



THE INTERNATIONAL REFERENCE
ALPHABET

William Stallings

Copyright 2003 William Stallings

A familiar example of data is **text** or character strings. While textual data are most convenient for human beings, they cannot, in character form, be easily stored or transmitted by data processing and communications systems. Such systems are designed for binary data. Thus a number of codes have been devised by which characters are represented by a sequence of bits. Perhaps the earliest common example of this is the Morse code. Today, the most commonly used text code is the International Reference Alphabet (IRA).¹ Each character in this code is represented by a unique 7-bit binary code; thus, 128 different characters can be represented. Table 1 lists all of the code values. In the table, the bits of each character are labeled from b_7 , which is the most significant bit, to b_1 , the least significant bit. Characters are of two types: printable and control (Table 2). Printable characters are the alphabetic, numeric, and special characters that can be printed on paper or displayed on a screen. For example, the bit representation of the character "K" is $b_7b_6b_5b_4b_3b_2b_1 = 1001011$. Some of the control characters have to do with controlling the printing or displaying of characters; an example is carriage return. Other control characters are concerned with communications procedures.

IRA-encoded characters are almost always stored and transmitted using 8 bits per character. The eighth bit is a parity bit used for error detection. The parity bit is the most significant bit and is therefore labeled b_8 . This bit is set such that the total number of binary 1s in each octet is always odd (odd parity) or always even (even parity). Thus a transmission error that changes a single bit, or any odd number of bits, can be detected.

¹ IRA is defined in ITU-T Recommendation T.50 and was formerly known as International Alphabet Number 5 (IA5). The U.S. national version of IRA is referred to as the American Standard Code for Information Interchange (ASCII).

Table 1 The International Reference Alphabet (IRA)

bit position

b ₇	0	0	0	0	1	1	1	1
b ₆	0	0	1	1	0	0	1	1
b ₅	0	1	0	1	0	1	0	1

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

Table 2 IRA Control Characters (page 1 of 2)

Format Control

BS (Backspace): Indicates movement of the printing mechanism or display cursor backward one position.

HT (Horizontal Tab): Indicates movement of the printing mechanism or display cursor forward to the next preassigned 'tab' or stopping position.

LF (Line Feed): Indicates movement of the printing mechanism or display cursor to the start of the next line.

VT (Vertical Tab): Indicates movement of the printing mechanism or display cursor to the next of a series or preassigned printing lines.

FF (Form Feed): Indicates movement of the printing mechanism or display cursor to the starting position of the next page, form, or screen.

CR (Carriage Return): Indicates movement of the printing mechanism or display cursor to the starting position of the same line.

Transmission Control

SOH (Start of Heading): Used to indicate the start of a heading, which may contain address or routing information.

STX (Start of Text): Used to indicate the start of the text and so also indicates the end of the heading.

ETX (End of Text): Used to terminate the text that was started with STX.

EOT (End of Transmission): Indicates the end of a transmission, which may have included one or more 'texts' with their headings.

ENQ (Enquiry): A request for a response from a remote station. It may be used as a 'WHO ARE YOU' request for a station to identify itself.

ACK (Acknowledge): A character transmitted by a receiving device as an affirmation response to a sender. It is used as a positive response to polling messages.

NAK (Negative Acknowledgment): A character transmitted by a receiving device as a negative response to a sender. It is used as a negative response to polling messages.

SYN (Synchronous/Idle): Used by a synchronous transmission system to achieve synchronization. When no data is being sent a synchronous transmission system may send SYN characters continuously.

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

Table 2 IRA Control Characters (page 2 of 2)

Information Separator

FS (File Separator) GS (Group Separator) RS (Record Separator) US (Unit Separator)	Information separators to be used in an optional manner except that their hierarchy shall be FS (the most inclusive) to US (the least inclusive)
---	--

Miscellaneous

NUL (Null): No character. Used for filling in time or filling space on tape when there are no data.	DLE (Data Link Escape): A character that shall change the meaning of one or more contiguously following characters. It can provide supplementary controls, or permits the sending of data characters having any bit combination.
BEL (Bell): Used when there is need to call human attention. It may control alarm or attention devices.	DC1, DC2, DC3, DC4 (Device Controls): Characters for the control of ancillary devices or special terminal features.
SO (Shift Out): Indicates that the code combinations that follow shall be interpreted as outside of the standard character set until a SI character is reached.	CAN (Cancel): Indicates that the data that precedes it in a message or block should be disregarded (usually because an error has been detected).
SI (Shift In): Indicates that the code combinations that follow shall be interpreted according to the standard character set.	EM (End of Medium): Indicates the physical end of a tape or other medium, or the end of the required or used portion of the medium.
DEL (Delete): Used to obliterate unwanted characters; for example by overwriting.	SUB (Substitute): Substituted for a character that is found to be erroneous or invalid.
SP (Space): A nonprinting character used to separate words, or to move the printing mechanism or display cursor forward by one position.	ESC (Escape): A character intended to provide code extension in that it gives a specified number of continuously following characters an alternate meaning.
