



Video Injection Attacks

Andrew Newell

Chief Scientific Officer

iProov

10th May 2023

Overview

- Understanding pace of risk-level change is critical for policymakers
- What are video injection attacks?
 - And how they are distinct from presentation attacks
- Why do they matter?
 - Observations from the current threat landscape
- Mitigations against injection attacks
 - A high level view of approaches
- Considerations on assessing the effectiveness of defences

Policymakers' focus is on risk & proportionality

The threat from synthetic imagery is increasingly well understood by policymakers...



"In 2030, non-state actors like criminal groups, hackers-for-hire as well as government actors will likely have the technological capabilities (e.g., deepfakes) to expand their disinformation efforts in the EU to manipulate communities." ENISA, 20023



"Deepfakes and the misuse of synthetic content pose a clear, present, and evolving threat to the public across national security, law enforcement, financial, and societal domains." US Dept of Homeland Sec, 2022

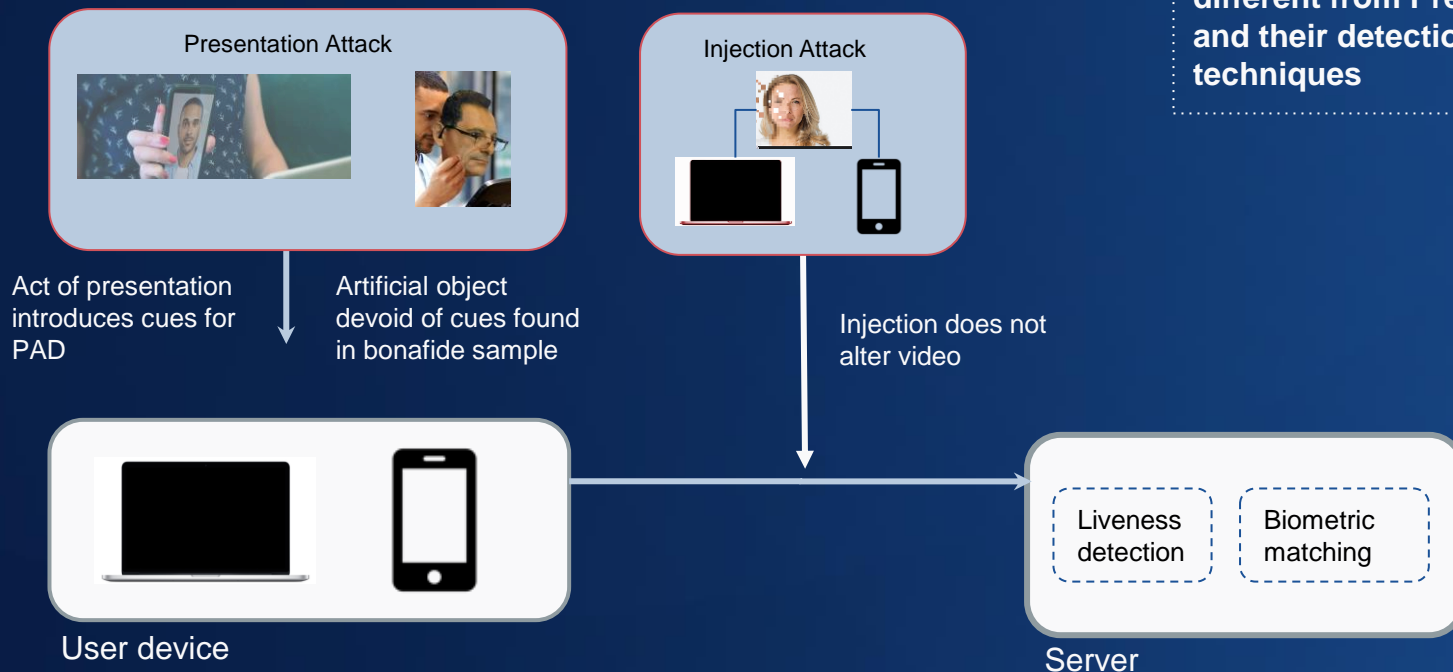
...and action is being taken

"European Digital Identity Wallets should ensure the highest level of security...taking into account the different levels of risk....Relying on the level of assurance "high", the European Digital Identity Wallets should benefit from the potential offered by tamperproof solutions..." (European Commission, eIDAS 2.0 proposal)

Proportionality requires a consideration of risk, systemic importance and threat complexity

Injection attacks vs presentation attacks

Injection attacks are fundamentally different from Presentation Attacks and their detection requires different techniques



Injection attacks: why they matter

Prevalence

- Injection attacks are a present threat (5x PA rate on web)
- They now present a threat to all platforms (149% increase H1->H2 2022 on mobile web, Android and iOS)
- Injection attacks the primary route for persistent threat actors

Evolution

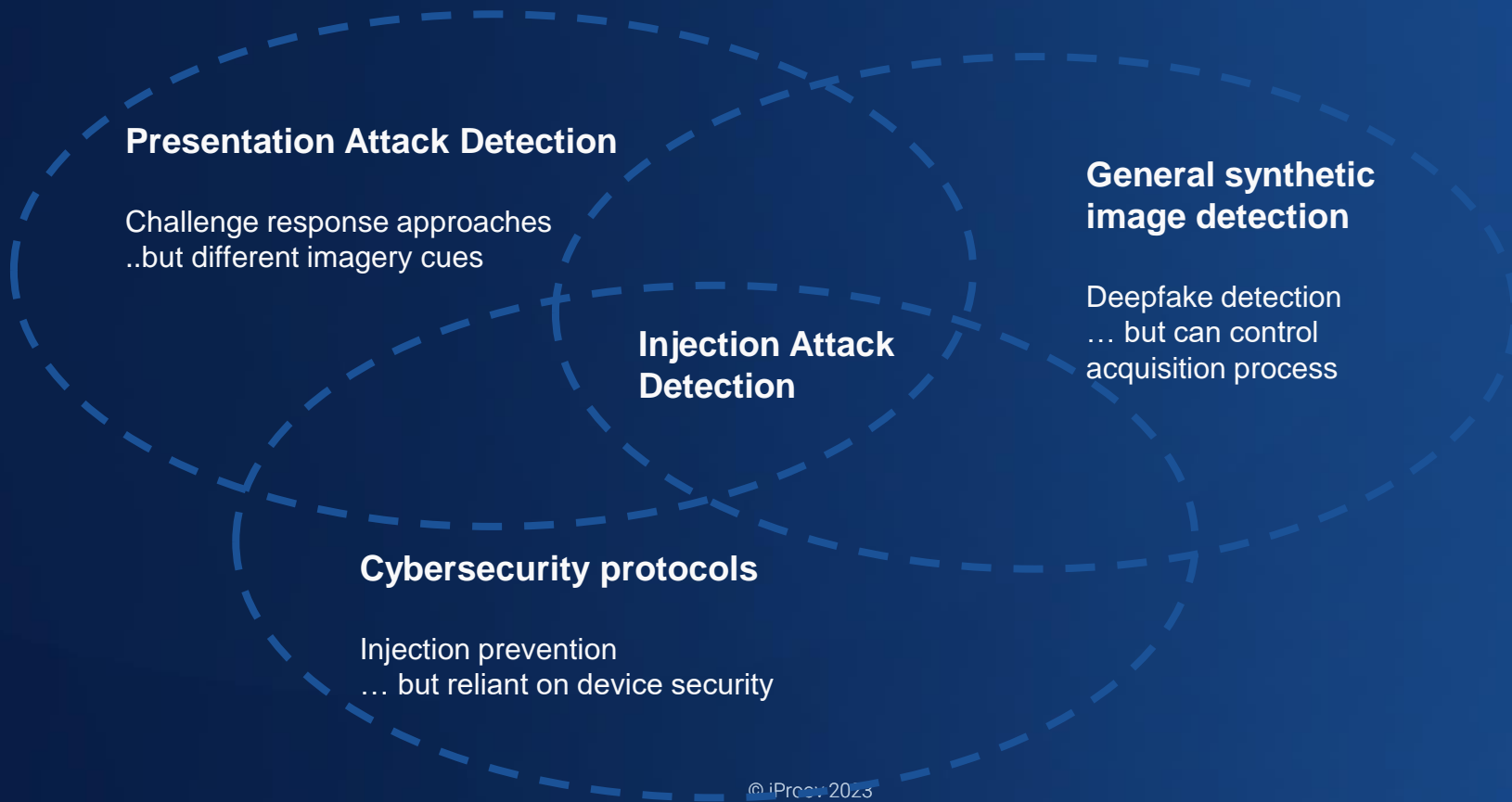
- Rapid of evolution of synthetic imagery methods (currently tracking >80 tools for faceswaps alone)
- Increased availability of injection and combined tools
- Example (295% increase in faceswap injection attacks H1->H2 2022)

Scalability

- Injection attacks can be launched by attack machines which can be fully automated
- Enables threat actors to explore areas of the threat landscape with minimal marginal cost per identity
- Current observation of bursts of IAs (00s or 000s) over short periods

Video injection attacks present a current threat which is highly scalable and evolving rapidly

Injection attack detection



Injection attack mitigations: high level approaches

Meta data approaches (non-biometric solution)

Detect whether an injection has occurred
Reliant on information that comes from the device
-> relies on obfuscation of the device code
-> can be perfectly forged

Significant advancement in meta data spoofing from threat actors in 2022 -> increasing number of emulators

Imagery-based approaches (biometric solution)

Determine whether the imagery comes from a bonafide user (regardless of whether an injection has occurred)
-> cannot be perfectly forged (anytime soon)
-> requires detection of synthetic imagery
-> aim is to make the imagery as hard to synthesise as possible and not repeatable (whilst ensuring system has high usability)

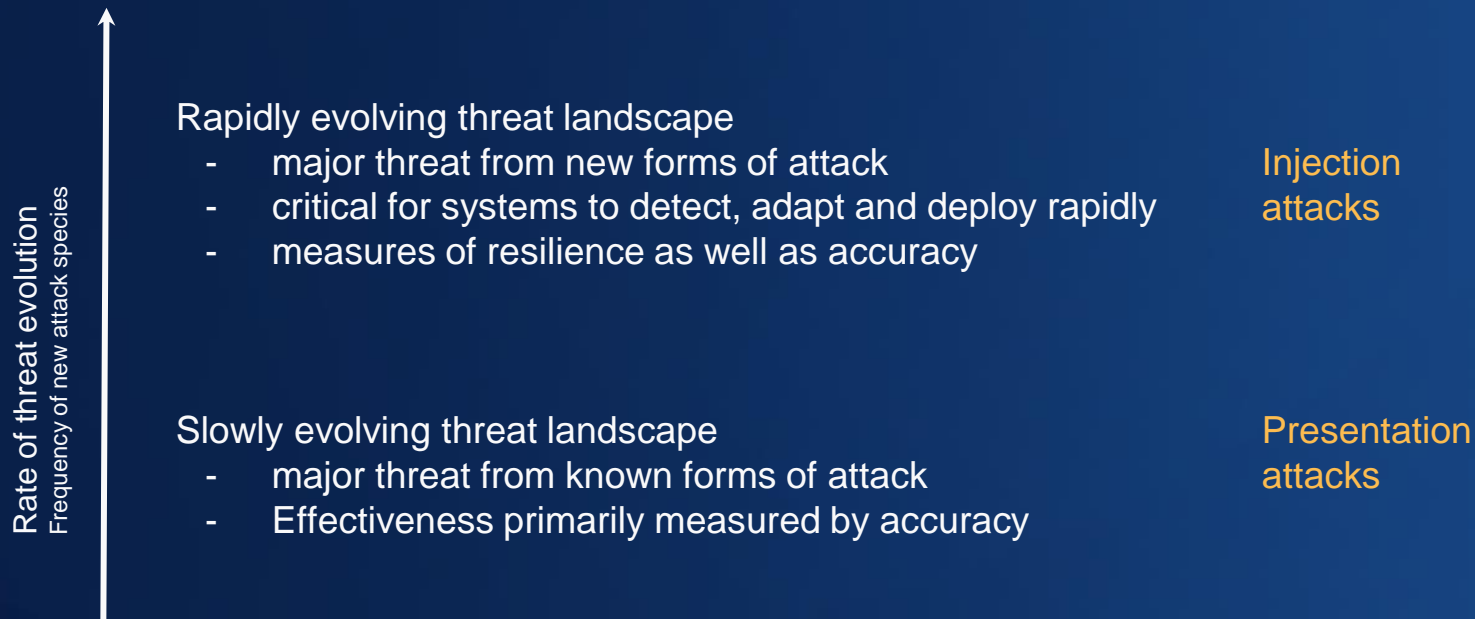
Evolution of image quality
Major challenge is generalisation

Both approaches can be used together - but this depends on the conceptual approach

Applicability of approaches

		No user action required	User action required
Repeated biometric signal	Repeated biometric signal	<p>E.g. single frame, passive video</p>  <p>Meta data approaches only</p>	<p>E.g. user blinking, head turning</p>  <p>Meta data approaches only (replay attack)</p>
	One-time biometric signal	<p>E.g. controlled illumination</p>  <p>Meta data approaches and Imagery-based approaches</p>	<p>E.g. user reading words, numbers, sequences of actions</p>  <p>Meta data approaches and Imagery-based approaches</p>

Injection attack mitigation: assessing effectiveness





Thank you