

Examples of incorrectly processed messages
<b>Message:</b> \rOtO carr0.K0nprare/0ld/trashed/crashed up/k.a.r.s/\callus now555O5O5O5O <b>Tokens:</b> ['rOtO carr0', 'K0nprare', '0ld', 'trashed', 'crashed', 'up', 'k', 'a', 'r', 's', 'callus', 'now555O5O5O5O'] <b>CTA:</b> [none] <b>Problem:</b> Incorrect tokenization and no CTA detection
<b>Message:</b> Hi, I want to meet you tonight, spamdomain.com.Support me <b>Tokens:</b> ['Hi', 'I', 'want', 'to', 'meet', 'you', 'tonight', 'spamdomain.com.Support', 'me'] <b>CTA:</b> ['spamdomain.com.Support'] <b>Problem:</b> Incorrect URL detection (wrong TLD)
<b>Message:</b> Ive had a lot of fun tonight.tk <b>Tokens:</b> ['Ive', 'had', 'a', 'lot', 'of', 'fun', 'tonight.tk'] <b>CTA:</b> ['tonight.tk'] <b>Problem:</b> Incorrect URL detection (no URL)

Table 1: Examples of standard tokenization and CTA detection on obfuscated messages.