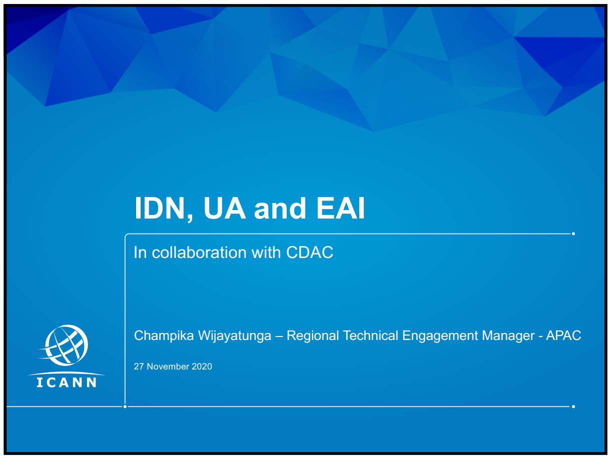
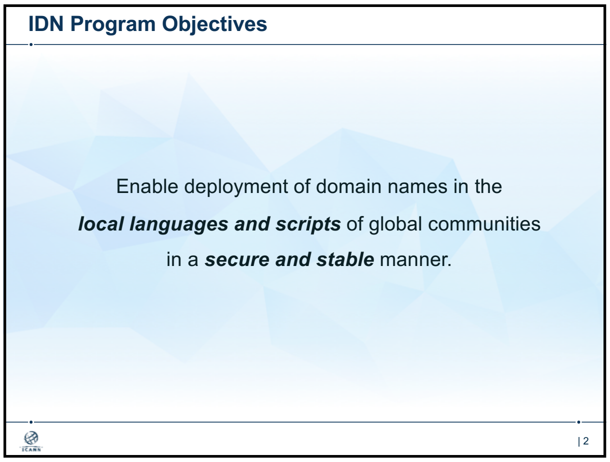


# IDN, UA and EAI

In collaboration with CDAC

Champika Wijayatunga – Regional Technical Engagement Manager - APAC

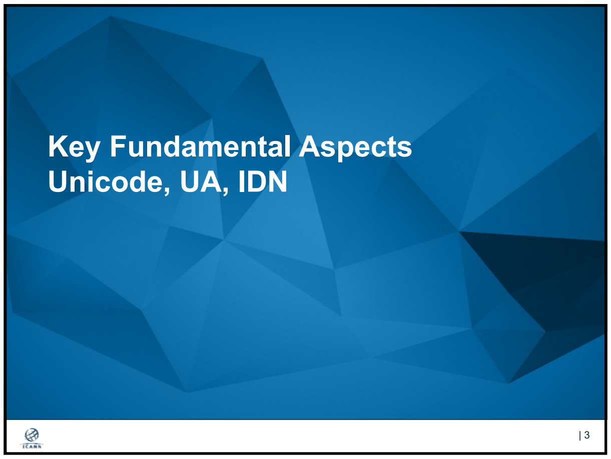
27 November 2020



## IDN Program Objectives

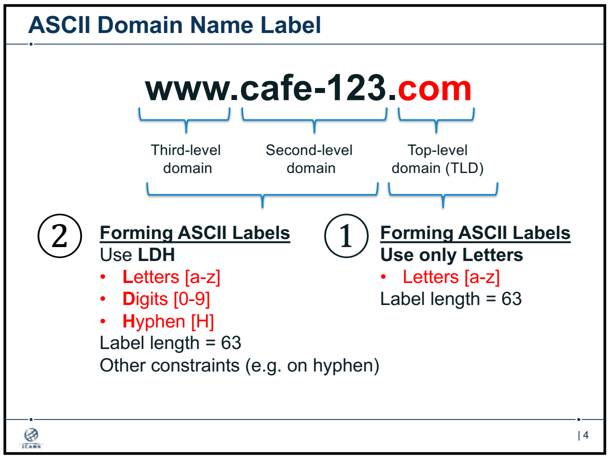
Enable deployment of domain names in the ***local languages and scripts*** of global communities in a ***secure and stable*** manner.

| 2



## Key Fundamental Aspects Unicode, UA, IDN

| 3



## ASCII Domain Name Label

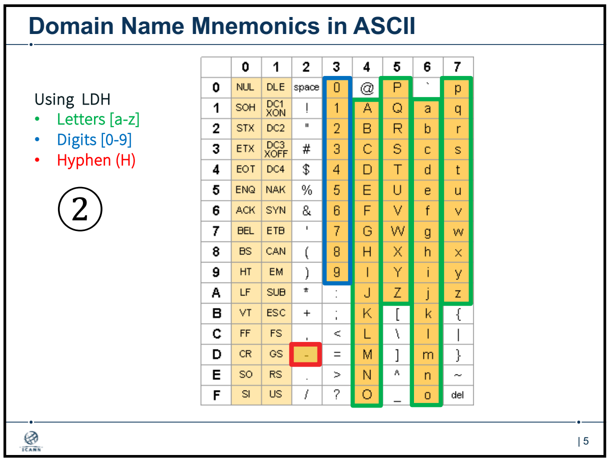
www.cafe-123.com

② **Forming ASCII Labels**
Use **LDH**
- **L**etters [a-z]
- **D**igits [0-9]
- **H**yphen [H]

Label length = 63
Other constraints (e.g. on hyphen)

① **Forming ASCII Labels**
**Use only Letters**
- Letters [a-z]

Label length = 63

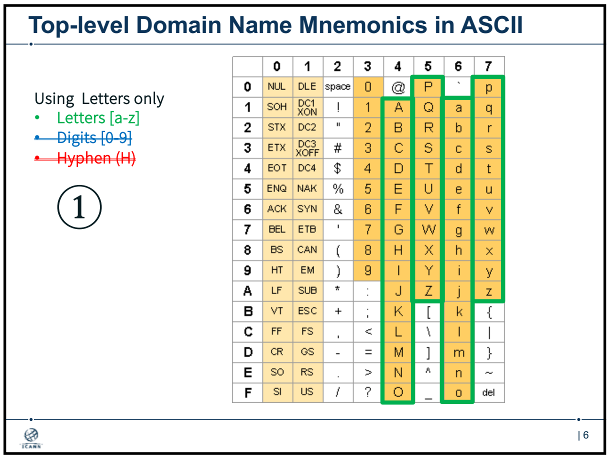


## Domain Name Mnemonics in ASCII

Using LDH
- Letters [a-z]
- Digits [0-9]
- Hyphen (H)

②

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | NUL | DLE | space | 0 | @ | P | ` | p |
| 1 | SOH | DC1<br>XON | ! | 1 | A | Q | a | q |
| 2 | STX | DC2 | " | 2 | B | R | b | r |
| 3 | ETX | DC3<br>XOFF | # | 3 | C | S | c | s |
| 4 | EOT | DC4 | $ | 4 | D | T | d | t |
| 5 | ENQ | NAK | % | 5 | E | U | e | u |
| 6 | ACK | SYN | & | 6 | F | V | f | v |
| 7 | BEL | ETB | ' | 7 | G | W | g | w |
| 8 | BS | CAN | ( | 8 | H | X | h | x |
| 9 | HT | EM | ) | 9 | I | Y | i | y |
| A | LF | SUB | * | : | J | Z | j | z |
| B | VT | ESC | + | ; | K | [ | k | { |
| C | FF | FS | , | < | L | \ | l | \| |
| D | CR | GS | - | = | M | ] | m | } |
| E | SO | RS | . | > | N | ^ | n | ~ |
| F | SI | US | / | ? | O | _ | o | del |



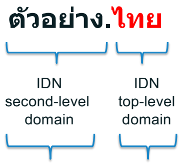
## Top-level Domain Name Mnemonics in ASCII

Using Letters only
- Letters [a-z]
- ~~Digits [0-9]~~
- ~~Hyphen (H)~~

①

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | NUL | DLE | space | 0 | @ | P | ` | p |
| 1 | SOH | DC1<br>XON | ! | 1 | A | Q | a | q |
| 2 | STX | DC2 | " | 2 | B | R | b | r |
| 3 | ETX | DC3<br>XOFF | # | 3 | C | S | c | s |
| 4 | EOT | DC4 | $ | 4 | D | T | d | t |
| 5 | ENQ | NAK | % | 5 | E | U | e | u |
| 6 | ACK | SYN | & | 6 | F | V | f | v |
| 7 | BEL | ETB | ' | 7 | G | W | g | w |
| 8 | BS | CAN | ( | 8 | H | X | h | x |
| 9 | HT | EM | ) | 9 | I | Y | i | y |
| A | LF | SUB | * | : | J | Z | j | z |
| B | VT | ESC | + | ; | K | [ | k | { |
| C | FF | FS | , | < | L | \ | l | \| |
| D | CR | GS | - | = | M | ] | m | } |
| E | SO | RS | . | > | N | ^ | n | ~ |
| F | SI | US | / | ? | O | _ | o | del |

## Internationalized Domain Name (IDN) Labels



ตัวอย่าง.ไทย

IDN second-level domain

IDN top-level domain

**Syntax of IDN Labels**
**Valid U-Label:** Unicode code points as constrained by the “LDH” scheme within IDNA 2008

②

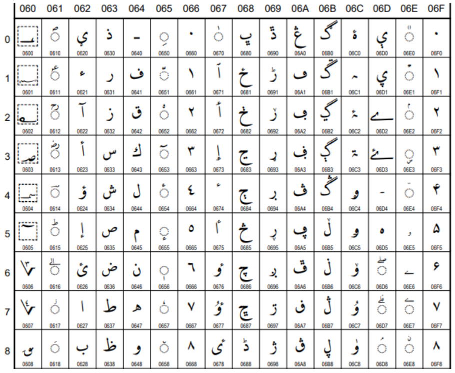
**Syntax of IDN Labels**
**Valid U-label,** further constrained by the “letter” principle for TLDs

①

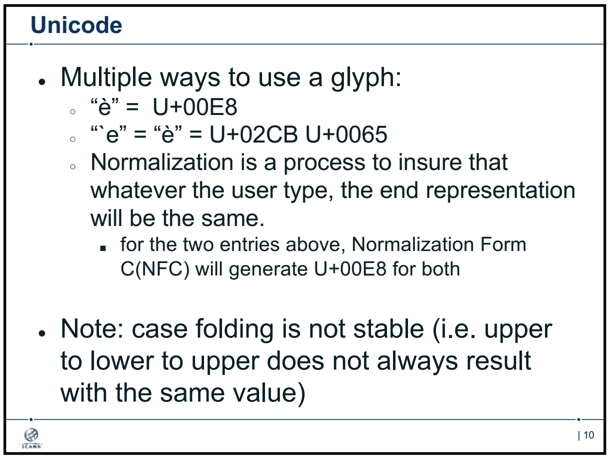
## Unicode

- Encoding glyphs into codepoints
- In specifications, codepoints are shown in hex using the U+XXXX notation
- Codepoints are typically carried using the UTF-8 (Unicode Transformation Format, 8 bit) format
  - variable number of bytes for a single codepoint.
  - ascii is used as is
  - gold standard for carrying Unicode codepoints, in web, protocols, etc...

## IDN Mnemonics

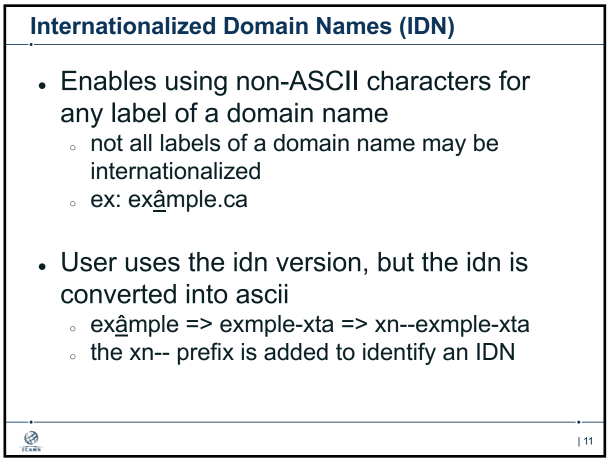


| | 060 | 061 | 062 | 063 | 064 | 065 | 066 | 067 | 068 | 069 | 06A | 06B | 06C | 06D | 06E | 06F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | ؀<br>0600 | ◌ؐ<br>0610 | ؠ<br>0620 | ذ<br>0630 | ـ<br>0640 | ◌ِ<br>0650 | ٠<br>0660 | ◌ٰ<br>0670 | ڀ<br>0680 | ڐ<br>0690 | ڠ<br>06A0 | ڰ<br>06B0 | ۀ<br>06C0 | ې<br>06D0 | ◌۠<br>06E0 | ۰<br>06F0 |
| 1 | ؁<br>0601 | ◌ؑ<br>0611 | ء<br>0621 | ر<br>0631 | ف<br>0641 | ◌ّ<br>0651 | ١<br>0661 | ٱ<br>0671 | ځ<br>0681 | ڑ<br>0691 | ڡ<br>06A1 | ڱ<br>06B1 | ہ<br>06C1 | ۑ<br>06D1 | ◌ۡ<br>06E1 | ۱<br>06F1 |
| 2 | ؂<br>0602 | ◌ؒ<br>0612 | آ<br>0622 | ز<br>0632 | ق<br>0642 | ◌ْ<br>0652 | ٢<br>0662 | ٲ<br>0672 | ڂ<br>0682 | ڒ<br>0692 | ڢ<br>06A2 | ڲ<br>06B2 | ۂ<br>06C2 | ے<br>06D2 | ◌ۢ<br>06E2 | ۲<br>06F2 |
| 3 | ؃<br>0603 | ◌ؓ<br>0613 | أ<br>0623 | س<br>0633 | ك<br>0643 | ◌ٓ<br>0653 | ٣<br>0663 | ٳ<br>0673 | ڃ<br>0683 | ړ<br>0693 | ڣ<br>06A3 | ڳ<br>06B3 | ۃ<br>06C3 | ۓ<br>06D3 | ◌ۣ<br>06E3 | ۳<br>06F3 |
| 4 | ؄<br>0604 | ◌ؔ<br>0614 | ؤ<br>0624 | ش<br>0634 | ل<br>0644 | ◌ٔ<br>0654 | ٤<br>0664 | ٴ<br>0674 | ڄ<br>0684 | ڔ<br>0694 | ڤ<br>06A4 | ڴ<br>06B4 | ۄ<br>06C4 | ۔<br>06D4 | ◌ۤ<br>06E4 | ۴<br>06F4 |
| 5 | ؅<br>0605 | ◌ؕ<br>0615 | إ<br>0625 | ص<br>0635 | م<br>0645 | ◌ٕ<br>0655 | ٥<br>0665 | ٵ<br>0675 | څ<br>0685 | ڕ<br>0695 | ڥ<br>06A5 | ڵ<br>06B5 | ۅ<br>06C5 | ە<br>06D5 | ۥ<br>06E5 | ۵<br>06F5 |
| 6 | ؆<br>0606 | ◌ؖ<br>0616 | ئ<br>0626 | ض<br>0636 | ن<br>0646 | ◌ٖ<br>0656 | ٦<br>0666 | ٶ<br>0676 | چ<br>0686 | ږ<br>0696 | ڦ<br>06A6 | ڶ<br>06B6 | ۆ<br>06C6 | ◌ۖ<br>06D6 | ۦ<br>06E6 | ۶<br>06F6 |
| 7 | ؇<br>0607 | ◌ؗ<br>0617 | ا<br>0627 | ط<br>0637 | ه<br>0647 | ◌ٗ<br>0657 | ٧<br>0667 | ٷ<br>0677 | ڇ<br>0687 | ڗ<br>0697 | ڧ<br>06A7 | ڷ<br>06B7 | ۇ<br>06C7 | ◌ۗ<br>06D7 | ◌ۧ<br>06E7 | ۷<br>06F7 |
| 8 | ؈<br>0608 | ◌ؘ<br>0618 | ب<br>0628 | ظ<br>0638 | و<br>0648 | ◌٘<br>0658 | ٨<br>0668 | ٸ<br>0678 | ڈ<br>0688 | ژ<br>0698 | ڨ<br>06A8 | ڸ<br>06B8 | ۈ<br>06C8 | ◌ۘ<br>06D8 | ◌ۨ<br>06E8 | ۸<br>06F8 |



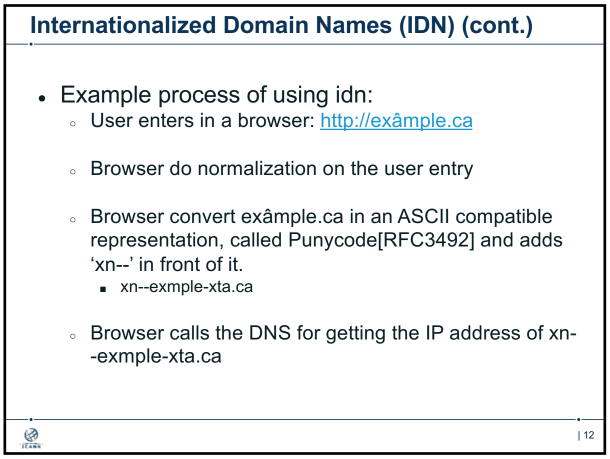
## Unicode

- Multiple ways to use a glyph:
  - “è” =  U+00E8
  - “ˋe” = “è” = U+02CB U+0065
  - Normalization is a process to insure that whatever the user type, the end representation will be the same.
    - for the two entries above, Normalization Form C(NFC) will generate U+00E8 for both
- Note: case folding is not stable (i.e. upper to lower to upper does not always result with the same value)



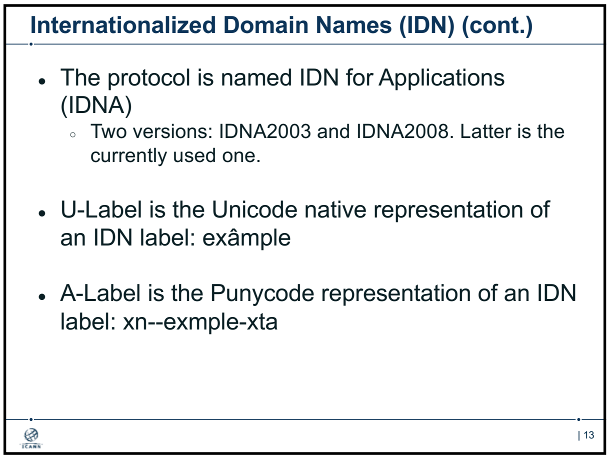
## Internationalized Domain Names (IDN)

- Enables using non-ASCII characters for any label of a domain name
  - not all labels of a domain name may be internationalized
  - ex: exâmple.ca
- User uses the idn version, but the idn is converted into ascii
  - exâmple => exmple-xta => xn--exmple-xta
  - the xn-- prefix is added to identify an IDN



## Internationalized Domain Names (IDN) (cont.)

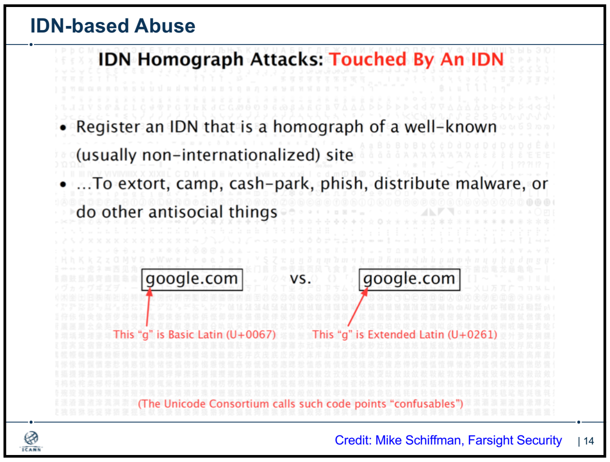
- Example process of using idn:
  - User enters in a browser: http://exâmple.ca
  - Browser do normalization on the user entry
  - Browser convert exâmple.ca in an ASCII compatible representation, called Punycode[RFC3492] and adds ‘xn--’ in front of it.
    - xn--exmple-xta.ca
  - Browser calls the DNS for getting the IP address of xn--exmple-xta.ca



## Internationalized Domain Names (IDN) (cont.)

- The protocol is named IDN for Applications (IDNA)
  - Two versions: IDNA2003 and IDNA2008. Latter is the currently used one.
- U-Label is the Unicode native representation of an IDN label: exâmple
- A-Label is the Punycode representation of an IDN label: xn--exmple-xta

| 13



## IDN-based Abuse

**IDN Homograph Attacks: Touched By An IDN**

- Register an IDN that is a homograph of a well-known (usually non-internationalized) site
- …To extort, camp, cash-park, phish, distribute malware, or do other antisocial things

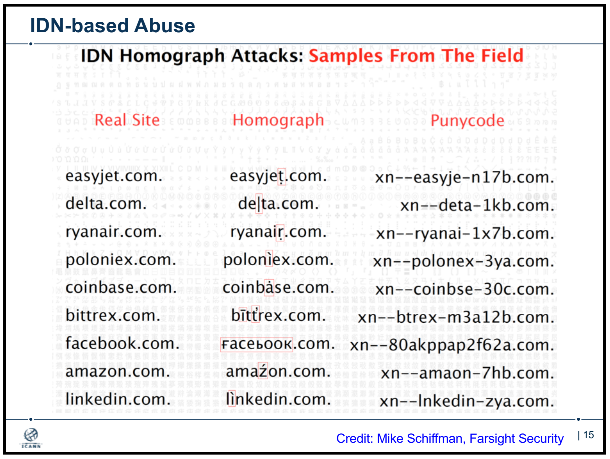
google.com vs. ɡoogle.com

This "g" is Basic Latin (U+0067)

This "g" is Extended Latin (U+0261)

(The Unicode Consortium calls such code points "confusables")

Credit: Mike Schiffman, Farsight Security | 14

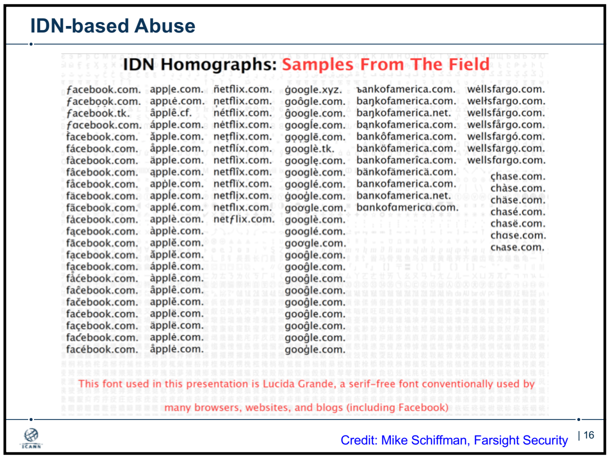


## IDN-based Abuse

**IDN Homograph Attacks: Samples From The Field**

| Real Site | Homograph | Punycode |
|---|---|---|
| easyjet.com. | easyjeṭ.com. | xn--easyje-n17b.com. |
| delta.com. | deǀta.com. | xn--deta-1kb.com. |
| ryanair.com. | ryanaiṛ.com. | xn--ryanai-1x7b.com. |
| poloniex.com. | polonìex.com. | xn--polonex-3ya.com. |
| coinbase.com. | coinbȧse.com. | xn--coinbse-30c.com. |
| bittrex.com. | bītťrex.com. | xn--btrex-m3a12b.com. |
| facebook.com. | ғасеьоок.com. | xn--80akppap2f62a.com. |
| amazon.com. | amaźon.com. | xn--amaon-7hb.com. |
| linkedin.com. | lìnkedin.com. | xn--lnkedin-zya.com. |

Credit: Mike Schiffman, Farsight Security | 15



## IDN-based Abuse

**IDN Homographs: Samples From The Field**

ƒacebook.com.
ƒacebọọk.com.
ƒacebook.tk.
ƒɑcebook.com.
ḟacebook.com.
fácebook.com.
fàcebook.com.
fâcebook.com.
fåcebook.com.
făcebook.com.
fãcebook.com.
fȧcebook.com.
fącebook.com.
fācebook.com.
fạcebook.com.
fa̧cebook.com.
fåćebook.com.
faĉebook.com.
fačebook.com.
faċebook.com.
façebook.com.
faćebook.com.
facébook.com.

appǀe.com.
appɩė.com.
âpplê.cf.
ápple.com.
ăpple.com.
åpple.com.
äpple.com.
ąpple.com.
apṗle.com.
appĺe.com.
applé.com.
applè.com.
àpplè.com.
applĕ.com.
ăpplĕ.com.
ápplê.com.
àpplê.com.
âpplê.com.
applĕ.com.
applë.com.
äpplë.com.
applė.com.
åpplė.com.

ñetflix.com.
ņetflix.com.
nétflix.com.
nėtflix.com.
neţflix.com.
netflíx.com.
netflìx.com.
netflîx.com.
netflïx.com.
netflịx.com.
netflıx.com.
netƒlix.com.

ġoogle.xyz.
goôgle.com.
ĝoogle.com.
googĺe.com.
gọọglē.com.
googlè.tk.
googlę.com.
googlè.com.
googlé.com.
ġooġle.com.
goơgle.com.
googlè.com.
googlé.com.
goơgle.com.
gooĝle.com.
gooĝle.com.
gooĝle.com.
gooĝle.com.
gooĝle.com.
gooĝle.com.
gooĝle.com.
gooĝle.com.
gooġle.com.

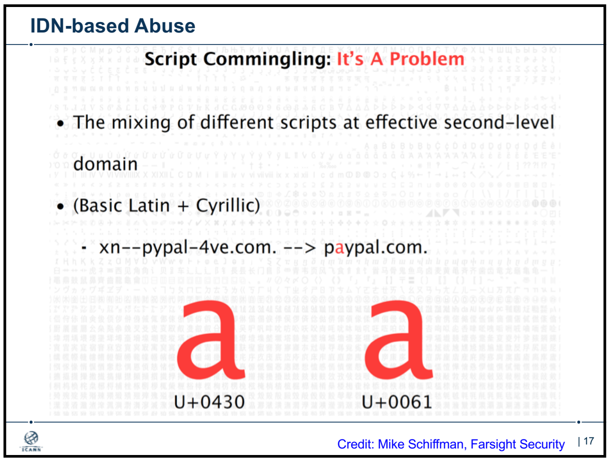
ъankofamerica.com.
baŋkofamerica.com.
baŋkofamerica.net.
bąnkofamerica.com.
bankôfamerica.com.
banköfamerica.com.
bankofamerîca.com.
bănkofămerică.com.
banĸofamerica.com.
banĸofamerica.net.
bɑnkofɑmericɑ.com.

wėllsfargo.com.
welłsfargo.com.
wellsfárgo.com.
wellsfårgo.com.
wellsfargó.com.
wellsfargọ.com.
wellsfɑrgo.com.

çhase.com.
chàse.com.
chäse.com.
chasé.com.
chasë.com.
chɑse.com.
cнase.com.

This font used in this presentation is Lucida Grande, a serif-free font conventionally used by many browsers, websites, and blogs (including Facebook)

Credit: Mike Schiffman, Farsight Security



## IDN-based Abuse

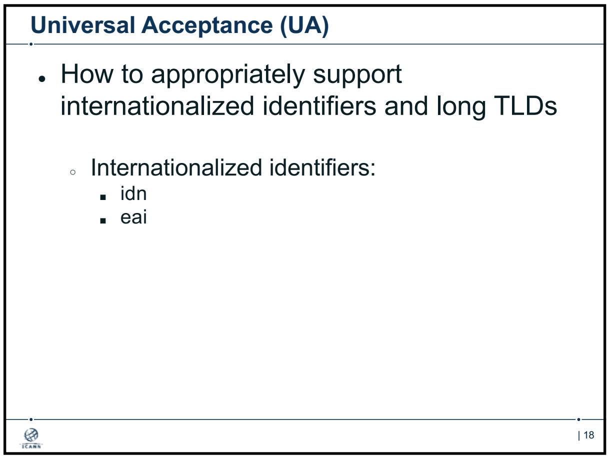
**Script Commingling: It's A Problem**

- The mixing of different scripts at effective second-level domain
- (Basic Latin + Cyrillic)
  - xn--pypal-4ve.com. --> paypal.com.

a U+0430

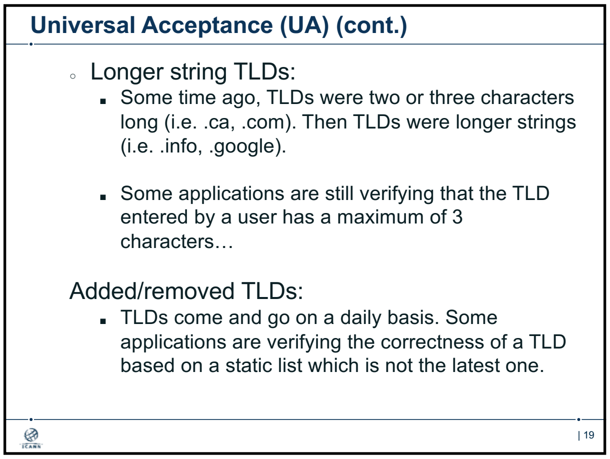
a U+0061

Credit: Mike Schiffman, Farsight Security



## Universal Acceptance (UA)

- How to appropriately support internationalized identifiers and long TLDs
  - Internationalized identifiers:
    - idn
    - eai



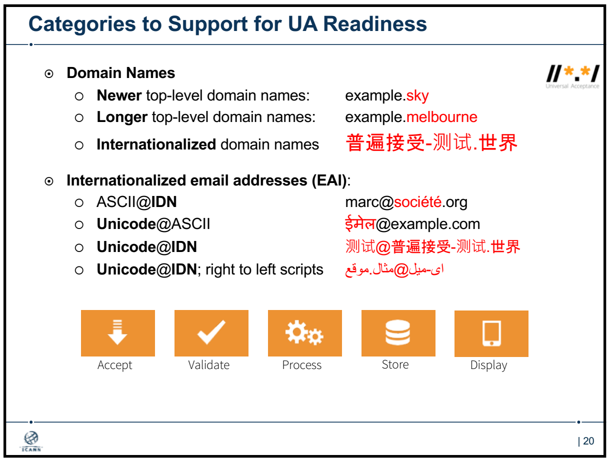
## Universal Acceptance (UA) (cont.)

- Longer string TLDs:
  - Some time ago, TLDs were two or three characters long (i.e. .ca, .com). Then TLDs were longer strings (i.e. .info, .google).
  - Some applications are still verifying that the TLD entered by a user has a maximum of 3 characters…

Added/removed TLDs:

- TLDs come and go on a daily basis. Some applications are verifying the correctness of a TLD based on a static list which is not the latest one.

| 19



## Categories to Support for UA Readiness

- **Domain Names**

| | |
|---|---|
| **Newer** top-level domain names: | example.sky |
| **Longer** top-level domain names: | example.melbourne |
| **Internationalized** domain names | 普遍接受-测试.世界 |

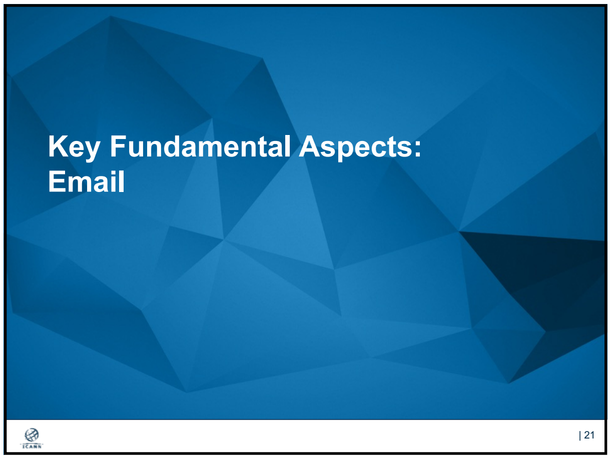
- **Internationalized email addresses (EAI)**:

| | |
|---|---|
| ASCII@**IDN** | marc@société.org |
| **Unicode**@ASCII | ईमेल@example.com |
| **Unicode**@**IDN** | 测试@普遍接受-测试.世界 |
| **Unicode**@**IDN**; right to left scripts | ای-میل@مثال.موقع |

Universal Acceptance

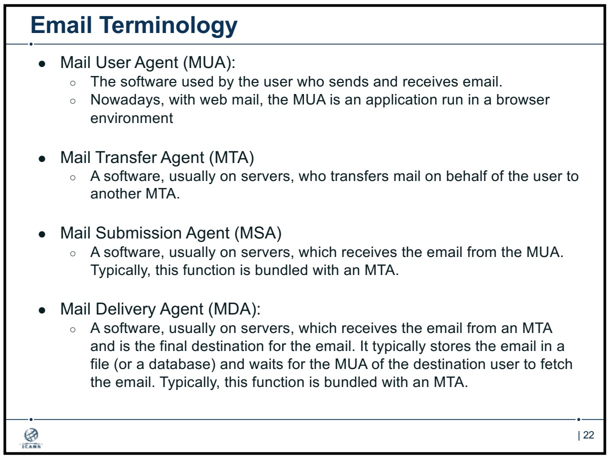
Accept Validate Process Store Display

| 20



## Key Fundamental Aspects: Email

| 21



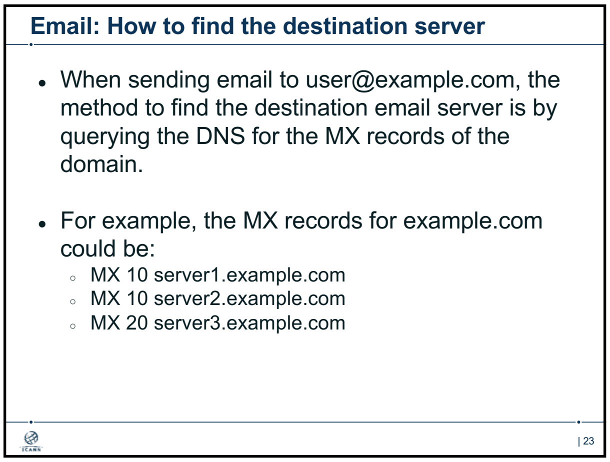
## Email Terminology

- Mail User Agent (MUA):
  - The software used by the user who sends and receives email.
  - Nowadays, with web mail, the MUA is an application run in a browser environment

- Mail Transfer Agent (MTA)
  - A software, usually on servers, who transfers mail on behalf of the user to another MTA.

- Mail Submission Agent (MSA)
  - A software, usually on servers, which receives the email from the MUA. Typically, this function is bundled with an MTA.

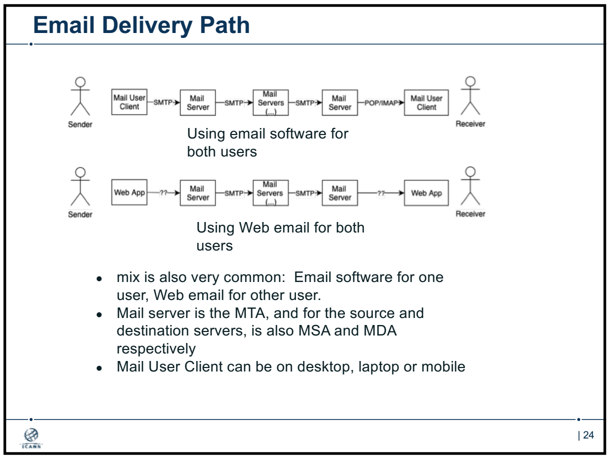
- Mail Delivery Agent (MDA):
  - A software, usually on servers, which receives the email from an MTA and is the final destination for the email. It typically stores the email in a file (or a database) and waits for the MUA of the destination user to fetch the email. Typically, this function is bundled with an MTA.



## Email: How to find the destination server

- When sending email to user@example.com, the method to find the destination email server is by querying the DNS for the MX records of the domain.

- For example, the MX records for example.com could be:
  - MX 10 server1.example.com
  - MX 10 server2.example.com
  - MX 20 server3.example.com

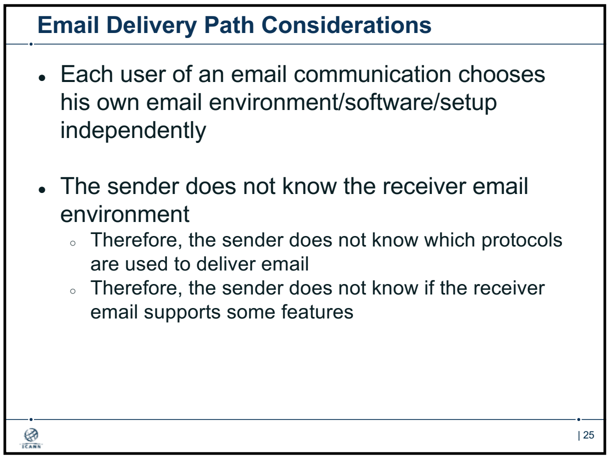


## Email Delivery Path

Using email software for both users

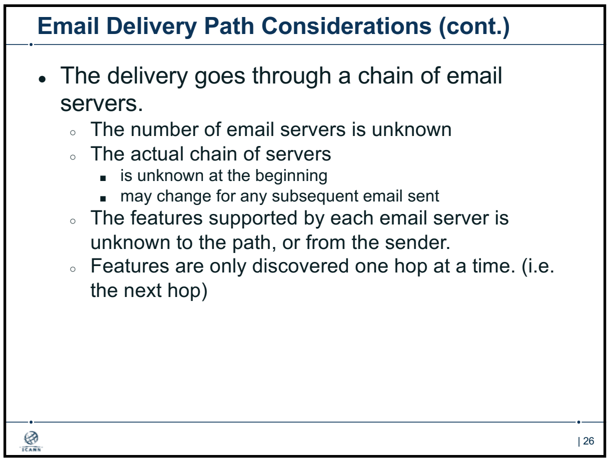
Using Web email for both users

- mix is also very common:  Email software for one user, Web email for other user.
- Mail server is the MTA, and for the source and destination servers, is also MSA and MDA respectively
- Mail User Client can be on desktop, laptop or mobile



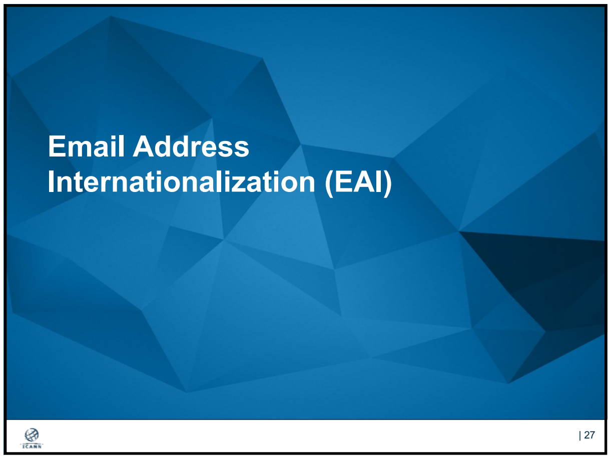
## Email Delivery Path Considerations

- Each user of an email communication chooses his own email environment/software/setup independently
- The sender does not know the receiver email environment
  - Therefore, the sender does not know which protocols are used to deliver email
  - Therefore, the sender does not know if the receiver email supports some features

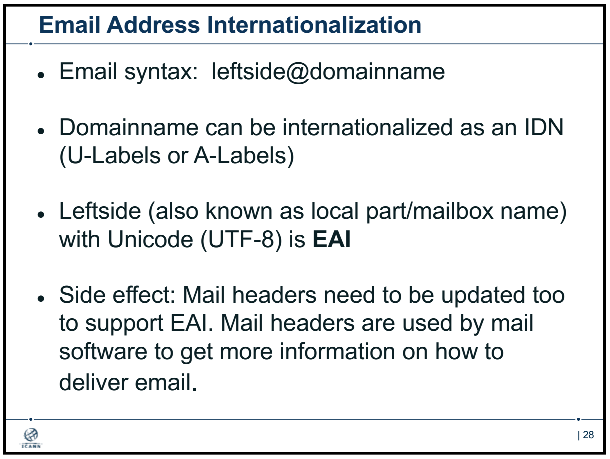


## Email Delivery Path Considerations (cont.)

- The delivery goes through a chain of email servers.
  - The number of email servers is unknown
  - The actual chain of servers
    - is unknown at the beginning
    - may change for any subsequent email sent
  - The features supported by each email server is unknown to the path, or from the sender.
  - Features are only discovered one hop at a time. (i.e. the next hop)

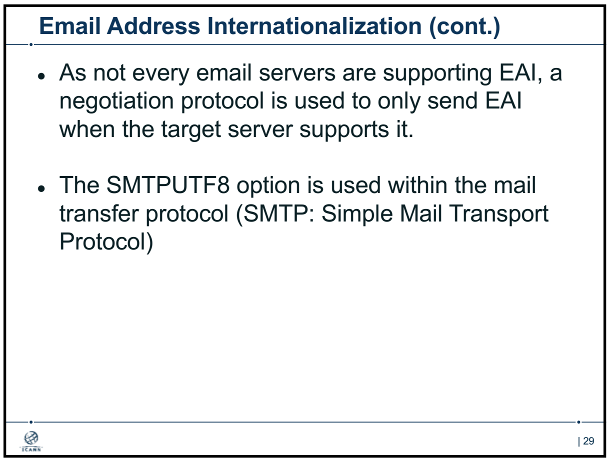


## Email Address Internationalization (EAI)



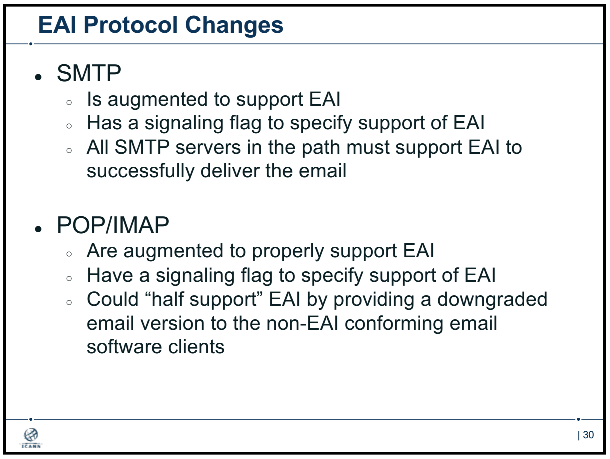
## Email Address Internationalization

- Email syntax: leftside@domainname
- Domainname can be internationalized as an IDN (U-Labels or A-Labels)
- Leftside (also known as local part/mailbox name) with Unicode (UTF-8) is **EAI**
- Side effect: Mail headers need to be updated too to support EAI. Mail headers are used by mail software to get more information on how to deliver email.



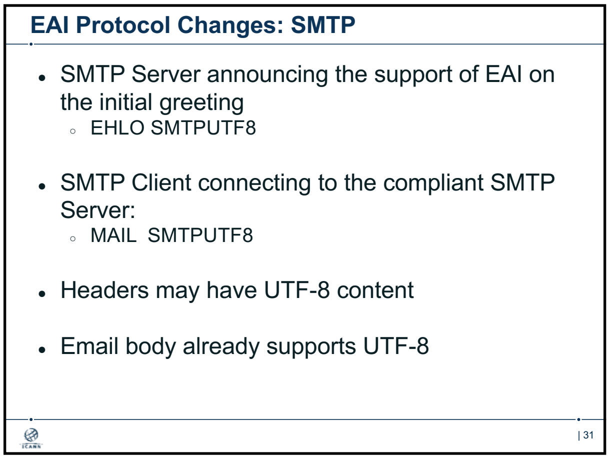
## Email Address Internationalization (cont.)

- As not every email servers are supporting EAI, a negotiation protocol is used to only send EAI when the target server supports it.
- The SMTPUTF8 option is used within the mail transfer protocol (SMTP: Simple Mail Transport Protocol)



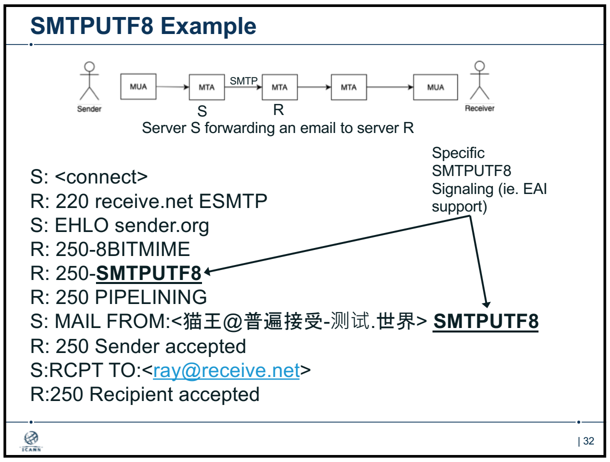
## EAI Protocol Changes

- SMTP
  - Is augmented to support EAI
  - Has a signaling flag to specify support of EAI
  - All SMTP servers in the path must support EAI to successfully deliver the email
- POP/IMAP
  - Are augmented to properly support EAI
  - Have a signaling flag to specify support of EAI
  - Could “half support” EAI by providing a downgraded email version to the non-EAI conforming email software clients



## EAI Protocol Changes: SMTP

- SMTP Server announcing the support of EAI on the initial greeting
  - EHLO SMTPUTF8
- SMTP Client connecting to the compliant SMTP Server:
  - MAIL  SMTPUTF8
- Headers may have UTF-8 content
- Email body already supports UTF-8

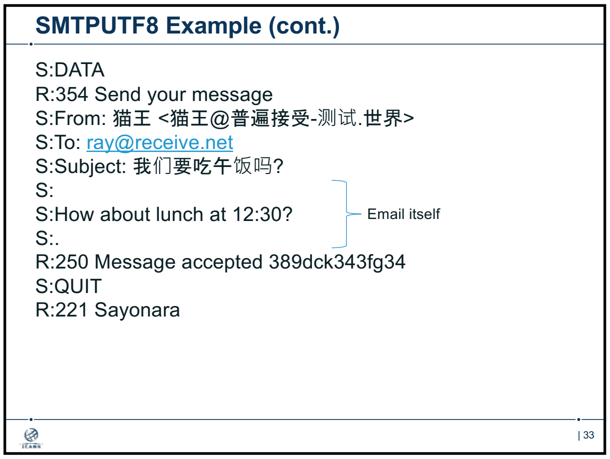


## SMTPUTF8 Example

Server S forwarding an email to server R

Specific SMTPUTF8 Signaling (ie. EAI support)

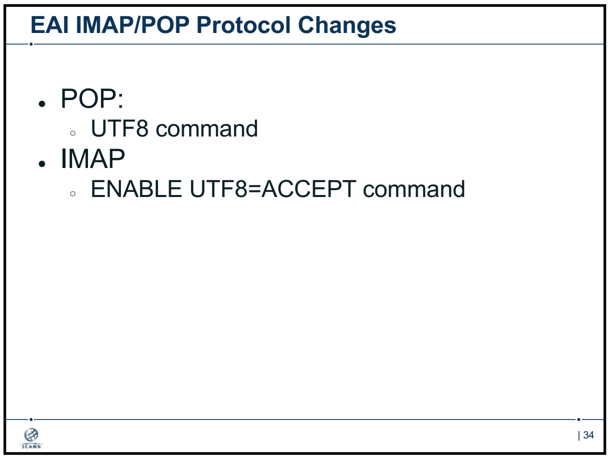
S: <connect>  
R: 220 receive.net ESMTP  
S: EHLO sender.org  
R: 250-8BITMIME  
R: 250-**SMTPUTF8**  
R: 250 PIPELINING  
S: MAIL FROM:<猫王@普遍接受-测试.世界> **SMTPUTF8**  
R: 250 Sender accepted  
S:RCPT TO:<ray@receive.net>  
R:250 Recipient accepted



## SMTPUTF8 Example (cont.)

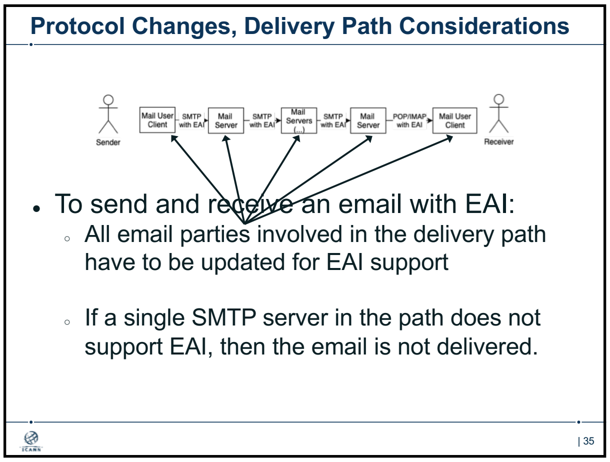
S:DATA  
R:354 Send your message  
S:From: 猫王 <猫王@普遍接受-测试.世界>  
S:To: ray@receive.net  
S:Subject: 我们要吃午饭吗?  
S:  
S:How about lunch at 12:30?  
S:.  
R:250 Message accepted 389dck343fg34  
S:QUIT  
R:221 Sayonara

Email itself



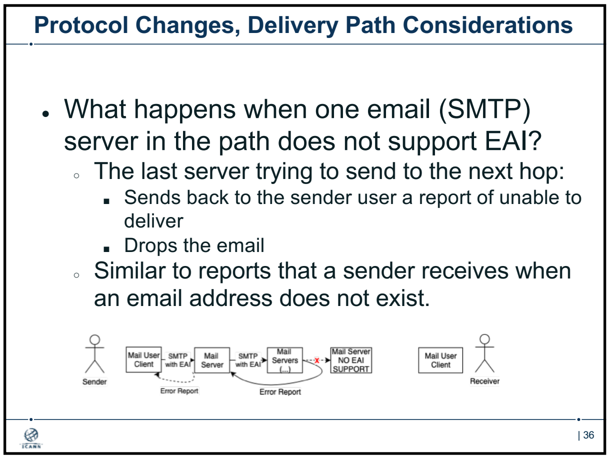
## EAI IMAP/POP Protocol Changes

- POP:
  - UTF8 command
- IMAP
  - ENABLE UTF8=ACCEPT command



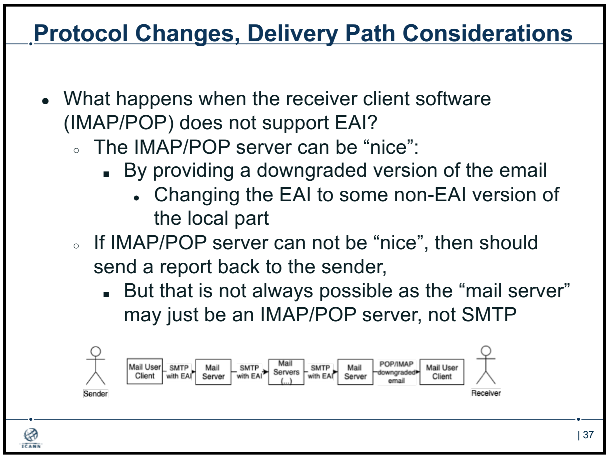
## Protocol Changes, Delivery Path Considerations

- To send and receive an email with EAI:
  - All email parties involved in the delivery path have to be updated for EAI support
  - If a single SMTP server in the path does not support EAI, then the email is not delivered.



## Protocol Changes, Delivery Path Considerations

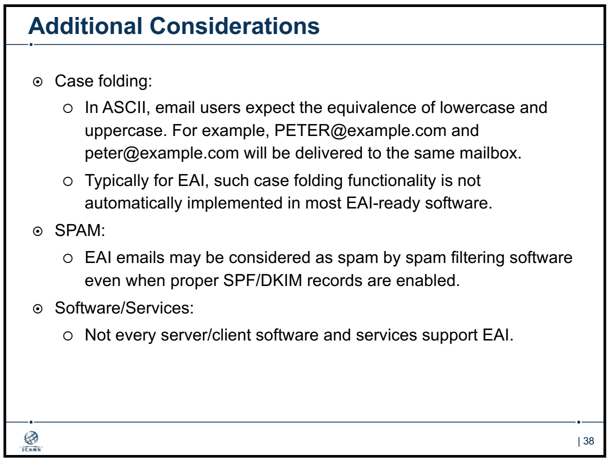
- What happens when one email (SMTP) server in the path does not support EAI?
  - The last server trying to send to the next hop:
    - Sends back to the sender user a report of unable to deliver
    - Drops the email
  - Similar to reports that a sender receives when an email address does not exist.



## Protocol Changes, Delivery Path Considerations

- What happens when the receiver client software (IMAP/POP) does not support EAI?
  - The IMAP/POP server can be "nice":
    - By providing a downgraded version of the email
      - Changing the EAI to some non-EAI version of the local part
  - If IMAP/POP server can not be "nice", then should send a report back to the sender,
    - But that is not always possible as the "mail server" may just be an IMAP/POP server, not SMTP

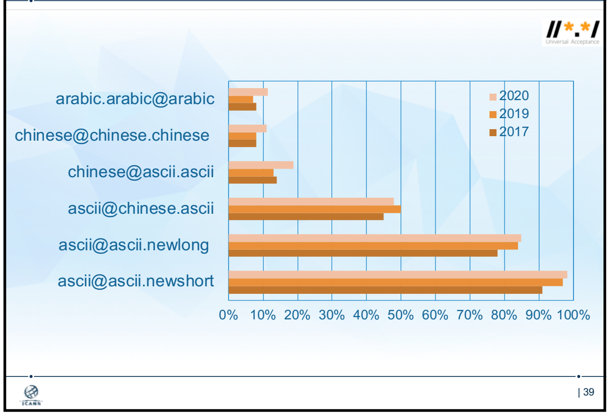
Sender → Mail User Client → SMTP with EAI → Mail Server → SMTP with EAI → Mail Servers (...) → SMTP with EAI → Mail Server → POP/IMAP downgraded email → Mail User Client → Receiver

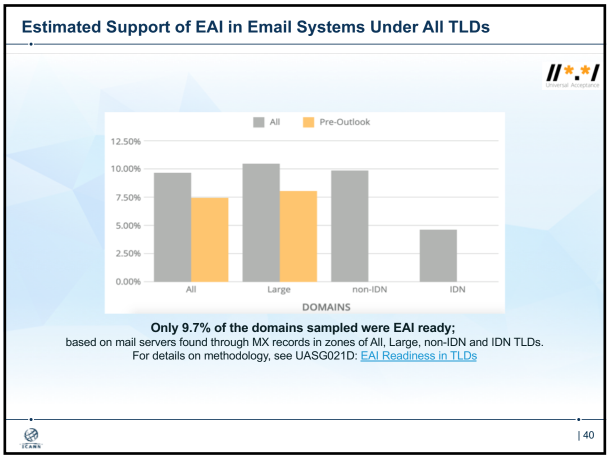


## Additional Considerations

- Case folding:
  - In ASCII, email users expect the equivalence of lowercase and uppercase. For example, PETER@example.com and peter@example.com will be delivered to the same mailbox.
  - Typically for EAI, such case folding functionality is not automatically implemented in most EAI-ready software.
- SPAM:
  - EAI emails may be considered as spam by spam filtering software even when proper SPF/DKIM records are enabled.
- Software/Services:
  - Not every server/client software and services support EAI.

## Acceptance of Email Addresses in Websites Globally

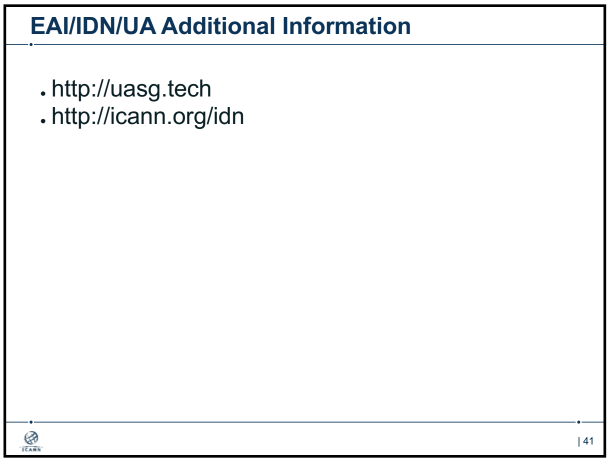




## Estimated Support of EAI in Email Systems Under All TLDs

**Only 9.7% of the domains sampled were EAI ready;**
based on mail servers found through MX records in zones of All, Large, non-IDN and IDN TLDs.
For details on methodology, see UASG021D: EAI Readiness in TLDs

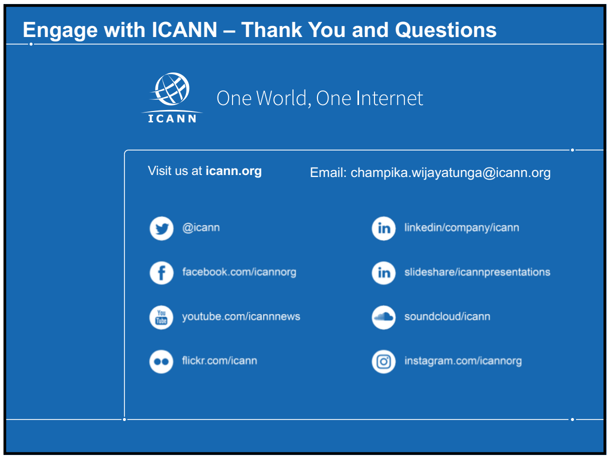
| 40



## EAI/IDN/UA Additional Information

- http://uasg.tech
- http://icann.org/idn

| 41



## Engage with ICANN – Thank You and Questions

One World, One Internet

Visit us at **icann.org**

Email: champika.wijayatunga@icann.org

@icann

linkedin/company/icann

facebook.com/icannorg

slideshare/icannpresentations

youtube.com/icannnews

soundcloud/icann

flickr.com/icann

instagram.com/icannorg