BlockShield: A TPM-Integrated Blockchainbased Framework for Shielding Against Deepfakes

VLSI-SoC 2024 Special Session: Security-by-Design (SbD)

Venkata K. Vishnu. V. Bathalapalli¹, A. Kumar², S. Mohanty³, E. Kougianos⁴, Venkata P. Yanambaka⁵

University of North Texas, Denton, TX, USA.^{1,2,3,4} and Texas Woman's University⁵.

Email: vb0194@unt.edu¹, aakarshankumar@my.unt.edu², saraju.mohanty@unt.edu³, elias.kougianos@unt.edu⁴, vyanambaka@twu.



Outline

- Introduction to Deepfake Techniques
- Deepfake Mitigation
- Introduction to BlockShield
- TPM-Video Attestation
- Experimental Validation
- Conclusion & Future Research Directions



Deepfake

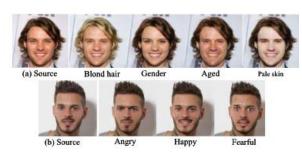


Al can be fooled by fake data



Al can create fake data (Deepfake)

Attribute Manipulation



Identity Swapping



Target image

Smart Electroni

Laboratory (S

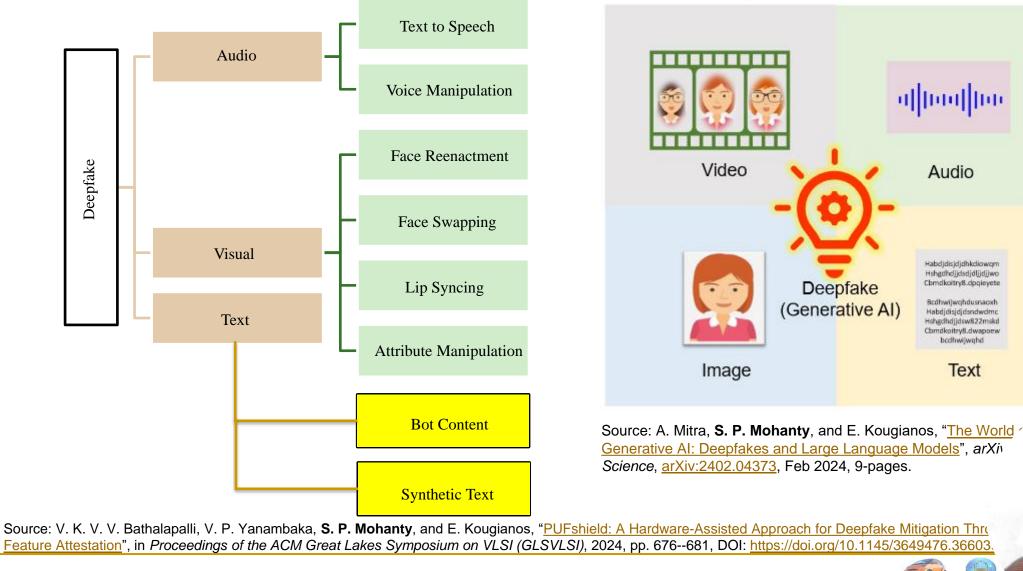
UNT DEPARTM

- Deepfake refers to super realistic, but fake images, sounds, 1. and videos generated by machine learning methods.
- Deepfake leverages a Generative adversarial network (GAN) 2. which enables the modification of human faces in a video or image.
- Deepfakes can be classified as Audio, Visual and Text 3.



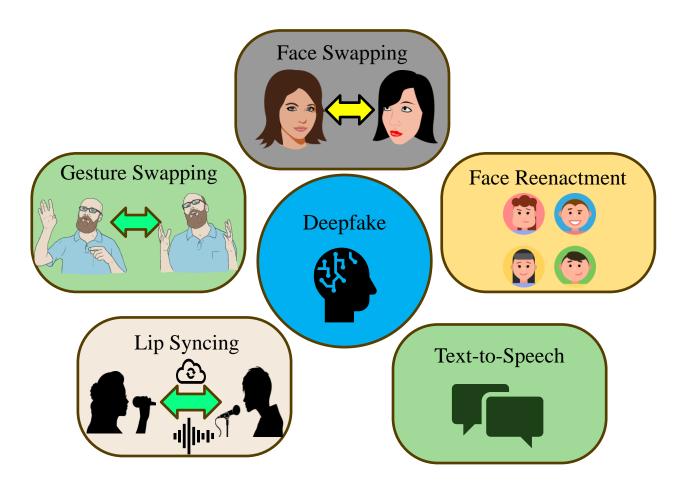
Source: A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Acc 18757-18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

Deepfake Techniques





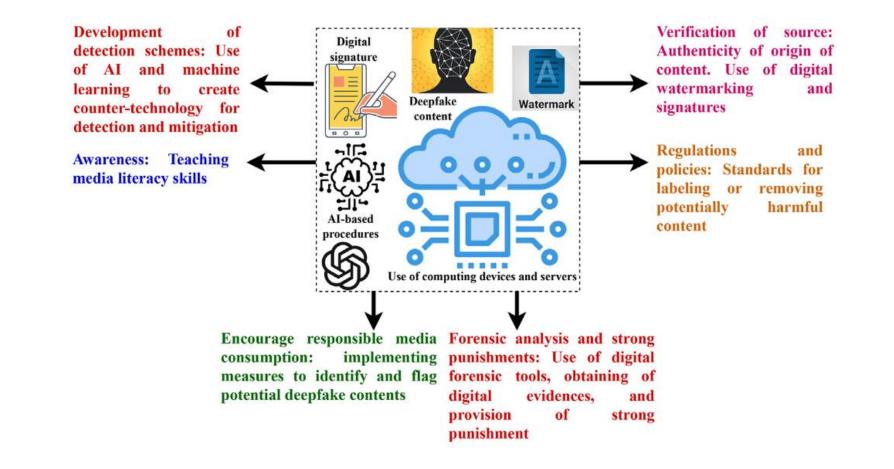
Visual Deepfake Techniques



Source: V. K. V. V. Bathalapalli, V. P. Yanambaka, **S. P. Mohanty**, and E. Kougianos, "<u>PUFshield: A Hardware-Assisted Approach for Deepfake Mitigation Through PUI</u> <u>Attestation</u>", in *Proceedings of the ACM Great Lakes Symposium on VLSI (GLSVLSI)*, 2024, pp. 676--681, DOI: <u>https://doi.org/10.1145/3649476.3660394</u>.



Deepfake Mitigation



Source: Wazid, M., Mishra, A. K., Mohd, N., & Das, A. K. (2024). A Secure Deepfake Mitigation Framework: Architecture, Issues, Challen Impact. *Cyber Security and Applications*, 100040.

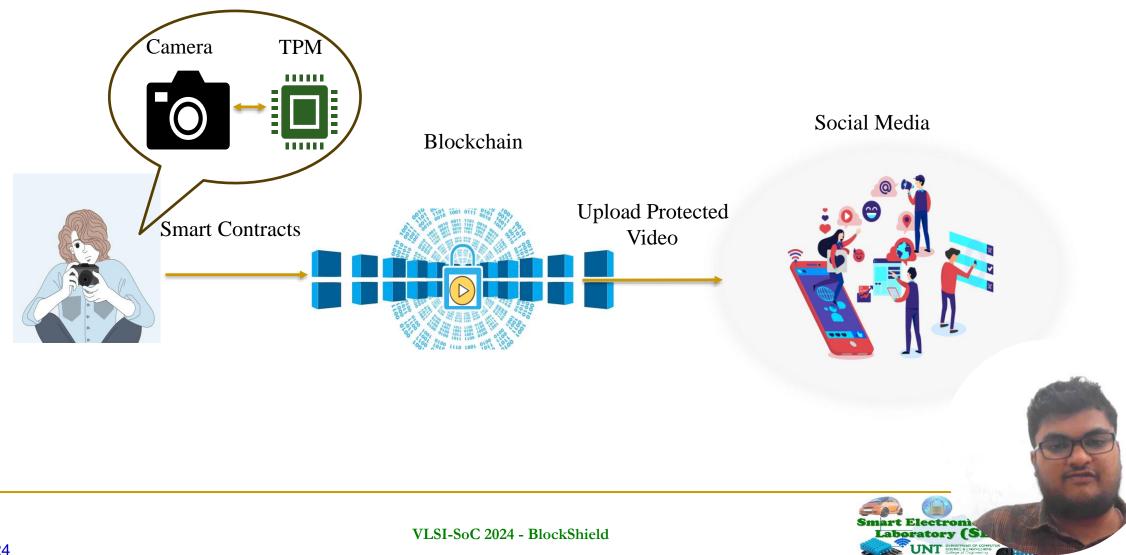
VLSI-SoC 2024 - BlockShield

Smart Electroni

Laboratory (S)

EST, 1890

BlockShield: Conceptual Overview



7

October 9, 2024

Related Research

Work	Technique	Methodology	Tools
Taeb et.al [8]	Detection	ML and Blockchain-Integrated Fake News Detection	Efficient Net, Smart Contracts
Bathalapalli et. al [14]	Mitigation(Image)	PUF and ML framework for facial feature attestation	Dlib 68 (Facial detection and keypoint prediction), PUF
Alattar et. al [10]	Fake news mitigation	Watermarking and Blockchain for Deepfake Video protection	IPFS, MTCNN algorithm, and Face Alignment Network (FAN) algorithm
Qureshi et. al [15]	Audio Deepfake Mitigation	Fragile speech watermarking with Blockchain	MTCNN, Wav2Lip
BlockShield	Visual Deepfake Mitigation	Blockchain and TPM-based video attestation	Hardware TPM, Sr Contracts

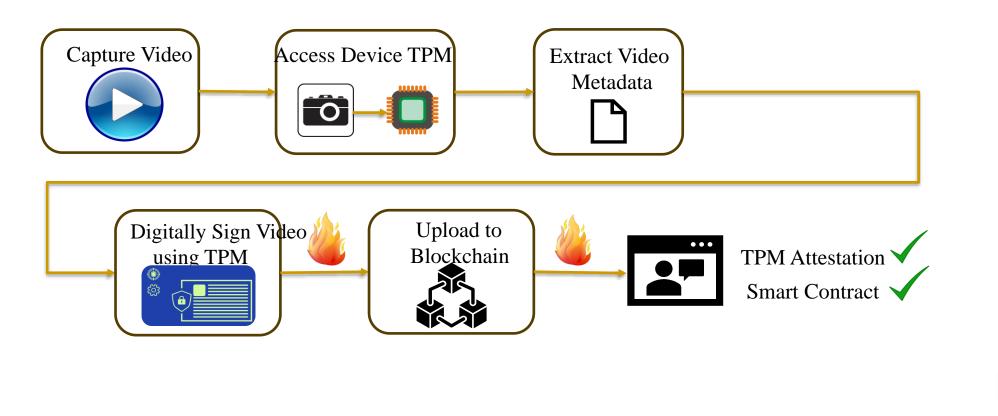


Novel contributions

- A sustainable Deepfake mitigation approach using state of-art TPM and Blockchain technologies.
- A secure visual Deepfake mitigation approach for individual content privacy and security on social media.
- An energy efficient solution that integrates TPM and Blockchain using smart contracts
- A secure digital content sharing framework using Blockchain to provide integrity and authenticity.
- An approach based on TPMs digital signature mechanism facilitating hardware root-of-trust for the video/image.

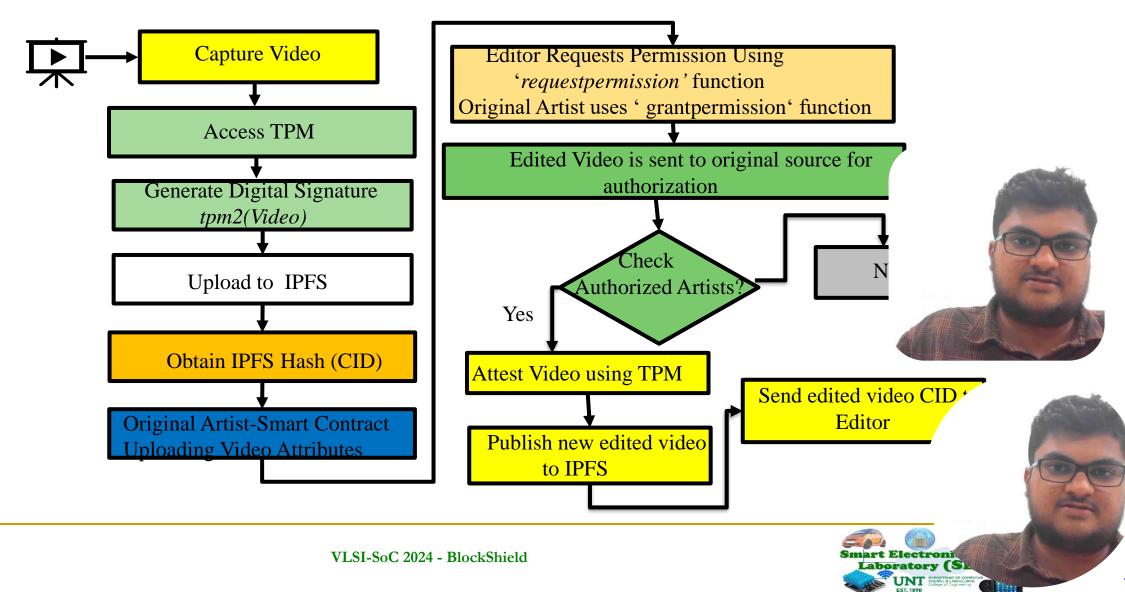


BlockShield: Proposed Deepfake Mitigation Technique



Smart Electroni Laboratory (S UNIT BURGENERS)

Working Flow



11

TPM Video Attestation Workflow

- 1: Access TPM hardware security module at the camera
- 2: tpm2 createprimary -C e -c primary.ctx

\\ Create Primary Key

- 3: tpm2 evictcontrol -C o -c primary.ctx 0x81010001
 \\Assign a unique identifier in TPM NV RAM to make it persistent
- 4: tpm2 create -G rsa -u rsa.pub -r rsa.priv -C 0x81010001
- 5: tpm2 load -C 0x81010001-u rsa.pub -r rsa.priv -c rsa.ctx.
- 6: tpm2 evictcontrol -C o -c rsa.ctx 0x81010002
 - \\Create RSA keys using primary key and make it persistent
- 7: Load the video file and Hash it Fi→ SHA256(Fi)→ Fi.hash
 \\Hash video File
- 8: tpm2 sign -c 0x81010002 -g sha256 -o sig.rssa Fi.hash
 \\Digitally sign the video hash file using TPM
- 9: tpm2 hash -C e -g sha 256 -o sig.rssa.hash -t ticket.sig.rssa sig.rssa
 \\Generate SHA 256 hash of Digital signature for video file



Smart Contract Validation Workflow

Input: Digital Signature of Video file D_{Fi} and Video File FiOutput: Digital content is securely stored in Blockchain and secure accessed using smart contract

- 1: for Each Primary artist do
- 2: Individual contract is called by Primary artist to manage access
- Primary Artist attested Video file information on to IPFS system
- 4: for Each Video File do
- 5: Upload video file Fi and Digital signature D_{Fi} on IPFS.
- 6: IPFSfile $Ii \leftarrow$ IPFS.upload(Fi, D_{Fi})
- 7: end for
- ArtistContract.addIPFSHash(*Ii*) Return hash is added as an attribute in the newly created Primary artist contract
- 9: end for
- 10: Share the contract address and IPFS hash of the video provide access to video file
- 11: Editor initiates a 'requestpermission' function to primary artist contract address which is accessible online.
- 12: Update list of artists at the primary artist side using 'grantpermission' function
- 13: Editor submits a new edited version of video file Fi to primary artist
- 14: Edited video file VFi is attested using TPM $VFi \rightarrow \text{TPM} \rightarrow F = D_{FVi}$
- 15: Upload Edited and attested video file VFi and D_{FVi} to IPFS and share it with secondary artist or editor.

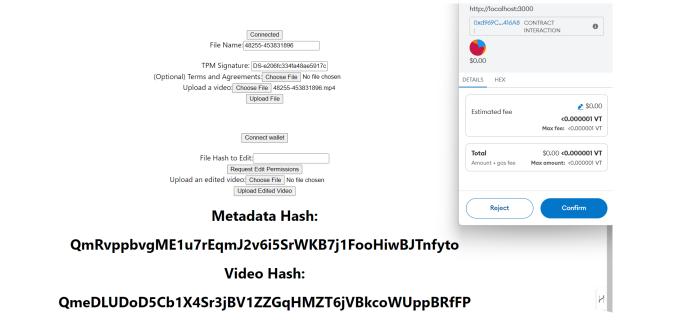
Smart Electroni

Laboratory (S

UNT SCIENCE

Experimental Validation of BlockShield

									pi@ı	rasp	berry	ypi: -	~				~ ^ X
File Edit	Tabs	H	elp														
			- \$	su	do	tpm:	2_Ve	erif	ysi	gnat	ture	e - (c 0)	x81(9100	902	-g sha256 -s sig.rssa▲
-m video			~	_		_											
bi@raspbe												~ 4	- 4	b 0	40	-14	
00000000							-	47		91	. –						
90000010				6f						33							eo3.\$
90000020								cb		ef							MM.u`
00000030		-						da		75							.QPuK.
00000040	9a							ed		3a							Q.(
00000050				e9						31							fFb01lwK.B.
00000060	· · ·	13		9d					_	7f							[J]
00000070				bb						6d							't.;"mT.
00000080				7f						4c							bp]."Lv6A
00000090				7c					57			e9				4d	aR\ y".8W>&M
000000a0	_	_		89			-			69							Vo0.i7PQ
000000b0				6a						d1							.>.j'.*.ti
000000c0	d1	80	11	3a	76	9a	15	d0	f2	1f	bb	76	0a	97	f4	56	VV
000000d0	d0	46		b4					e0	6a							.F]>jJ.]\$
000000e0		59		f3					83			3b					.Y0Hw;
000000f0		-	_		_		1e	36	28	f7	36	48	3e	d5	62	ff	.;/6(.6H>.b.
00000100	d4	db	75	c 8	70	d4											u.p.
00000106																	
pi@raspbe	rry	oi:∽	- \$	su	do 1	tpm:	2_Ve	erif	ysi	gnat	ture	e -(c 0)	x81(9100	902	-g sha256 -s sig.rssa
-m video	3.ha	ash		_													
pi@raspbe	rry	oi:∙	- \$	Π													





Smart Contract Validation

Upload Media

Transaction Details

Transaction Hash	0x61r2017a43e62r5a89ecf1c904475febde34d32770d05a2e912dafcd11a078d5 @	
Result	⊘ Success	
Status	Confirmed Confirmed by 11.968	
Block	28585005	
Timestamp	Q 23 hours ago July-18-2024 01:40:40 PM +-5 UTC Confirmed within <= 6.869 seconds	
From	0x6c27c94191c630438ace12e123164a1b628882a6 👩	
Interacted With (To)	0xd969cd7c368cdec700f85c0a12f16079263416a8 @	
O Value	0 VT	
Transaction Fee	0.000000000385344 VT	
Gas Price	0.00000008 Gwei	
Transaction Type	2 (EIP-1559)	
🖲 Gas Limit	48,168	
Max Fee per Gas	0.00000008 Gwei	
Max Priority Fee per Gas	0.00000008 Gwei	
Priority Fee / Tip	0.000000000048168 VT	
Transaction Burnt Fee	© 0.0000000000337176 VT	
Gas Used by Transaction	48,168 100%	
Nonce Position	15 0	
Raw Input	Hex (Default) •	đ
		*

Grant Permission

Transaction Details

Transaction Hash	0x61/2017a43e62/5a89ecf1c904475febde34d32770d05a2e912dafcd11a078d5 g	
Result	@ Success	
Status	Confirmed Dy 11,968	
Block	28585005	
Timestamp	() 23 hours ago July-18-2024 01:40:40 PM +-5 UTC Confirmed within <= 6.869 seconds	
From	0x6c27c94191c630438ace12e123164a1b628882a6 g	
Interacted With (To)	0xd969cd7c368cdec700f85c0a12f16079263416a8 👩	
Value	0 VT	
Transaction Fee	0.000000000385344 VT	
Gas Price	0.000000008 Gwei	
Transaction Type	2 (EIP-1559)	
🚯 Gas Limit	48,168	
Max Fee per Gas	0.00000008 Gwei	
Max Priority Fee per Gas	0.00000008 Gwel	
Priority Fee / Tip	0.000000000048168 VT	
Transaction Burnt Fee	© 0.000000000337176 VT	
Gas Used by Transaction	48,168 100%	
Nonce Position	15 0	
Raw Input	Hex (Default) +	Ø
	8-1.8.2.2.7003/0000000000000000000000000000000000	

Request Permission Transaction Details Transaction Hast 0x61/2017a43e62/5a89ecf1c904475febde34d32770d05a2e912dafcd11a078d5 ⊘ Success Result Status Confirmed Confirmed by 11.968 Block 28585005 @ 23 hours ago | July-18-2024 01:40:40 PM +-5 UTC | Confirmed within <= 6.869 seconds @ Timestamp @ From 0x6c27c94191c630438ace12e123164a1b628882a6 @ 0xd969cd7c368cdec700/85c0a12/16079263416a8 @ Interacted With (To) Value 0 VT Transaction Fee 0.00000000000385344 VT Gas Price 0.000000008 Gwei Transaction Type 2 (EIP-1559) Gas Limit 48,168 0.000000008 Gwei Max Fee per Gar @ 0.0000000000337176 V1 48,168 | 100% Nonce Position 15 0 Raw Input lex (Default) Edit Media Transaction Details Transaction Hash 0xb38f11cbfe13a90fc6ffc0a51129fa60cc1cf4f6a167f55244cc4256645916dc Result ⊘ Success Status Confirmed Down Confirmed by 11,994 Block 28585008 Timestamp () 23 hours ago | July-18-2024 01:41:05 PM +-5 UTC | Confirmed within <= 6.919 seconds 0x6c27c94191c630438ace12e123164a1b628882a6 🗗 From Interacted With (To) Ouri969rri7r368cdec700f85c0a12f16079263416a8 Value 0 VT

5-0-4

Smart Electroni

Laboratory (S

EST 1890

VLSI-SoC 2024 - BlockShield

Transaction Fee

Transaction Type

Nonce Position

Raw Input

Gas Price

0.000000000003842 VT

0.00000000000048025 ∨T © 0.000000000000336175 ∨ 48.0251100%

Hex (Default) +

0.000000008 Gwei

2 (EIP-1559)

48,025 0.00000008 Gwe

16 2

Max Priority Fee per Gas 0.00000008 Gwe

Performance Analysis

Video	Duration (s)	Frame Rate	Bitrate	TPM-Signature	TX Hash	IPFS CID
48255- 453831896.mp4	16	24	1633.922	e206fc334fa 48ae5917cac93dff 260d0fc0f0535f4e2 25c932466c 2291833df9	0x4c99d2c8f26 5498b09c53b372 e94f3cefc89d17be 12b401de25bf2b db892609b	QmQjxbmrjFbS7Xz4WXSig tkP msAKRi5FEXdDDpPK1Zn5 g1
61299- 498228517.mp4	26.56	29.97	3526.986	fa60e6faf0f64 c50846ac74ca185ffc d83d89fbd68fb 9d2985a6bb5a454eab1a	0x631719a0744dd 4d880924ac7ad 57b98d5d 385a73af7e8e5 4e55039d 4612e723b	QmS5PjDzYqGtfTCYYCm9 QrW Fam6ZHZVKU2uw5vJgN6 abqf
61706- 500316063.mp4	15.65	29.97	937.89	e206fc334fa48a e5917cac93dff260d0fc0f0535f4e 225c93 2466c2291833df9	0x539dfe6fad 4015a6b4ed84 85717614f b9091f7ec0ff1 aa5fc7c93 3a76821b0fe	QmTtS4J4GHG4n2tMEPsf NbVT MuTEmw3uKni5x4djUPEq SN
73711- 549547411.mp4	25.20	29.27	1098.274	9c1f7e38f1528cc18765b79e28fa 76f3fab662d0163cd309260939B 208d7dcee	0xbc61d12a 52c0815799da10 e0ea8806 28c287a538eb aae65fe857e8 aa0b1435cc	QmNw2PUDAtZ Mt8twnVQ4Kjw7Py 1P4NsZ7dWnak71eVHAK
44645- 439940290.mp4	10.88	25	5596.436	9bee3bef81c8ac 3f596a 6f4c44b b218cb713d2 ba2541e89c487a 59362729f60f	0x067ee279182e24 372e49ecd7e5 22551169d23 1a30da152f7 651c1c4ac2 945a6d	QmWV/ 7NHtV RyqC



Conclusion and Future Research

- This research work presented and experimentally validated a Blockchain and TPM integrated approach for Deepfake mitigation through TPM-based hardware digital content attestation.
- The proposed work with state-of-art TPM-digital signature approach ensures hardware based digital content source attestation facilitated through Blockchain smart contract-based access control approach ensuring digital content authenticity.
- This is a novel work with TPM attestation and blockchain smart contract for access control and digital content sharing with the substantial performance indicators showcasing the robustness of the proposed Deepfake mitigation approach.
- Furthermore, proposed research work could be further applied to Deepfake mitigation for images with effective mechanism for facial feature and biometric-based user authentication.
- Additionally, this work could be extended to smart cities surveillance applications which work in untrusted environments to guarantee privacy, security and traceability to digital content.



Thank You!

