

Security Implications of IDNs and EAI, and The Role of ICANN

Internationalized Domain Names and Email Address Internationalization: the risks they introduce, the controls that manage them, and what ICANN does about both

Walaa Jamaluddin

Master's student in Computer Engineering

October 2026

BACKGROUND

IDNs and EAI in Brief

Internationalized Domain Names (IDNs)

Domain labels in non-ASCII scripts. People read the Unicode U-label; the DNS carries an ASCII A-label encoded with Punycode.

U-label	مثال. إختبار
A-label	xn--mgbh0fb.xn--kgbechtv

Email Address Internationalization (EAI)

UTF-8 in both the local part and the domain of an email address, carried over SMTP with the SMTPUTF8 extension.

Address	مستخدم@مثال. إختبار
Local part	UTF-8 only, with no ASCII form

Universal Acceptance (UA): every valid domain name and email address works in every application.

OVERVIEW

Effect of UA on Security

What UA improves

Readable names. People can check a name written in the script they read.

Less transliteration. Fewer improvised Latin spellings for attackers to imitate.

Consistent handling. Every system treats the same name the same way.

What UA exposes

More lookalikes. Letters that look like ASCII but are different code points.

A larger namespace. More names to monitor, register and block.

More processing steps. Validation, storage and display must each handle Unicode.

The balance depends on implementation: partial or inconsistent support carries the most risk.

THREATS

The Homograph Problem

REAL

adobe.com

FAKE

adobe.com

IN THE DNS

xn--adobe-x34a.com

One letter is different

The fake uses ɔ̣ (U+1E05), a b with a dot below, in place of b (U+0062).

Easy to miss

In an address bar the dot is tiny, and people read the name they expect to see.

Used in a real attack

In 2017 this name offered a fake Flash Player download that installed the Betabot malware.

THREATS

IDN Threat Surface

Threat	How it works	Example	Consequence
Homographs	Letters from another script imitate a trusted name	apple.com and apple.com in Cyrillic	Phishing, malware and brand abuse
Variant labels	One name has several valid spellings in the same script	يبريد and بريد: the third letter is ي (Arabic yeh) or ی (Persian yeh)	Variants held by different registrants split traffic and trust
Bidirectional text	Right-to-left and left-to-right labels reorder on screen	شبكة.مثال.com: the label beside .com is the subdomain	The order on screen can mislead the reader
Inconsistent display	Apps show either the Unicode or the xn-- form	مثال in one app, xn--mgbh0fb in another	Users have no stable visual cue to rely on

THREATS

A Larger Namespace

79.1%

of 15,276 candidate accented forms of popular site names were unregistered

34.6%

of the registered forms belonged to a party other than the site owner

RISK

More names to defend

Each name has many plausible forms across scripts, and defensive registration does not scale.

RISK

Tools built for ASCII

Blocklists, filters and logs have to match the Unicode and Punycode forms of the same name.

BENEFIT

More legitimate choice

Organizations can register their real name in their own script and avoid improvised transliterations.

THREATS

EAI Threat Surface

Confusable local parts

Homographs reach the mailbox name too. Local parts are case-sensitive, so forcing lowercase cannot normalize them.

Normalization gaps

Different UTF-8 sequences can display as the same address. Systems that compare without normalizing see different users.

Downgrade and ASCII aliases

Falling back to an ASCII address must not assume it is bound to the same authenticated mailbox.

Filter and client mismatch

If a filter and the recipient's mail client read a UTF-8 message differently, content can pass that should be rejected.

THREATS

Attack Points in a UA-Ready System

Stage	What can go wrong	Defence
Accept	Valid names are rejected, or raw Unicode reaches code that expects ASCII	Accept UTF-8 input and treat it as text throughout
Validate	Pattern and length checks miss diacritics, invisible joiners and mixed digit sets	Validate with IDNA2008 and UTS #39 libraries
Store	é can be one code point (U+00E9) or two (e U+0065 + ´ U+0301), so one address is treated as two	Store UTF-8, normalize to NFC, keep one canonical form
Process	Comparison treats Arabic yeh ي and Persian yeh ی as different, so lookalike accounts coexist	Compare canonical forms and apply variant rules
Display	Right-to-left reordering, lookalike glyphs and missing fonts hide the real name	Show Punycode or a warning for suspicious names

UNEVEN ADOPTION

29%

of domains have UA-ready mail servers, up from about 20% in 2022

With partial support, an internationalized address can be accepted by one system and rejected or mishandled by the next.

When the Local Script Is Not Supported

Someone who cannot read Latin script cannot tell a genuine ASCII name from a fake one.

Names cannot be checked

An unfamiliar script is a row of shapes, so a spoofed name looks as valid as the real one.

Transliteration is ambiguous

A name may have several Latin spellings, and an attacker can register the unused ones.

Reliance on intermediaries

Users ask someone else to read or type addresses for them, which opens the door to impersonation.

MITIGATIONS

Mitigations by Layer

Standards

IETF and Unicode

Allow only approved characters in names and publish lists of look alike characters.

Registries and registrars

ICANN

Apply rules for each script, limit script mixing, and keep variants with one owner.

Browsers and mail clients

Outside ICANN

Show the xn-- form when a name looks suspicious, and flag look alike senders.

Mail operators

Outside ICANN

Verify the sender using the name's ASCII form, so email authentication works for IDNs.

Only one of these four layers sits with ICANN.

Security and Inclusion

The risks are real, but going back to ASCII-only names would not remove them.

01

ASCII names are spoofed too

example.com

The digit 1 stands in for the letter l. The answer was better defences, not fewer characters.

02

Blocking shuts people out

Software that rejects IDNs and EAI excludes everyone who reads and writes in another script.

03

The aim is both

Every valid name works everywhere, with imitation handled at registration, display and authentication.

Role of the ICANN Community

SSAC	Advice on confusability, variants and string similarity.
GNSO and ccNSO	Policy on IDN top-level domains and how their variants are managed.
Generation panels	Script communities that write the Label Generation Rules for the root zone.
Registries and registrars	Apply the IDN Implementation Guidelines when names are registered.
UA community	Measures UA-readiness and guides developers and mail operators.
End users	Test the software, report failures, and ask for UA-ready support.

What ICANN Does

Root zone rules

Sets which characters and variants a top-level domain may use in each script.

Registry guidelines

Gives registries rules for registering IDNs that limit cybersquatting and confusion.

Variant policy

In 2024 the GNSO Council approved recommendations that all variants of a name go to the same registrant.

UA measurement and outreach

Reports each year on how UA-ready software is and runs UA Day and training.

Where ICANN's Remit Ends

Within ICANN's remit

The root zone. Which top-level labels and variants are valid.

Generic top-level domains. Policy and guidelines for IDN registration.

Coordination. Measuring UA-readiness and convening the community.

Outside ICANN's remit

Protocols. Standards come from the IETF, character data from Unicode.

Browsers and mail clients. How a name is displayed to the user.

Mail servers and apps. Whether software accepts an internationalized address.

“ICANN shall not regulate ... services that use the Internet’s unique identifiers or the content that such services carry or provide, outside the express scope of Section 1.1(a).”

SUMMARY

Key Takeaways

- 01 UA changes security in both directions: people can verify names in their own script, while attackers gain look alikes and a larger namespace.
- 02 Uneven adoption of an internationalized addresses can work in one system and fail in the next.
- 03 The alternative is not to abandon UA but to strengthen what we have: ICANN's rules for names at the registry, and correct handling in software and mail systems.

SOURCES

References

IETF STANDARDS

RFC 3492: Punycode (2003)

RFC 5890–5894: Internationalized Domain Names for Applications, IDNA2008 (2010)

RFC 6530, 6531, 6532, 6533: internationalized email framework, SMTP extension, headers and delivery notifications (2012)

RFC 8616: Email Authentication for Internationalized Mail (2019)

RESEARCH

X. Zheng: Phishing with Unicode Domains, 14 April 2017

M. Mimoso: IDN Homograph Attack Spreading Betabot Backdoor, Threatpost, 6 September 2017

V. Le Pochat, T. Van Goethem, W. Joosen: Funny Accents: Exploring Genuine Interest in Internationalized Domain Names, 2019

ICANN AND COMMUNITY

ICANN: Root Zone Label Generation Rules, version 5, May 2022, including the Arabic script LGR

ICANN: IDN Implementation Guidelines, version 4.1 (2022)

ICANN: Bylaws, Section 1.1 (Mission)

ICANN: IDN Implementation and UA Adoption Report 2026

ICANN: Universal Acceptance programme

GNSO: EPDP on IDNs, Phase 2 Recommendations Report, December 2024

SSAC: IDN Homograph Attacks, presentation, 22 October 2018

ICANNWiki: IDN Variant Labels

UNICODE

UTS #39: Unicode Security Mechanisms

UTR #36: Unicode Security Considerations (last revised 2014)

THANK YOU

Contact

Walaa Jamaluddin

EMAIL

w.jamaluddin@outlook.com

LINKEDIN

linkedin.com/in/walaa-jamaluddin



SCAN FOR LINKEDIN