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

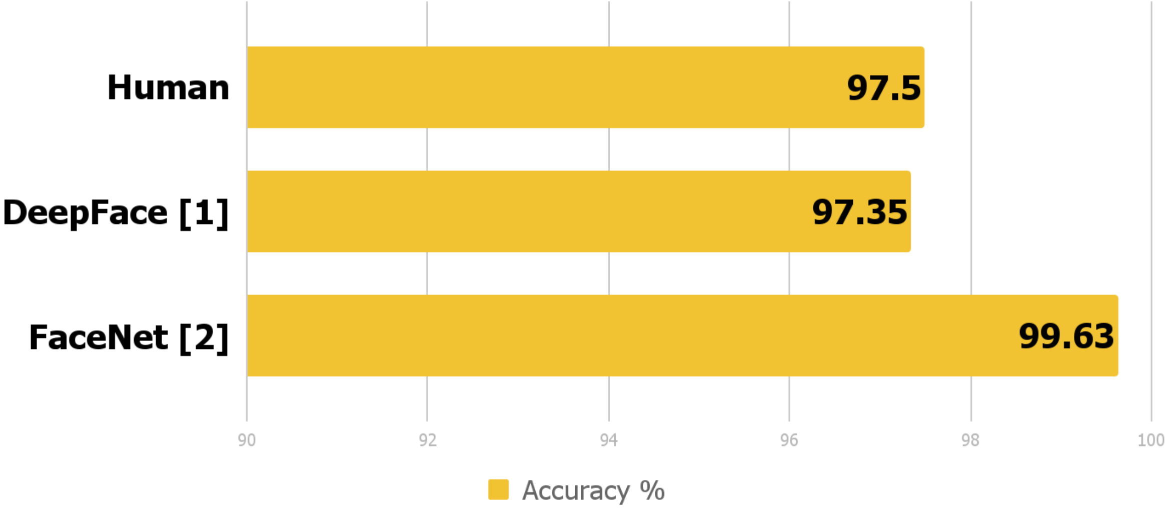
1

Face Authentication Systems

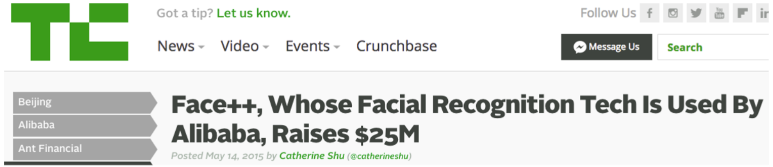
Background

Deep Learning Outperforms

Face recognition performance on LFW dataset

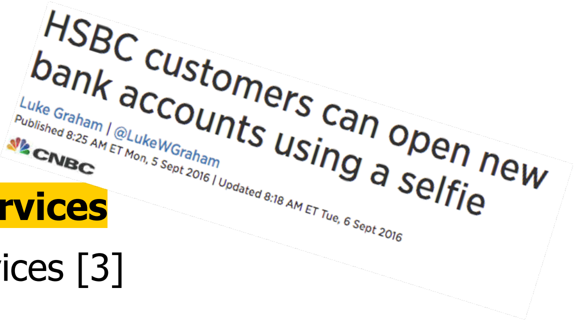


Deployed by Major Companies



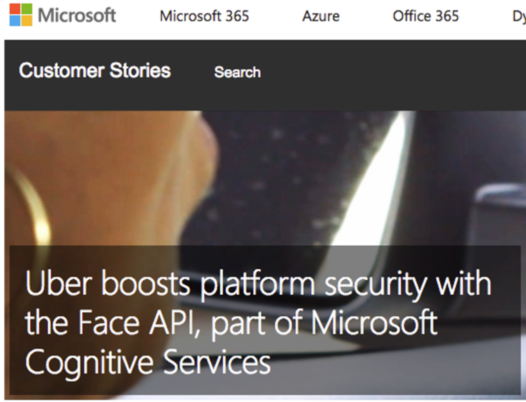
Face++, Whose Facial Recognition Tech Is Used By Alibaba, Raises $25M

Posted May 14, 2015 by Catherine Shu (@catherineshu)


HSBC customers can open new bank accounts using a selfie

Luke Graham | @LukeWGraham

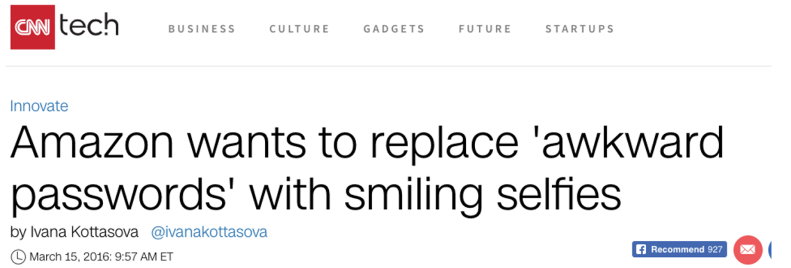
Published 8:25 AM ET Mon, 5 Sept 2016 | Updated 8:18 AM ET Tue, 6 Sept 2016



Uber boosts platform security with the Face API, part of Microsoft Cognitive Services

Face Verification Cloud Services

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

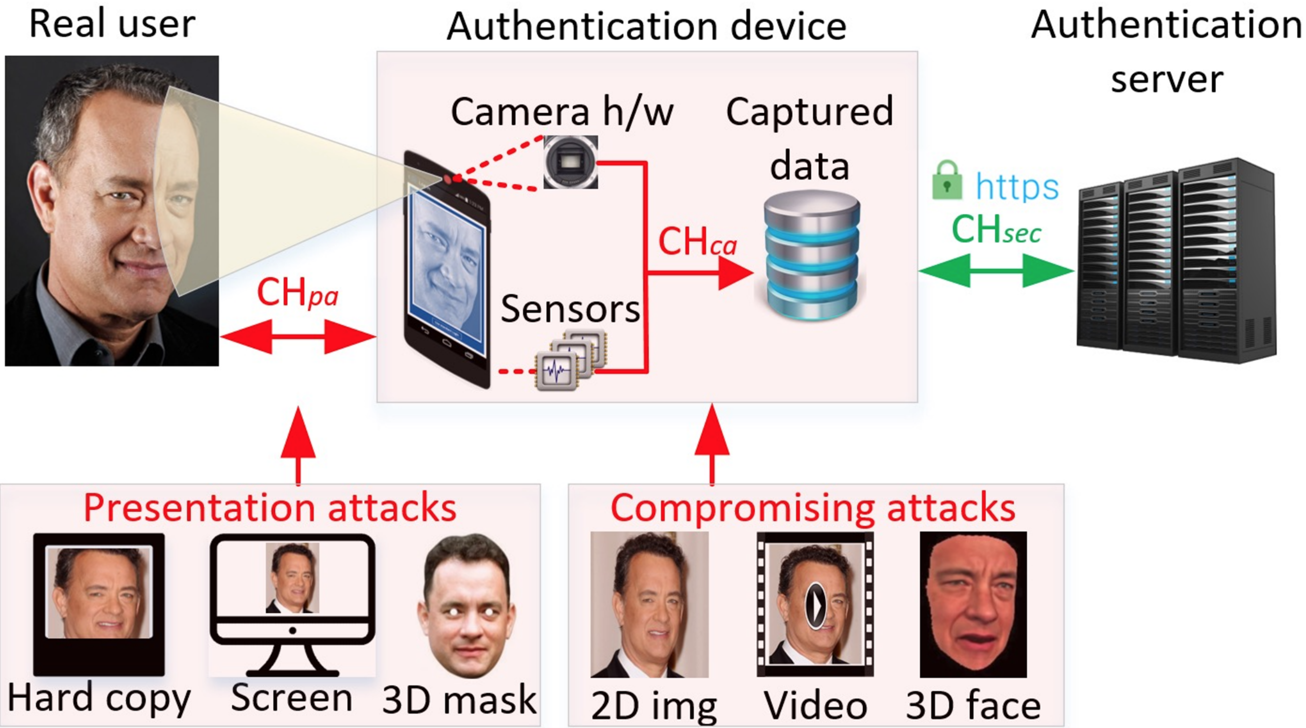


Amazon wants to replace 'awkward passwords' with smiling selfies

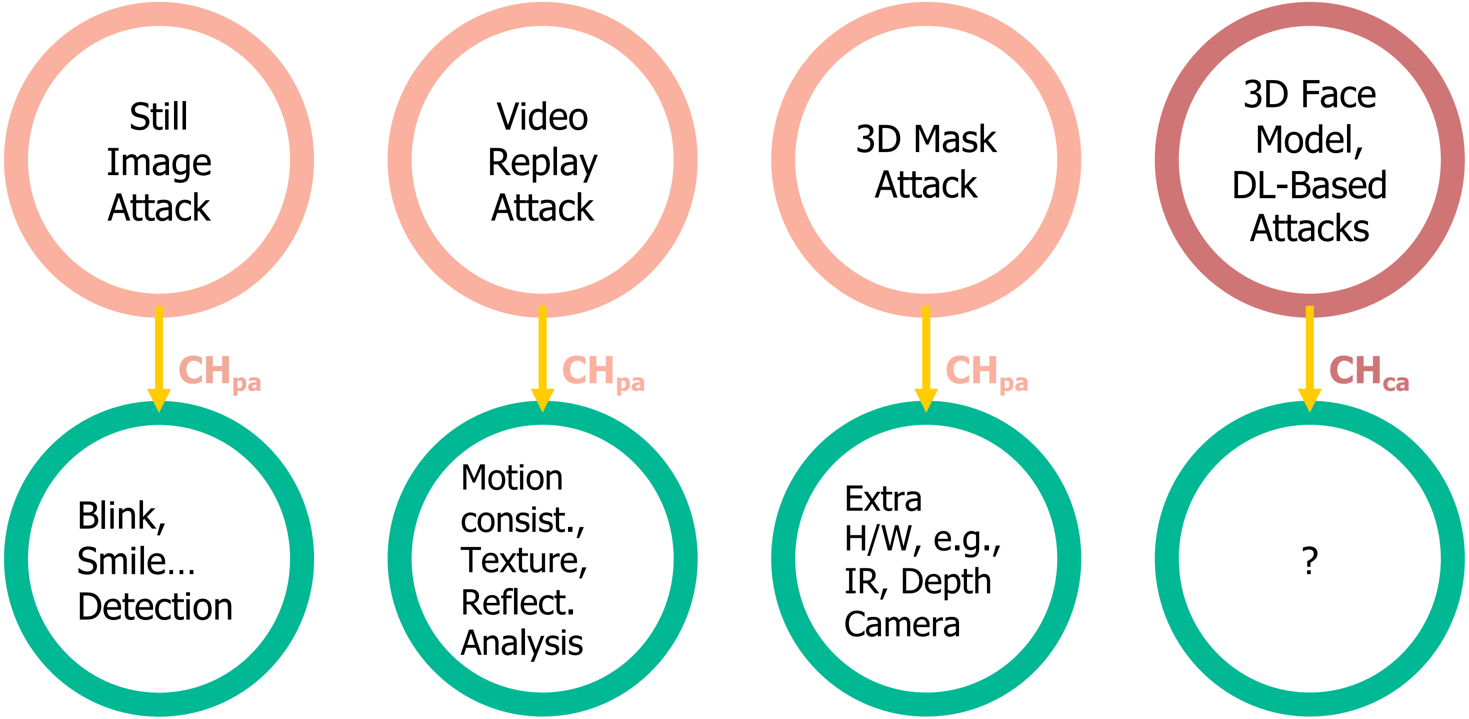
by Ivana Kottasova @ivanakottasova

March 15, 2016: 9:57 AM ET

Attack Channels of Biometric Authentication



Adversarial Models vs Defense Systems



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.

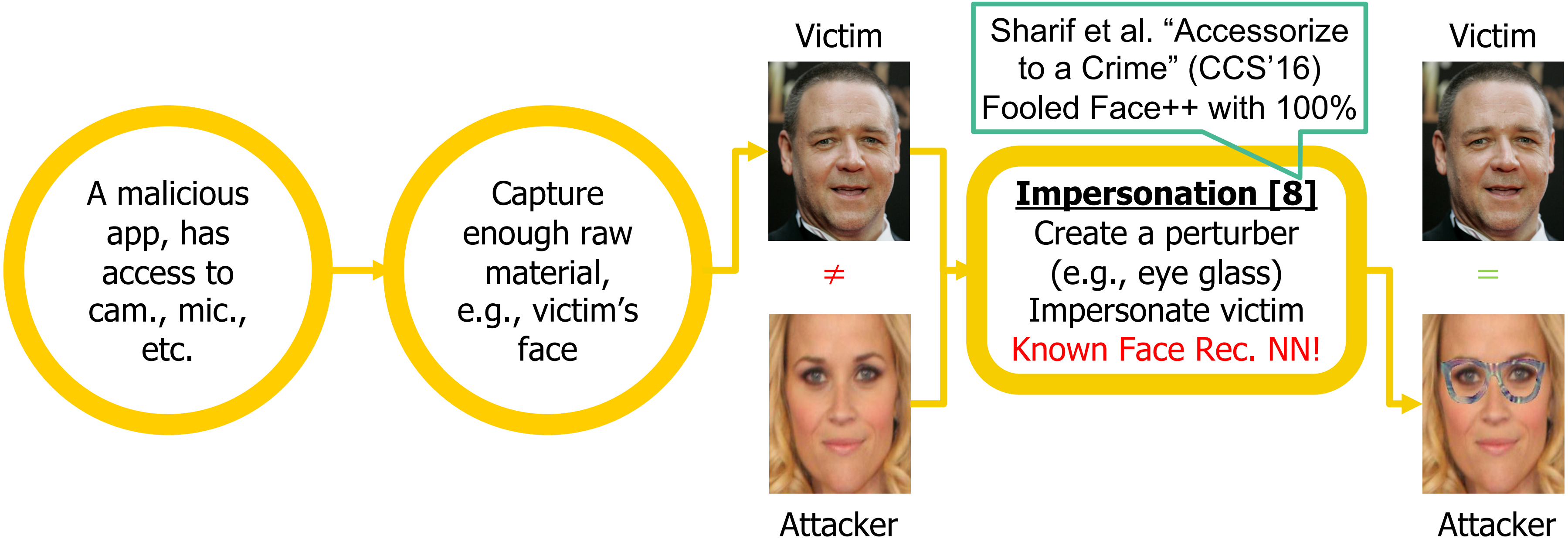
Capture enough raw material, e.g., victim's face

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

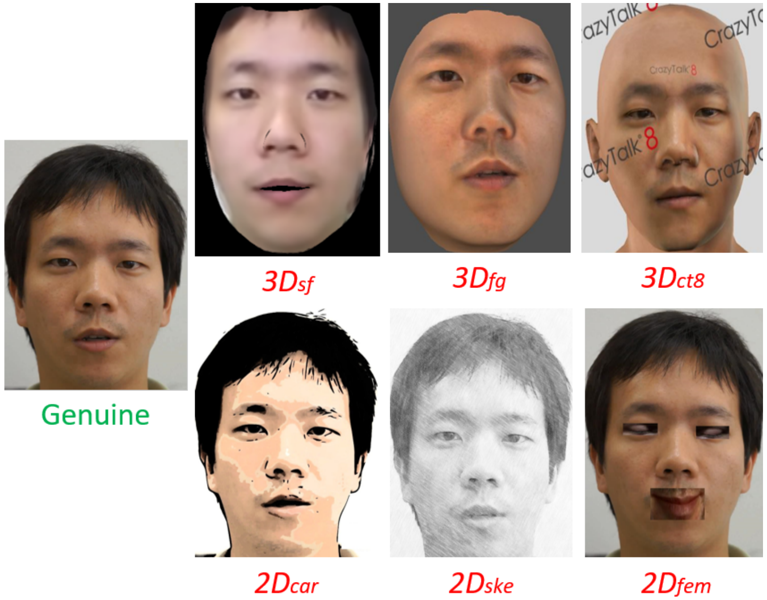
Face Verification Cloud Services

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

Database

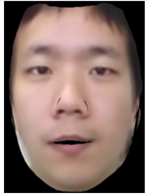
- First 10 subjects of CASIA Face Anti-Spoofing Database [9].
- Six attack images are generated for each subject.

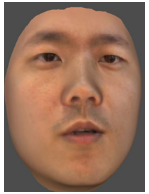
Attack Vector

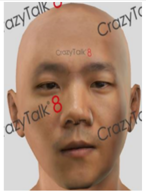


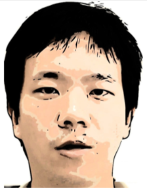
Security of Cloud Systems (cont'd)

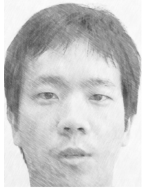
Cognitive Service	Baseline/Conf. (%)		Spoofed/Overall Confidence (%)					
	TP	TN	$3D_{sf}$	$3D_{fg}$	$3D_{ct8}$	$2D_{car}$	$2D_{ske}$	$2D_{fem}$
MS Cognitive	100/78	100/65	100/70	100/75	100/70	100/82	100/84	100/86
Amazon	100/97	100/82	100/89	80/77	90/67	70/84	60/84	90/89
Face++	100/87	100/83	100/86	100/71	100/72	90/77	70/80	70/75
Kairos	100/80	80/58	100/75	100/78	100/73	100/91	100/83	100/80

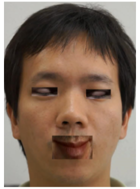




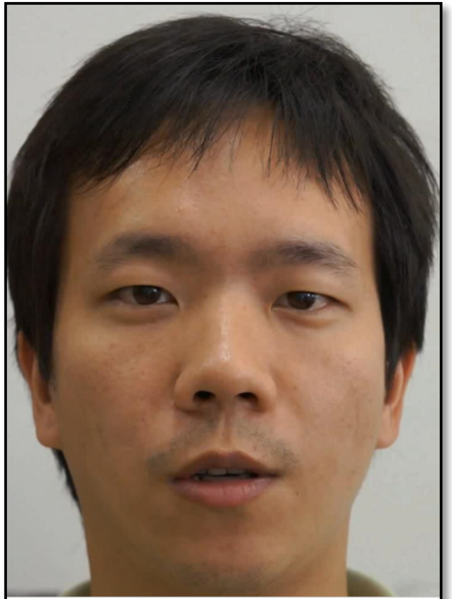






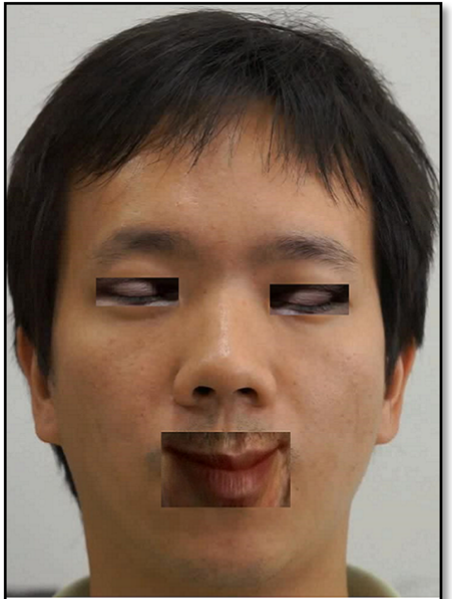


Security of Cloud Systems (cont'd)



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		
V_{asv}^{3}	47.5	V_{asv}^{8}	33.8		

2 Proposed System

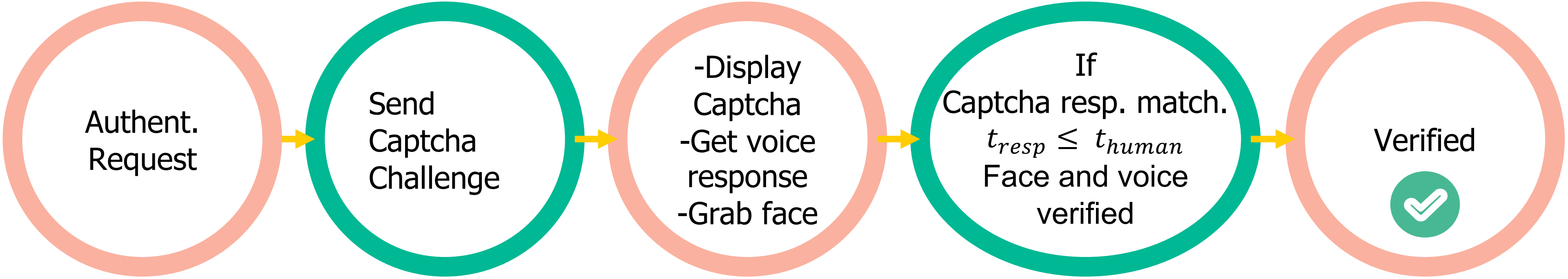
Fundamental Problem of Existing Schemes

- 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.
- Captcha provides two types of randomness: 1) Challenge semantic, 2) Captcha scheme

System Overview



Authent. Request

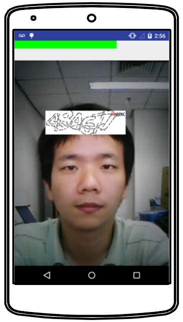
Send Captcha Challenge

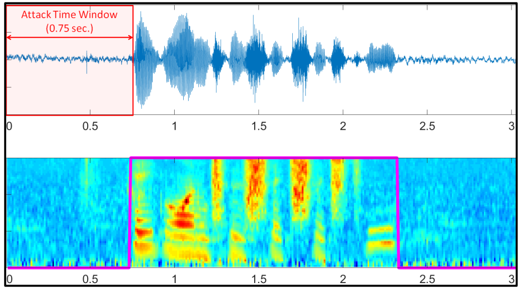
-Display Captcha
-Get voice response
-Grab face

If Captcha resp. match.
$t_{resp} \le t_{human}$
Face and voice verified

Verified ✓

qDm3jNq
dBpyHyz6
NOAH
ebyspan
19 7161

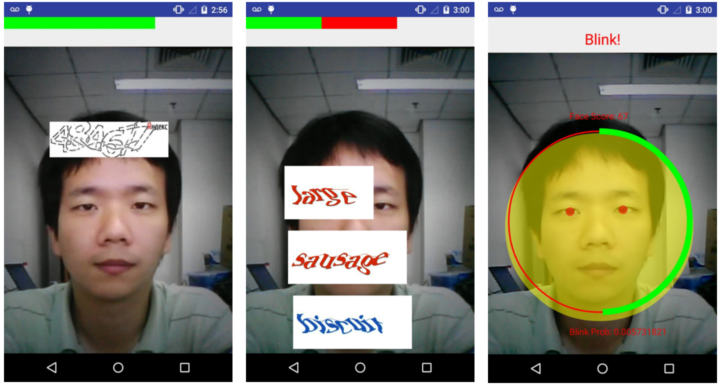




User Study

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

$\mathbf{Hum_{aud}}$: Users in our user study.

$\mathbf{Atc_{typ}}$: Man-powered Captcha solving services [12].

$\mathbf{Atc_{ocr}}$: OCR-based Captcha decoding services [13].

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

Captcha Sample	Captcha Scheme	Recognition Accuracy (%)				Response Time (seconds)			
		$\mathbf{Hum_{aud}}$	$\mathrm{Atc_{typ}}$	$\mathrm{Atc_{ocr}}$	$\mathrm{Atc_{best}}$	$\mathbf{Hum_{aud}}$	$\mathrm{Atc_{typ}}$	$\mathrm{Atc_{ocr}}$	$\mathrm{Atc_{best}}$
	$\mathbf{reCaptcha_{numeric}}$	**87.1**	96.7	0	77.2	**0.90**	22.11	2.98	10.27
	$\mathbf{Ebay_{numeric}}$	**94.1**	100	0	58.8	**0.73**	12.33	2.79	5.98
	$\mathbf{Yandex_{numeric}}$	**87.7**	96.7	0	2.2	**0.89**	15.05	3.30	15.50
	$\mathbf{reCaptcha_{phrase}}$	**88.0**	91.5	0	N/A	**1.02**	20.88	3.03	N/A

Captcha Breaking/Solving Attacks

$\mathbf{Atc_{typ}}$: Man-powered Captcha solving services.

Reported Avg. Accuracy (%) and Response Time (sec.) of Man-Powered Captcha Solving Services					
Service	**Acc.(%)**	**Time**	**Service**	**Acc.(%)**	**Time**
anti-captcha	99.0	7	2captcha	96.6	10
captchaboss	99.9	8	imagetyperz	99.0	12
deathbycaptcha	95.8	10	9kw.eu	N/A	30

Captcha Breaking/Solving Attacks

$\mathrm{Atc}_{\mathrm{best}}$: ML-Based Captcha Breaking Tools.

Captcha Scheme	Gao et al. [14]		Burzstein et al. [15]		Captcha Scheme	Gao et al. [14]		Burzstein et al. [15]	
	Acc.(%)	Time(s)	Acc.(%)	Time(s)		Acc.(%)	Time(s)	Acc.(%)	Time(s)
reCaptcha (old)	7.8	8.06	21.74	7.16	**Microsoft**	16.2	12.59	N/A	N/A
reCaptcha	77.2	10.27	19.22	4.59	**Amazon**	25.8	13.18	N/A	N/A
Yahoo!	5	28.56	3.67	7.95	**Taobao**	23.4	4.64	N/A	N/A
Baidu	44.2	2.81	54.38	1.9	**Sina**	9.4	4.83	N/A	N/A
Wikipedia	23.8	3.74	28.29	N/A	**Ebay**	58.8	5.98	47.92	2.31
QQ	56	4.95	N/A	N/A	**Yandex**	2.2	15.5	N/A	N/A

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.

Future Work

- rtCaptcha is only a part of a bigger umbrella project to make facial recognition based authentication **both usable and secure**.
- To ease adoption, we've also implemented the OpenID Connect protocol to make our face-based authentication a single sign on service.
- Currently working on the **privacy** issue of biometrics-based authentication: you want to log in with your biometrics, but you don't want the server to know what you look/sound like.

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.

[15] Bursztein, Elie, et al. "The End is Nigh: Generic Solving of Text-based CAPTCHAs." WOOT. 2014.

Thanks!

Any questions ?