Expedited Policy Development Process on Internationalized Domain Names (EPDP on IDNs)

Update & Consultation with At-Large CPWG Re: Role of Variants in String Similarity Review

Satish Babu
Justine Chew
Lianna Galstyan
Hadia Elminiawi
Abdulkarim Oloyede

12 October 2022

ICANN
AT-LARGE

Agenda

- **Recap – Source Label, Allocatable & Blocked Variant Labels**
- **End-User Interest:**
  - Good/consistent experience
  - Security
- **String Similarity Review: Role of Allocatable & Blocked Variants**
  - Charter Questions e1, e3, e3a (also b4a, e4)
- **EPDP on IDNs String Similarity Small Group**
  - Assignment – Narrow Remit
  - Recommendation
  - Implications for implementation
- **Straw Poll to ascertain support for Recommendation**

Recap – Source Label, Allocatable & Blocked Variant Labels

A real example of RZ-LGR output for an Arabic label

Valid means available for application and delegation
Allocatable means available for request and activation

#	Type	U-label	A-label	Disposition	Code point sequence
1	original	شبكة	xn--ngbc5azd	valid	U+0634 U+0628 U+0643 U+0629
2	varlabel	شبكه	xn--ngbx0cq	allocatable	U+0634 U+0628 U+0643 U+0647
3	varlabel	شبكهٖ	xn--ngbx0c15a	blocked	U+0634 U+0628 U+0643 U+06BE
4	varlabel	شبكهٔ	xn--ngbx0c95a	blocked	U+0634 U+0628 U+0643 U+06C0
5	varlabel	شبكهٕ	xn--ngbx0cy6a	blocked	U+0634 U+0628 U+0643 U+06C1
6	varlabel	شبكهٗ	xn--ngbx0c26a	blocked	U+0634 U+0628 U+0643 U+06C2
7	varlabel	شبكهٙ	xn--ngbx0c66a	allocatable	U+0634 U+0628 U+0643 U+06C3
8	varlabel	شبكهٕ	xn--ngbx0c31b	blocked	U+0634 U+0628 U+0643 U+06D5
9	varlabel	شبكةٔ	xn--ngbc5az1b	allocatable	U+0634 U+0628 U+06A9 U+0629
10	varlabel	شبكهٔ	xn--ngbx2d5u	allocatable	U+0634 U+0628 U+06A9 U+0647
11	varlabel	شبكهٖ	xn--ngbx66ayc	blocked	U+0634 U+0628 U+06A9 U+06BE
12	varlabel	شبكهٔ	xn--ngbx66a6c	blocked	U+0634 U+0628 U+06A9 U+06C0
13	varlabel	شبكهٕ	xn--ngbx66agd	blocked	U+0634 U+0628 U+06A9 U+06C1
14	varlabel	شبكهٗ	xn--ngbx66akd	blocked	U+0634 U+0628 U+06A9 U+06C2
15	varlabel	شبكهٙ	xn--ngbx66aod	allocatable	U+0634 U+0628 U+06A9 U+06C3
16	varlabel	شبكهٕ	xn--ngbx66a0f	blocked	U+0634 U+0628 U+06A9 U+06D5
17	varlabel	شبكةٔ	xn--ngbc5a31b	allocatable	U+0634 U+0628 U+06AA U+0629
18	varlabel	شبكهٔ	xn--ngbx2d9u	allocatable	U+0634 U+0628 U+06AA U+0647
19	varlabel	شبكهٖ	xn--ngbx96asc	blocked	U+0634 U+0628 U+06AA U+06BE
20	varlabel	شبكهٔ	xn--ngbx96a0c	blocked	U+0634 U+0628 U+06AA U+06C0
21	varlabel	شبكهٕ	xn--ngbx96a4c	blocked	U+0634 U+0628 U+06AA U+06C1
22	varlabel	شبكهٗ	xn--ngbx96a8c	blocked	U+0634 U+0628 U+06AA U+06C2
23	varlabel	شبكهٙ	xn--ngbx96ahd	allocatable	U+0634 U+0628 U+06AA U+06C3
24	varlabel	شبكهٕ	xn--ngbx96arf	blocked	U+0634 U+0628 U+06AA U+06D5

String Similarity Review for gTLDs

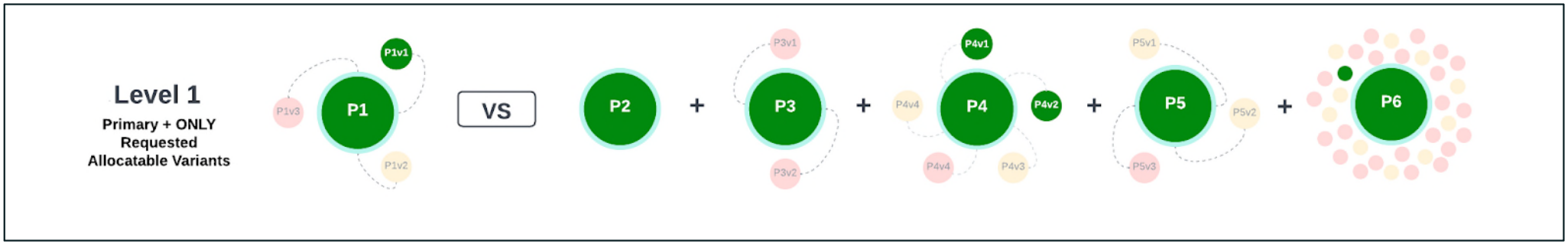
- **String similarity review focuses on visual confusability, conducted with String Similarity Review Panel**

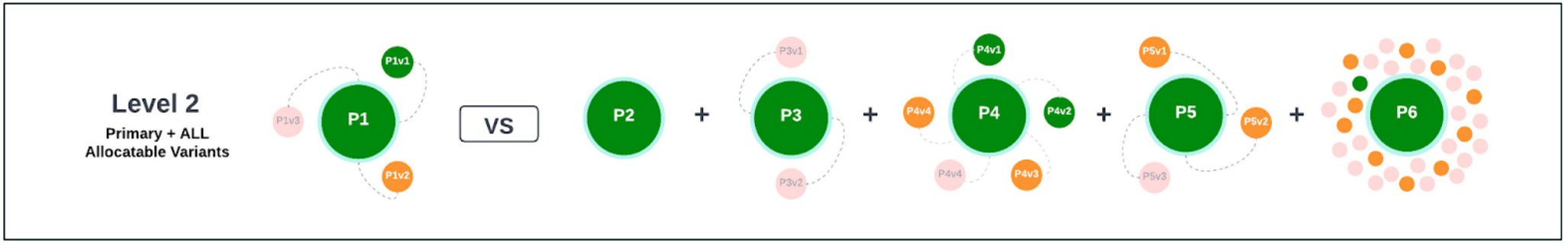
- **EPDP on IDNs Charter:**
  - What potential adjustments are needed to string similarity review due to variant implementation?
  - Specifically, what role should Allocatable & Blocked Variant Labels have in string similarity review?

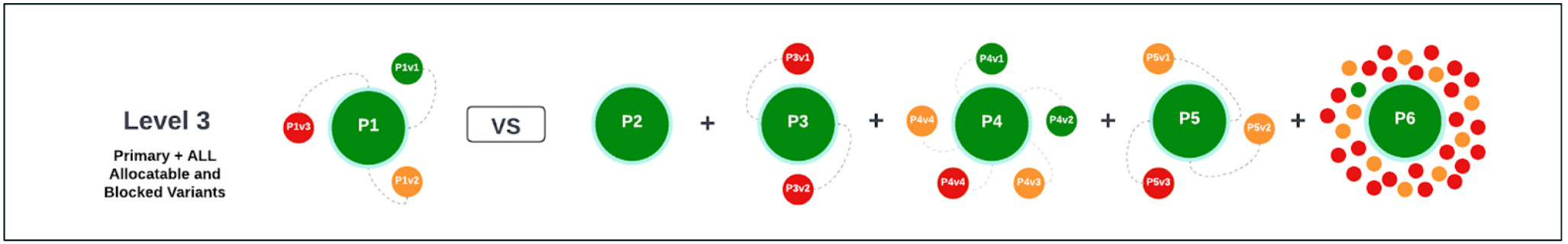
- **3 possible levels of comparison**
  - Level 1: Primary + only requested allocatable variants
  - Level 2: Primary + all allocatable variants
  - Level 3: Primary + all valid variants (blocked + allocatable**)**

  Staff Paper on variant management advocated for Level 3 – maximally conservative approach

Comparison Matrix – Consolidated View







EPDP on IDNs String Similarity Small Group

- Task 1: Develop concrete examples of strings that have blocked and/or allocatable variant labels and may be visually confusable with other strings in the same script or across scripts

- **Task 2: Demonstrate how these examples would be compared against each other in the string similarity review according to the three levels – showcase impact and potential consequences**

- Task 3: Demonstrate how these examples would undergo the objection process according to the three levels – showcase impact and potential consequences

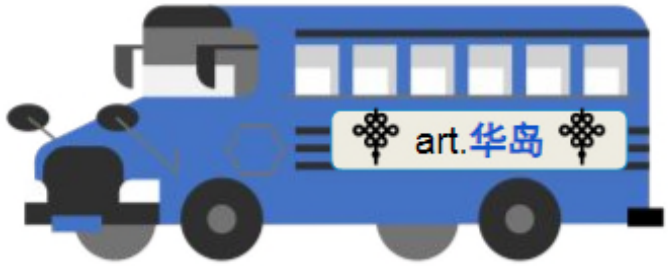
- **Exclusion:** Complexity implementation for Task 2 (and Task 3) is out of scope – defer to full EPDP Team.

Small Group Recommendation: Hybrid Model

- **A mixed-level approach between level 2 and level 3**
- Goal is to mitigate possibility of confusing similarity leading to two failure modes –
  - (i) Denial of Service (NOT DDOS!) and
  - (ii) Misconnection
- Considered
  - **RFC 5891**: Any domain name registry, including that of the root zone, should develop and apply **additional restrictions** as needed **to reduce confusion and other problems** (part of IDNA2008 standard)
  - **RFC 6921**: Zones higher in the DNS tree tend to have **more restrictive rules**...the context is that the root zone serves the entire Internet population
  - **SAC089**: **Confusability cannot be considered in isolation from other issues related to security**. **Phishing and other social engineering attacks based on domain name confusion are a security problem for end users**
  - **Staff Paper**: Variant implementation must be done in a way that operation and maintenance of the DNS not be adversely impacted by the introduction of variants; it should **avoid including variant TLDs in a manner that would create user vulnerabilities or a probability of confusion**

Denial of Service Example & Consequence

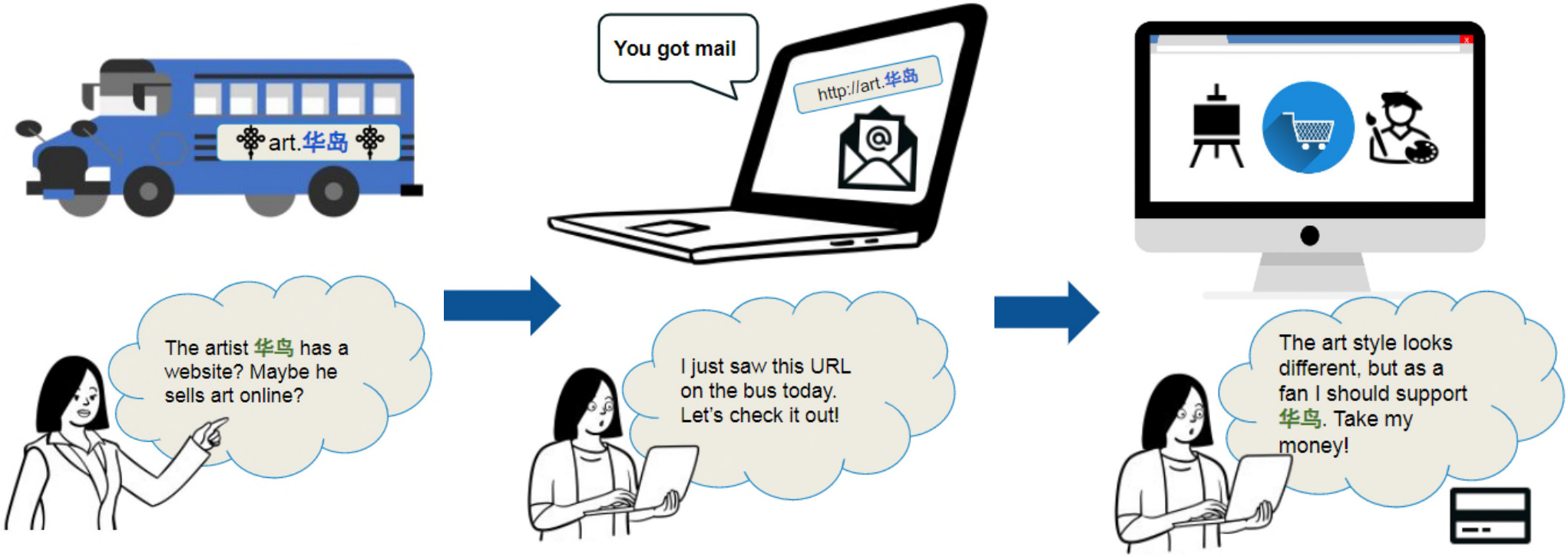
- A user attempts to visit http://example.X, reading it as being the same as the http://example.Y that, for example, he or she saw in an advertisement. After typing the address (http://example.X), the connection does not work as http://example.X is not registered.



Denial of service will likely cause user confusion and frustration but not harm

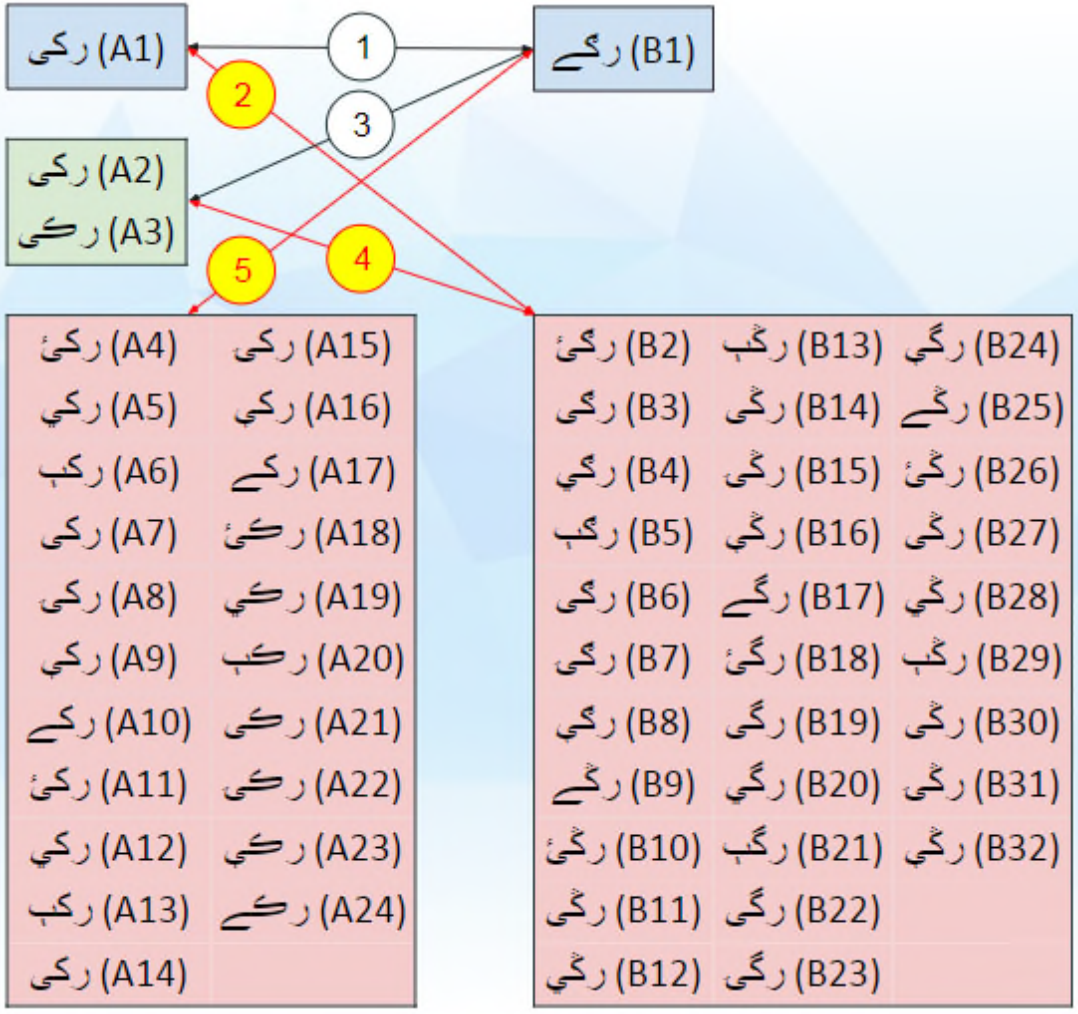
Misconnection Example & Potential Consequences

- A user attempts to visit http://example.X, reading it as being the same as the http://example.Y that, for example, he or she saw in an advertisement. After clicking on http://example.Y, the user arrives at a site controlled by a registrant different to http://example.X.



- Misconnection **may** be more problematic than denial of service, **cause more harm to end-user beyond confusion and frustration**
- Arriving at the wrong site, even if legitimate**, can result in credential compromise and accidental exposure of information**
- If confusing similarity is maliciously leveraged, it **can be a DNS abuse vector. When confusion is at the TL, the possibility of DNS abuse is much greater** than that at the SL

Example 6 – impact, potential consequences



String Similarity Review may find the following confusingly similar strings

2 ركى (A1) & رگى (B3) & رگى (B6)

4 ركى (A2) & رگى (B3) & رگى (B6)

4 ركے (A3) & رگى (B3) & رگى (B6)

5 رگے (B1) & ركے (A10) & ركے (A17) & ركے (A24)

Potential Outcome of the String Similarity Review

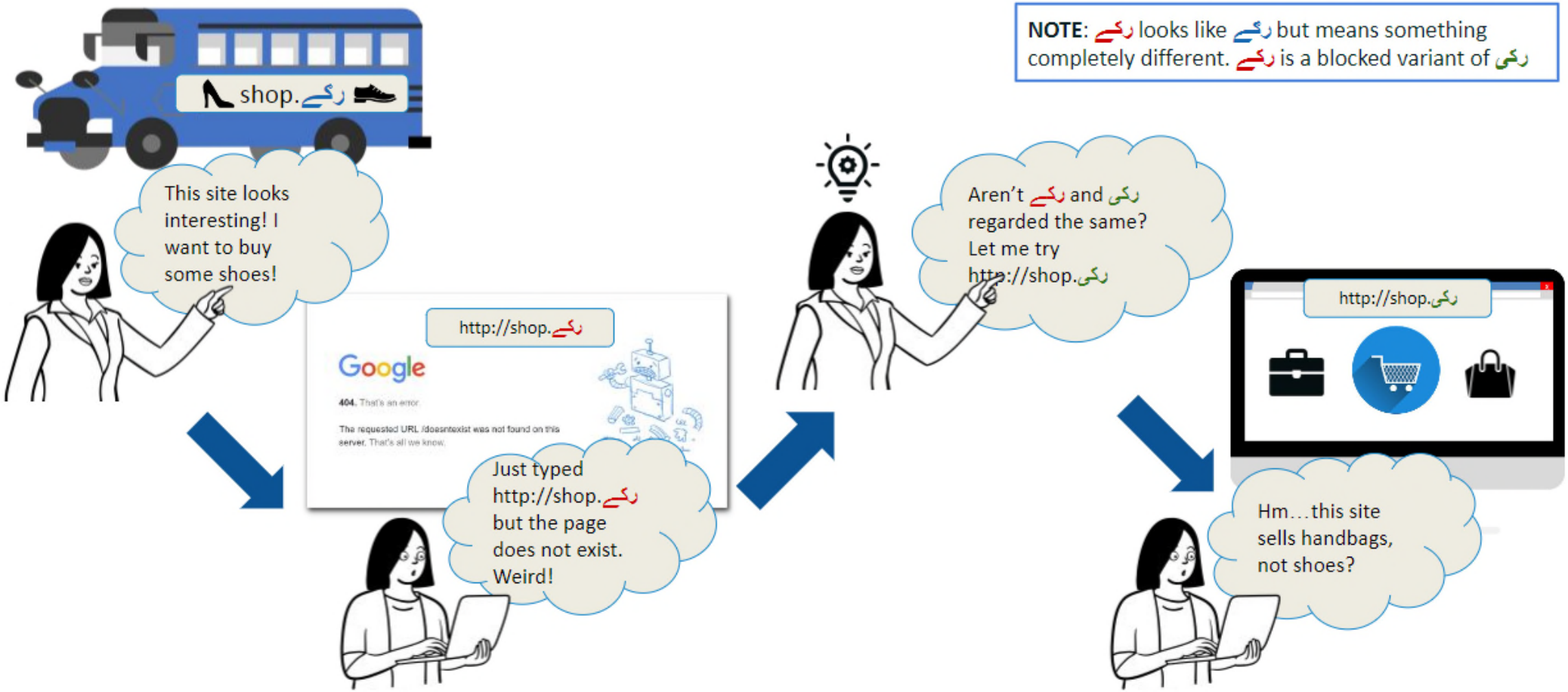
ركى (A1) & its variants A2-A24 AND رگے (B1) & its variants B2-B32 get processed in a contention set

If the hybrid model were not used and blocked variants were not taken into account in String Similarity Review

ركى (A1) and رگے (B1) would have been both delegated with the misconnection risk. E.g., a user may mistake ركى (A1) as رگى (B3), a blocked variant of رگے (B1), but arrive at site controlled by a registrant different to رگے (B1).

Misconnection Involving Blocked Variants

- A label may be a blocked variant label by RZ-LGR calculation, but end-users may still perceive and intend to access a blocked variant label domain name without knowing that it does not exist in the root



Small Group Recommendation: Hybrid Model

- **In practice, proposes to modify string similarity review …**

From existing	To add Levels 2+3 manifestation
1/ Compare an applied-for IDN gTLD Against: • Existing TLDs • Strings requested as IDN ccTLDs • Other applied-for gTLDs in the same round • Reserved Names • Any other 2-char ASCII strings	**1/** Compare an applied-for source IDN gTLD and all its allocatable variant label(s) Against: • Existing TLD and all their allocatable and blocked variant labels • Strings requested as IDN ccTLDs and all their allocatable and blocked variant labels • Other applied-for gTLDs in the same round and all their allocatable and blocked variant labels • Reserved Names; and • Any other 2-char ASCII strings and all their allocatable and blocked variant labels (if the applied-for source IDN gTLD is a 2-char string) **2/** Also compare all the blocked variant label(s) of an applied-for primary IDN gTLD Against: • Existing TLDs and all of their allocatable variant labels But do not compare an IDN TLD's blocked variant labels against blocked variants of another IDN TLD

Reactions to Hybrid Model

- As at early Oct 2022,
  - Nominating groups in EPDP on IDNs asked re: level of support
    - RySG – yes, with some refinement
    - NCSG – yes
    - GAC – yes
    - RrSG – yes, probably
  - Thus, need for risk analysis exercise possibly averted
- Clarifying questions?
- **STRAW POLL – Do you support the logic of the Hybrid Model as summarily explained?**
  - Noting that the ALAC Team may need to exercise discretion to consider refinements or other factors arising from EPDP deliberations eg. risk analysis, operational impact, complexity in implementation, cost & benefit of model

End

Thank you for your input.