Base62x: An Alternative Approach to Base64 for non-Alphanumeric Characters

Dr. Lu Liu, **ICNC/FSKD'11**

July 2011, Shanghai, China

Contents

- The problems
- Some remedies
- This approach: Base62x definition
- Benefits of Base62x

The Problems of Base64

- Length of encoded text gets longer than that of original
  - 3 bytes * 8 bits = 24 bits
  - 4 bytes * 6 bits = 24 bits
- Using symbols to make up the shortage of letters(a-z, A-Z, 0-9), which only have 62.
  - ‘+’ stands for 62
  - ‘/’ stands for 63
  - ‘=’ serves as a pad

The Problems of Base64 -2

- Symbols in Base64 raise problems, e.g.
  - Prevent Base64 to be used in arithmetic due to '+' is defined as 'plus/add' in arithmetic
  - Conflict with the scheme URL, where '+' switches itself with a space or '%20'. '/' is reserved
  - Unsafe used in file systems, where '/' is denoted a layer of directories
  - '=', like '+' hinders Base64 in arithmetic as we know in general that '=' means 'equal to'
  - Can not be used in some other symbol-sensitive environments

Some remedies of Original Base64

- Modified Base64 for filenames (non standard)
  - ‘+’ for 62, ‘/’ is replaced by ‘-’ for 63, hereinafter
- Modified Base64 for URL applications ('base64url' encoding)
  - ‘-’, ‘_’
- Modified Base64 for XML name tokens (Nmtoken)
  - ‘.’, ‘-’
- Modified Base64 for XML identifiers (Name)
  - ‘_’, ‘:’
- Modified Base64 for Program identifiers
  - ‘_’, ‘-’
- Modified Base64 for Regular expressions
  - ‘!’, ‘-’

Base62x: Definition

- Base62x is an improved implementation of Base64 algorithm.
  - Not use any symbols in its representation.
  - Only the alphabetical (a-z, A-Z) and numeric characters (0-9) are used in this scheme.
- Symbols “+”, “/” and “=” are discarded.
  - Alphabet “x” (or any other one amongst the group of 0-9, a-z and A-Z) is a special tag
  - x1 represents number 61, x2 for 62 and x3 for 63.

Base62x: Code index comparisons

Base62x				***Base64***			
Value	***Enc***	***Value***	***Enc***	***Value***	***Enc***	***Value***	***Enc***
0	0	.		0	A	.	
1	1	.		1	B	.	
2	2	.		2	C	.	
3	3	60	z	3	D	60	8
4	4	61	x1	4	E	61	9
.		62	x2	.		62	+
.		63	x3	.		63	/
.		(tag)	x	.		(pad)	=

Base62x: Encoding Algorithm

Base62x Encoding

```
for each byte in input
        newbyte = bits operations on byte
        if newbyte > 60 then
                newbyte = 'x' + (newbyte – 60)
        else
                newbyte
        end if
end loop
output
```

Base62x: Examples

No.	*Original text*	*Encoded text*
1	A	GG
2	Ab	GM8
3	aBC	OK93
4	A__B*	GLx1VGYe
5	COMPSAC 2011	Gqx1DK5D1Go0oC34n
6	中文简体	vBYjvfQ7vww0vBsJ
7	メインページ	uuEXuuAauuEpuuEQuuEzuuAu

Base62x: comparison of length between Base62x and Base64

Original text	*Base62x encoding*	*length*	*Base64 encoding*	*length*
aBC	OK93	4	YUJD	4
A__B*	GLx1VGYe	8	QV9fQio=	8
all subsets of A	OMni87DrOdDbT7CWRsOWGG	22	YWxsIHN1YnNldHMgb2YgQQ==	24
Unix - Live Free or Die.	LMvfU20j84nfTcKWHd9bPI1lSY14QMKk	32	VW5peCAtIExpdmUgRnJlZSBvciBEaWUu	32

Base62x: Potential Applications

- All the remedies in the previous slide 5 can be realized and standardized in Base62x
  - in file systems, in URL applications, in XML, in program identifiers, in regular expressions …
- Numeric Arithmetic
  - B + + = / → 1 + x2 = x3

    Base64 → *Base62x*
- Some other symbol-sensitive environments, e.g. IPv6 address notation.

Summary

- What we have exploded:
  - Problems current Base64
  - Some remedies to these problems
  - Our approach: Base62x
    - Definition
    - Algorithm
    - Examples
    - Comparisons
    - Implementations

Thank you!

- Base62x online
  - http://ufqi.com/dev/base62x
- Questions?