February 21, 2005

Title: UNICODE UTF-16LE for String Descriptors

Applies to: Universal Serial Bus Specification, Revision 2.0

Summary of ECN

This ECN more completely specifies the encoding scheme for string descriptors. String descriptors must use UNICODE UTF-16LE encodings.

Reasons for ECN

To ensure consistent coding and decoding of string descriptors, the spec has to be clear on the exact encoding to use. The original spec said to use UNICODE encodings, but there are multiple encoding schemes that make up UNICODE. UTF-16LE was the intended encoding scheme, and the one assumed by most developers.

Impact on Existing Peripherals and Systems

USB compliance testing checks for proper encoding of string descriptors on peripherals. There are no known cases of devices using an encoding other than UTF-16LE. There is at least one OS that has been using UTF-8 encoding and ignoring null bytes. Since most devices have been encoding ASCII characters resulting in the upper byte being NULL, this has not been a problem.

Hardware Implications

None expected.

Software Implications

Any software not using UTF-16LE will need to change.

Compliance Testing Implications

None. USB compliance testing already checks string descriptors for proper UTF-16LE encoding.

Specification Changes

Section 9.6.7 String, p. 273
Change This:

…
String descriptors use UNICODE encodings as defined by *The Unicode Standard, Worldwide Character Encoding, Version 3.0*, The Unicode Consortium, Addison-Wesley Publishing Company, Reading, Massachusetts (URL: http://www.unicode.com).

…
To this:

…
String descriptors use UNICODE UTF-16LE encodings as defined by *The Unicode Standard, Worldwide Character Encoding, Version 4.0*, The Unicode Consortium, Addison-Wesley Publishing Company, Reading, Massachusetts (URL: http://www.unicode.org).

…

Table 9-16, p. 274
Change Description of bString from this:

UNICODE encoded string

To this:

UNICODE UTF-16LE encoded string