

He Keke, He Kēkē Ranei?
Is it a Cake or an Armpit?



Tim Valentine
Founder, Nē.nz

He Keke, He Kēkē Ranei?

- What this talk is/isn't...
 - ✗ An expert analysis of IDNs
 - ✓ A nerds ideas on increasing the use of IDNs

What are IDNs?

- Allows use of non-ASCII characters in domain names

Bits b₇ b₆ b₅ b₄ b₃ b₂ b₁ Column Row					0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b ₄	b ₃	b ₂	b ₁	Row	0	1	2	3	4	5	6	7
					0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	/	7	G	W	g	w
1	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	1	9	HT	EM	)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[	k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	
1	1	0	1	13	CR	GS	—	=	M	]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

What are Internationalised Domain Names (IDNs)?

- Enabled through IDNA2003 by ICANN in 2003
 - RFC 3490: Internationalizing Domain Names in Applications (IDNA)
 - RFC 3492: Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)
- Further refined in IDNA2008 in 2010
 - RFC5890, RFC5891, RFC5892, RFC5893, RFC4894
- Uses “punycode” to encode non-ASCII unicode characters



Punycode

U-Lable: māori.nz

A-Lable: xn--mori-qa.nz

- xn-- → ACE Prefix / “Prefix”
- mori → Basic String / “Scaffold”
- - → Delimiter
- qa → Extended String / “Instructions”

Nē

Extended String / “Instructions”

The non-ASCII characters are sorted by Unicode value, lowest first (if a character occurs more than once, they are sorted by position). Each is then encoded as a single number. This single number defines both the location to insert the character at and which character to insert.

- An index into the result to insert the code at, starting at 0 (for insertion at the start).
- The number of insertionPoints (current length of the result plus one).
- The reducedCodepoint is the Unicode code point to insert minus 127.

The encoded number is $\text{insertionPoints} \times \text{reducedCodepoint} + \text{index}$. By dividing by insertionPoints and also getting the remainder, a decoder can determine reducedCodepoint and index.

The number is encoded using the letters a through z and the digits 0 through 9. It is not base-36 but a more complex scheme, generalized variable-length integers, which allows the numbers to be concatenated with nothing separating them.

Punycode (2025) Wikipedia. Available at: <https://en.wikipedia.org/wiki/Punycode> (Accessed: 24 November 2025).



Scripts & Label Generation Rules

- Scripts are defined by the Unicode Character Database
 - Latin, Greek, Armenian, Hebrew, Arabic, etc.
 - Te Reo Māori is Latin script
- Label Generation Rules (LGRs)
 - Define rules for TLDs and can be defined for second level domains
 - Root Zone Label Generation Rules Version 6 (RZ-LGR-6) defines rules for TLDs
 - .nz has an “IDN Table” which is defined as “nz-latn-1.0”
 - It allows a-z, 0-9, ā, ē, ī, ō, ū and hyphens (-)

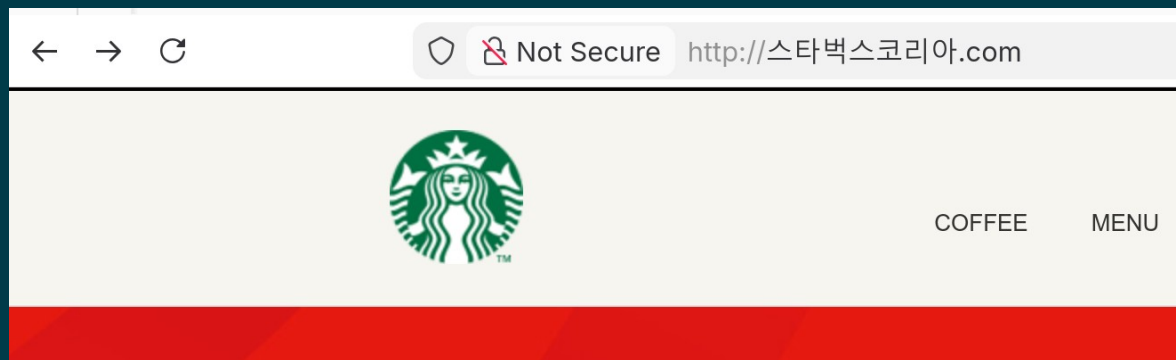


Real World Examples

스타벅스코리아 .com

xn--oy2b35ckwhba574atvuzkc.com

Starbucks Korea



Real World Examples

президент.рф

xn--d1abbgf6aiiy.xn--p1ai

The Kremlin (redirect to kremlin.ru)



Security Implications

- Homograph Attacks
 - Exploits the fact that there are many unicode characters that look like ASCII characters
- Basic String Collision
 - When changing a single non-ASCII character in an IDN the scaffold remains but the instructions change, i.e...
 - māori.nz → xn--mori-qa.nz
 - mēori.nz → xn--mori-bva.nz

Homograph Attacks

- In 2017 Xudong Zheng registered...

apple.com

- Instead of...

apple.com



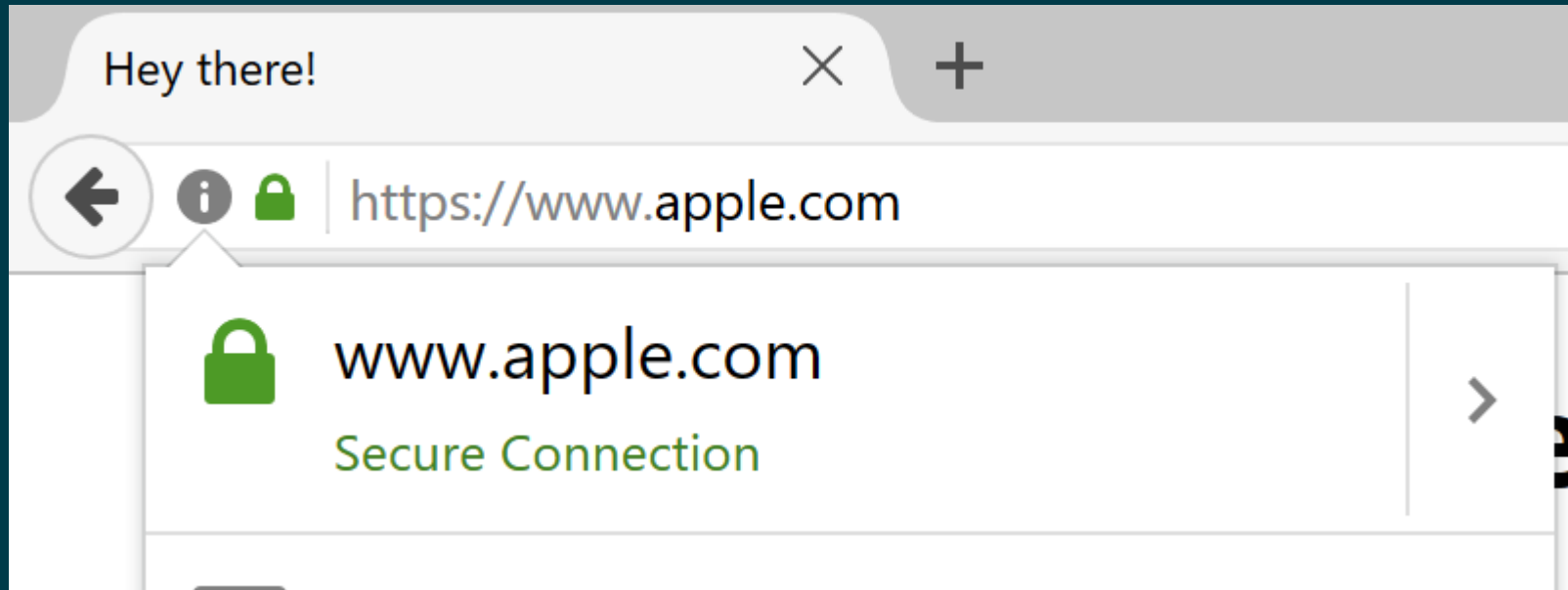
Homograph Attacks

- It worked as it only contained characters from one script (cyrillic)

apple.com

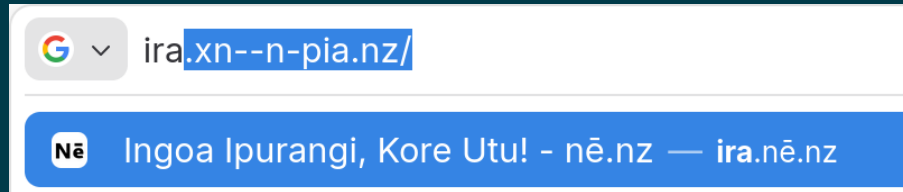
xn--pple-43d.com

Homograph Attacks

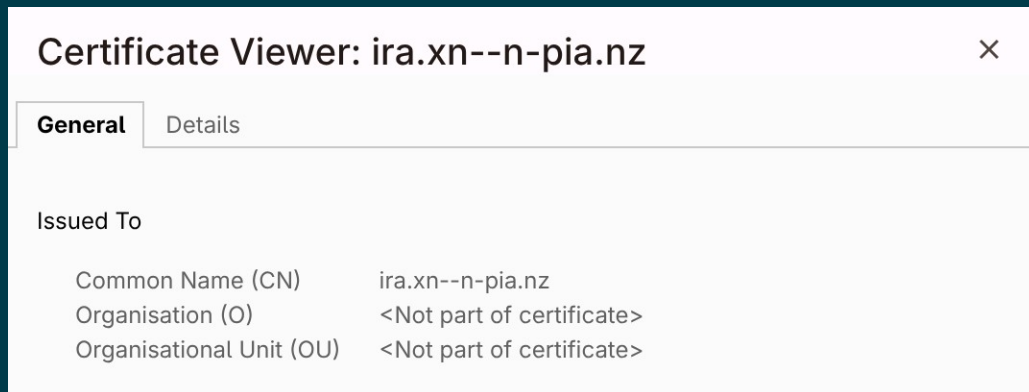


Software Issues

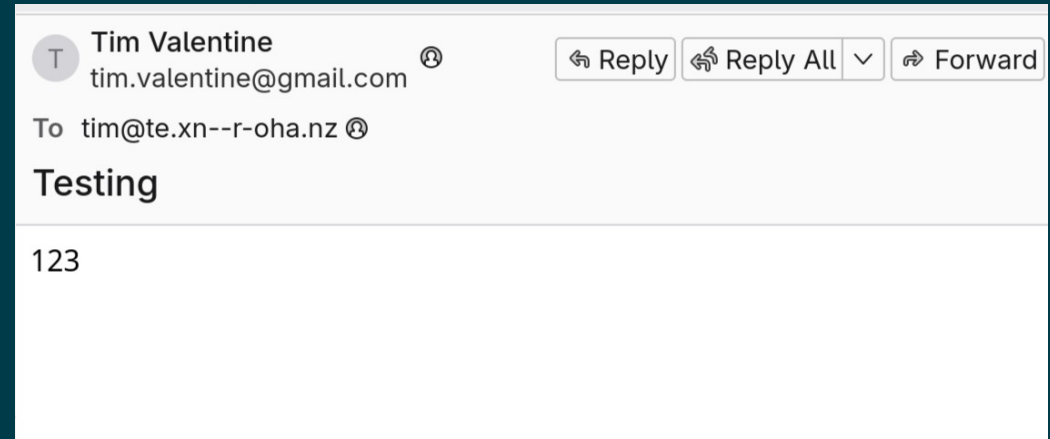
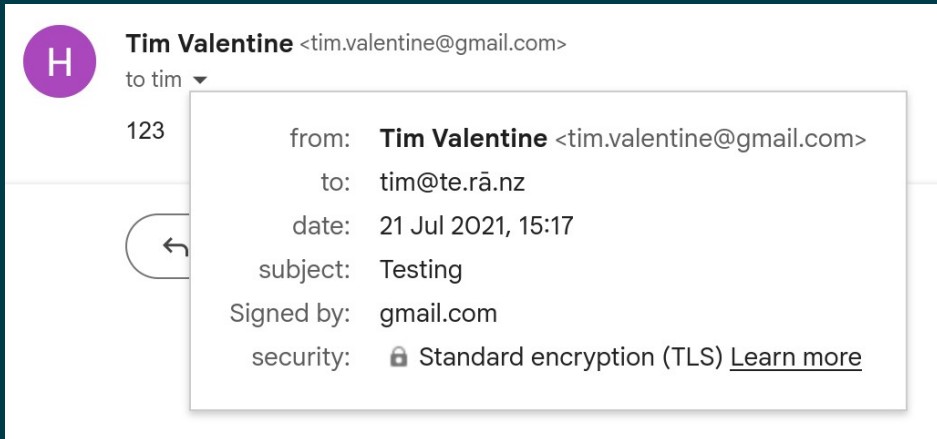
- Firefox shows punycode and will not show the U-Label in autocomplete...



- Chrome displays U-Label in address bar but A-Label on TLS cert...



Software Issues



Linguistic Necessity

- Getting language correct is important to avoid confusion and embarrassment....



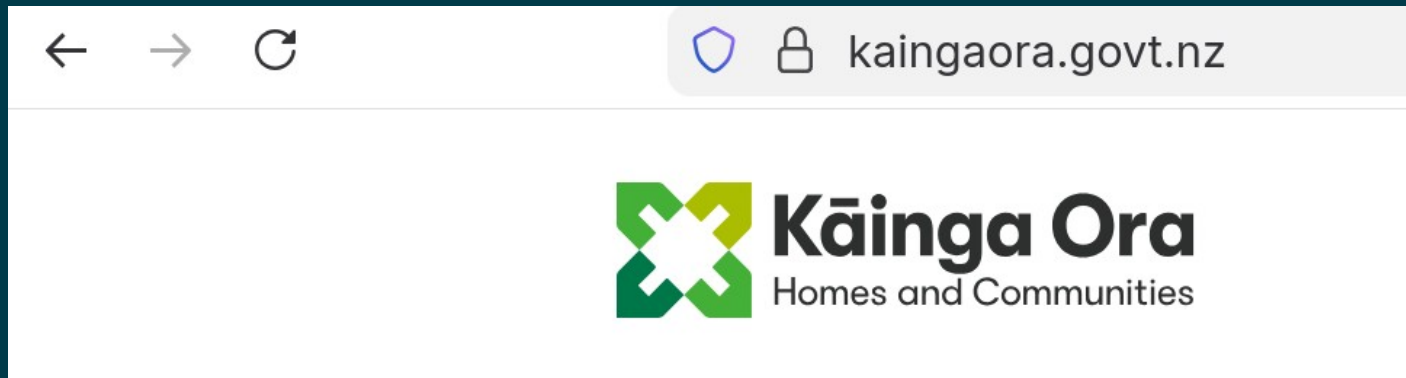
Nē

Linguistic Necessity

- Keke – Cake
- Kēkē – Armpit
- Kainga – Command to eat
- Kāinga - Home

A Real World Example

- kaingaora.govt.nz – main site
- kāingaora.govt.nz – redirects to main site
- kaaingaora.govt.nz – redirects to main site



Universal Acceptance

ICANN Internationalized Domain Names Report (June 2025)...

"However, to fully harness the power of IDNs, the ICANN community must continue to champion Universal Acceptance (UA) to ensure that all domain names and email addresses, regardless of script, language, or character length, can be used seamlessly in all Internet-enabled applications, devices, and systems."



Nē.nz

nē

1.is that so? won't you? won't we? isn't it? -
interrogative emphasising a question, request or
proposal.

Māu ngā rīhi e horoi, nē? / You'll wash the dishes,
won't you?

Nē

Nē.nz

- Free subdomains
- Full registry with RDAP and whois
- Security rules to prevent issues
 - Verifiable New Zealanders only
 - Homoglyph bundling
 - Open to ideas?



Kua Mutu Tāku Kōrero

Tim Valentine – Founder, Nē.nz



<https://www.linkedin.com/in/mr-tim-valentine>

