td>
<td>file://CSR-CD-ROMS/cd1/4.0</td>
</tr>
<tr>
<td>BSD-4_1_snap</td>
<td>file://CSR-CD-ROMS/cd1/4.1_snap</td>
</tr>
<tr>
<td>BSD-4_1c_2</td>
<td>file://CSR-CD-ROMS/cd1/4.1c_2</td>
</tr>
<tr>
<td>BSD-4_2</td>
<td>file://CSR-CD-ROMS/cd1/4.2</td>
</tr>
<tr>
<td>BSD-4_3</td>
<td>file://CSR-CD-ROMS/cd1/4.3</td>
</tr>
<tr>
<td>BSD-4_3_Tahoe</td>
<td>file://CSR-CD-ROMS/cd2/4.3tahoe</td>
</tr>
<tr>
<td>BSD-4_3_Net_1</td>
<td>file://CSR-CD-ROMS/cd2/net.1</td>
</tr>
<tr>
<td>BSD-4_3_Reno</td>
<td>file://CSR-CD-ROMS/cd2/4.3reno</td>
</tr>
<tr>
<td>BSD-4_3_Net_2</td>
<td>file://CSR-CD-ROMS/cd2/net.2</td>
</tr>
<tr>
<td>BSD-4_4</td>
<td>file://CSR-CD-ROMS/cd3/4.4</td>
</tr>
<tr>
<td>BSD-4_4_Lite1</td>
<td>file://CSR-CD-ROMS/cd2/4.4BSD-Lite1</td>
</tr>
<tr>
<td>BSD-4_4_Lite2</td>
<td>file://CSR-CD-ROMS/cd3/4.4BSD-Lite2</td>
</tr>
<tr>
<td>BSD-SCCS</td>
<td>file://CSR-CD-ROMS/cd4</td>
</tr>
<tr>
<td>386BSD-0.0</td>
<td><a href="http://www.oldlinux.org/Linux.old/distributions/386BSD/386bsd-0.0/floppies/3in/src/">http://www.oldlinux.org/Linux.old/distributions/386BSD/386bsd-0.0/floppies/3in/src/</a></td>
</tr>
<tr>
<td>386BSD-0.1</td>
<td><a href="http://www.oldlinux.org/Linux.old/distributions/386BSD/0.1/386BSD/">http://www.oldlinux.org/Linux.old/distributions/386BSD/0.1/386BSD/</a></td>
</tr>
<tr>
<td>FreeBSD-release/1.0</td>
<td><a href="http://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/ISO-IMAGES/1.0/1.0-disc1.iso">http://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/ISO-IMAGES/1.0/1.0-disc1.iso</a></td>
</tr>
<tr>
<td>FreeBSD-release/1.1.5</td>
<td><a href="http://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/ISO-IMAGES/FreeBSD-1.1.5.1/cd1.iso">http://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/ISO-IMAGES/FreeBSD-1.1.5.1/cd1.iso</a></td>
</tr>
<tr>
<td>FreeBSD-release/2...</td>
<td><a href="https://github.com/freebsd/freebsd">https://github.com/freebsd/freebsd</a></td>
</tr>
</tbody>
</table>
# "Text- and data-processing utilities:
# spell, diff, sort, join, graph, speak, etc."
usr/src/cmd/diff.*   doug
usr/src/cmd/graph\c   doug
usr/src/cmd/join\c    doug
usr/src/cmd/spell/.*  doug
bin/spell            doug

# 3. [Morris] was also the author of the series of crypt programs
# that came with early Unix, including the final one distributed with the
# Seventh Edition
# http://cm.bell-labs.com/cm/cs/who/dmr/crypt.html
usr/man/man1/crypt\1   rhm
usr/man/man3/crypt\3   rhm
usr/src/cmd/crypt\c    rhm
usr/src/libc/gen/crypt\c rhm

# 5. Volume 2 of the manual (supplementary documents)
# Based on the authors listed in each document
usr/doc/adb/.*         jfm,srb
usr/doc/adv.ed/.*      bwk
usr/doc/assembler      dmr
usr/doc/awk            aho,pjw,bwk
author Bill Joy <wnj@ucbvax.Berkeley.EDU> 315830189 -0800
commiter Bill Joy <wnj@ucbvax.Berkeley.EDU> 315830189 -0800
data 75
BSD 3 development
Work on file usr/src/cmd/ex/ex_addr.c
(Synthetic commit)
M 100644 :3 usr/src/cmd/ex/ex_addr.c
 [...] 

# Release
commit refs/heads/BSD-Release
mark :3700
author Bill Joy <wnj@ucbvax.Berkeley.EDU> 315928541 -0800
commiter Bill Joy <wnj@ucbvax.Berkeley.EDU> 315928541 -0800
data 78
BSD 3 release
Snapshot of the completed development branch
(Synthetic commit)
from :3699
merge Bell-32V
merge BSD-2
D .ref-Bell-32V
D .ref-BSD-2
tag BSD-3
from :3700
tagger Bill Joy <wnj@ucbvax.Berkeley.EDU> 315928541 -0800
data 91
Tagged 3 release snapshot of BSD with 3
Source directory: ../archive/3bsd
(Synthetic tag)
done
Extending the Data Set

• Add data download Makefile rule
• Add authorship information
• Add non-import file list
• Add tree graft import statement
• Rebuild the history repository.
• Verify checked out version matches original data
• Verify git blame / log, branches / merges
• Add corresponding verification rules
Action Items

- Use the repository for research
- Improve repository
  - Authors and maps
  - Merge concurrent SCCS, CVS commits
  - 2.* BSD
  - NetBSD, OpenBSD
- Lobby to open the code of System V
- Improve Git’s performance and accuracy
Thank you!

[GitHub repository link]

[Email address]
[Website link]
[Twitter handle]
Image Credits

- VAX 11/780: Joe Mabel
- Pentium: lorsh
- Sun: Thomas Kaiser