

Improving the SNIA Trace Repository

Andrew Glass, Tyler Wolf, Megan Campbell, & Geoff Kuenning

A trace is a series of records recorded by an instrument or program.

```

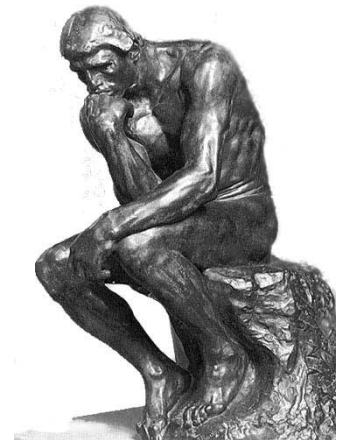
28 UID 0 PID 16 ?? A 842994410.696819 unlink("/etc/mtab") = 0
64 UID 0 PID 16 ?? S 842994410.698442 unlink("/var/run/utmp") = 0
104 UID 0 PID 16 ?? S 842994410.698700 exit(0) = 0
132 UID 0 PID 17 ?? A 842994410.843222 unlink("/etc/mtab-") = 0
168 UID 0 PID 17 ?? S 842994410.843424 exit(0) = 0
196 UID 0 PID 28 ?? A 843019612.475530 rename("/etc/modt.tmp",
"/etc/modt") = 0
248 UID 0 PID 28 ?? S 843019612.475798 exit(0) = 0
276 UID 0 PID 35 ?? G 843019614.361109 chdir("/usr/local/lib/cron") = 0
320 UID 0 PID 35 ?? A 843019614.361109 unlink(".ci.35") = 0
352 UID 0 PID 48 ?? A 843019615.159634 unlink("/etc/mtab-") = 0
388 UID 0 PID 48 ?? S 843019615.159837 exit(0) = 0
416 UID 0 PID 50 ?? S 843019615.279373 unlink("/dev/log") = -1 (2)
444 UID 0 PID 50 ?? S 843019620.207199 unlink("/dev/printer") = -1 (2)
492 UID 0 PID 71 ?? A 843019626.556377 unlink("/usr/lib/libc.so.1") = 0
536 UID 0 PID 71 ?? A 843019627.079503 unlink("/etc/ld.so.cache-") = -1 (2)
580 UID 0 PID 71 ?? A 843019627.089002 rename("/etc/ld.so.cache-",
"/etc/ld.so.cache") = 0
644 UID 0 PID 71 ?? S 843019627.089110 exit(0) = 0
672 UID 0 PID 97 ?? A 843019629.098477 rename("/etc/resolv.new",
"/etc/resolv.conf") = 0
732 UID 0 PID 97 ?? S 843019629.098745 exit(0) = 0
760 UID 0 PID 104 ?? A 843019632.039537 unlink("/tmp/00104aaa") = 0
800 UID 0 PID 104 ?? A 843019632.072698 unlink("/tmp/00104baa") = 0
840 UID 0 PID 104 ?? A 843019632.074664 unlink("/tmp/00104caa") = 0
880 UID 921 PID 159 ?? A 843019645.562265 chdir("/h/users/awang") = 0
920 UID 921 PID 159 ?? B 843019645.568375 execve("/bin/tcsh") = 0
956 UID 921 PID 159 /bin/tcsh A 843019645.612017 execve("") = 0
980 UID 921 PID 159 /bin/tcsh A 843019645.626720 stat("/etc/ld.so.cache") = 0
1000 UID 921 PID 159 /bin/tcsh A 843019645.627721 open("/etc/ld.so.cache",
O_RDONLY) = 3
1072 UID 921 PID 159 /bin/tcsh A 843019645.628764 close(3) = 0
1100 UID 921 PID 159 /bin/tcsh A 843019645.630253
open("/lib/libtermcap.so.2.0.8", O_RDONLY) = 3
1156 UID 921 PID 159 /bin/tcsh A 843019645.635437 close(3) = 0
1184 UID 921 PID 159 /bin/tcsh A 843019645.637080 open("/lib/libc.so.5.2.18",
O_RDONLY) = 3
    
```

- Worked with SNIA (Storage Networking Industry Association)
- Built a Trace Repository Website:
 - A single website that will hold many different traces for CS research
 - We needed to find new traces, convert them to DataSeries, and upload them to the repository
- DataSeries
 - An efficient, self-describing trace format developed by HP labs
 - Intended to become standard format for all traces
 - Needed interface improvements.
- Semantics
 - We needed to define *how* traces should be stored in DataSeries
 - Had to define rules for over 100 different record types



What We Accomplished

- Defined semantics for over fifty different types of records
- Vastly improved the DataSeries interface
- Improved the repository website functionality and UI
- Converted and uploaded over 30 gigabytes of traces
- Defined a new 128-bit time type for use in traces



Trace Name	Year Recorded	Time Span	Record Count	Description	File Size	SHA1 Checksum	Trace Download
LASR Traces	2000/2001	1 Year	566 Million	System-call I/O traces of thirteen machines taken in 2000 and 2001 Details	3.2 GB	Download	Download
Seer Traces	1996/1997	1 Year	72 Million	System-call traces (excluding read/write) of nine laptops, taken in 1996 and 1997. Details	576 MB	Download	Download
TraceFS Sample Traces	2007	16 minutes	1 million	Sample traces collected with the TraceFS system Details	58 MB	Download	Download

Future Work

- Continue improving the DataSeries interface (perhaps even a complete rewrite)
- Finish defining all of the necessary semantics
- Find and convert more traces into the DataSeries format



Funding:
NSF, Award Number 0627856