ASCII format for *Network Interchange*

Vint Cerf

UCLA

For concreteness, we suggest the use of standard 7-bit ASCII embedded in an 8 bit byte whose high order bit is always 0. This leads to the standard code given on the attached page, copies from USAS X3, 4-1968. This code will be used over HOST-HOST primary connections. Break characters will be defined by the receiving remote host, e.g. SRI uses "." (ASCII X'2E' or 2/14) as the end-of-line character, where as UCLA uses X'OD' or 0/13 (carriage return).

USA Standard Code for Information Interchange

1. Scope

This coded character set is to be used for the general interchange of information among information processing systems, communication systems, and associated equipment.

2. Standard Code

```
|-----------------------------------------------------------------|
  B  \ b7 ----------->| 0   | 0   | 0   | 0   | 1   | 1   | 1   | 1   |
   I  \  b6 ---------->| 0   | 0   | 1   | 1   | 0   | 0   | 1   | 1   |
    T  \   b5 -------->| 0   | 1   | 0   | 1   | 0   | 1   | 0   | 1   |
     S                 |-----------------------------------------------|
                COLUMN→| 0   | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|b4 |b3 |b2 |b1 | ROW |     |     |     |     |     |     |     |     |
+---------------------+-----------------------------------------------+
| 0 | 0 | 0 | 0 | 0   | NUL | DLE | SP  | 0   | @   | P   |  `  |  p  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 0 | 0 | 1 | 1   | SOH | DC1 | !   | 1   | A   | Q   |  a  |  q  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 0 | 1 | 0 | 2   | STX | DC2 | "   | 2   | B   | R   |  b  |  r  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 0 | 1 | 1 | 3   | ETX | DC3 | #   | 3   | C   | S   |  c  |  s  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 1 | 0 | 0 | 4   | EOT | DC4 | $   | 4   | D   | T   |  d  |  t  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 1 | 0 | 1 | 5   | ENQ | NAK | %   | 5   | E   | U   |  e  |  u  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 1 | 1 | 0 | 6   | ACK | SYN | &   | 6   | F   | V   |  f  |  v  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | 1 | 1 | 1 | 7   | BEL | ETB | '   | 7   | G   | W   |  g  |  w  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 0 | 0 | 8   | BS  | CAN | (   | 8   | H   | X   |  h  |  x  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 0 | 1 | 9   | HT  | EM  | )   | 9   | I   | Y   |  i  |  y  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 1 | 0 | 10  | LF  | SUB | *   | :   | J   | Z   |  j  |  z  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 0 | 1 | 1 | 11  | VT  | ESC | +   | ;   | K   | [   |  k  |  {  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 0 | 0 | 12  | FF  | FS  | ,   | <   | L   | \   |  l  |  |  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 0 | 1 | 13  | CR  | GS  | -   | =   | M   | ]   |  m  |  }  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 1 | 0 | 14  | SO  | RS  | .   | >   | N   | ^   |  n  |  ~  |
|---|---|---|---|---|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 1 | 1 | 15  | SI  | US  | /   | ?   | O   | _   |  o  | DEL |
+---------------------+-----------------------------------------------+
```

3. Character Representation and Code Identification

The standard 7-bit character representation, with b7 the high-order bit and b1 the low-order bit, is shown below:

EXAMPLE: The bit representation for the character "K," positioned in column 4, row 11, is

```
b7 b6 b5 b4 b3 b2 b1
 1  0  0  1  0  1  1
```

The code table position for the character "K" may also be represented by the notation "column 4, row 11" or alternatively as "4/11." The decimal equivalent of the binary number formed by bits b7, b6, and b5, collectively, forms the column number, and the decimal equivalent of the binary number formed by bits b4, b3, b2, and b1, collectively, forms the row number.

The standard code may be identified by the use of the notation ASCII or USASCII.

The notation ASCII (pronounced as'-key) or USASCII (pronounced you-sas'-key) should ordinarily be taken to mean the code prescribed by the latest issue of the standard. To explicitly designate a particular (perhaps prior) issue, the last two digits of the year of issue may be appended, as, "ASCII 63" or "USASCII 63".

4. Legend

4.1 Control Characters

```
NUL Null                              DLE Data Link Escape (CC)
SOH Start of Heading (CC)             DC1 Device Control 1
STX Start of Text (CC)                DC2 Device Control 2
ETX End of Text (CC)                  DC3 Device Control 3
EOT End of Transmission (CC)          DC4 Device Control 4 (Stop)
ENQ Enquiry (CC)                      NAK Negative Acknowledge (CC)
ACK Acknowledge (CC)                  SYN Synchronous Idle (CC)
BEL Bell (audible or                  ETB End of Transmission
    attention signal)                     Block (CC)
BS Backspace (FE)                     CAN Cancel
HT Horizontal Tabulation              EM End of Medium
   (punched card skip) (FE)
LF Line Feed (FE)                     SUB Substitute
VT Vertical Tabulation (FE)           ESC Escape
FF Form Feed (FE)                     FS File Separator IS)
CR Carriage Return (FE)               GS Group Separator (IS)
SO Shift Out                          RS Record Separator (IS)
SI Shift In                           US Unit Separator (IS)
                                      DEL Delete [1]
```

```
_________
NOTE: (CC) Communication Control
      (FE) Format Effector
      (IS) Information Separator
[1] In the strict sense, DEL is not a control character.  (See 5.2)
```

4.2 Graphic Characters

```
Column/Row  Symbol     Name

2/0         SP         Space (Normally Non-Printing)
2/1         !          Exclamation Point
2/2         "          Quotation Marks (Diaeresis [2])
2/3         #          Number Sign [3,4]
2/4         $          Dollar Sign
2/5         %          Percent
2/6         &          Ampersand
2/7         '          Apostrophe (Closing Single Quotation Mark
                       Acute Accent [2])
2/8         (          Opening Parenthesis
2/9         )          Closing Parenthesis
2/10        *          Asterisk
2/11        +          Plus
2/12        ,          Comma (Cedilla [2])
2/13        -          Hyphen (Minus)
2/14        .          Period (Decimal Point)
2/15        /          Slant
3/10        :          Colon
3/11        ;          Semicolon
3/12        <          Less Than
3/13        =          Equals
3/14        >          Greater Than
3/15        ?          Question Mark
4/0         @          Commercial At [3]
5/11        [          Opening Bracket [3]
5/12        \          Reversal Slant [3]
5/13        ]          Closing Bracket [3]
5/14        ^          Circumflex [2,3]
5/15        _          Underline
6/0         `          Grave Accent [2,3] (Opening Single Quotation
                                           Mark)
7/11        {          Opening Brace [3]
7/12        |          Vertical Line [3]
7/13        }          Closing Brace [3]
7/14        ~          Overline [3] (Tilde [2]; General Accent [2])
```

2 The use of the symbols in 2/2, 2/7, 2/12, 5/14, /6/0, and 7/14 as diacritical marks is described in Appendix A, A5.2

3 These characters should not be used in international interchange without determining that there is agreement between sender and recipient. (See Appendix B4.)

4 In applications where there is no requirement for the symbol #, the symbol (Pounds Sterling) may be used in position 2/3.

5. Definitions

5.1 General

(CC) Communication Control: A functional character intended to control or facilitate transmission of information over communication networks.

(FE) Format Effector: A functional character which controls the layout or positioning of information in printing or display devices.

(IS) Information Separator: A character which is used to separate and qualify information in a logical sense. There is a group of four such characters, which are to be used in a hierarchical order.

5.2 Control Characters

NUL (Null): The all-zeros character which may serve to accomplish time fill and media fill.

SOH (Start of Heading): A communication control character used at the beginning of a sequence of characters which constitute a machine-sensible address or routing information. Such a sequence is referred to as the "heading." An STX character has the effect of terminating a heading.

STX (Start of Text): A communication control character which precedes a sequence of characters that is to be treated as an entity and entirely transmitted through to the ultimate destination. Such a sequence is referred to as "text." STX may be used to terminate a sequence of characters started by SOH.

ETX (End of Text): A communication control character used to terminate a sequence of characters started with STX and transmitted as an entity.

EOT (End of Transmission): A communication control character used to indicate the conclusion of a transmission, which may have contained one or more texts and any associated headings.

ENQ (Enquiry): A communication control character used in data communication systems as a request for a response from a remote station. It may be used as a "Who Are You" (WRU) to obtain identification, or may be used to obtain station status, or both.

ACK (Acknowledge): A communication control character transmitted by a receiver as an affirmative response to a sender.

BEL (Bell): A character for use when there is a need to call for human attention. It may control alarm or attention devices.

BS (Backspace): A format effector which controls the movement of the printing position one printing space backward on the same printing line. (Applicable also to display devices.)

HT (Horizontal Tabulation): A format effector which controls the movement of the printing position to the next in a series of predetermined positions along the printing line. (Applicable also to display devices and the skip function on punched cards.)

LF (Line Feed): A format effector which controls the movement of the printing position to the next printing line. (Applicable also to display devices.) Where appropriate, this character may have the meaning "New Line" (NL), a format effector which controls the movement of the printing point to the first printing position on the next printing line. Use of this convention requires agreement between sender and recipient of data.

VT (Vertical Tabulation): A format effector which controls the movement of the printing position to the next in a series of predetermined printing lines. (Applicable also to display devices.)

FF (Form Feed): A format effector which controls the movement of the printing position to the first pre-determined printing line on the next form or page. (Applicable also to display devices.)

CR (Carriage Return): A format effector which controls the movement of the printing position to the first printing position on the same printing line. (Applicable also to display devices.)

SO (Shift Out): A control character indicating that the code combinations which follow shall be interpreted as outside of the character set of the standard code table until a Shift In character is reached.

SI (Shift In): A control character indicating that the code combinations which follow shall be interpreted according to the standard code table.

DLE (Data Link Escape): A communication control character which will change the meaning of a limited number of contiguously following characters. It is used exclusively to provide supplementary controls in data communication networks.

DC1, DC2, DC3, DC4 (Device Controls): Characters for the control of ancillary devices associated with data processing or telecommunication systems, more especially switching devices "on" or "off." (If a single "stop" control is required to interrupt or turn off ancillary devices, DC4 is the preferred assignment.)

NAK (Negative Acknowledge): A communication control character transmitted by a receiver as a negative response to the sender.

SYN (Synchronous Idle): A communication control character used by a synchronous transmission system in the absence of any other character to provide a signal from which synchronism may be achieved or retained.

ETB (End of Transmission Block): A communication control character used to indicate the end of a block of data for communication purposes. ETB is used for blocking data where the block structure is not necessarily related to the processing format.

CAN (Cancel): A control character used to indicate that the data with which it is sent is in error or is to be disregarded.

EM (End of Medium): A control character associated with the sent data which may be used to identify the physical end of the medium, or the end of the used, or wanted, portion of information recorded on a medium. (The position of this character does not necessarily correspond to the physical end of the medium.)

SUB (Substitute): A character that may be substituted for a character which is determined to be invalid or in error.

ESC (Escape): A control character intended to provide code extension (supplementary characters) in general information interchange. The Escape character itself is a prefix affecting the interpretation of a limited number of contiguously following characters.

FS (File Separator), GS (Group Separator), RS (Record Separator), and US (Unit Separator): These information separators may be used within data in optional fashion, except that their hierarchical relationship shall be: FS is the most inclusive, then GS, then RS, and US is least inclusive. (The content and length of a File, Group, Record, or Unit are not specified.)

DEL (Delete): This character is used primarily to "erase" or "obliterate" erroneous or unwanted characters in perforated tape. (In the strict sense, DEL is not a control character.)

5.3 Graphic Characters

SP (Space): A normally non-printing graphic character used to separate words. It is also a format effector which controls the movement of the printing position, one printing position forward. (Applicable also to display devices.)

6. General Considerations

6.1 This standard does not define the means by which the coded set is to be recorded in any physical medium, nor does it include any redundancy or define techniques for error control. Further, this standard does not define data communication character structure, data communication formats, code extension techniques, or graphic representation of control characters.

6.2 Deviations from the standard may create serious difficulties in information interchange and should be used only with full cognizance of the parties involved.

6.3 The relative sequence of any two characters, when used as a basis for collation, is defined by their binary values.

6.4 No specific meaning is prescribed for any of the graphics in the code table except that which is understood by the users. Furthermore, this standard does not specify a type style for the printing or display of the various graphic characters. In specific applications, it may be desirable to employ distinctive styling of individual graphics to facilitate their use for specific purposes as, for example, to stylize the graphics in code positions 2/1 and 5/15 into those frequently associated with logical OR (|) and logical NOT (252), respectively.

6.5 The appendixes to this standard contain additional information on the design and use of this code.

This RFC was put into machine readable form for entry into the online RFC archives by Robbie Bennet 9/99 doi:10.17487/RFC0020