



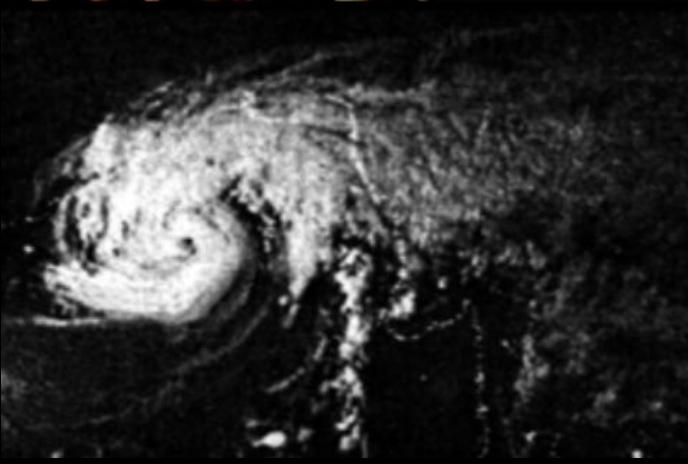
# The Unix history in a repository

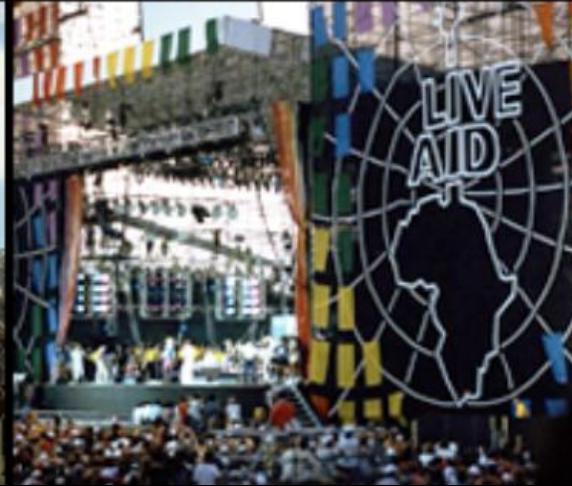
## 44 years of Unix history in one Git repo

Diomidis Spinellis

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

[www.spinellis.gr](http://www.spinellis.gr)  
[dds@auerb.gr](mailto:dds@auerb.gr)  
[@CoolSWEEng](https://twitter.com/CoolSWEEng)









LIVE FREE OR DIE

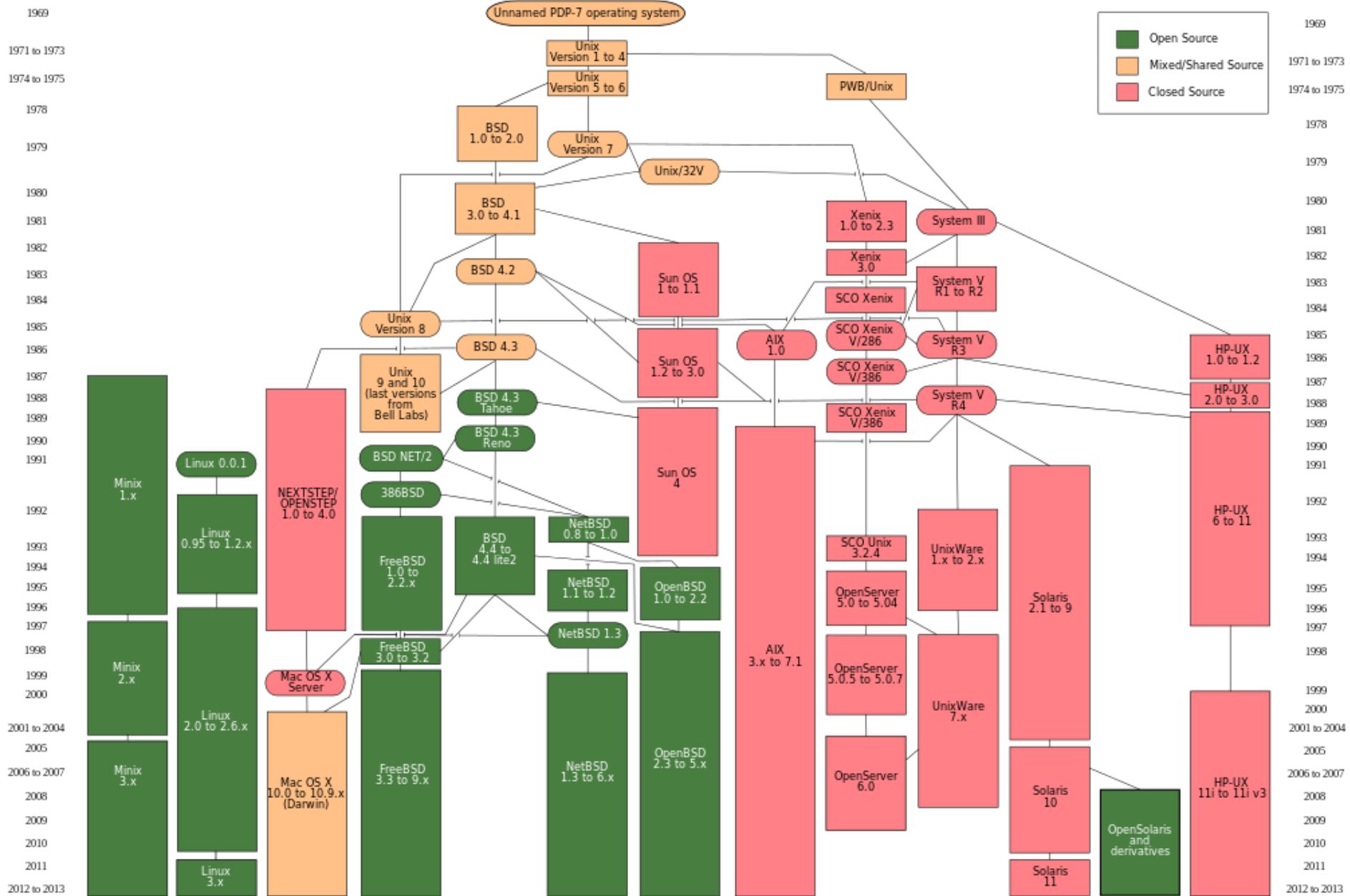
UNIX\*

TRADEMARK OF BELL LABS\*











240 West Center Street  
Orem, Utah 84057  
801-765-4999 Fax 801-765-4481

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

# 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

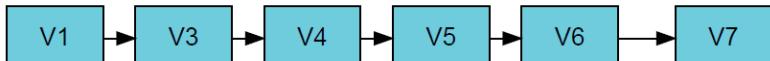
# In Numbers ...

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

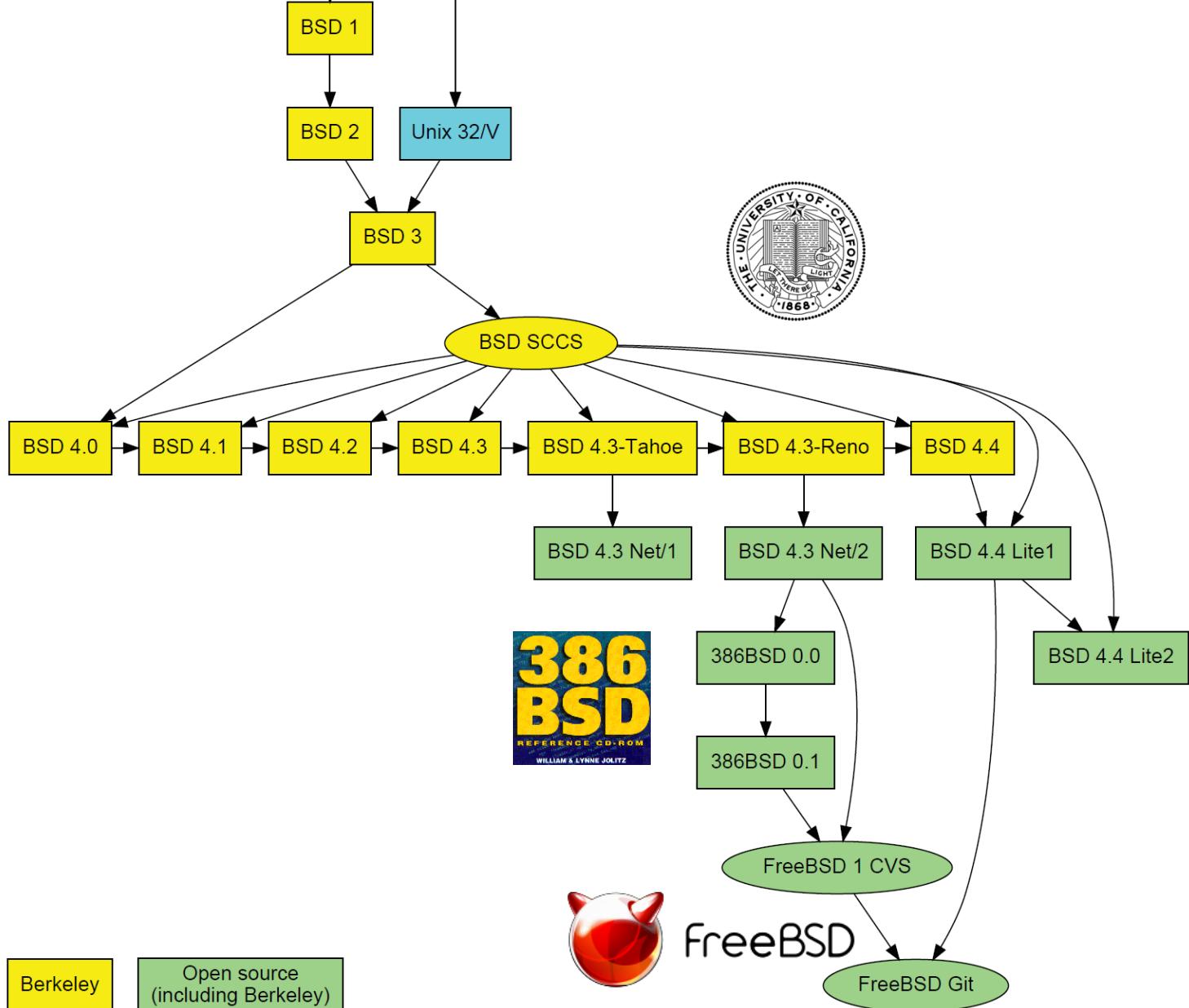
# Repository Contents

- 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





**Bell Laboratories**







Bell Laboratories

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

Copy to  
Mr. G. L. Baldwin

T. R. Bashkow

*TR Bashkow*

/ initialize inodes for special files (inodes 1 to 40.)

1:  
    mov     \$40.,r1 / set r1=i-node-number 40.  
    jsr     r0,iget / 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,seitimod / 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.

```

/u0 -- unix

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

rkda = 177412 / disk address reg
rkds = 177400 / driv status reg
rkc5 = 177404 / control status reg
rcsr = 174000 / receiver status reg
rcbr = 174002 / receiver buffer reg
tcsr = 174004 / xmtr status reg
tcbr = 174006 / xmtr buffer reg
tcst = 177340 / dec tape control status
tccm = 177342 / dec tape command reg
tcwc = 177344 / word count
tcba = 177346 / bus addr
tcdt = 177350 / data reg
dcs = 177460 / drum control status
dae = 177470 / drum address extension
lks = 177546 / clock status reg
prs = 177550 / papertape reader status
prb = 177552 / buffer
pps = 177554 / punch status
ppb = 177556 / punch buffer
/lps = 177514 line printer status
/lpb = 177516 line printer buffer
tks = 177560 / console read status
tkb = 177562 / read buffer
tps = 177564 / punch status
tpb = 177566 / punch buffer
ps = 177776 / processor status

halt = 0
wait = 1
rti = 2

nproc = 16. / number of processes
nfiles = 50.
ntty = 8+1
nbuf = 6
.if cold / ignored if cold = 0
nbuf = 2
.endif

core = orig+40000 / specifies beginning of user's core
ecore = core+20000 / specifies end of user's core (4096 w)
/freq 4;4 init by copy
/0;2 unkni;0 bus error
/4;6 unkni;0 illg in tr
/10;11 ffpsym;0 illg in tr
/14;16 unkni;0 trace and trap (see Sec. B.1 page )
/20;22 unkni;0 / trap
/24;26 panic;0 / pwr
/30;32 rtssym;0 / emt
/34;36 sysent;0 / sys

```

```

mov $idata,r0 / r0=base addr. of assembled directories.
mov $u.off,u.ofop / pointer to u.off in u.ofop (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
       jsr r0,imap / locate the inode map bit for i-node 'r1'
       bisb md,(r2) / set the bit to indicate the i-node is not
                     / available
       jsr r0,iget / read inode 'r1' from disk into inode area of
                     / core and write modified i-node on drum (if any)
       mov (r0)+,i.flgs / 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
       jsr r0,setimod / set imod=1 to indicate inode modified: also,
                     / stuff time of modification into i-node
       mov (r0)+,u.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
       clr u.off / clear file offset used in 'seek' and 'tell'
       add u.count,r0 / r0 points to the header of the next directory
       jsr 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,sb0 / sb0 I/O queue entry for superblock on drum;
                     / set bit 10 to 1
     jsr r0,ppoke / read drum superblock

1:      tstb sb0+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
2:      </etc/init> / UNIX looks for strings term, noted by null\0
     (but first)
     this file#17 luid on EO,d see EO,10

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

```

## 10 44-5 sysfork

**FUNCTION:** sysfork creates a new process. This process is referred to as the child process. This new processes core image is a copy of that of the caller of "sysfork". The only distinction is the return location and the fact that (viro) in the old process (parent) contains the process id (ppid) of the new process (child). This id is used by "syswait". "sysfork" 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 "puthi".
- 5) A new process name is gotten from mpid. (actually it's a unique number) and is put in the child's unique identifier; the process id (ppid).
- 6) The process name of the parent is then obtained and placed in the unique identifier of the parent process of the child. (p.ppid) The parent process name is then put in (viro).
- 7) The child process is then written out on disk by "wswap", i.e., the parent process is copied onto disk and the child is born.
- 8) The parent process number is then restored to (viro).
- 9) The child process name is put in (viro).
- 10) The PC on the stack SP + 18 is incremented by 2 to create the return address for the parent process.
- 11) The ufp list is then searched to see what files the parent has opened. For each file the parent has opened, the corresponding fsp entry must be updated to indicate that the child process also has opened the file. A branch to sysret is then made.

### CALLING SEQUENCE

#### ARGUMENTS

##### INPUTS

from shell?

p.stat	status of a process	active, dead, unused
viro	parent process number	
vttyp	pointer to parents process	control tty buffer
mpid	process name generator	
ufp	- list of index into fsp table	
fsp	- 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.

## TABLE OF CONTENTS

### I. COMMANDS

:	place label
acct	get connect-time accounting
ar	archive (combine) files
as	assembler
bas	BASIC dialect
cat	concatenate (or print) files
cc	compile C program
chdir	change working directory
check	check consistency of file system
chmod	change access mode of files
chown	change owner of files
cmp	compare file contents
cp	copy file
date	get date and time of day
db	symbolic debugger
dc	desk calculator
df	find free disk space
dpd	spawn data-phone daemon
ds	verify directory hierarchy
dsw	delete files interactively
du	find disk usage
echo	print command arguments
ed	text editor
exit	end command sequence
fc	compile Fortran program
fed	form-letter editor
find	find file with given name
form	generate form letter
goto	command transfer
if	conditional command
istat	file status by i-number
ld	link editor (loader)
ln	link to file
login	log on to system
ls	list contents of directory
mail	send mail to another user
man	run off manual section
mesg	permit or deny messages
mkdir	create directory
mount	mount detachable file system
mt	save/restore files on magtape
mv	move or rename file
m6	macroprocessor
nm	print namelist
nroff	format text for printing
od	octal dump of file
opr	print file off-line
ov	page overlay file print
pr	print file with headings
rew	rewind DECTape
rm	remove (delete) file

rmdir ..... remove (delete) directory  
 roff ..... format text for printing  
 salv ..... repair damaged file system  
 sh .. command interpreter  
 sort ..... sort ASCII file  
 stat ..... get file status  
 strip ..... remove symbols, relocation bits  
 stty ..... set typewriter modes  
 su .. become super-user  
 sum ..... sum file  
 tacct ..... connect-time accounting  
 tap ..... manipulate DECTape  
 tm .. get time information  
 tss ..... communicate with MH-TSS (GCOS)  
 tty ..... find name of terminal  
 type ..... print file page-by-page  
 umount ..... dismount removable file system  
 un ..... find undefined symbols  
 wc .. get (English) word count  
 who ..... who is on the system  
 write ..... write to another user

## II. SYSTEM CALLS

break ..... set program break  
 cemt ..... catch EMT traps  
 chdir ..... change working directory  
 chmod ..... change mode of file  
 chown ..... change owner of file  
 close ..... close open file  
 creat ..... create file  
 exec ..... execute program file  
 exit ..... terminate execution  
 fork ..... create new process  
 fstat ..... status of open file  
 getuid ..... get user ID  
 gtty ..... get typewriter mode  
 hog ..... set low-priority status  
 ilgins ..... catch illegal instruction trap  
 intr ..... catch or inhibit interrupts  
 link ..... link to file  
 kill ..... destroy process  
 makdir ..... create directory  
 mdate ..... set date modified of file  
 mount ..... mount file system  
 open ..... open file  
 quit ..... catch or inhibit quits  
 read ..... read file  
 rele ..... release processor  
 seek ..... move read or write pointer  
 setuid ..... set user ID  
 sleep ..... delay execution  
 stat ..... get file status  
 stime ..... set system time  
 stty ..... set mode of typewriter

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

## III. SUBROUTINES

atan ..... arctangent  
 atof ..... convert ASCII to floating  
 atoi ..... convert ASCII to integer  
 const ..... floating-point constants  
 ctime ..... convert time to ASCII  
 exp ..... exponential function  
 ftrap ..... floating-point simulator  
 ftoa ..... convert floating to ASCII  
 gerts ..... communicate with GCOS  
 getc ..... get character  
 hypot ..... compute hypotenuse  
 itoa ..... convert integer to ASCII  
 log ..... logarithm base e  
 mesg ..... print string on typewriter  
 nlist ..... read name list  
 ptime ..... print time  
 putc ..... write character or word  
 qsort ..... quicker sort  
 salloc ..... storage allocator  
 sin ..... sine, cosine  
 sqrt ..... square root  
 switch ..... transfer depending on value

## IV. SPECIAL FILES

dn0 ..... 801 ACU  
 dp0 ..... 201 Dataphone  
 lpr ..... line printer  
 mem ..... core memory  
 mt0 ..... magtape  
 ppt ..... punched paper tape  
 rf0 ..... RF disk  
 rk0 ..... RK disk  
 rp0 ..... RP disk  
 tap0 ..... DECTape  
 tty ..... console typewriter  
 tty0 ..... remote typewriter

## V. FILE FORMATS

a.out ..... assembler and loader output  
 archive ..... archive file  
 core ..... core image file

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

## VI. USER MAINTAINED PROGRAMS

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

## VII. MISCELLANEOUS

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

# Research Editions

- 1<sup>st</sup>: (Nov 1971) Printed PDP-11 kernel
- 2<sup>nd</sup>: (Jun 1972) Dump DECtape fragments of programs
- 3<sup>rd</sup>: (**Feb 1973**) 90% C kernel
- 4<sup>th</sup>: (**Nov 1973**) only troff manual
- 5<sup>th</sup>: (**June 1974**): No manual source
- 6<sup>th</sup>: (**May 1975**): Complete, widely distributed



# 32/V

- 1978
- By John Raiser and Tom London
- Bell Labs Holmdel
- VAX as a large PDP-11
  - swapping, not paging

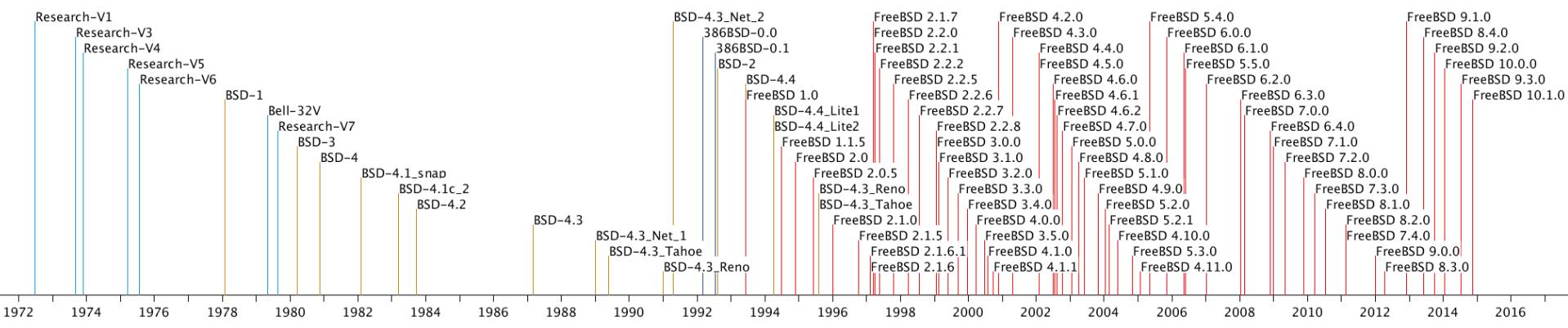
# 15 Berkeley Snapshots

- 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 enahancements

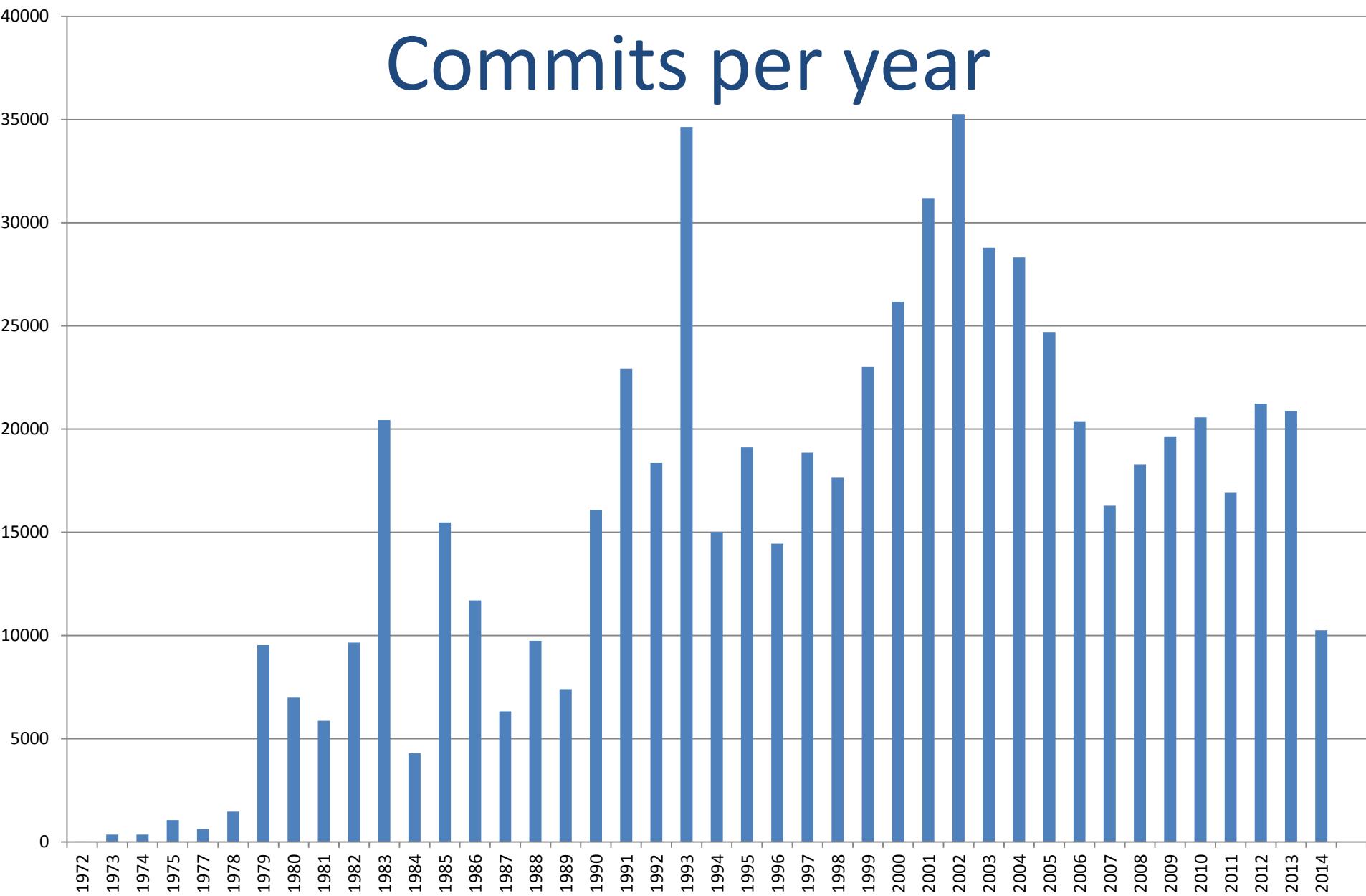




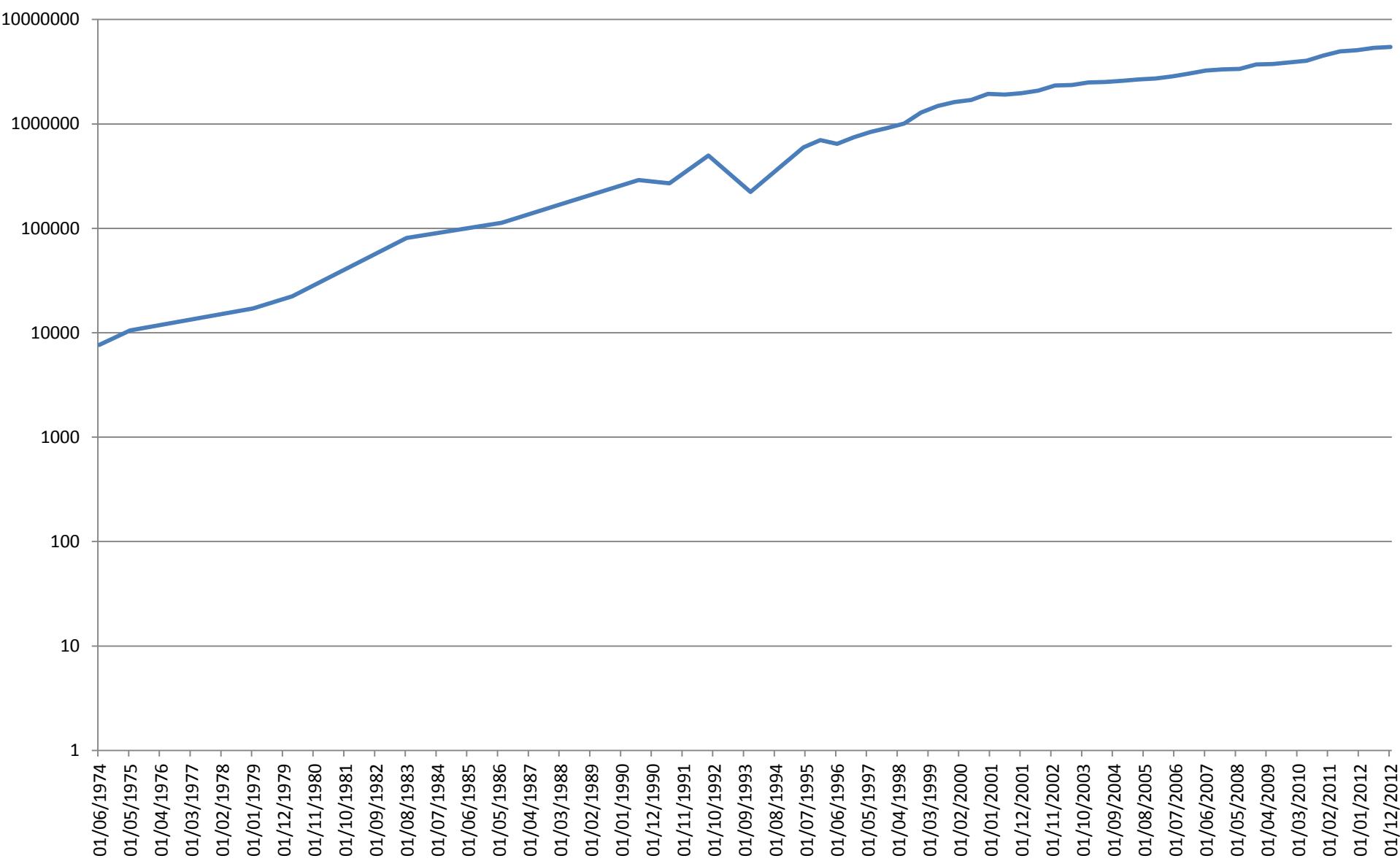
FreeBSD®



# Commits per year



# Unix Kernel (Research, BSD, FreeBSD) LOC



# Metadata

- Date
- Author
- Commit parents



COMMIT 94A21182365EBB258EEEE2AA41C5FBCB1F7FD566  
\\ AUTHOR: \\ KEN \\ THOMPSON AND \\ DENNIS \\ RITCHIE <RESEARCH !{ KEN, DMR } >  
\\ DATE: \\ TUE \\ JUN 20 04:00:00 1972 -0500

\\ RESEARCH \\ V1 DEVELOPMENT

\\ WORK ON FILE U5.S

[ ... ]

COMMIT B7B2640B9E27415D453A8FBE975A87902A01849D  
\\ AUTHOR: \\ KEN \\ THOMPSON <RESEARCH !KEN>  
\\ DATE: \\ TUE \\ NOV 26 18:13:21 1974 -0500

\\ RESEARCH \\ V5 DEVELOPMENT

\\ WORK ON FILE USR/SYS/KEN/SLP.C

[ ... ]

COMMIT 3D19667A65D35A411D911282ED8B87E32A56A349  
\\ AUTHOR: \\ DENNIS \\ RITCHIE <RESEARCH !DMR>  
\\ DATE: \\ MON \\ DEC 2 18:18:02 1974 -0500

\\ RESEARCH \\ V5 DEVELOPMENT

\\ WORK ON FILE USR/SYS/DMR/KL.C

[ ... ]

COMMIT 171931A3F6F28CE4D196C20FDEC6A4413A843F89  
\\ AUTHOR: \\ BRIAN \\ W. \\ KERNIGHAN <RESEARCH !BWK>  
\\ DATE: \\ TUE \\ MAY 13 19:43:47 1975 -0500

\\ RESEARCH \\ V6 DEVELOPMENT

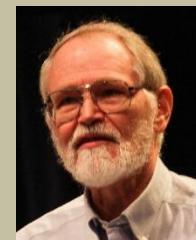
\\ WORK ON FILE RAT/R.G

[ ... ]

COMMIT AC4B13BCA433A44A97689AF10247970118834696  
\\ AUTHOR: \\ S. \\ R. \\ BOURNE <RESEARCH !SRB>  
\\ DATE: \\ FRI \\ JAN 12 02:17:45 1979 -0500

\\ 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



Google

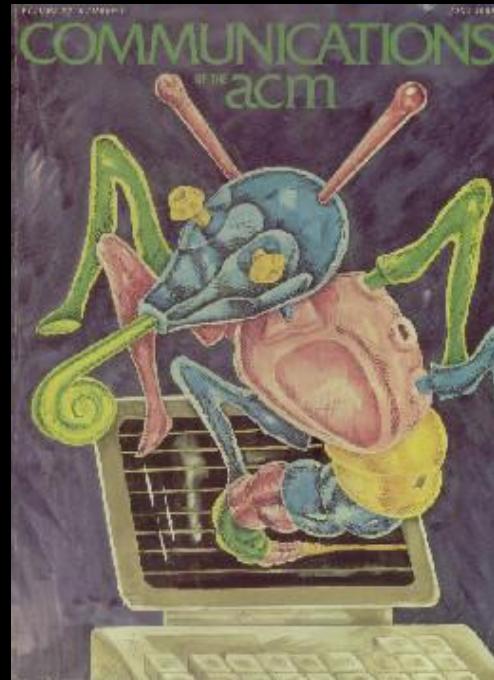
commit a7436288bb7f734dfe0b44d3c5d58f9d57fbb6c0  
Author: Eric Allman <eric@ucbvax.Berkeley.EDU>  
Date: Sat Jun 26 02:38:29 1982 -0800

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

        case CMDDBGDEBUG: /* set debug mode */
        Debug = atoi(p);
        message("200", "Debug = %d", Debug);
        break;
# endif DEBUG
```



```
$ git blame -C -C usr/sys/sys/pipe.c
```

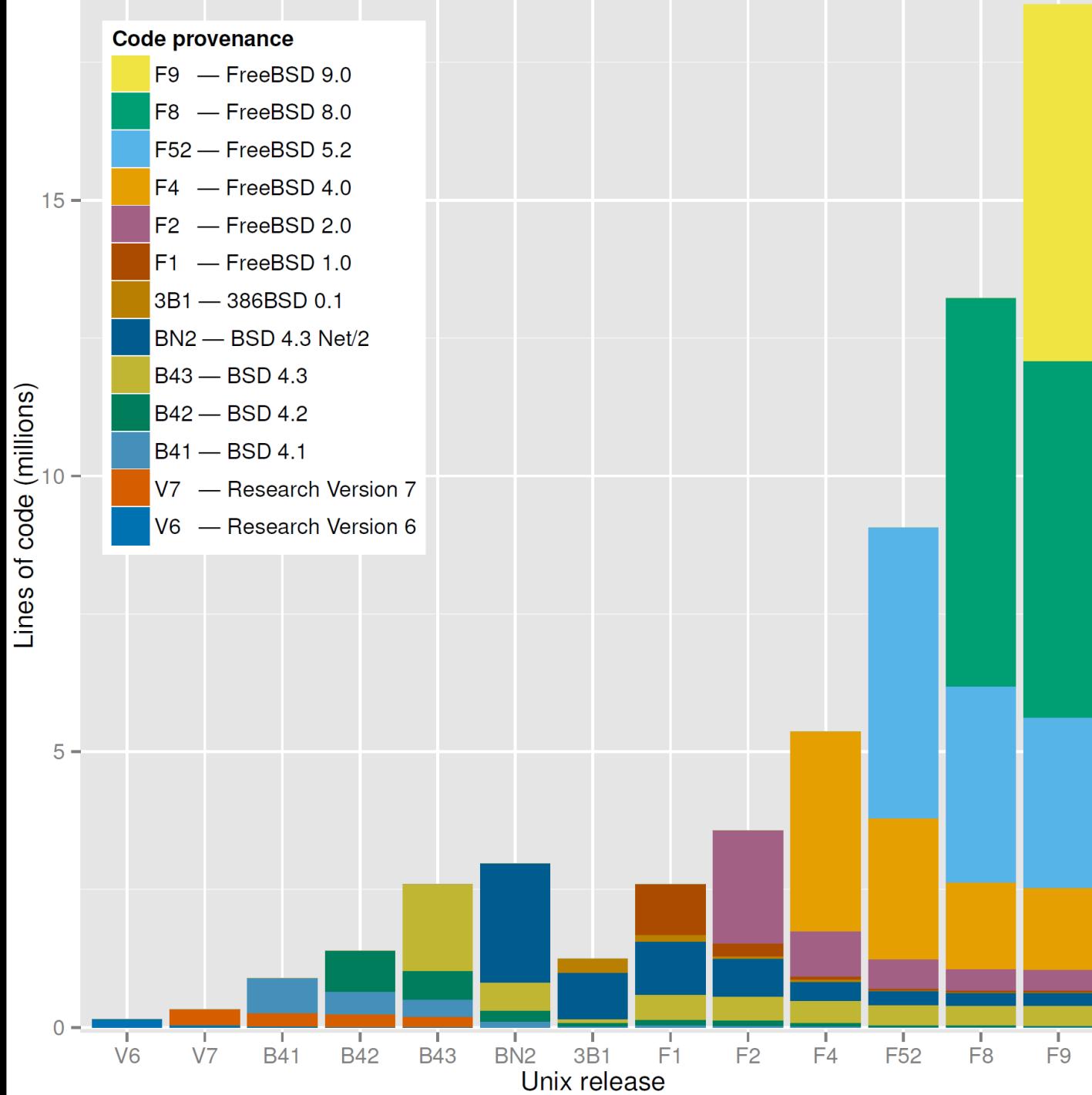
```
3cc1108b usr/sys/ken/pipe.c (Ken Thompson 1974-11-26 18:13:21 -0500 53) rf->f_flag = FREAD|FPIPE;
3cc1108b usr/sys/ken/pipe.c (Ken Thompson 1974-11-26 18:13:21 -0500 54) rf->f_inode = ip;
3cc1108b usr/sys/ken/pipe.c (Ken Thompson 1974-11-26 18:13:21 -0500 55) ip->i_count = 2;
[...]
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 122) register struct inode *ip;
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 123)
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 124)
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 125)
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 126)
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 127) loop:
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 128)
1f183be2 usr/sys/sys/pipe.c (Ken Thompson 1979-01-10 15:19:35 -0500 129)
9a9f6b22 usr/src/sys/sys/pipe.c (Bill Joy 1980-01-05 05:51:18 -0800 130) /*
9a9f6b22 usr/src/sys/sys/pipe.c (Bill Joy 1980-01-05 05:51:18 -0800 131) * If error or all done, return.
9a9f6b22 usr/src/sys/sys/pipe.c (Bill Joy 1980-01-05 05:51:18 -0800 132)
9a9f6b22 usr/src/sys/sys/pipe.c (Bill Joy 1980-01-05 05:51:18 -0800 133)
9a9f6b22 usr/src/sys/sys/pipe.c (Bill Joy 1980-01-05 05:51:18 -0800 134)
6d632e85 usr/sys/ken/pipe.c (Ken Thompson 1975-07-17 10:33:37 -0500 135)
6d632e85 usr/sys/ken/pipe.c (Ken Thompson 1975-07-17 10:33:37 -0500 136)
6d632e85 usr/sys/ken/pipe.c (Ken Thompson 1975-07-17 10:33:37 -0500 137)
6d632e85 usr/sys/ken/pipe.c (Ken Thompson 1975-07-17 10:33:37 -0500 138)
6d632e85 usr/sys/ken/pipe.c (Ken Thompson 1975-07-17 10:33:37 -0500 139)
6d632e85 usr/sys/ken/pipe.c (Ken Thompson 1975-07-17 10:33:37 -0500 140)
```

```
$ git checkout FreeBSD-release/10.0.0
$ git blame -M -M -C ./lib/libc/gen/timezone.c
```

```
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 76) static struct zone {
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 77)     int offset;
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 78)     char *stdzone;
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 79)     char *dlzone;
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 80) } zonetab[] = {
lib/libc/gen/timezone.c (Jordan K. Hubbard 1996-07-12 18:57:58 +0000 81)     {-1*60, "MET", "MET DST"}, [...]
lib/libc/gen/timezone.c (Jordan K. Hubbard 1996-07-12 18:57:58 +0000 96)     {-1}
usr/src/lib/libc/gen/timezone.c (Bill Joy 1980-12-22 00:40:25 -0800 97) };
usr/src/lib/libc/gen/timezone.c (Bill Joy 1980-12-22 00:40:25 -0800 98)
usr/src/lib/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) {
lib/libc/gen/timezone.c (David E. O'Brien 2002-02-01 01:08:48 +0000 109)     struct zone *zp;
lib/libc/gen/timezone.c (David E. O'Brien 2002-02-01 01:08:48 +0000 110)     char sign;
usr/src/lib/libc/gen/timezone.c (Bill Joy 1980-12-22 00:40:25 -0800 111)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 112)     for (zp = zonetab; zp->offset != -1; ++zp) /* static tables */
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 113)
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 114)
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 115)
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 116)
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 117)
usr/src/libc/gen/timezone.c (Dennis Ritchie 1979-01-10 14:58:45 -0500 118)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 119)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 120) /* create one */
usr/src/lib/libc/gen/timezone.c (Bill Joy 1980-12-22 00:40:25 -0800 121)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 122)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 123)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 124)
usr/src/lib/libc/gen/timezone.c (Keith Bostic 1987-03-28 19:27:07 -0800 125)
lib/libc/gen/timezone.c (Warner Losh 1998-01-21 21:46:36 +0000 126)
sizeof(czone),
lib/libc/gen/timezone.c (Warner Losh 1998-01-21 21:46:36 +0000 127)
60,zone % 60);
lib/libc/gen/timezone.c (Rodney Grimes 1994-05-27 05:00:24 +0000 128)
lib/libc/gen/timezone.c (Rodney Grimes 1994-05-27 05:00:24 +0000 129) }

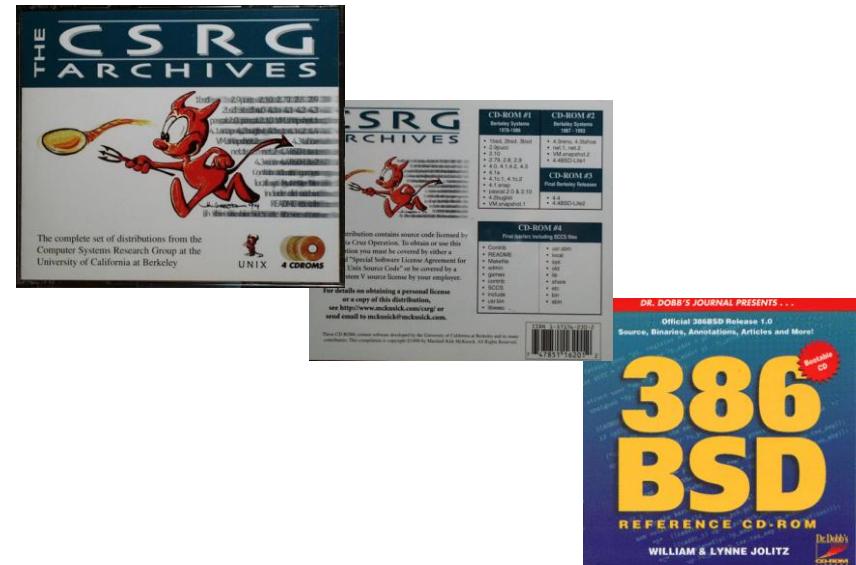
    static struct zone {
        int offset;
        char *stdzone;
        char *dlzone;
    } zonetab[] = {
        {-1*60, "MET", "MET DST"}, [...]
    };

    _tztab(int zone, int dst) {
        struct zone *zp;
        char sign;
        for (zp = zonetab; zp->offset != -1; ++zp) {
            if (zp->offset == zone) {
                if (dst && zp->dlzone)
                    return(zp->dlzone);
                if (!dst && zp->stdzone)
                    return(zp->stdzone);
            }
        }
        if (zone < 0) {
            zone = -zone;
            sign = '+';
        } else
            sign = '-';
        (void)snprintf(czone,
                      "GMT%c%d:%02d", sign, zone /
                      60, zone % 60);
        return(czone);
    }
}
```



# 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

Tag	Data source(s)
Research-V1	<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>
Research-V3	<a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v3/nsys.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v3/nsys.tar.gz</a>
Research-V4	<a href="http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v4/v4man.tar.gz">http://www.tuhs.org/Archive/PDP-11/Distributions/research/Dennis_v4/v4man.tar.gz</a>
Research-V5	<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>
Research-V6	<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> <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> <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>
BSD-1	<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>
BSD-2	<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>
Research-V7	<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> <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>
Bell-32V	<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>
BSD-3	<a href="http://www.tuhs.org/Archive/4BSD/Distributions/3bsd.tar.gz">http://www.tuhs.org/Archive/4BSD/Distributions/3bsd.tar.gz</a>
BSD-4	<a href="file:///CSRG-CD-ROMs/cd1/4.0">file:///CSRG-CD-ROMs/cd1/4.0</a>
BSD-4_1_snap	<a href="file:///CSRG-CD-ROMs/cd1/4.1.snap">file:///CSRG-CD-ROMs/cd1/4.1.snap</a>
BSD-4_1c_2	<a href="file:///CSRG-CD-ROMs/cd1/4.1c.2">file:///CSRG-CD-ROMs/cd1/4.1c.2</a>
BSD-4_2	<a href="file:///CSRG-CD-ROMs/cd1/4.2">file:///CSRG-CD-ROMs/cd1/4.2</a>
BSD-4_3	<a href="file:///CSRG-CD-ROMs/cd1/4.3">file:///CSRG-CD-ROMs/cd1/4.3</a>
BSD-4_3_Tahoe	<a href="file:///CSRG-CD-ROMs/cd2/4.3tahoe">file:///CSRG-CD-ROMs/cd2/4.3tahoe</a>
BSD-4_3_Net_1	<a href="file:///CSRG-CD-ROMs/cd2/net.1">file:///CSRG-CD-ROMs/cd2/net.1</a>
BSD-4_3_Reno	<a href="file:///CSRG-CD-ROMs/cd2/4.3reno">file:///CSRG-CD-ROMs/cd2/4.3reno</a>
BSD-4_3_Net_2	<a href="file:///CSRG-CD-ROMs/cd2/net.2">file:///CSRG-CD-ROMs/cd2/net.2</a>
BSD-4_4	<a href="file:///CSRG-CD-ROMs/cd3/4.4">file:///CSRG-CD-ROMs/cd3/4.4</a>
BSD-4_4_Lite1	<a href="file:///CSRG-CD-ROMs/cd2/4.4BSD-Lite1">file:///CSRG-CD-ROMs/cd2/4.4BSD-Lite1</a>
BSD-4_4_Lite2	<a href="file:///CSRG-CD-ROMs/cd3/4.4BSD-Lite2">file:///CSRG-CD-ROMs/cd3/4.4BSD-Lite2</a>
BSD-SCCS	<a href="file:///CSRG-CD-ROMs/cd4">file:///CSRG-CD-ROMs/cd4</a>
386BSD-0.0	<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>
386BSD-0.1	<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>
FreeBSD-release/1.0	<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>
FreeBSD-release/1.1	<a href="http://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/ISO-IMAGES/FreeBSD-1.1-RELEASE/cd1.iso">http://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/ISO-IMAGES/FreeBSD-1.1-RELEASE/cd1.iso</a>
FreeBSD-release/1.1.5	<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>
FreeBSD-release/2...	<a href="https://github.com/freebsd/freebsd">https://github.com/freebsd/freebsd</a>

# Authorship Specification

```
# 2. http://www.cs.bell-labs.com/who/doug/index.html
```

```
# "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
```

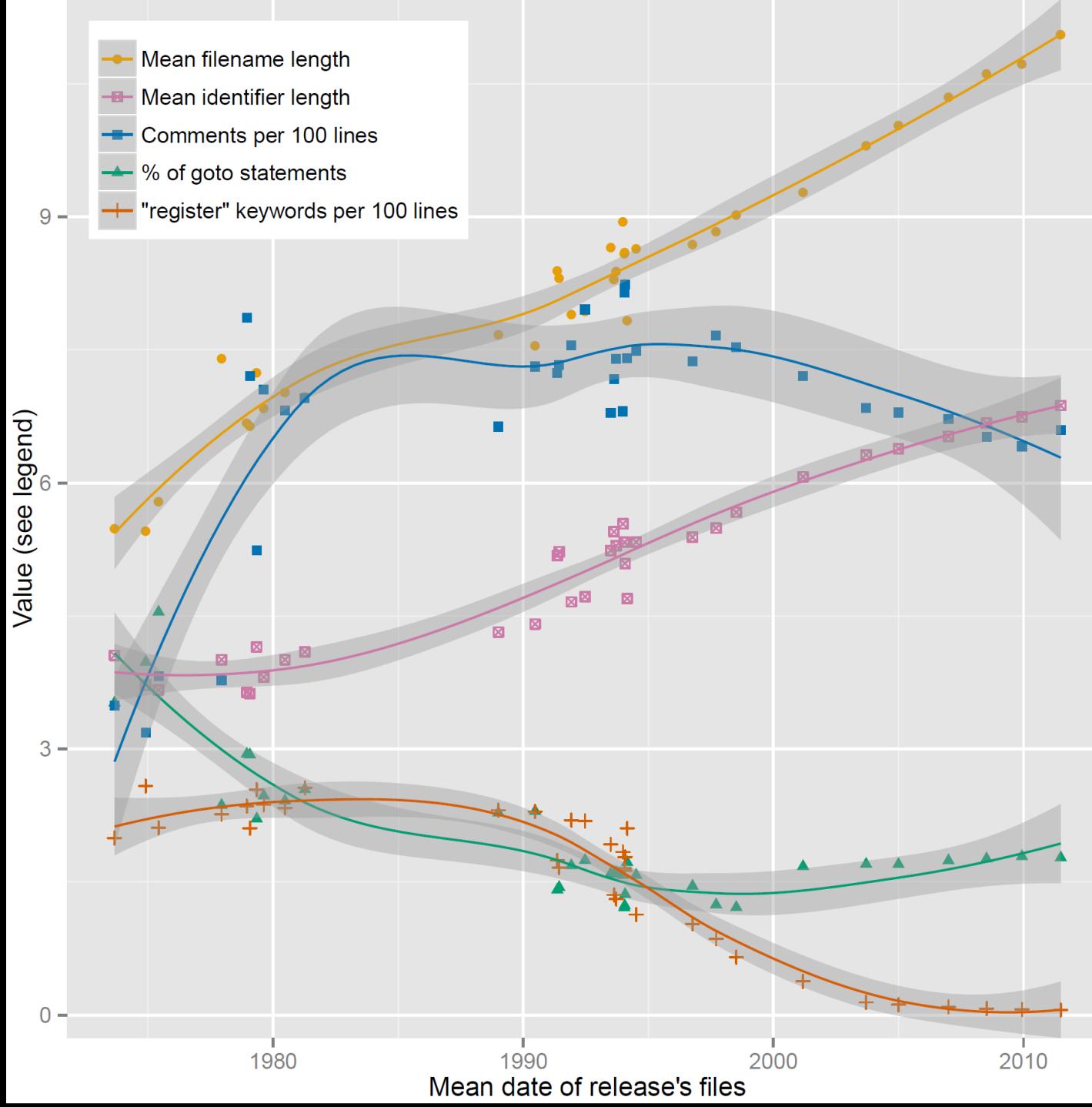
# Git Fast Import

```
# 315830189 .../archive/3bsd/usr/src/cmd/ex/ex_addr.c
blob
mark :3
data 5190
/* Copyright (c) 1979 Regents of the University of California */
#include "ex.h"
#include "ex_re.h"
[...]

# Start development commits from a clean slate
commit refs/heads/BSD-3-Snapshot-Development
mark :10
author Bill Joy <wnj@ucbvax.Berkeley.EDU> 287674317 -0800
committer Bill Joy <wnj@ucbvax.Berkeley.EDU> 287674317 -0800
data 99
Start development on BSD 3
Create reference copy of all prior development files
(Synthetic commit)
merge Bell-32V
merge BSD-2
M 100644 1468bde18e292c07e5d30ecbd7fd2b91a60e4626 .ref-Bell-
    32V/usr/include/stat.h
M 100644 1468bde18e292c07e5d30ecbd7fd2b91a60e4626 .ref-Bell-
    32V/usr/include/sys/stat.h
M 100644 816685f1f60f44dfa7e673294b9d07a12114e5 .ref-Bell-
    32V/usr/man/man2/open.2
[...]
# 315830189 .../archive/3bsd/usr/src/cmd/ex/ex_addr.c
commit refs/heads/BSD-3-Snapshot-Development
mark :13
author Bill Joy <wnj@ucbvax.Berkeley.EDU> 315830189 -0800
committer 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
committer 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!

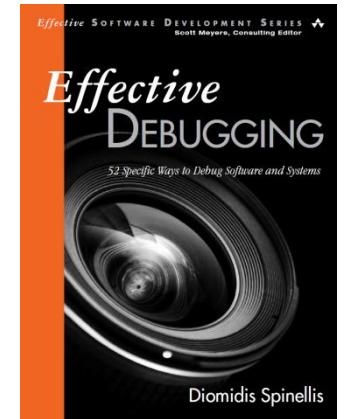
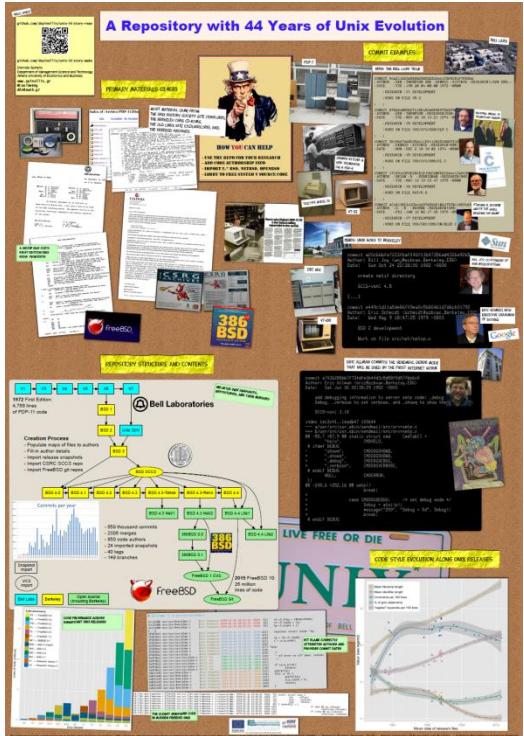
[github.com/dspinellis/unix-history-repo](https://github.com/dspinellis/unix-history-repo)



dds@aeub.gr

[www.spinellis.gr](http://www.spinellis.gr)

 @CoolSWEng



# Image Credits

- Decades: Wikipedia collages  
(1970,80,80,2000)
- VAX 11/780: [Joe Mabel](#)
- Pentium: [Iorsh](#)
- Sun: [Thomas Kaiser](#)