Codeset independent ldterm(7M) and stty(1)

1. Introduction

1.1. Project/Component Working Name:

- *Codeset independent ldterm(7M) and stty(1) to fix 4089374*

1.2. Name of Document Author/Supplier:

- *Ienup Sung (is@eng.sun.com)*

1.3. Date this Document was Created:

- *03/30/1999*

1.4. Name of Major Document Customer(s)/Consumer(s):

- *SOESC, ONSC, PSARC*
- *I-teams of OS I18N group, GE&IS I18N/L10N Engineering groups (ALC/CLE/ELC/JLC)*

1.5. Reference:

- */home/is/WWW/ldterm/ldterm_csi_projplan.ps*
  (or, http://nirvana.eng/ldterm/ldterm_csi_projplan.pdf)
- */home/is/WWW/ldterm/ldterm_csi_testplan.ps*
  (or, http://nirvana.eng/ldterm/ldterm_csi_testplan.pdf)
- */home/is/WWW/ldterm/onepager-ldterm-csi.txt*
  (or, http://nirvana.eng/ldterm/onepager-ldterm-csi.txt)
- */home/is/WWW/ldterm/ldterm_csi_boundary_slides.ps*
  (or, http://nirvana.eng/ldterm/ldterm_csi_boundary_slides.pdf)

2. Project Summary

2.1. Project Description

- Currently we are supporting many non-EUC codeset based locales like Shift JIS for Japan, BIG5 for Taiwan, GBK for China and Unicode/UTF-8 locales for European and Asian countries including Germany, France, Italy, Spain, Sweden, Japan, Korea, and, also domestic customers.

Codeset independent ldterm(7M) and stty(1)

- Since current ldterm(7M) and stty(1) are EUC representation dependent implementations, we cannot directly support any line discipline from terminal (and terminal emulators) without having a workaround solution, in other words, a pair of (code conversion) STREAMS modules that surround the ldterm(7M) module. Also, to use the workaround solution, we have to ask customers go through many steps of instructions like following Unicode locale example that is quite cumbersome and also easy to make mistake:

```
The dtterm(1) and any terminal that supports input and output of
UTF-8 codeset should have following Streams configuration:

         head <-> ttcompat <-> u8euc <-> ldterm <-> eucu8 <-> pseudo-TTY
                               -----                -----

Note: Among above, u8euc and eucu8 are the modules the en_US.UTF-8 locale is
supporting.

Make sure you already loaded the STREAMS modules into the kernel as specified in
the previous section.

To set up the above Streams configuration, use strchg(1) like below:

         system% cat > /tmp/mystreams
         ttcompat
         u8euc
         ldterm
         eucu8
         ptem
         ^D
         system% strchg -f /tmp/mystreams

Be sure that you are either superuser or the owner of the device when you are
using strchg(1) as like above. To see the current configuration of the Streams,
use strconf(1) as like below:

         system% strconf
         ttcompat
         u8euc
         ldterm
         eucu8
         ptem
         pts
         system%

To reset to the original configuration, you can again set the Streams
configuration as like below:
```

```
system% cat > /tmp/orgstreams
ttcompat
ldterm
ptem
^D
system% strchg -f /tmp/orgstreams
```

- This project is to make the ldterm(7M) and stty(1) codeset independent to solve such problems so that a simple one time use of stty(1) command will setup proper line discipline inside the ldterm(7M). Following is an example use of stty(1):

```
system% stty -defeucw
```

2.2. Key Requirements & Boundary Conditions

2.2.1. Feature Boundary

- This feature will make ldterm(7M) and stty(1) codeset independent such that customers will not have to workaround the problem to do proper line editing.
- We assume that there are only three different kind of codesets in the world:
  — EUC codeset
  — PC environment originated codeset (e.g., ShiftJIS, Big5, Johap, ...)
  — Unicode/UTF-8 codeset
  Other codesets like UCS-2 and UTF-16 that are not file system safe and also not preserving NULL byte notion of C programming language are not going to be supported

2.2.2. Quality Boundary

- To ensure there is no regression, we will perform vigorous pre-integration test on following areas:
  — File concatenation through cat(1) and more(1).
  — A series of shell interaction that will include ed(1) and vi(1) executions.
  — Line editing check, especially, reprint, erase, werase, and, kill.
  — File transfer testing by using zmodem and kermit on various files.
  — Testing with actual terminal.
- Above testing will be done on following five locales:
  — C

— de (German EUC single byte locale)
— zh (Simplified Chinese EUC multibyte locale)
— ja_JP.PCK (Japanese PCK/ShiftJIS multibyte locale)
— en_US.UTF-8 (American English Unicode/UTF-8 multibyte locale)

- All P1-P3 bugs will be resolved before the integration.

2.2.3. Resource Boundary

- The project is fully staffed for development and testing.
- Once buyback is done, the maintenance will be provided by ON. If necessary, however, GE&IS I18N group will work on any necessary maintenance works on the deliverables from the project on the request of the ON group.

2.2.4. Schedule Boundary

- We will deliver output from this project to Solaris 8 as soon as we get the approvals.
- Detail schedule can also be found at §5.

2.3. Reference to SAC Review

- Case number: PSARC/1999/140
- Inception review scheduled at: 5/26/1999

2.4. Risks & Assumptions

2.4.1. Assumption

We assume that there are only three different kind of codesets in the world:
— EUC codeset
— PC environment originated codeset (e.g., ShiftJIS, Big5, Johap, ...)
— Unicode/UTF-8 codeset
Other codesets like UCS-2 and UTF-16 that are not file system safe and also not preserving NULL byte notion of C programming language are not going to be supported.

Since we will not support any other codesets that are not superset of ASCII, not file system safe, and, not preserving NULL byte notion of C language, we believe there is no need to have risk management/contingency plan on this matter.

2.4.2. Risks

Risk: Due to lack of PSARC slots, the PSARC inception review is scheduled at 5/26. Thus, it is possible that PSARC may require changes and thus delay the schedule.

Likelihood of Risk: Med.

Management of Risk: We have asked to bump up the PSARC slot if there are any cancellations. We are also looking into a possible fast-track.

2.5. Dependencies

- None.

2.6. Design Review

- A design review will take place before May 7, 1999 and will include engineers from OS I18N group (Kenjiro Tsuji, Robbin Kawabata), GE&IS I18N (Ienup Sung), and, Internet Engineering Group (Randall Fishel).

3. Business Summary

3.1. Business Justification

- This project makes our ldterm(7M) and stty(1) codeset independent like other Unix vendors.
- There were various requesters:

  Domestic and China, Europe, Japan, Korea, Taiwan field offices and customers, especially, developer communities. Foreign Language Committee (FLC) of U.S. Government, Computer Systems' Market Development Engineering group, Oracle Corp. test and development teams.

3.2. Competitive Analysis

- It seems other Unix vendors either do not care about this particular issue or they have already resolved this problem in their own way. Microsoft limits their console application to a certain codeset and also the console or terminal emulator does not have a concept of STREAM like ours but provides correct line editing functionalities without any special settings.

4. Deliverables

- We will deliver the following to the WOS:

/kernel/strmod/ldterm	32-bit CSI ldterm
/kernel/strmod/sparcv9/ldterm	64-bit sparcv9 CSI ldterm
/usr/bin/stty and /usr/xpg4/bin/stty	CSI stty commands

- stty(1) is the only man page that needs modification as a result of this project. A man page request number is 990416-01.
- About one month before of actual integration, we will also provide a localization guide to the localization centers so that they will be prepared for their locale data file and also documentation changes. This guide will be written by a technical writer in GE&IS I18N engineering group and delivered to the localization centers through email.

5. Schedule

Milestone	Original Date (mm/dd/yyyy)	Current/Target Date (mm/dd/yyyy)	Completion Date (mm/dd/yyyy)
Requirements	1/29/1999		1/29/1999
Design Review	2/26/1999	5/7/1999	
PSARC Inception Review	5/26/1999		
One Pager Review	2/22/1999		2/22/1999
ONSC Sponsor Assigned	4/1999		4/9/1999
ONSC Inception/Boundary Review	4/30/1999		
Prototype Complete	4/30/1999		
Draft Project Plan	2/22/1999		4/1/1999
Draft Test Plan	2/22/1999		4/1/1999
Man Request and Docs Submitted	4/15/1999		4/16/1999
Test Development Complete	5/10/1999		
Development Complete	5/10/1999		
Forward changed ldterm & stty to all L10N centers for verification	5/10/1999		
PSARC Final Approval	5/26/1999		
ONSC Commitment Review for S8	4/30/1999		
Completed Code Review	5/14/1999		
C-Team Review	5/17/1999		
Man Pages, Doc Complete	6/7/1999		

Milestone	Original Date (mm/dd/yyyy)	Current/Target Date (mm/dd/yyyy)	Completion Date (mm/dd/yyyy)
Testing Complete	6/3/1999		
Ready To Integrate for S8	6/4/1999		

6. Resources

- Staffing:

Ienup Sung	Lead engineer/part-time designer/developer (30%)
Prabhat Hegde	Part-time developer (30%)
Suresh Chandrasekharan	Test engineer (30%)
Robin Hodgkin	Technical writer (10%)

- Post-integration maintenance will be provided by ON group. However, if necessary, GE&IS I18N group will work on any necessary maintenance works on the deliverables from the project on the request of the ON group.
- The ON CTE organization will support this product after FCS. GE&IS I18N/L10N CTE group will provide help on request from the ON CTE group.

7. Acceptance Requirements

7.1. Quality

- Successful completion of test plan execution.
- All known P1-P3 bugs will be fixed before RTI.

7.2. Performance

- There will be no performance degradation to existing control path of the ldterm(7M) for any codeset locales. There will be no performance degradation to existing control path of the stty(1) for any codeset locales. The method on which we plan to measure performance is outlined in the test plan (§3.2).

7.3. Functionality

- Correct line discipline for any locale/codeset that we are supporting and we are going to support in future will be provided and supported. The codesets, however, should be a ASCII-superset, file system safe, and, preserves NULL byte notion of C language.

7.4. Size Footprint

- We are expecting the size of 32-bit sparc version of ldterm(7M) increase to 34KB (currently 32KB) and stty(1) increase to 28KB (currently 27KB). About same amount of size increase is expected at both sparcv9 and i386 version of binaries.

7.5. Legal Documentation

- There is no relevant legal documentation required to this project.

7.6. Completion Criteria

- Approval of PSARC case.
- Completion of man page change.
- Forwarding of localization guide to the localization centers.
- Successful verifications from the localization centers on the project deliverables.
- Successful completion of test plan execution.
- Successful delivery of components to ON gate.
- All new/modified code must pass cstyle, hdrchk, lint, keywords.

8. Internationalization and Year 2000

- We will make sure this internationalization project meets guidelines of software internationalization. We will also make sure delivering localization guide to the localization centers for proper localizations of their locales.
- The deliverables of the project is Y2000 conforming since there is no date/time related functionalities being used, changed, or, created.

9. Other Information

- ISO/IEC, ISO/IEC 10646-1:1993(E), "Information Technology—Universal Multiple-Octet Coded Character Set (UCS)—Part 1: Architecture and Basic Multilingual Plane," May 1993, (International Standard).
- The Unicode Consortium, "The Unicode Standard: Worldwide Character Encoding Version 1.0, Volume 1 and 2," October 1991.
- The Unicode Consortium, "The Unicode Standard: Version 2.0," July 1996.
- Sun Microsystems, Inc., STREAMS Programmer's Guide (Part No. 801-6679-10), 1994.
- Sun Microsystems, Inc., Solaris Internationalization Guide For Developers (http://docs.sun.com), 1998.
- http://nirvana.eng/ldterm/onepager-ldterm-csi.txt
- http://nirvana.eng/ldterm/ldterm-csi.txt (design memo)
- http://nirvana.eng/ldterm/ldterm-euc-dependency.txt and http://nirvana.eng/ldterm/stty-euc-dependency.txt (investigation memos)

10. Solaris Update (if applicable)

- N/A.