RFC 9226
Bioctal: Hexadecimal 2.0

Abstract

The prevailing hexadecimal system was chosen for congruence with groups of four binary digits, but its design exhibits an indifference to cognitive factors. An alternative is introduced that is designed to reduce brain cycles in cases where a hexadecimal number should be readily convertible to binary by a human being.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

1.1. The Pernicious Advance of Hexadecimal

Octal has long been used to represent groups of three binary digits as single characters, and that system has the considerable merit of not requiring any digits other than those already familiar from decimal numbers. Unfortunately, the increasing use of 16-bit machines and other machines that have word lengths that are evenly divisible by four (but not by three) has led to the widespread adoption of hexadecimal. Table 1 presents the digits of the hexadecimal alphabet.

Value	Digit
0	0
1	1
2	2
3	3
4	4

Value	Digit
5	5
6	6
7	7
8	8
9	9
10	A
11	B
12	C
13	D
14	E
15	F

Table 1: The Hexadecimal Alphabet

The choice of alphabet is clearly arbitrary: On the exhaustion of the decimal digits, the first letters of the Latin alphabet are used in sequence for the remaining hexadecimal digits. An arbitrary alphabet may be acceptable on an interim or experimental basis. However, given the diminishing likelihood of a return to 18-bit computing, a review of this choice of alphabet is merited before its use, like that of the QWERTY keyboard, becomes too deeply established to permit the easy adoption of a more logical alternative.

1.2. Problems with Hexadecimal

One problem with the hexadecimal alphabet is well known: It contains two vowels, and numbers expressed in hexadecimal have been found to collide with words offensive to vegetarians and other groups.

Imposing a greater constraint on the solution space, however, is the difficulty of mentally converting a number expressed in hexadecimal to (or from) binary. Consider the hexadecimal digit 'D', for example. First, one must remember that 'D' represents a value of 13 -- and, while it may be easy to recall that 'F' is 15 with all bits set, for digits in the middle of the non-decimal range, such as 'C' and 'D', one may resort to counting ("A is ten, B is eleven, ..."). Next, one must subtract eight from that number to arrive at a number that is in the octal range. Thus, the benefit of representing one additional bit incurs the cost of two additional mental operations before one arrives at the position where the task that remains reduces to the difficulty of converting the remaining three digits to binary.

These mental steps are not difficult per se, since a child could do them, but if it is possible to avoid employing children, then it should be avoided. An appeal to the authority of cognitive psychology is perhaps also due here, in particular to the "seven plus or minus two" principle [Miller] -- either because octal is within the upper end of that range (nine) and hexadecimal is not, or else because the difference in the size of the alphabets is greater than the lower end of that range (five). Either way, it is almost certainly relevant.

1.3. Other Proposals

Various alternatives have already been suggested. Some of these are equally arbitrary, e.g., in selecting the last six letters of the Latin alphabet rather than the first six letters.

The scheme that comes closest to solving the main problem to date is described by Bruce A. Martin [Martin] who proposes new characters for the entire octal alphabet. While his principal motivation is to distinguish hexadecimal numbers from decimals, the design of each character uses horizontal lines to directly represent the "ones" of the corresponding binary number, making mental translation to binary a trivial task.

Unfortunately for this and other proposals involving new symbols, proposals to change the US-ASCII character set [USASCII] might no longer be accepted. Also, it seems unrealistic to expect keyboards or printer type elements (whether of the golf ball or daisy wheel kind) to be replaced to accommodate new character designs.

2. Bioctal

Table 2 presents the hexadecimal alphabet once again, this time in a sequence of two octaves with values increasing left to right and top to bottom.

0	1	2	3	4	5	6	7
8	9	A	B	C	D	E	F

Table 2: The Hexadecimal Alphabet in Sequential Octaves

Arranged thus, the binary representation of each digit in the second octave is the same as the digit above it, but with the most significant of the four bits set to '1' instead of '0'.

The incongruity of two decimal digits in the second octave also suggests that, in blindly aligning with four bits, hexadecimal (six plus ten, neither of which are powers of two) misses an opportunity to align also with three bits.

Bioctal restores congruence by replacing the second row with characters mnemonically related to the corresponding character in the first octave.
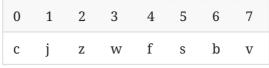
Table 3 shows the compelling result.

0	1	2	3	4	5	6	7
c	j	z	w	f	s	b	v

Table 3: Bioctal in Sequential Octaves

The mnemonic basis is the shape of the lowercase character. This is seen directly for '2', '5', and '6'. For '3', '4', and '7', the corresponding letters are the result of a quarter-turn clockwise (assuming an "open" '4'). The choice of 'c' and 'j' for '0' and '1' avoids vowels and lowercase 'L', the latter being confusable with '1' in some fonts.

With this choice of letters, it is immediately evident that both problems with hexadecimal are solved. Mental conversion is now straightforward: if the digit is a letter, then the most significant of the four binary bits is '1', and the remaining three bits are the same as for the Arabic numeral with the same shape in the first octave.

3. Objections to Be Dismissed

Several objections can be anticipated, the first of which concerns the name. The term "bioctal" is already used to refer to the combination of two octal characters into a single field on, for example, paper tape (e.g., [UNIVAC]). However, if the word "bioctal" must be disadvantaged relative to words such as "biannual" in the number of meanings it is allowed to have, then it is the paper tapers who must give way: in that context, the "octal" part of "bioctal" refers to the number of distinct values that three bits can have, while the "bi" refers to a doubling of the number of bits, not values. A meaning depending on such a discordant etymology does not deserve to endure.

Second, it may be argued that the use of hexadecimal has already become too entrenched to be changed in the short term: Bioctal should be introduced only after those working in the industry who have grown accustomed to hexadecimal have retired. Such a dilatory contention cannot be allowed to impede the march of progress. Instead, any data entry technician who claims to have difficulty with bioctal may be reassigned to duties involving only binary numbers.

A third possible objection is that numbers in bioctal do not sort numerically. However, this assumes a sort based on the US-ASCII order of symbols; it is quite possible that bioctal numbers sort naturally in some lesser known variety of EBCDIC. Further, resistance to numeric sorting may be an indicator of virtue, being suggestive of an alphabet with a certain strength of character.

One difficulty remains: Not all computers support lowercase letters. While this is indeed true, it should be confirmed in any particular instance: the author has observed that in many cases a machine having a keyboard with buttons marked only with uppercase letters also supports lowercase letters. In any case, it is permissible to use uppercase letters instead of the lowercase ones of Table 3; the morphology mnemonic continues to work for most bioctal digits in uppercase, although an extra mental cycle is required for 'B'.

4. Security Considerations

The letters 'b' and 'f' appear in both the bioctal and hexadecimal alphabets, which makes potential misinterpretation a concern. A case of particular hazard arises where two embedded systems engineers work to develop a miniature lizard detector designed to be worn like a wristwatch. One engineer works on the lizard proximity sensor and the other on a minimal two-character display. The interface between the circuits is 14 bits. To make things easier, the engineer working on the display arranges for these bits to be set in a pattern that allows them to be used directly as two seven-bit US-ASCII characters indicating the most significant lacertilian species detected in the vicinity of the device. Due to the use of an old US-ASCII table (i.e., one in hex, not bioctal) and human error, some of the values specified as outputs for the detection subsystem are in hexadecimal, not the bioctal the engineer developing that subsystem expects -- including, in the case of one type of lizard, "4b 4f". The result is that the detector displays "NL" (No Lizards) when it should display "KO" (Komodo dragon). This may be considered prejudicial to the security of the user of the device.

Extensive research has uncovered no other security-related scenarios to date.

5. IANA Considerations

This document has no IANA actions.

6. Conclusion

Bioctal is a significant advance over hexadecimal technology and promises to reduce the small (but assuredly non-zero) contribution to anthropogenic global warming of mental hex-to-binary conversions. Since the mnemonic basis of the alphabet is independent of English or any other particular natural language, there is no reason that it should not be adopted immediately around the world, excepting perhaps certain islands of Indonesia to which *Varanus komodoensis* is native.

7. Informative References

[Martin]   Martin, B. A., "Letters to the editor: On binary notation", Communications of the ACM, Vol. 11, No. 10, DOI 10.1145/364096.364107, October 1968, <https://doi.org/10.1145/364096.364107>.

[Miller]   Miller, G. A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review, Vol. 101, No. 2, 1956.

[UNIVAC]   Sperry Rand Corporation, "Programmers Reference Manual for UNIVAC 1218 Computer", Revision C, Update 2, November 1969, <http://bitsavers.computerhistory.org/pdf/univac/military/1218/PX2910_Univac1218PgmrRef_Nov69.pdf>.

[USASCII]   American National Standards Institute, "Coded Character Set -- 7-bit American Standard Code for Information Interchange", ANSI X3.4, 1986.

Acknowledgments

Author's Address

Michael Breen
mbreen.com
Email: rfc@mbreen.com