```
**********************************************************
   15032 Tue Sep 22 09:13:08 2015
new/usr/src/uts/common/exec/elf/elf_notes.c
5780 Truncated coredumps
Reviewed by: Yuri Pankov <yuri.pankov@nexenta.com>
**********************************************************
______unchanged_portion_omitted_

 166 int
 167 write_elfnotes(proc_t *p, int sig, vnode_t *vp, offset_t offset,
 168     rlim64_t rlimit, cred_t *credp, core_content_t content)
 169 {
 170         union {
 171                 psinfo_t        psinfo;
 172                 pstatus_t       pstatus;
 173                 lwpsinfo_t      lwpsinfo;
 174                 lwpstatus_t     lwpstatus;
 175 #if defined(__sparc)
 176                 gwindows_t      gwindows;
 177                 asrset_t        asrset;
 178 #endif /* __sparc */
 179                 char            xregs[1];
 180                 aux_entry_t     auxv[__KERN_NAUXV_IMPL];
 181                 prcred_t        pcred;
 182                 prpriv_t        ppriv;
 183                 priv_impl_info_t prinfo;
 184                 struct utsname  uts;
 185         } *bigwad;

 187         size_t xregsize = prhasx(p)? prgetprxregsize(p) : 0;
 188         size_t crsize = sizeof (prcred_t) + sizeof (gid_t) * (ngroups_max - 1);
 189         size_t psize = prgetprivsize();
 190         size_t bigsize = MAX(psize, MAX(sizeof (*bigwad),
 191             MAX(xregsize, crsize)));

 193         priv_impl_info_t *prii;

 195         lwpdir_t *ldp;
 196         lwpent_t *lep;
 197         kthread_t *t;
 198         klwp_t *lwp;
 199         user_t *up;
 200         int i;
 201         int nlwp;
 202         int nzomb;
 203         int error;
 204         uchar_t oldsig;
 205         uf_info_t *fip;
 206         int fd;
 207         vnode_t *vroot;

 209 #if defined(__i386) || defined(__i386_COMPAT)
 210         struct ssd *ssd;
 211         size_t ssdsize;
 212 #endif  /* __i386 || __i386_COMPAT */

 214         bigsize = MAX(bigsize, priv_get_implinfo_size());

 216         bigwad = kmem_alloc(bigsize, KM_SLEEP);

 218         /*
 219          * The order of the elfnote entries should be same here
 220          * and in the gcore(1) command.  Synchronization is
 221          * needed between the kernel and gcore(1).
 222          */
```



```
 224         /*
 225          * Get the psinfo, and set the wait status to indicate that a core was
 226          * dumped.  We have to forge this since p->p_wcode is not set yet.
 227          */
 228         mutex_enter(&p->p_lock);
 229         prgetpsinfo(p, &bigwad->psinfo);
 230         mutex_exit(&p->p_lock);
 231         bigwad->psinfo.pr_wstat = wstat(CLD_DUMPED, sig);

 233         error = elfnote(vp, &offset, NT_PSINFO, sizeof (bigwad->psinfo),
 234             (caddr_t)&bigwad->psinfo, rlimit, credp);
 235         if (error)
 236                 goto done;

 238         /*
 239          * Modify t_whystop and lwp_cursig so it appears that the current LWP
 240          * is stopped after faulting on the signal that caused the core dump.
 241          * As a result, prgetstatus() will record that signal, the saved
 242          * lwp_siginfo, and its signal handler in the core file status.  We
 243          * restore lwp_cursig in case a subsequent signal was received while
 244          * dumping core.
 245          */
 246         mutex_enter(&p->p_lock);
 247         lwp = ttolwp(curthread);

 249         oldsig = lwp->lwp_cursig;
 250         lwp->lwp_cursig = (uchar_t)sig;
 251         curthread->t_whystop = PR_FAULTED;

 253         prgetstatus(p, &bigwad->pstatus, p->p_zone);
 254         bigwad->pstatus.pr_lwp.pr_why = 0;

 256         curthread->t_whystop = 0;
 257         lwp->lwp_cursig = oldsig;
 258         mutex_exit(&p->p_lock);

 260         error = elfnote(vp, &offset, NT_PSTATUS, sizeof (bigwad->pstatus),
 261             (caddr_t)&bigwad->pstatus, rlimit, credp);
 262         if (error)
 263                 goto done;

 265         error = elfnote(vp, &offset, NT_PLATFORM, strlen(platform) + 1,
 266             platform, rlimit, credp);
 267         if (error)
 268                 goto done;

 270         up = PTOU(p);
 271         for (i = 0; i < __KERN_NAUXV_IMPL; i++) {
 272                 bigwad->auxv[i].a_type = up->u_auxv[i].a_type;
 273                 bigwad->auxv[i].a_un.a_val = up->u_auxv[i].a_un.a_val;
 274         }
 275         error = elfnote(vp, &offset, NT_AUXV, sizeof (bigwad->auxv),
 276             (caddr_t)bigwad->auxv, rlimit, credp);
 277         if (error)
 278                 goto done;

 280         bcopy(&utsname, &bigwad->uts, sizeof (struct utsname));
 281         if (!INGLOBALZONE(p)) {
 282                 bcopy(p->p_zone->zone_nodename, &bigwad->uts.nodename,
 283                     _SYS_NMLN);
 284         }
 285         error = elfnote(vp, &offset, NT_UTSNAME, sizeof (struct utsname),
 286             (caddr_t)&bigwad->uts, rlimit, credp);
 287         if (error)
 288                 goto done;
```

```
290         prgetcred(p, &bigwad->pcred);

292         if (bigwad->pcred.pr_ngroups != 0) {
293                 crsize = sizeof (prcred_t) +
294                     sizeof (gid_t) * (bigwad->pcred.pr_ngroups - 1);
295         } else
296                 crsize = sizeof (prcred_t);

298         error = elfnote(vp, &offset, NT_PRCRED, crsize,
299             (caddr_t)&bigwad->pcred, rlimit, credp);
300         if (error)
301                 goto done;

303         error = elfnote(vp, &offset, NT_CONTENT, sizeof (core_content_t),
304             (caddr_t)&content, rlimit, credp);
305         if (error)
306                 goto done;

308         prgetpriv(p, &bigwad->ppriv);

310         error = elfnote(vp, &offset, NT_PRPRIV, psize,
311             (caddr_t)&bigwad->ppriv, rlimit, credp);
312         if (error)
313                 goto done;

315         prii = priv_hold_implinfo();
316         error = elfnote(vp, &offset, NT_PRPRIVINFO, priv_get_implinfo_size(),
317             (caddr_t)prii, rlimit, credp);
318         priv_release_implinfo();
319         if (error)
320                 goto done;

322         /* zone can't go away as long as process exists */
323         error = elfnote(vp, &offset, NT_ZONENAME,
324             strlen(p->p_zone->zone_name) + 1, p->p_zone->zone_name,
325             rlimit, credp);
326         if (error)
327                 goto done;


330         /* open file table */
331         vroot = PTOU(p)->u_rdir;
332         if (vroot == NULL)
333                 vroot = rootdir;

335         VN_HOLD(vroot);

337         fip = P_FINFO(p);

339         for (fd = 0; fd < fip->fi_nfiles; fd++) {
340                 uf_entry_t *ufp;
341                 vnode_t *fvp;
342                 struct file *fp;
343                 vattr_t vattr;
344                 prfdinfo_t fdinfo;

346                 bzero(&fdinfo, sizeof (fdinfo));

348                 mutex_enter(&fip->fi_lock);
349                 UF_ENTER(ufp, fip, fd);
350                 if (((fp = ufp->uf_file) == NULL) || (fp->f_count < 1)) {
351                         UF_EXIT(ufp);
352                         mutex_exit(&fip->fi_lock);
353                         continue;
354                 }
```



```
356                 fdinfo.pr_fd = fd;
357                 fdinfo.pr_fdflags = ufp->uf_flag;
358                 fdinfo.pr_fileflags = fp->f_flag2;
359                 fdinfo.pr_fileflags <<= 16;
360                 fdinfo.pr_fileflags |= fp->f_flag;
361                 if ((fdinfo.pr_fileflags & (FSEARCH | FEXEC)) == 0)
362                         fdinfo.pr_fileflags += FOPEN;
363                 fdinfo.pr_offset = fp->f_offset;


366                 fvp = fp->f_vnode;
367                 VN_HOLD(fvp);
368                 UF_EXIT(ufp);
369                 mutex_exit(&fip->fi_lock);

371                 /*
372                  * There are some vnodes that have no corresponding
373                  * path.  Its reasonable for this to fail, in which
374                  * case the path will remain an empty string.
375                  */
376                 (void) vnodetopath(vroot, fvp, fdinfo.pr_path,
377                     sizeof (fdinfo.pr_path), credp);

379                 if (VOP_GETATTR(fvp, &vattr, 0, credp, NULL) != 0) {
379                 error = VOP_GETATTR(fvp, &vattr, 0, credp, NULL);
380                 if (error != 0) {
380                         VN_RELE(fvp);
381                         VN_RELE(vroot);
382                         continue;
383                         goto done;
383                 }

385                 if (fvp->v_type == VSOCK)
386                         fdinfo.pr_fileflags |= sock_getfasync(fvp);

388                 VN_RELE(fvp);

390                 /*
391                  * This logic mirrors fstat(), which we cannot use
392                  * directly, as it calls copyout().
393                  */
394                 fdinfo.pr_major = getmajor(vattr.va_fsid);
395                 fdinfo.pr_minor = getminor(vattr.va_fsid);
396                 fdinfo.pr_ino = (ino64_t)vattr.va_nodeid;
397                 fdinfo.pr_mode = VTTOIF(vattr.va_type) | vattr.va_mode;
398                 fdinfo.pr_uid = vattr.va_uid;
399                 fdinfo.pr_gid = vattr.va_gid;
400                 fdinfo.pr_rmajor = getmajor(vattr.va_rdev);
401                 fdinfo.pr_rminor = getminor(vattr.va_rdev);
402                 fdinfo.pr_size = (off64_t)vattr.va_size;

404                 error = elfnote(vp, &offset, NT_FDINFO,
405                     sizeof (fdinfo), &fdinfo, rlimit, credp);
406                 if (error) {
407                         VN_RELE(vroot);
408                         goto done;
409                 }
410         }

412         VN_RELE(vroot);

414 #if defined(__i386) || defined(__i386_COMPAT)
415         mutex_enter(&p->p_ldtlock);
416         ssdsize = prnldt(p) * sizeof (struct ssd);
417         if (ssdsize != 0) {
418                 ssd = kmem_alloc(ssdsize, KM_SLEEP);
```

```
419                 prgetldt(p, ssd);
420                 error = elfnote(vp, &offset, NT_LDT, ssdsize,
421                     (caddr_t)ssd, rlimit, credp);
422                 kmem_free(ssd, ssdsize);
423         }
424         mutex_exit(&p->p_ldtlock);
425         if (error)
426                 goto done;
427 #endif  /* __i386 || defined(__i386_COMPAT) */

429         nlwp = p->p_lwpcnt;
430         nzomb = p->p_zombcnt;
431         /* for each entry in the lwp directory ... */
432         for (ldp = p->p_lwpdir; nlwp + nzomb != 0; ldp++) {

434                 if ((lep = ldp->ld_entry) == NULL)      /* empty slot */
435                         continue;

437                 if ((t = lep->le_thread) != NULL) {     /* active lwp */
438                         ASSERT(nlwp != 0);
439                         nlwp--;
440                         lwp = ttolwp(t);
441                         mutex_enter(&p->p_lock);
442                         prgetlwpsinfo(t, &bigwad->lwpsinfo);
443                         mutex_exit(&p->p_lock);
444                 } else {                                /* zombie lwp */
445                         ASSERT(nzomb != 0);
446                         nzomb--;
447                         bzero(&bigwad->lwpsinfo, sizeof (bigwad->lwpsinfo));
448                         bigwad->lwpsinfo.pr_lwpid = lep->le_lwpid;
449                         bigwad->lwpsinfo.pr_state = SZOMB;
450                         bigwad->lwpsinfo.pr_sname = 'Z';
451                         bigwad->lwpsinfo.pr_start.tv_sec = lep->le_start;
452                 }
453                 error = elfnote(vp, &offset, NT_LWPSINFO,
454                     sizeof (bigwad->lwpsinfo), (caddr_t)&bigwad->lwpsinfo,
455                     rlimit, credp);
456                 if (error)
457                         goto done;
458                 if (t == NULL)          /* nothing more to do for a zombie */
459                         continue;

461                 mutex_enter(&p->p_lock);
462                 if (t == curthread) {
463                         /*
464                          * Modify t_whystop and lwp_cursig so it appears that
465                          * the current LWP is stopped after faulting on the
466                          * signal that caused the core dump.  As a result,
467                          * prgetlwpstatus() will record that signal, the saved
468                          * lwp_siginfo, and its signal handler in the core file
469                          * status.  We restore lwp_cursig in case a subsequent
470                          * signal was received while dumping core.
471                          */
472                         oldsig = lwp->lwp_cursig;
473                         lwp->lwp_cursig = (uchar_t)sig;
474                         t->t_whystop = PR_FAULTED;

476                         prgetlwpstatus(t, &bigwad->lwpstatus, p->p_zone);
477                         bigwad->lwpstatus.pr_why = 0;

479                         t->t_whystop = 0;
480                         lwp->lwp_cursig = oldsig;
481                 } else {
482                         prgetlwpstatus(t, &bigwad->lwpstatus, p->p_zone);
483                 }
484                 mutex_exit(&p->p_lock);
```



```
485                 error = elfnote(vp, &offset, NT_LWPSTATUS,
486                     sizeof (bigwad->lwpstatus), (caddr_t)&bigwad->lwpstatus,
487                     rlimit, credp);
488                 if (error)
489                         goto done;

491 #if defined(__sparc)
492                 /*
493                  * Unspilled SPARC register windows.
494                  */
495                 {
496                         size_t size = prnwindows(lwp);

498                         if (size != 0) {
499                                 size = sizeof (gwindows_t) -
500                                     (SPARC_MAXREGWINDOW - size) *
501                                     sizeof (struct rwindow);
502                                 prgetwindows(lwp, &bigwad->gwindows);
503                                 error = elfnote(vp, &offset, NT_GWINDOWS,
504                                     size, (caddr_t)&bigwad->gwindows,
505                                     rlimit, credp);
506                                 if (error)
507                                         goto done;
508                         }
509                 }
510                 /*
511                  * Ancillary State Registers.
512                  */
513                 if (p->p_model == DATAMODEL_LP64) {
514                         prgetasregs(lwp, bigwad->asrset);
515                         error = elfnote(vp, &offset, NT_ASRS,
516                             sizeof (asrset_t), (caddr_t)bigwad->asrset,
517                             rlimit, credp);
518                         if (error)
519                                 goto done;
520                 }
521 #endif /* __sparc */

523                 if (xregsize) {
524                         prgetprxregs(lwp, bigwad->xregs);
525                         error = elfnote(vp, &offset, NT_PRXREG,
526                             xregsize, bigwad->xregs, rlimit, credp);
527                         if (error)
528                                 goto done;
529                 }

531                 if (t->t_lwp->lwp_spymaster != NULL) {
532                         void *psaddr = t->t_lwp->lwp_spymaster;
533 #ifdef _ELF32_COMPAT
534                         /*
535                          * On a 64-bit kernel with 32-bit ELF compatibility,
536                          * this file is compiled into two different objects:
537                          * one is compiled normally, and the other is compiled
538                          * with _ELF32_COMPAT set -- and therefore with a
539                          * psinfo_t defined to be a psinfo32_t.  However, the
540                          * psinfo_t denoting our spymaster is always of the
541                          * native type; if we are in the _ELF32_COMPAT case,
542                          * we need to explicitly convert it.
543                          */
544                         if (p->p_model == DATAMODEL_ILP32) {
545                                 psinfo_kto32(psaddr, &bigwad->psinfo);
546                                 psaddr = &bigwad->psinfo;
547                         }
548 #endif

550                         error = elfnote(vp, &offset, NT_SPYMASTER,
```

```
 551                             sizeof (psinfo_t), psaddr, rlimit, credp);
 552                         if (error)
 553                                 goto done;
 554                 }
 555         }
 556         ASSERT(nlwp == 0);

 558 done:
 559         kmem_free(bigwad, bigsize);
 560         return (error);
 561 }
```

_____***unchanged_portion_omitted_***