Topics of Discussion

- Encoding and encryption
- What is base64 used for?
- Usage of base64 encoding
- The algorithm
- Drawbacks
- Alternatives

Encoding and Encryption

- People often refer to base64 (and similar methods) as a form of encryption.
- Encoding is the process of transforming information from one format into another.
  - Transfer
  - Storage
  - Formating (eg. LaTeX – semantics encoding)
- Encryption is a type of encoding that obscures information to make it unreadable without special knowledge.

What is base64?

- Positional notation – a quadrosexagesimal number system
- Numerals 0–9, alphabetical characters a–z and A–Z plus two special characters (all printable ASCII)
- VGhlIHF1aWNrIGJyb3duIGZveCByYW4gb3ZlciB0aGUgbGF6eSBkb2c=

Value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Base64	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

Value	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Base64	Q	R	S	T	U	V	W	X	Y	Z	a	b	c	d	e	f

Value	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
Base64	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v

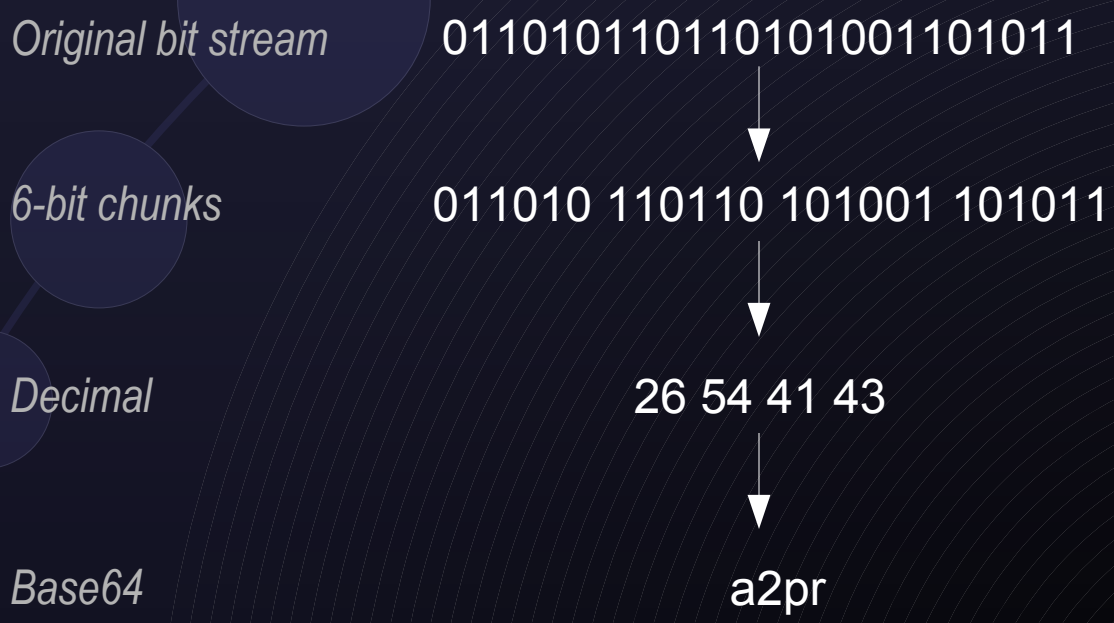
Value	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
Base64	w	x	y	z	0	1	2	3	4	5	6	7	8	9	+	/

What is base64 used for?

- “Binary to text” encoding.
- Primarily used in SMTP – transfer protocol allows only 7-bit ASCII characters to be used.
- Base64 will encode any bit stream as a sequence of 7-bit ASCII characters (ie. Binary data to plaintext).
- Multipurpose Internet Mail Extensions
- UTF-7
- Anti-spam evasion
- Binary data embedded in XML

The Encoding Algorithm

- Binary data is split into groups of 24 bits (3 bytes), then split into 6-bit chunks and converted to the corresponding ASCII characters.



The Encoding Algorithm

- What if we don't have a multiple of 3 bytes in our bit stream?

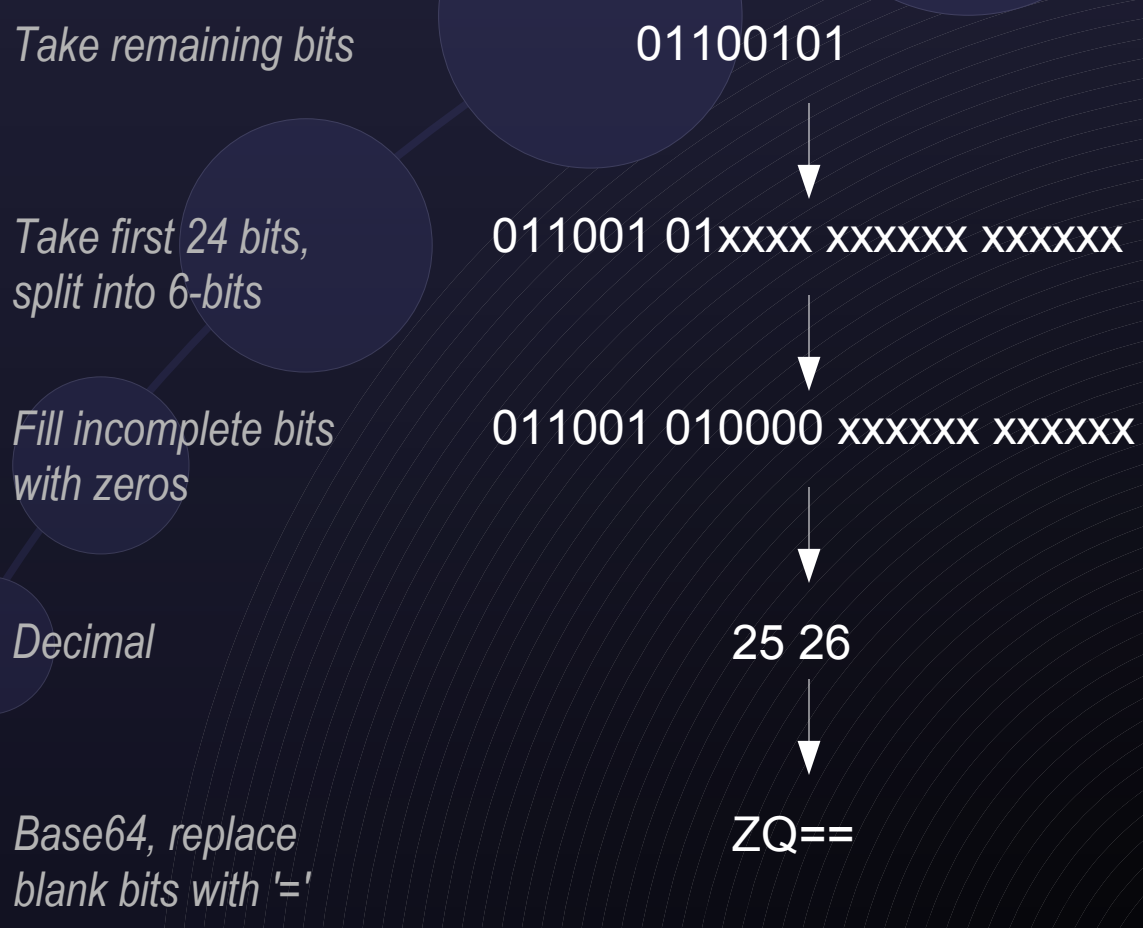
Original bit stream 011000100110110001110101011001 01

Take first 24 bits, split into 6-bits 011000 100110 110001 110101

Decimal 24 38 49 53

Base64 Ymx1

The Encoding Algorithm



Other Notes

- 33% increase in size
- ASCII85, Quoted-printable, 8BITMIME
- Privacy-Enhanced Electronic Mail was the first place where base64 was used
- Headers usually describe if a message is encoded in base64, it is up to the client to translate