I can't recognize (yet):
Delayed Rendering to Defeat Visual Phishing Detectors

Ying Yuan[1,5], Cristiano Alex Rado[2], Giovanni Apruzzese[3,4], Mauro Conti[2,5], Luigi Vincenzo Mancini[1]

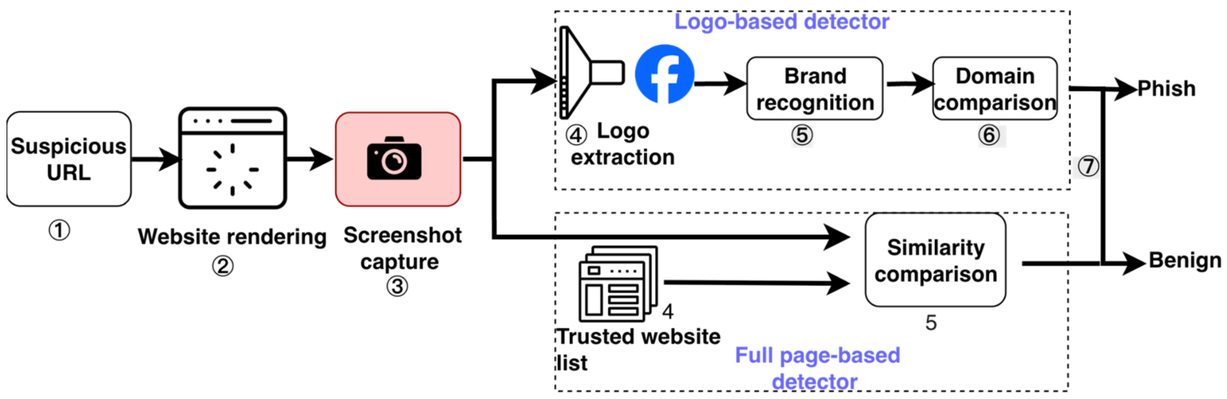
[1]Sapienza University of Rome, [2]University of Padua, [3]Reykjavik University, [4]University of Liechtenstein, [5]Örebro University

paper

attack

Motivation

Visual phishing detectors, which identify phishing websites based on the visual similarity of logos or whole website screenshots, have proven highly effective.



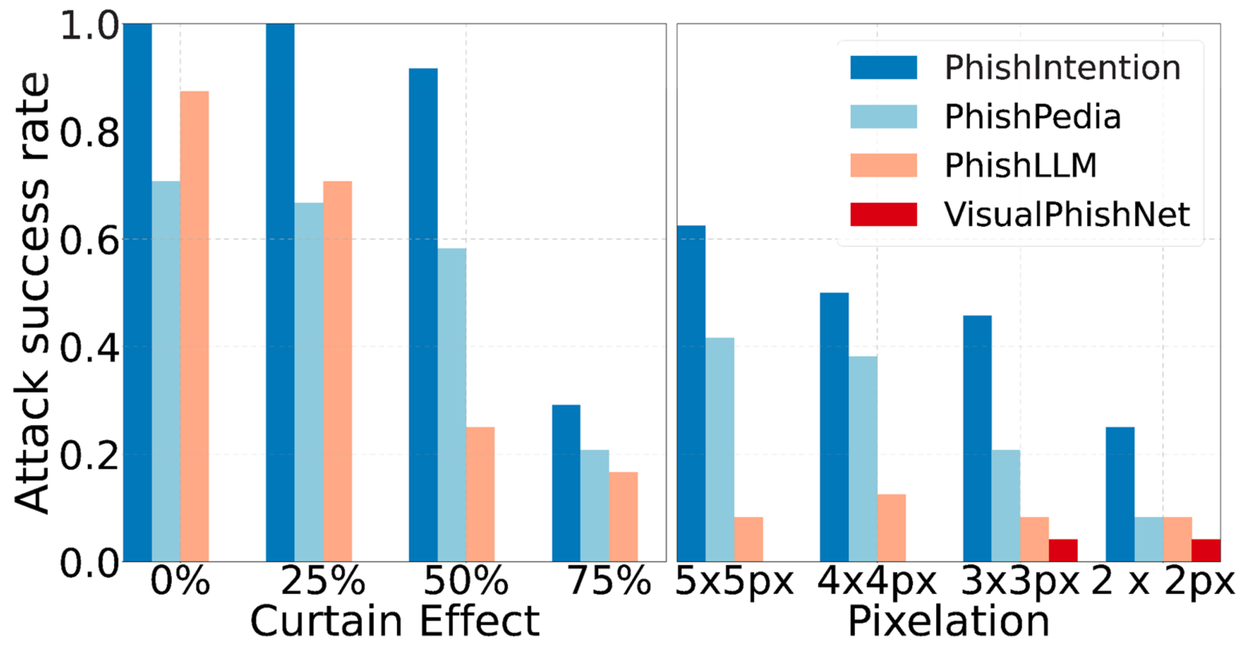
Hidden assumption: *webpage is fully rendered before screenshot capture*
Reality: *screenshot capture may happen before the webpage is fully rendered*

Research questions:

1. *How can delayed-rendering techniques be exploited to bypass visual phishing detectors?*
2. *What is the impact of delayed-rendering attacks on state-of-the-art visual phishing detectors?*
3. *Can humans spot 'rendering-based perturbations' and, if so, when?*
4. *What practical countermeasures can mitigate this attack?*

Impacts on SOTA detectors

We test 880 adversarial webpages against four SOTA visual phishing detectors: PhishIntention, PhishPedia, PhishLLM, and VisualPhishNet.



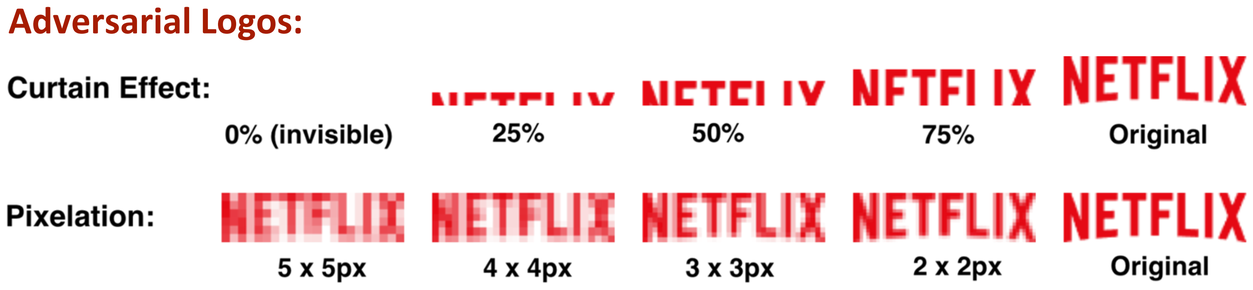
- curtain effect and pixelation attacks significantly degrade visual detectors (PhishIntention achieves 100% attack success rate).
- pixelation is statistically weaker than the curtain effect ($p < .05$).
- Full-page based detector (VisualPhishNet), is less affected.
- The most effective attack window occurs during the initial stage of page rendering.

Delayed-rendering Attacks

We implement two practical delayed-rendering attacks:

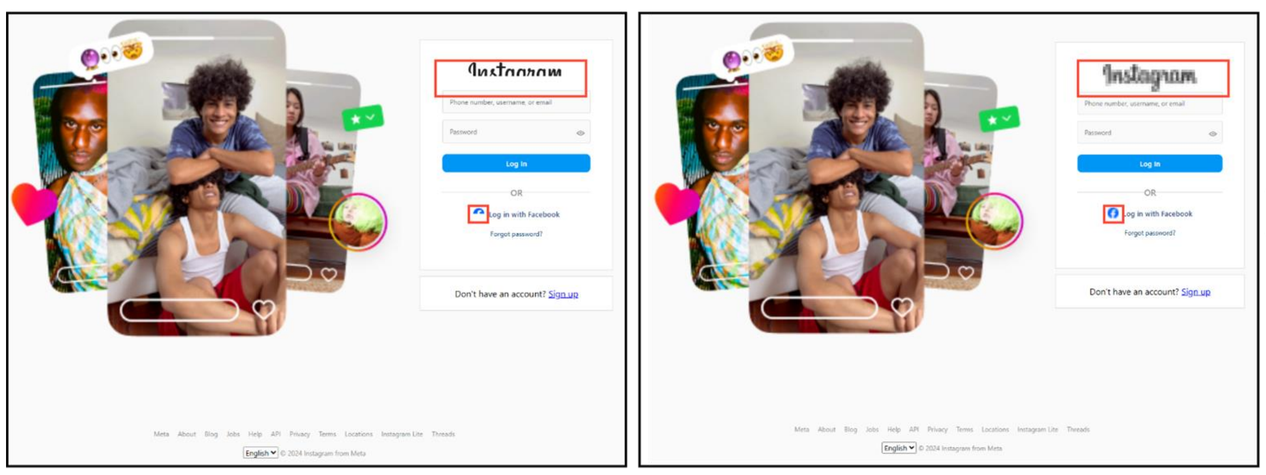
- Curtain effect: UI elements gradually appear from top to bottom, from 0% to 100% visibility.
- Pixelation: Images gradually transition from blurred to clear as the pixel blocks shrink from 5×5 to 1×1 pixels.

Adversarial Logos:



PhishMe.js automatically generates 60 attack variants by applying Curtain Effect, Pixelation, or both to the logo, background, or both, producing 880 adversarial webpages.

Adversarial webpages:



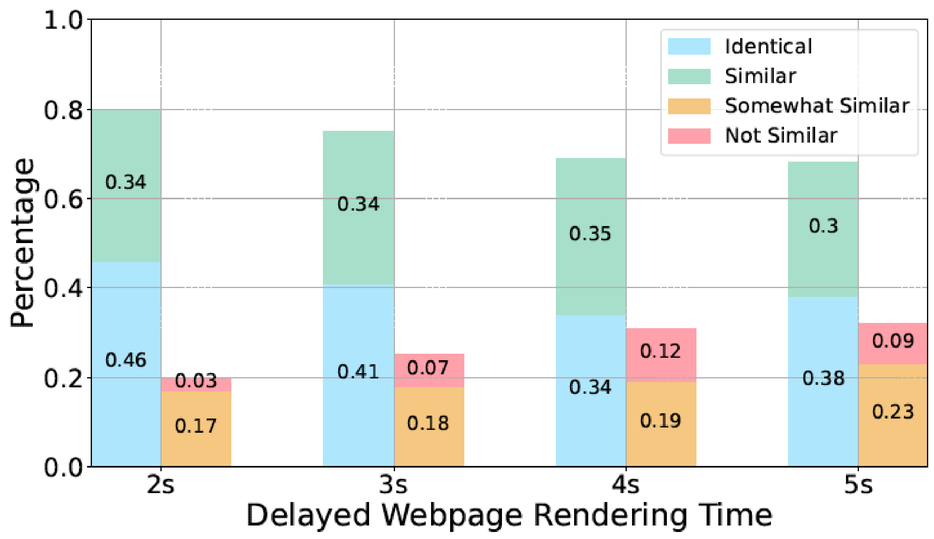
(a) Curtain Effect

(b) Pixelation

Human perception

We recruit 247 participants across two study settings:

- **Realistic**: watch pairs of videos (normal-speed vs. delayed rendering) and judge their similarity — without prior knowledge of phishing
- **Primed**: watch videos with rendering delays of 2-5 seconds and rate their trust — informed that phishing sites may change the loading process



- delayed-rendering attacks also evade human detection (around 73% of adversarial sites were rated identical/similar to normal-speed ones).
- Participants seem to notice attacks when the delay reaches 4 seconds.

Takeaway

- Delayed-rendering attacks in problem-space can completely defeat SOTA visual phishing detectors (ASR=100%).
- Users cannot reliably spot these attacks ($p < .05$), but notice something different when delay reaches 4 seconds.
- Our prototype browser extension flagged 23 out of 24 phishing sites.

EuroS&P 2026

Resources: https://github.com/joanyy/eurosp26_delayphish
Contact: ying.yuan@oru.se