rtCaptcha: A Real-Time CAPTCHA Based Liveness Detection System

Erkam Uzun, Simon Pak Ho Chung, Irfan Essa and Wenke Lee
Department of Computer Science
Georgia Institute of Technology, USA

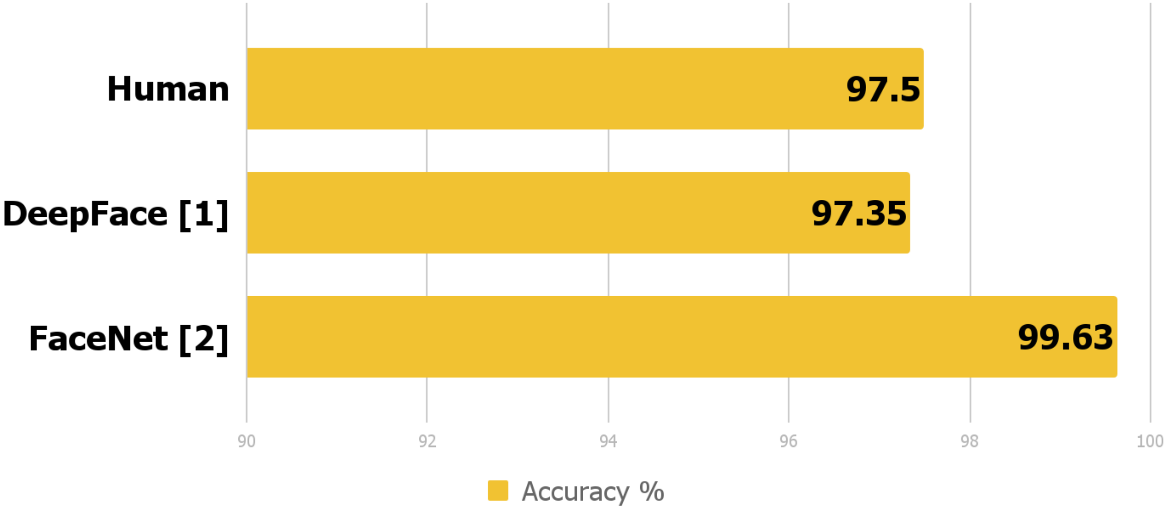
1

Face Authentication Systems

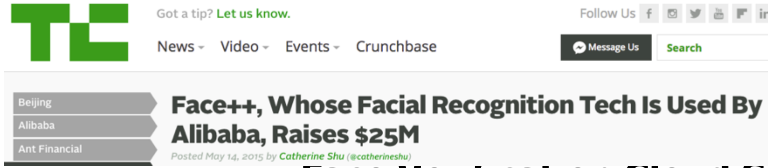
Background

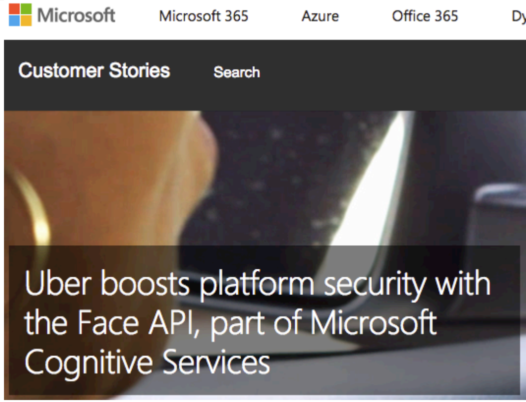
Deep Learning Outperforms

Face recognition performance on LFW dataset



Deployed by Major Companies

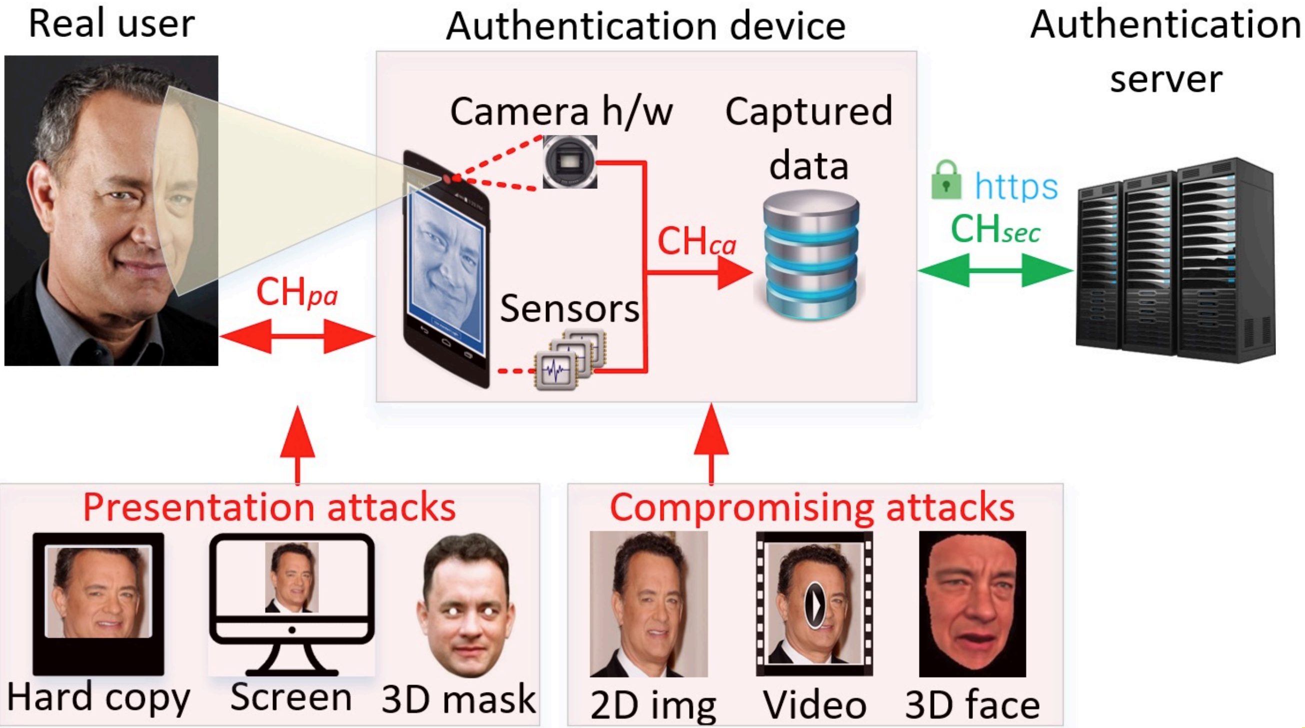




Face Verification Cloud Services

- Microsoft Cognitive Services [3]
- Amazon Rekognition [4]
- Face++ [5]
- Kairos Human Analytics [6]

Attack Channels of Biometric Authentication



Adversarial Models vs Defense Systems

Attack	Challenge	Defense
Still Image Attack	CH_{pa}	Blink, Smile… Detection
Video Replay Attack	CH_{pa}	Motion consist., Texture, Reflect. Analysis
3D Mask Attack	CH_{pa}	Extra H/W, e.g., IR, Depth Camera
3D Face Model, DL-Based Attacks	CH_{ca}	?

Threat Model

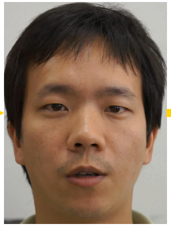
Automated compromising attacks.

- Camera, microphone and device kernel are compromised.
- No form of attestation.
- Known client-server protocol.
- State-of-the art synthesizers and Captcha breaking tools.
- Authentication server is NOT compromised.

Compromising Attack: Example-1

A malicious app, has access to cam., mic., etc.

Capture enough raw material, e.g., victim's face



3D Model Fitting [7]

- Fit face model on a 3DMM.
- Synthesize photorealistic facial texture.
- Transfer 3D face to a VR environment.
- Answer challenge at real-time.

Applied by Xu et al. "VirtualU" (Usenix'16)

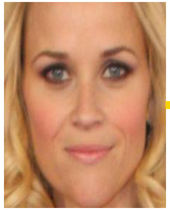
Compromising Attack: Example-2

A malicious app, has access to cam., mic., etc.

Capture enough raw material, e.g., victim's face

Victim

≠



Attacker

Sharif et al. "Accessorize to a Crime" (CCS'16) Fooled Face++ with 100%

Impersonation [8]
Create a perturber (e.g., eye glass)
Impersonate victim
Known Face Rec. NN!

Victim

=

Attacker

2

Security of Industry Leading Solutions (Face Authentication)

Do we need sophisticated attacks?

Security of Cloud Systems

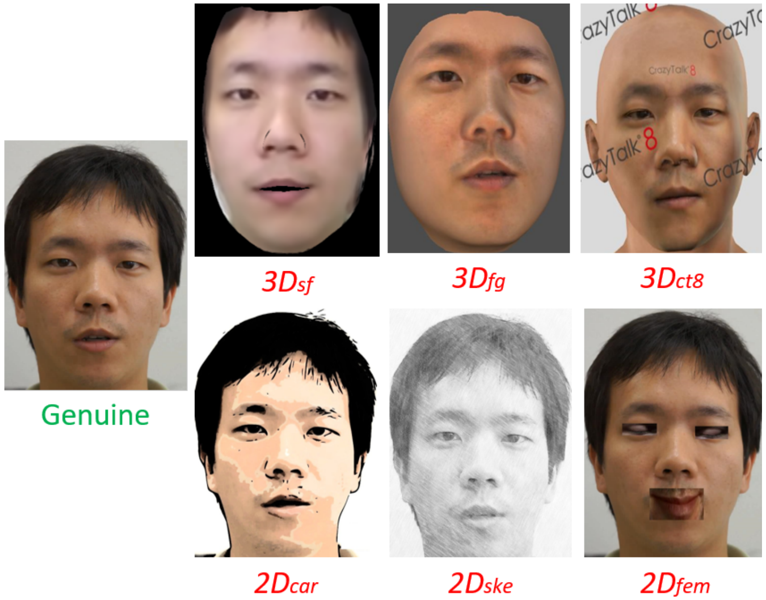
Face Verification Cloud Services

- Microsoft Cognitive Services
- Amazon Rekognition
- Face++
- Kairos Human Analytics

Database

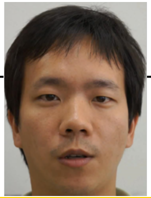
- First 10 subjects of CASIA Face Anti-Spoofing Database [9].
- Six attack images are generated for each subject.

Attack Vector



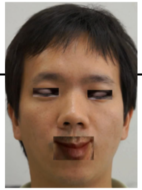
Security of Cloud Systems (cont'd)

Cognitive Service	**Baseline/Conf. (%)**		**Spoofed/Overall Confidence (%)**					
	TP	**TN**	$3D_{sf}$	$3D_{fg}$	$3D_{ct8}$	$2D_{car}$	$2D_{ske}$	$2D_{fem}$
MS Cognitive	100/78	100/65	100/70	100/75	100/70	100/82	100/84	100/86
Amazon	100/97	100/82	100/89	80/77	90/67	70/84	60/84	90/89
Face++	100/87	100/83	100/86	100/71	100/72	90/77	70/80	70/75
Kairos		80/58						





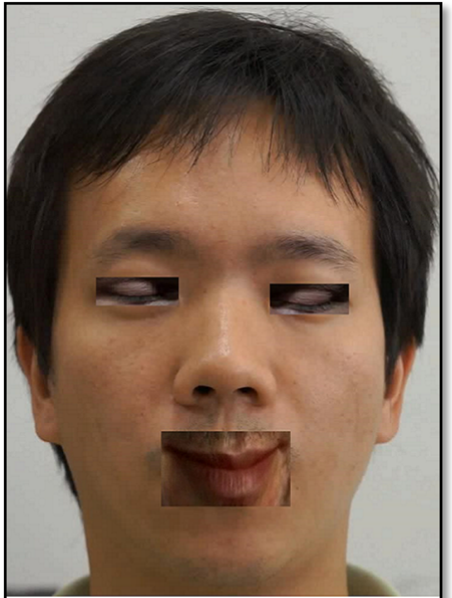




Security of Cloud Systems (cont'd)

Genuine. Smile:0.001

MS Cognitive Service



Fake. Smile:0.421

3

Security of Industry Leading Solutions (Speaker Authentication)

Do they also vulnerable to spoof?

Security of Cloud Systems (cont'd)

Speaker Verification Cloud Services

- Microsoft Cognitive Services

Database

- V_{dnn}^{1-7} : Contain 7 different DL-based synthesized version of genuine samples from two subjects, both female and male [10].
- V_{asv}^{1} to V_{asv}^{10} : Contain genuine samples and their voice converted (7) and synthesized (3) versions of randomly selected 8 subjects from ASV Spoofing Challenge database [11].

Methodology

- 30 seconds of genuine samples are enrolled for each subject. Hence, a group with 10 people in MS Cognitive Service is created.
- Randomly selected different samples for genuine and spoofed voices are tested.

Security of Cloud Systems (cont'd)

Test Sample	Detected as Original (%)	Test Sample	Detected as Original (%)	Test Sample	Detected as Original (%)
Original	**97.0**	V_{asv}^{4}	60.0	V_{asv}^{9}	71.3
V_{dnn}^{1-7}	100	V_{asv}^{5}	77.5	V_{asv}^{10}	91.3
V_{asv}^{1}	81.3	V_{asv}^{6}	77.5		
V_{asv}^{2}	28.8	V_{asv}^{7}	50.0		

2 Proposed System

Fundamental Problem of Existing Schemes

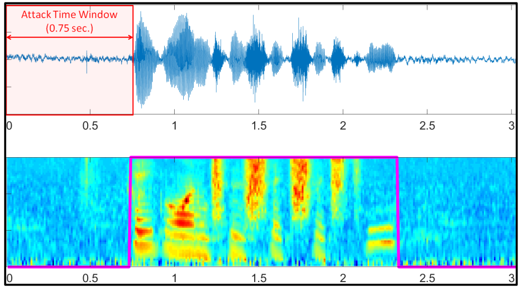
- Predictable challenges.
- Security relies on audio/face analysis, which has endless improvement in adversarial settings.

Real-Time Captcha (rtCaptcha)

- Randomized challenges.
- Security relies on an existing liveness detection mechanism.

System Overview

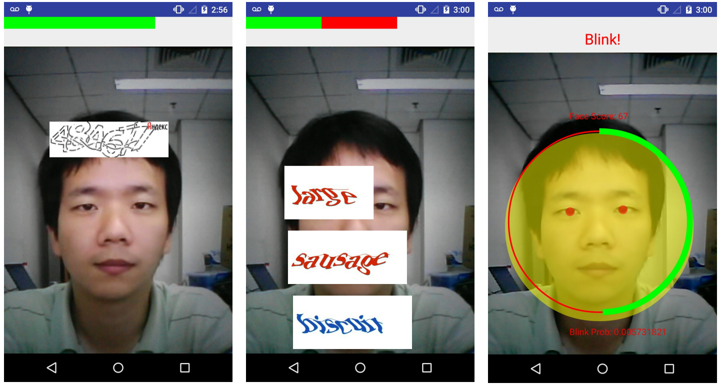
Authent. Request → Send Captcha Challenge → -Display Captcha -Get voice response -Grap face → If Captcha resp. match. $t_{resp} \leq t_{human}$ Face and voice verified → Verified



User Study

Challenges

- Plaintext – Numeric and Phrases
- Numeric Captchas – reCaptcha, Ebay, Yandex
- Animated Phrase Captchas – reCaptcha
- Blink/Smile



Challenge	Accuracy (%) (1 trial)	Accuracy (%) (2 trials)	Response Time (seconds)
Plain-text	90.3	100	0.77
Captcha	88.8	98.4	0.93
Smile/Blink	85.5	100	5.01

Captcha Breaking/Solving Attacks

Hum_{aud} **:** Users in our user study.

Atc_{typ} **:** Man-powered Captcha solving services [12].

Atc_{ocr} **:** OCR-based Captcha decoding services [13].

Atc_{best} **:** State-of-the art Captcha breaking tool [14].

Captcha Sample	Captcha Scheme	Recognition Accuracy (%)				Response Time (seconds)			
		Hum_{aud}	Atc_{typ}	Atc_{ocr}	Atc_{best}	Hum_{aud}	Atc_{typ}	Atc_{ocr}	Atc_{best}
 	$reCaptcha_{numeric}$	**87.1**	96.7	0	77.2	**0.90**	22.11	2.98	10.27
	$Ebay_{numeric}$	**94.1**	100	0	58.8	**0.73**	12.33	2.79	5.98
	$Yandex_{numeri}$	**87.7**	96.7	0	2.2	**0.89**	15.05	3.30	15.50

Conclusions

- Smile/blink etc. detection is weak against spoofing.
- rtCaptcha: Audio/image analysis → CAPTCHA
- rtCaptcha: Very limited time to;
  - Break Captcha
  - Synthesize voice/face of the victim.
- Limitation: rtCaptcha needs audible response, which could NOT be usable in certain environments.

References

[1] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification.” *IEEE CVPR*. 2014.

[2] Schroff, Florian, et al. "Facenet: A unified embedding for face recognition and clustering.” *IEEE CVPR*. 2015.

[3] https://azure.microsoft.com/en-us/services/cognitive-services/

[4] http://ws.amazon.com/rekognition

[5] https://www.faceplusplus.com/

[6] http://kairos.com/

[7] Jackson, Aaron S., et al. "Large pose 3D face reconstruction from a single image via direct volumetric CNN regression." *IEEE ICCV.* 2017.

[8] Sharif, Mahmood, et al. "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition." *ACM CCS.* 2016.

[9] Zhang, Zhiwei, et al. "A face antispoofing database with diverse attacks." *IEEE ICB*. 2012.

[10] Wu, Zhizheng, et al. "A study of speaker adaptation for DNN-based speech synthesis." *INTERSPEECH*. 2015.

[11] Wu, Zhizheng, et al. "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge." *INTERSPEECH*. 2015.

[12] https://anti-captcha.com/

[13] http://www.captchatronix.com/

[14] Gao, Haichang, et al. "A Simple Generic Attack on Text Captchas." *NDSS*. 2016.

Thanks!

Any questions ?