9931
Rcvd 30May72

NETWORK WORKING GROUP
REQUEST FOR COMMENTS #338
NIC #9931

R. T. BRADEN
UCLA/CCN
MAY 17, 1972

EBCDIC/ASCII MAPPING FOR NETWORK RJE

A. INTRODUCTION

Under NETRJS(1), CCN's Network rje protocol(2), a virtual remote batch terminal may be either EBCDIC or ASCII. CCN operates an IBM 360/91 which performs all of its normal processing in EBCDIC. When a virtual ASCII terminal signs onto NETRJS, CCN translates the "card reader" stream to EBCDIC and translates the "printer" stream back to ASCII(3).

In recent months, a number of ASCII hosts (RAND PDP-10, Utah PDP-10, Illinois PDP-11) have completed user processes for NETRJS. Several users at these sites have noted deficiencies in the ASCII/EBCDIC mapping rules originally implemented in NETRJS. Since their objections were well founded, we have altered the existing mapping and added a new one.

This RFC has three purposes:

(1) to make all users of NETRJS aware of the changed ASCII mapping;

(2) to call this problem to the attention of the Network RJE Protocol Committee;

(3) to knowledge and support Joel Winett's pioneering work(4) in this area.

B. THE EBCDIC CHIMERA

A year ago, Joel Winett published RFC #183, containing the results of his careful research into just what EBCDIC really means. He sounded a clarion call for all EBCDIC sites to join in defining a Network standard mapping. At that time, we at CCN were primarily absorbed in the timely implementation of the NETRJS protocol to serve an EBCDIC (!) user site, RAND, so we were not very supportive of his efforts.

RFC #183 is a valuable document; we hope a copy falls into the hands of Armonk. It is clear from RFC #183 that EBCDIC consists of a standard ("basic") set of characters, combined with a number of overlapping ad-hoc character happenings. Fortunately, if we exclude special-purpose text composition programs, IBM 360 programs use only the 89 "basic" EBCDIC graphics(5) shown in RFC #183 as well as in Figure 1. An IBM 029 "EBCDIC" keypunch can create 63 graphics: the 89 basic EBCDIC graphics less the 26 lower case letters. In fact, OS/360 requires an even smaller subset of EBCDIC, 60 characters commonly called the "PL/1 character set". The PL/1 set consists of the 89 basic graphics, less the 26 lower case letters as well as the three graphics ¢ ! " (cent sign, exclamation point, and quotation).

C. CHARACTER MAPPING IN NETRJS

We consider now the requirements of a ASCII/EBCDIC mapping for NETRJS or any rje protocol. These requirements are as follows:

Efficiency:

The translation should be character-to-character, so that the CPU operation "translate" can be used and character scans obviated. This is important because a significant volume of character data may be moved during rje operations.

Usability:

(1) All of the 89 EBCDIC graphics should be mapped into corresponding ASCII characters.

(2) The mapping should be as nearly transparent as possible, i.e., whenever the same graphic appears in both sets, it should map onto itself.

(3) To minimize the adaptation required of an EBCDIC-oriented programmer, the ASCII graphics should evoke the corresponding EBCDIC graphic, when they are not identical.

Theses considerations led us to incorporate Winett's rules III (a) and III (b) (see page 4 of RFC #183) into NETRJS:

ASCII	EBCDIC
|	|
~	¬
\	¢

This defines all 89 basic EBCDIC graphics in terms of ASCII. However, there is still a question of how to map the 6 "maverick" ASCII characters ([]{}^`   ) which are not in EBCDIC and not in the list above.

We could (and did) take the view that all CCN users are concerned only with writing and executing normal 360 programs using EBCDIC and that they would enter one of the maverick ASCII graphics only in error. Our original choice, therefore, was to map the mavericks in the input into EBCDIC question marks. We also assumed that, if a user needs to access a larger subset of EBCDIC than the basic 89 , he should do so by doing his rje directly in EBCDIC.

We now realize that there were two deficiencies in the original mapping rules.

1. The 360 program may be intended to manipulate ASCII text from the Network. In that case, the Network user needs to have all ASCII characters, including the mavericks, uniquely mapped into EBCDIC in some (standard) manner.

2. The present mapping is convenient only if a user has a full ASCII terminal. In particular, a user at an AT&T Model 33 /35 Teletype (or simulator thereof) needs a different mapping for ease of use.

For the first case, we have changed the mapping of the 6 maverick ASCII characters from "?", using instead Winett's rules III (c) and III (d):

ASCII	EBCDIC
[	X'AD'
]	X'BD'
{	X'8B'
}	X'9B'
^	X'71'
`	X'79'

For the user with a Model 33/35 Teletype, we have expanded the set of virtual remote batch terminal types, adding "TTY" to "ASCII" and "EBCDIC". A user establishes his virtual remote batch terminal as type TTY by either doing his initial ICP to socket 15 (vs. 11 for EBCDIC, 13 for ASCII), or by doing an ICP to Socket 1 and entering the command "TTYRJS" (vs. "RJS" for EBCDIC, "ARJS" for ASCII). The mapping used by NETRJS for a TTY remote site is:

Model 33 Graphic	Corresponding ASCII	EBCDIC
\	\	¬
↑	^	|
←	_	_
[	[	¢
]	]	X'BD'

This is ugly, but it is probably the best we can do.

D. CONCLUSIONS

It is obvious that one pair of translation tables won't do the job; NETRJS needs (at least) two mappings for each direction. How long will it be before an important set of users appears with a different terminal character set, requiring yet a different mapping?(6) An rje server site needs to be prepared to provide a variety of translation tables, and perhaps to allow a user to specify his own table(s); this mini-subset of "Date Reconfiguration Service" might be necessary to prevent translation-table-proliferation. The tendency in Network discussions has been to put the burden upon the user sites to adapt to different conventions. In the real world of users and servers, the server will often have to do the adapting.

E. NOTES AND REFERENCES

1. R. T. Braden, Interim NETRJS Specifications, RFC #189 (NIC #7133), July 15, 1971.
2. Please note that "RJS" is the proper name of a particular rje package written at CCN; the generic name for remote batch service is "rje".
3. Notice that the mapping question discussed in this RFC is significant only for the virtual card reader and printer connections in NETRJS. The punch connection is always transparent, i.e., never translated. The remote operator connections use the extended EBCDIC/ASCII mapping including the maverick characters, but this is not important since operator commands require only a limited character set.
4. Joel Winett, "The EBCDIC Codes and their Mapping to ASCII", RFC #183 (NIC #7127), July 21, 1971.

5. Winett lists only 88 basic EBCDIC graphics, excluding the space which he regards as a control character. This is a matter of taste, but we find it less confusing to include the space as a graphic.

6. CCN recently received a new Teletype-replacement terminal. This machine has a bastardized graphic character set—mostly ASCII, with a sprinkling of both (!) EBCDIC and TTY.

RTB:gjm
5/9/72
Rev. 5/17/72

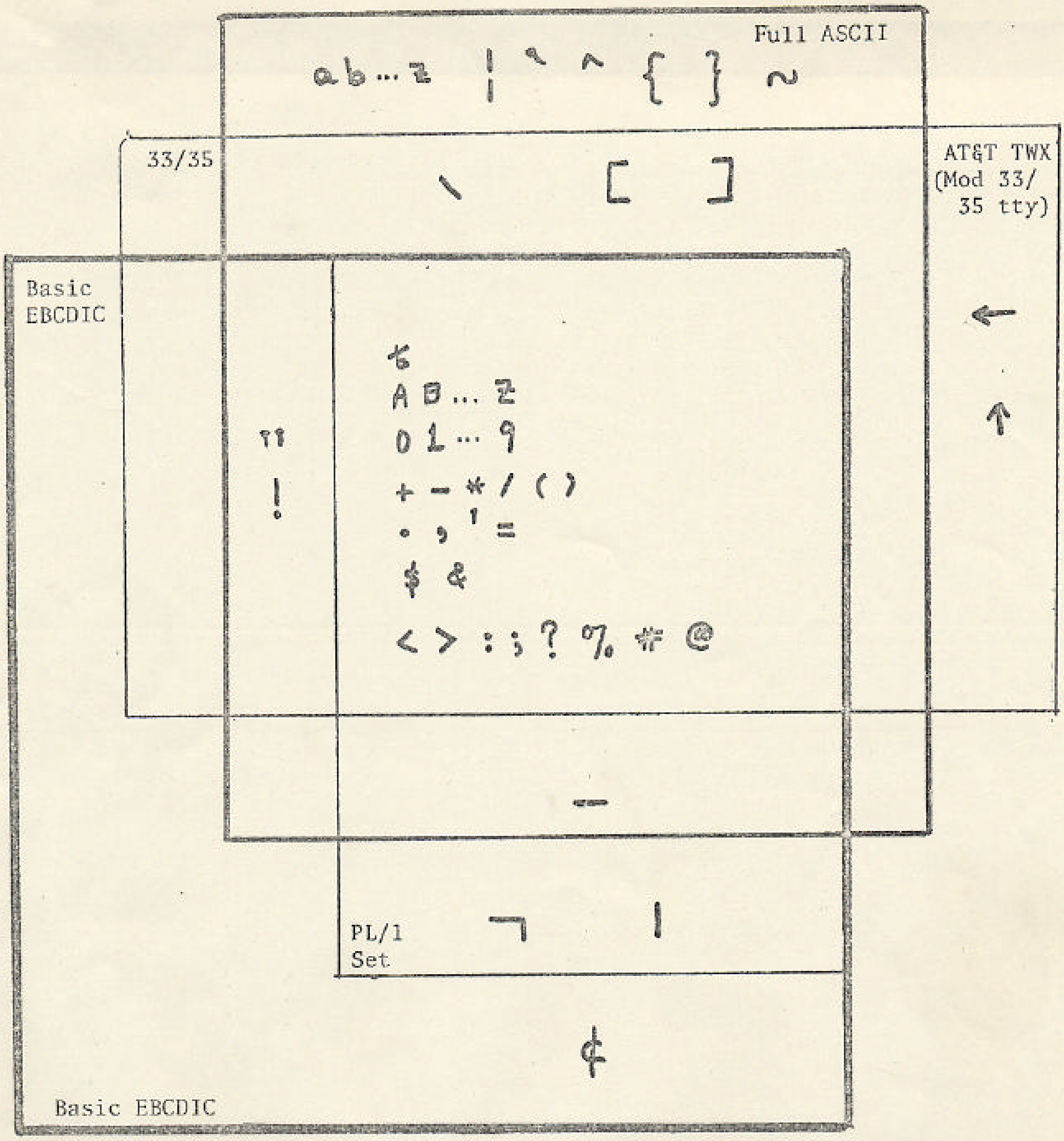

Figure 1. Character Sets Commonly Abused

rtb 5/17/72