base64 인코딩 & 디코딩 원리

By **Bbolmin**, bbolmin.tistory.com
6월 13일, 2012
원본 보기

Base64 인코딩은 64개의 문자를 이용하여 바이너리 데이터를 아스키 텍스트 데이터로 표현하기 위해 사용됩니다.

base64는 8bit의 데이터(바이너리)를 6bit의 크기로 표현합니다. 따라서 24bit를 단위로 하여 3개의 문자에서 4개의 문자를 얻게 되는 것입니다.

Byte character	a (97)								b (98)								c (99)							
8 bit value	0	1	1	0	0	0	0	1	0	1	1	0	0	0	1	0	0	1	1	0	0	0	1	1
6 bit value	0	1	1	0	0	0	0	1	0	1	1	0	0	0	1	0	0	1	1	0	0	0	1	1
6 bit character	Y (24)						W (22)						J (9)				j (35)							

위와 같이 abc를 base64 인코딩 하여 YWJj를 얻을 수 있게 됩니다. 여기서 6bit의 이진수는 아래의 base64 table을 이용하여 문자로 바꿔줍니다.

Value	Char	Value	Char	Value	Char	Value	Char
0	A	16	Q	32	g	48	w
1	B	17	R	33	h	49	x
2	C	18	S	34	i	50	y
3	D	19	T	35	j	51	z
4	E	20	U	36	k	52	0
5	F	21	V	37	l	53	1
6	G	22	W	38	m	54	2
7	H	23	X	39	n	55	3
8	I	24	Y	40	o	56	4
9	J	25	Z	41	p	57	5
10	K	26	a	42	q	58	6
11	L	27	b	43	r	59	7
12	M	28	c	44	s	60	8
13	N	29	d	45	t	61	9
14	O	30	e	46	u	62	
15	P	31	f	47	v	63	/

그리고 base64 인코딩 24bit 단위인데 인코딩할 문자가 3개(24bit) 단위가 아닐 때는 어떻게 되는지 알아보겠습니다.

Byte character	a (97)																							
8 bit value	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 bit value	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 bit character	Y (24)						Q (16)						=				=							

위의 결과를 보면 a라는 문자 하나를 넣었을 때는 YQ==으로 base64 table에 없는 '=' 문자가 추가된 것을 알 수 있습니다. '='은 bit수를 맞춰주기 위해 0으로 채워주는 패딩이라는 것 입니다.

그럼 문자의 개수가 3n+1개 일 때는 '='이 2개가 될 것이고, 3n+2개 일 때는 '='이 1개가 되는 것을 생각해 볼 수 있습니다.

※ secuinside2012 IU 문제를 보고 base64의 decoding시 충돌이 있는 것을 알게 되었는데요 -

예를 들어 위에서 'a'를 base64인코딩 하여 YQ==를 얻었습니다. 따라서 YQ==를 디코딩하면 'a'를 얻게 됩니다.

그런데 YR==, YS==, YT== 등 을 디코딩 해보면 모두 'a'가 나오는 것을 볼 수 있습니다.

아래 그림을 보면서 생각해보면 왜 그런지 알수 있습니다.

Byte character	a (97)																							
8 bit value	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 bit value	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 bit character	Y (24)						Q (16)				=				=									

여기서 '='이 2개 이므로 디코딩시 문자가 3n+1개 라는 것을 알 수 있습니다. 그러면 Q(16)을 디코딩시 010000에서 패딩으로 채워진 0000의 4bit는 0이든 1이든 관계가 없다는 것 입니다.

따라서 3n+1개의 문자일 때는 충돌 가지수가 2의 4승인 16개가 생길 것이고, 3n+2개의 문자일 때는 2의 2승인 4개가 생기고, 3n개의 문자일 때는 충돌이 생기지 않게 됩니다.

▶ base64 인코딩, 디코딩 사이트 :http://ostermiller.org/calc/encode.html