

ARCHANGEL: Trusted Archives of Digital Public Records

EPSRC

John Collomosse | Centre for Vision Speech and Signal Processing, University of Surrey









Securing our National Archives for Future Generations

ARCHANGEL will deliver **long-term sustainability of digital archives** through new **technologies** that will ensure both **accessibility and integrity of digital archives** – whilst **maximizing their impact** through new models for **commodification and open access**.



Use Cases

UC1: Research Dataset Integrity



UC2: Supreme Court Video Records



UC3: Delayed Disclosure Public Records



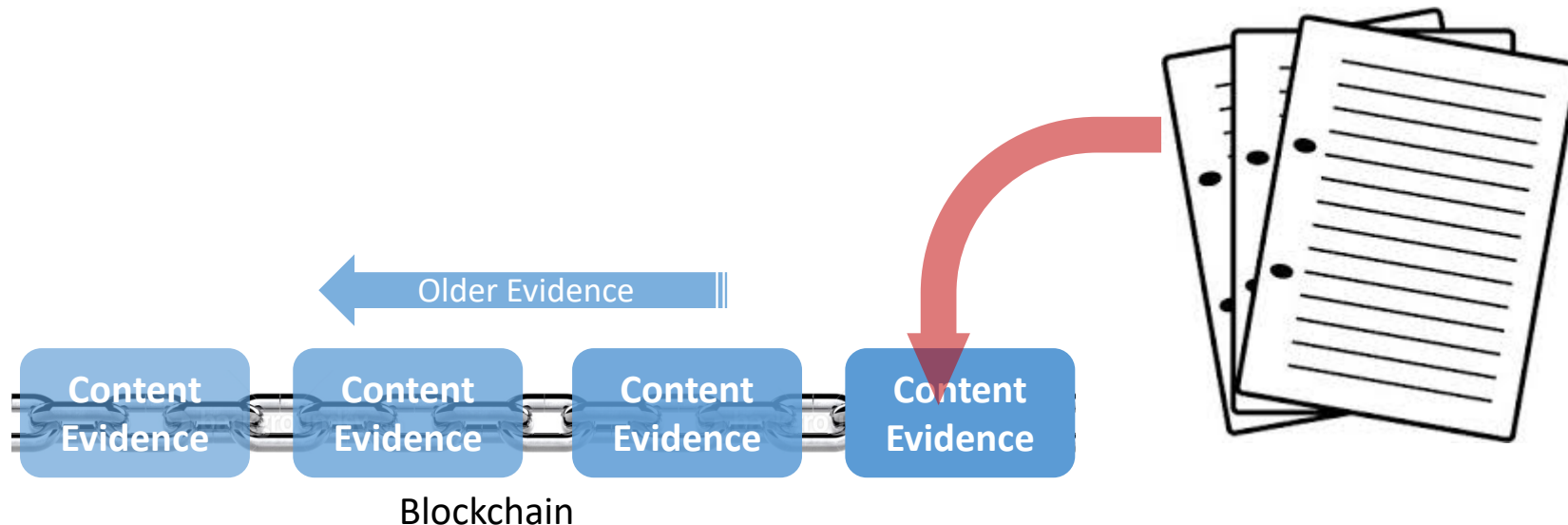


21C - Blockchain (Distributed Ledger)

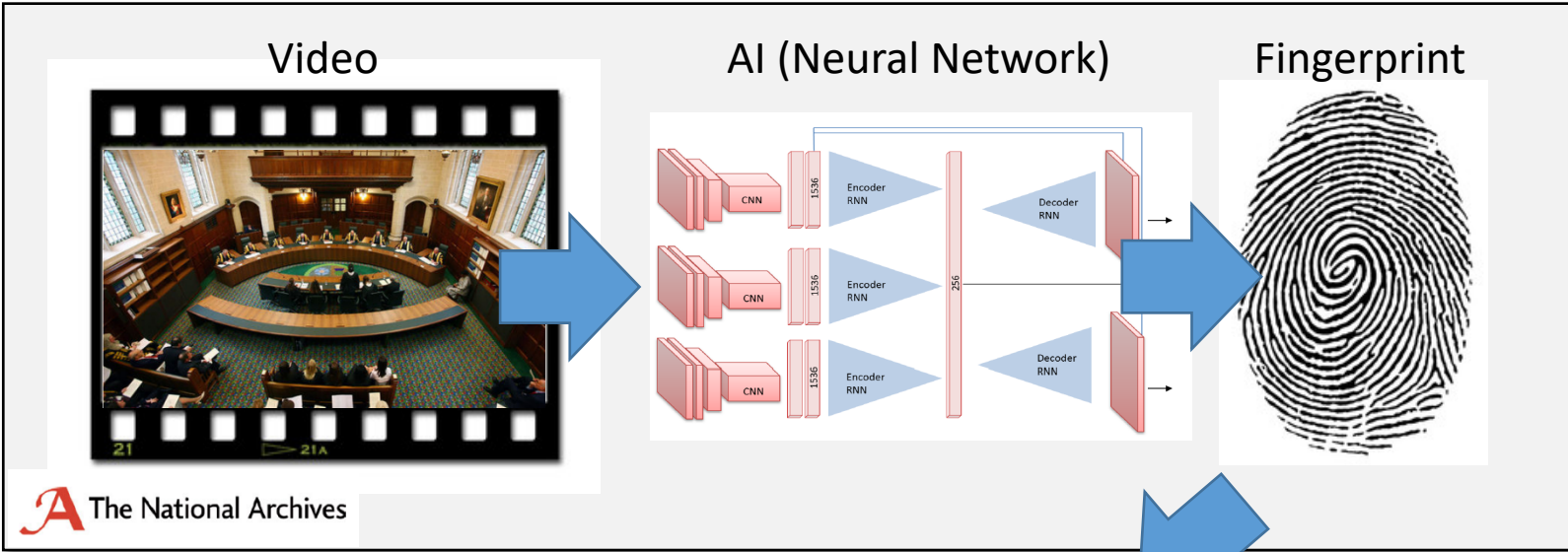
A **Blockchain** is a way of **storing data** (a kind of database). It has unique properties:

- ✓ It can **guarantee the provenance of data** within it (**tamper-proof data**)
- ✓ This guarantee can be made over data distributed anywhere **without reliance on a centralised authority**

Data is stored in '**blocks**'. Each block contains a **hash (fingerprint)** of the **previous block**.



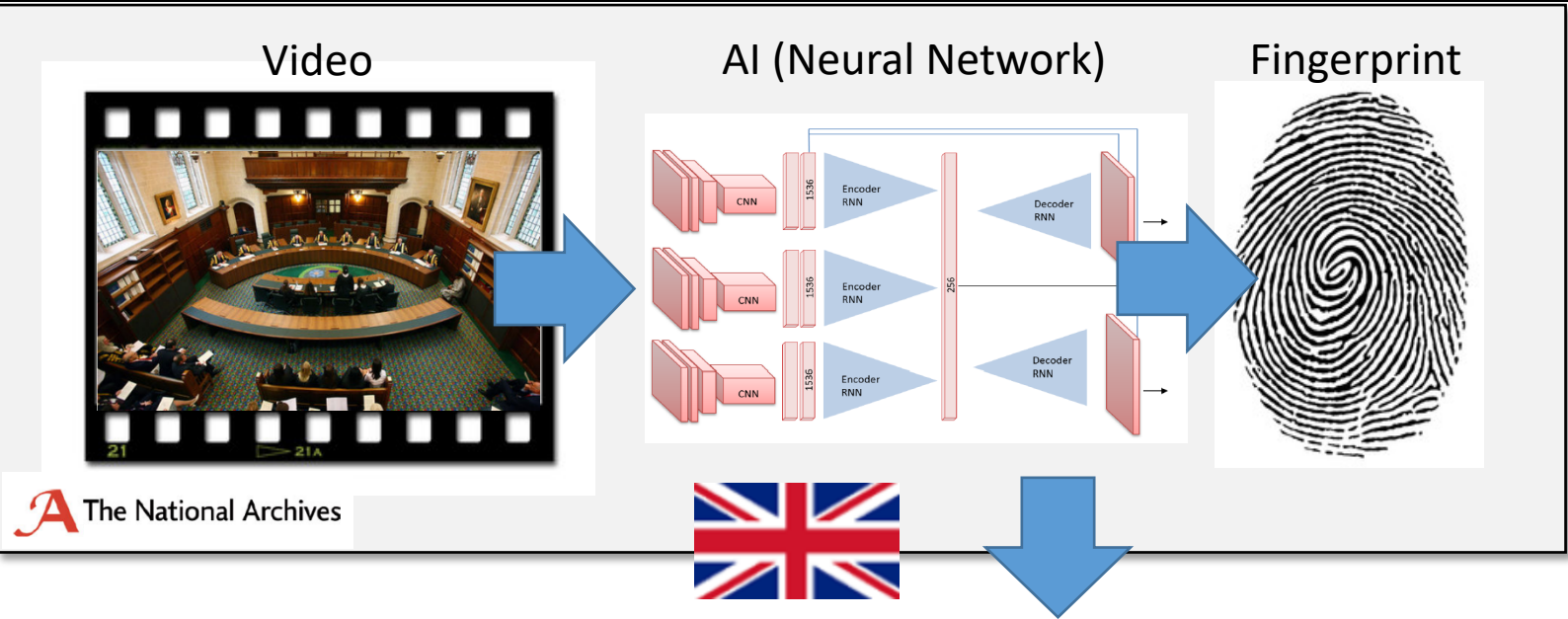
ARCHANGEL – Fusing AI and Blockchain for Archive Integrity



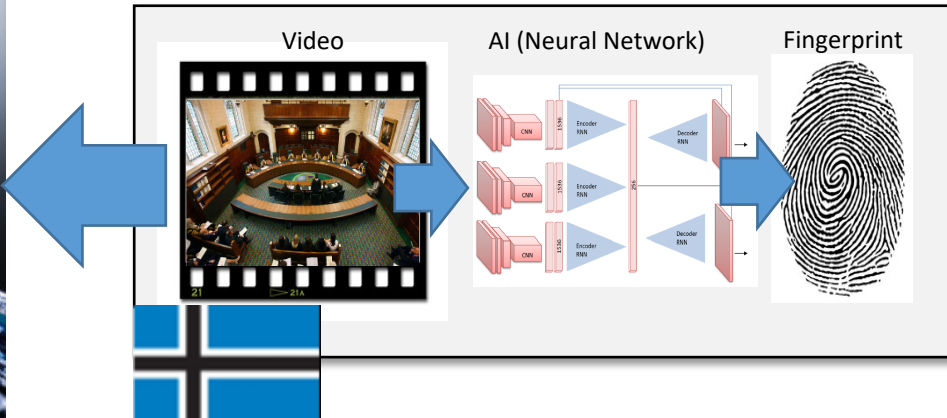
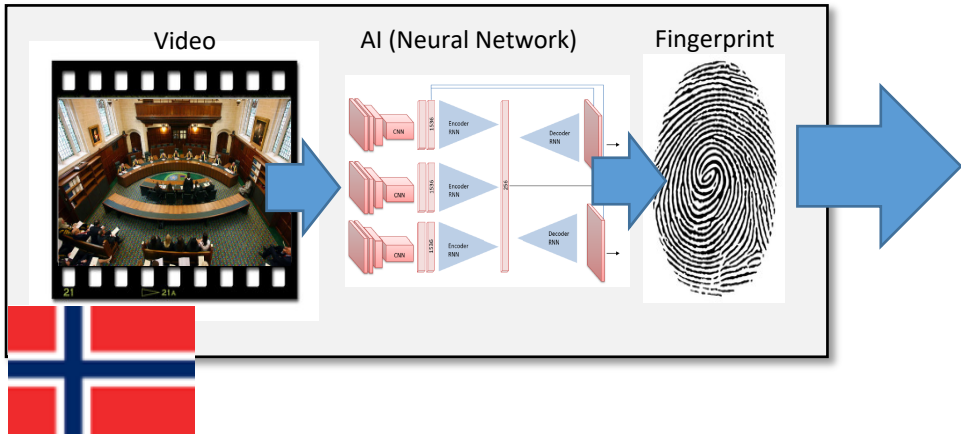
Fingerprints stored in Blockchain



ARCHANGEL – Fusing AI and Blockchain for Archive Integrity

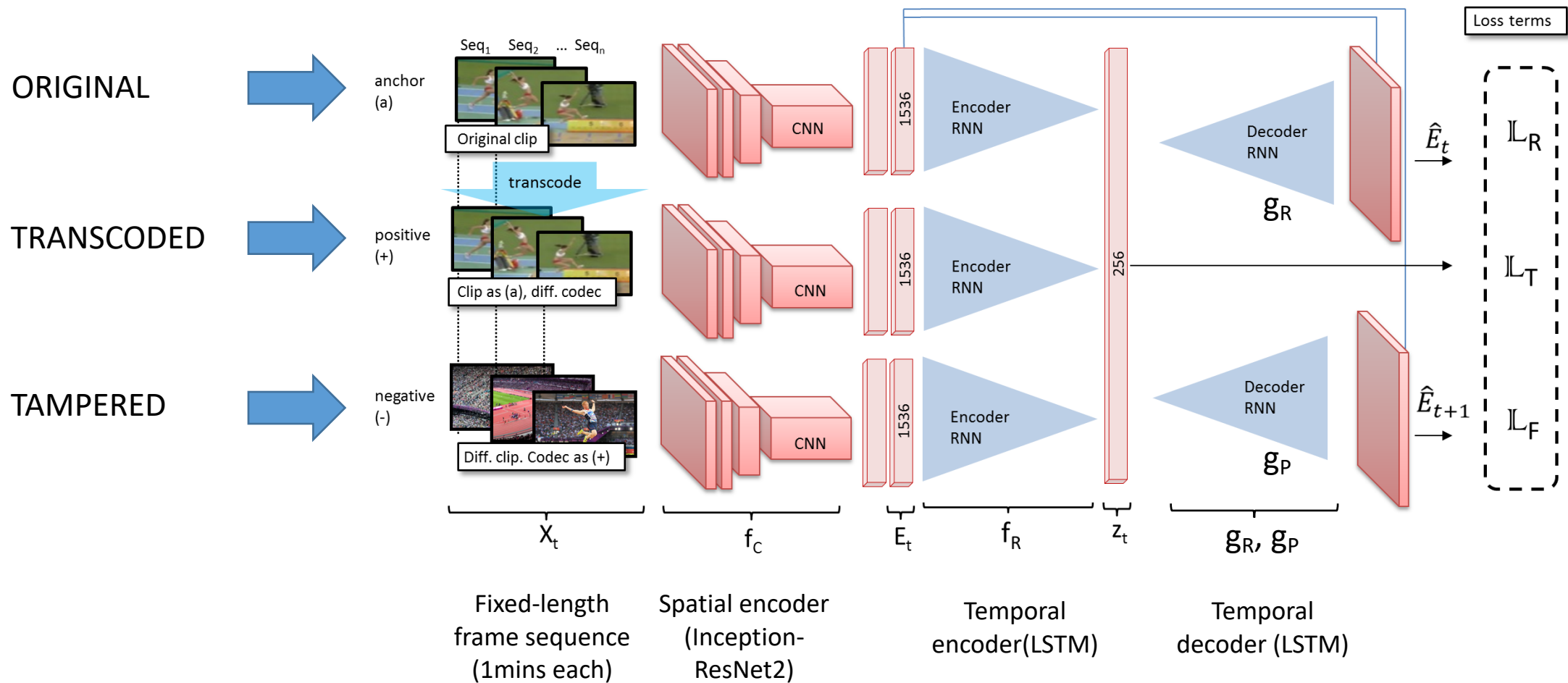


Fingerprints stored in Blockchain



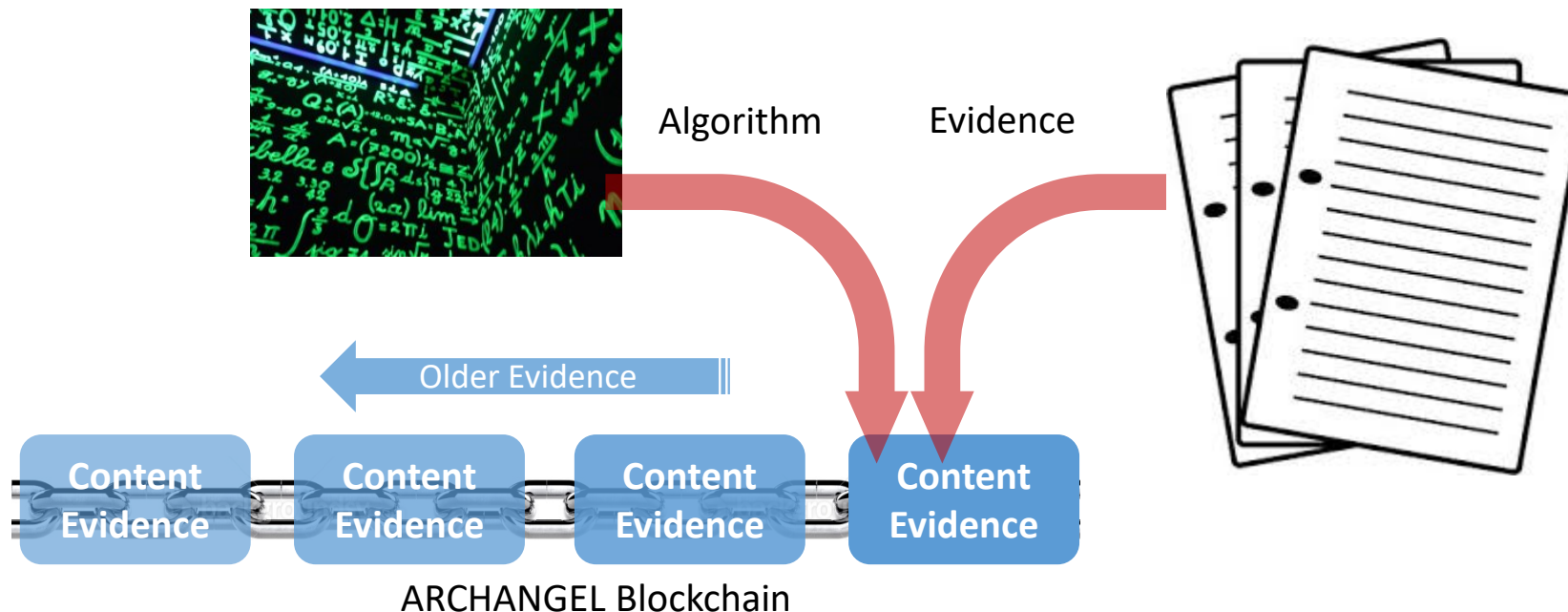
ARCHANGEL – Fusing AI and Blockchain for Archive Integrity

We train a neural network to tell the difference between tampered and format shifted video



Securing the Neural Network

The network used to compute the content evidence must also be hashed and stored.

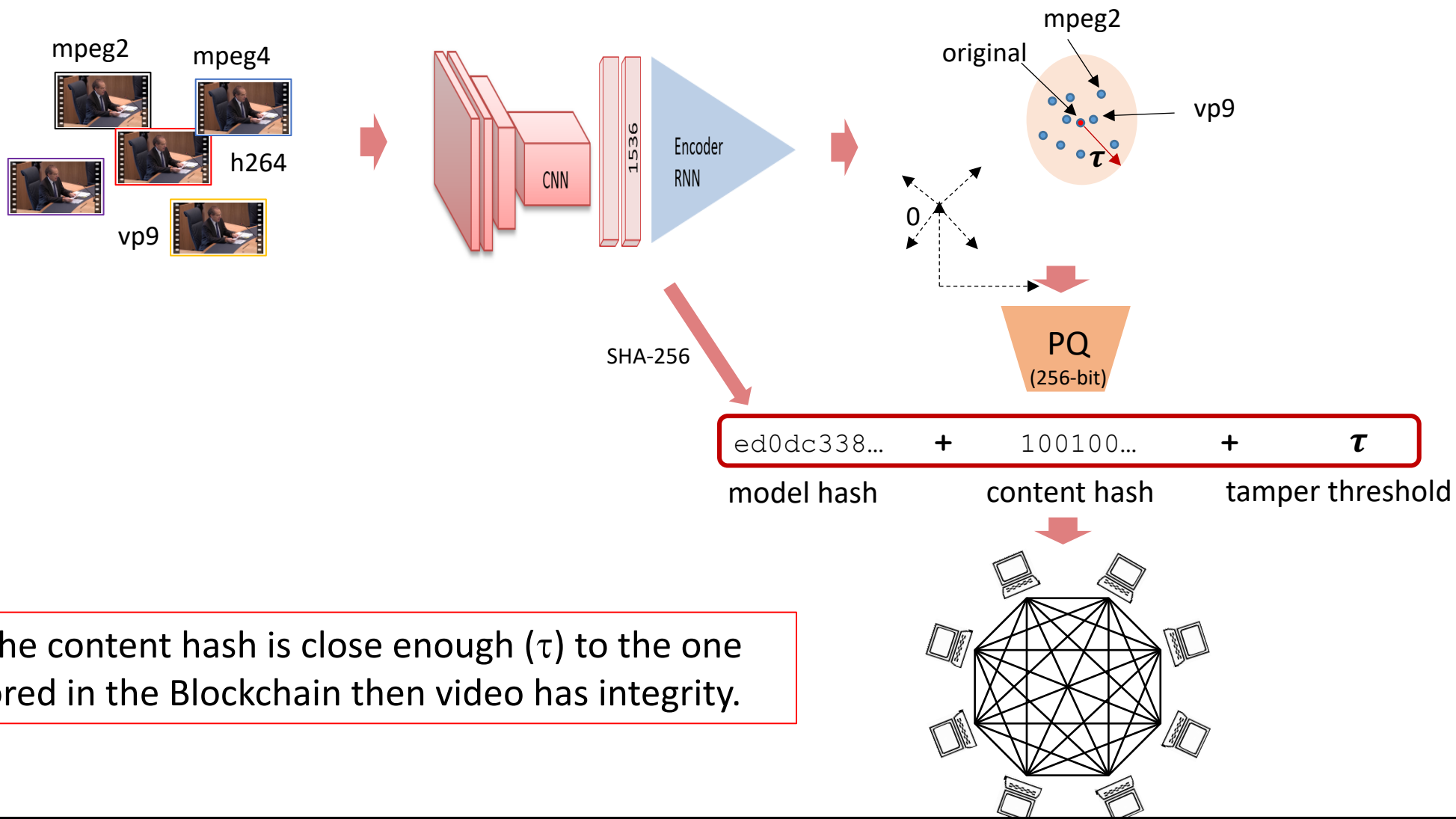


Video has been protected by a neural network ('model') that computes a **content hash**

The **model** is protected using a regular **cryptographic/bit hash** (SHA-256)

ARCHANGEL - Verification Process

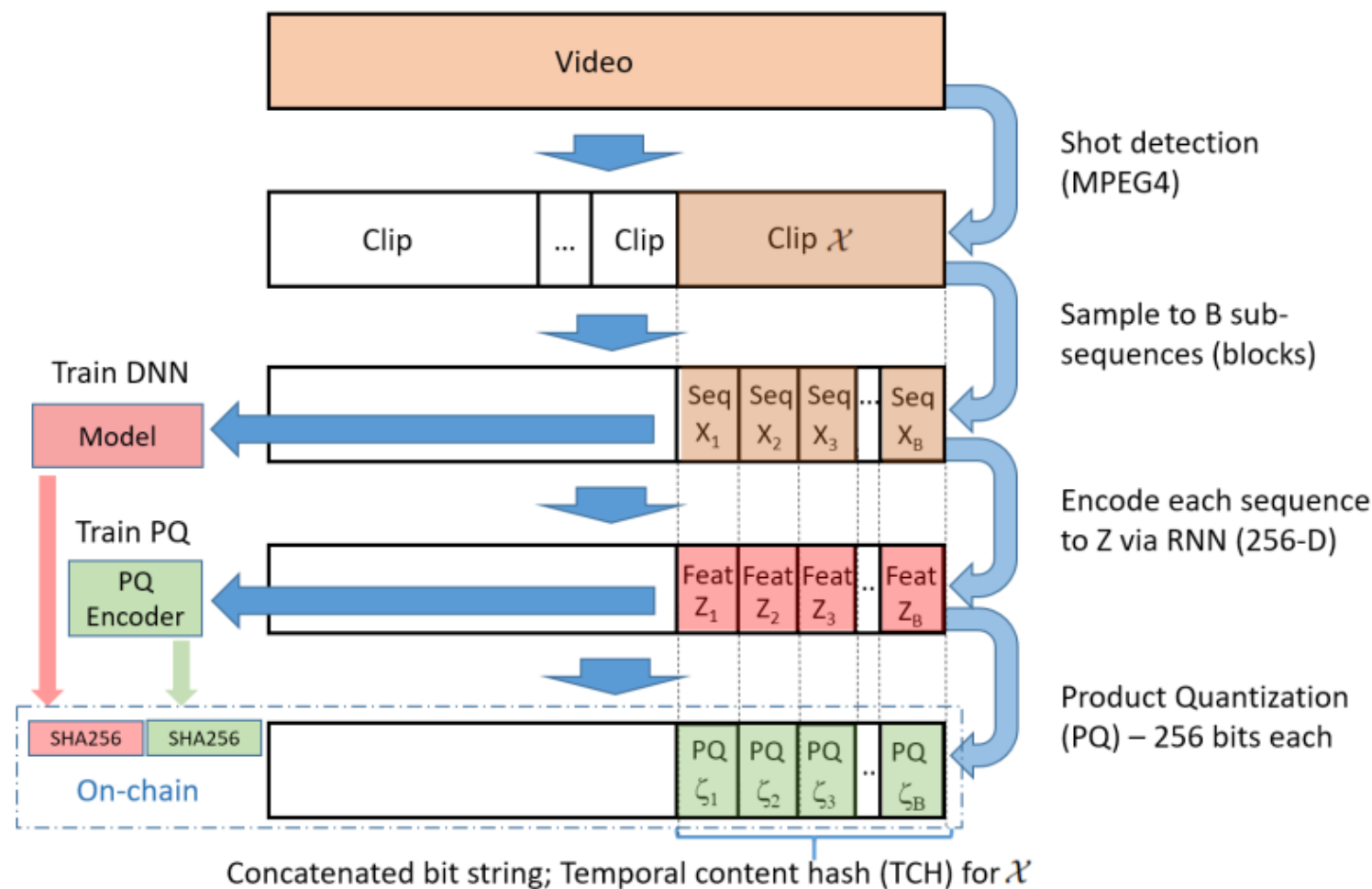
Video in its current format is 'played' through the model (inference) to generate content hash



If the content hash is close enough (τ) to the one stored in the Blockchain then video has integrity.

ARCHANGEL - What is stored and where?

Video (*GB*) and its model (*MB*) are off-chain. Content and model hashes (*KB*) are on-chain.



ARCHANGEL – Evaluation (Tamper)

Three datasets were used to characterize the tamper-proofing afforded by ARCHANGEL

Dataset	No. videos	Total length (h:m:s)	Min-max (avg)
ASSAVID	21	1:29:11	12s-39m (4m)
TNA	7	6:23:54	4s-2h (54m)
OLYMPICS	5	0:48:43	8m-13m (9m)

Each dataset has:

- ✓ Control set: transcoded using different codecs.
- ✓ Temporal tamper set: randomly remove 1-10s.
- ✓ Spatial tamper set: replace frame/audio in random 1-10s section with white noise.

The National Archives
(TNA, UK)

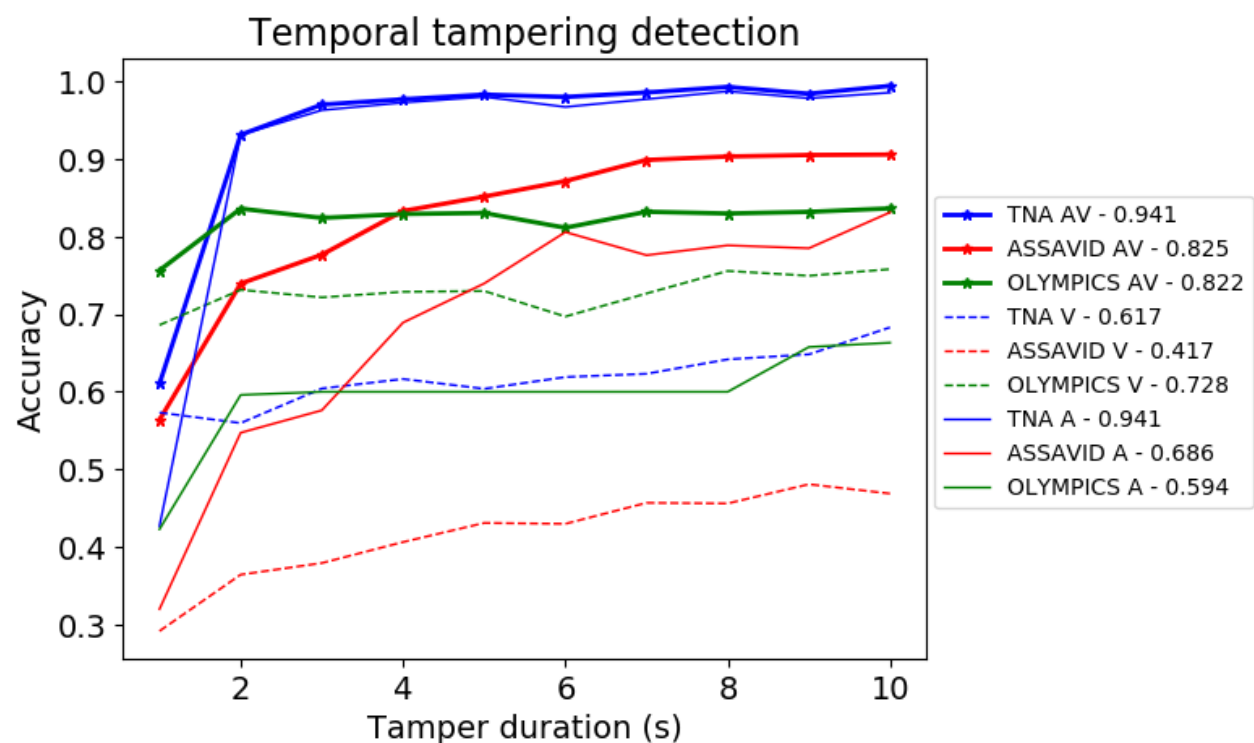


BBC Archives

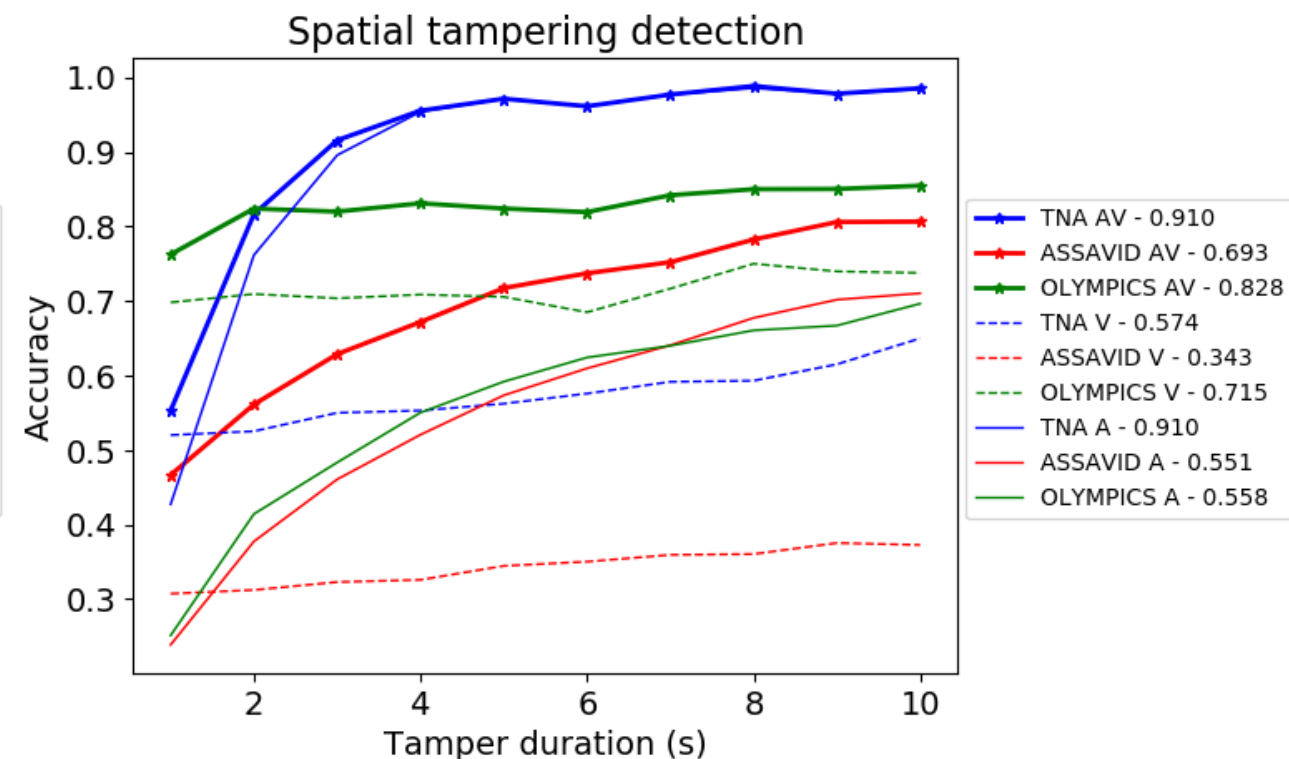


ARCHANGEL – Evaluation (Tamper)

ARCHANGEL can detect >3 seconds tamper within ~2 hour video clips at 95% accuracy (TNA)



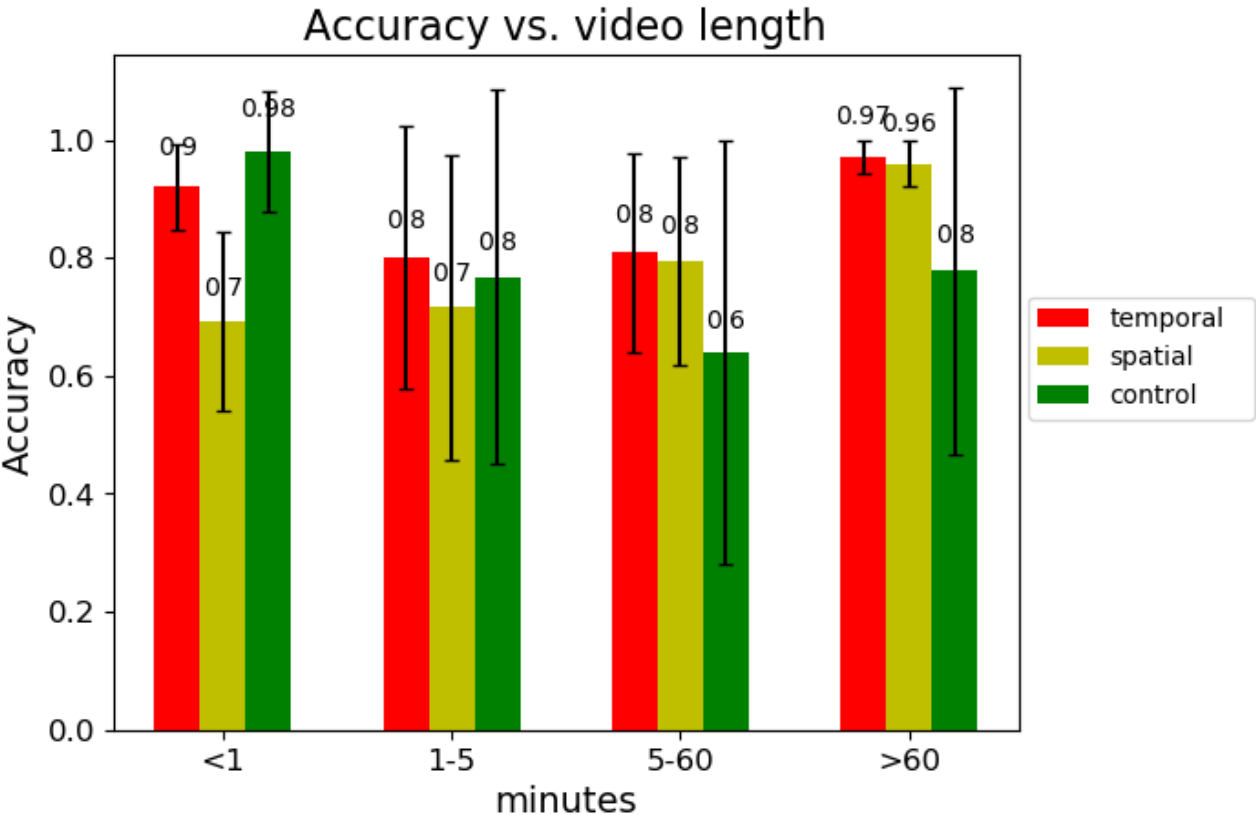
Temporal tampering detection



Spatial tampering detection

ARCHANGEL – Evaluation (Tamper)

Tamper detection accuracy does not degrade significantly with video clip length



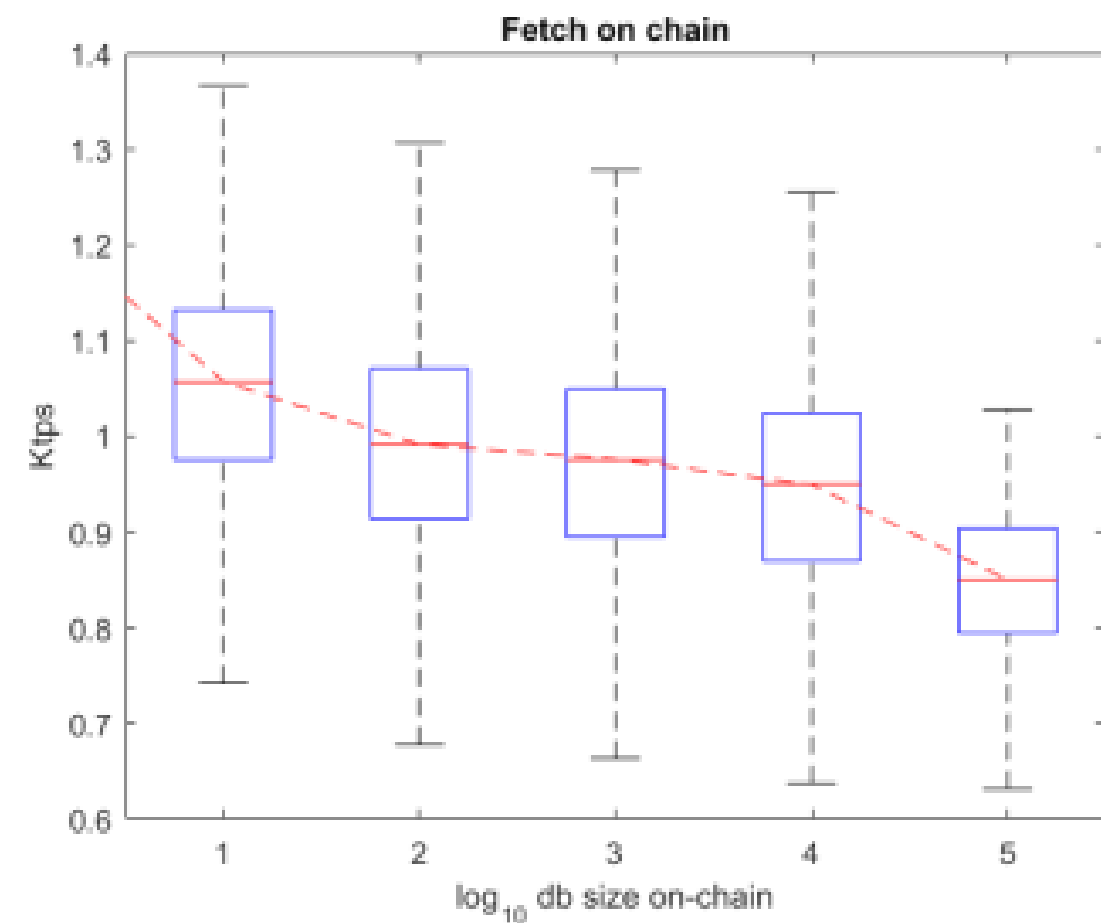
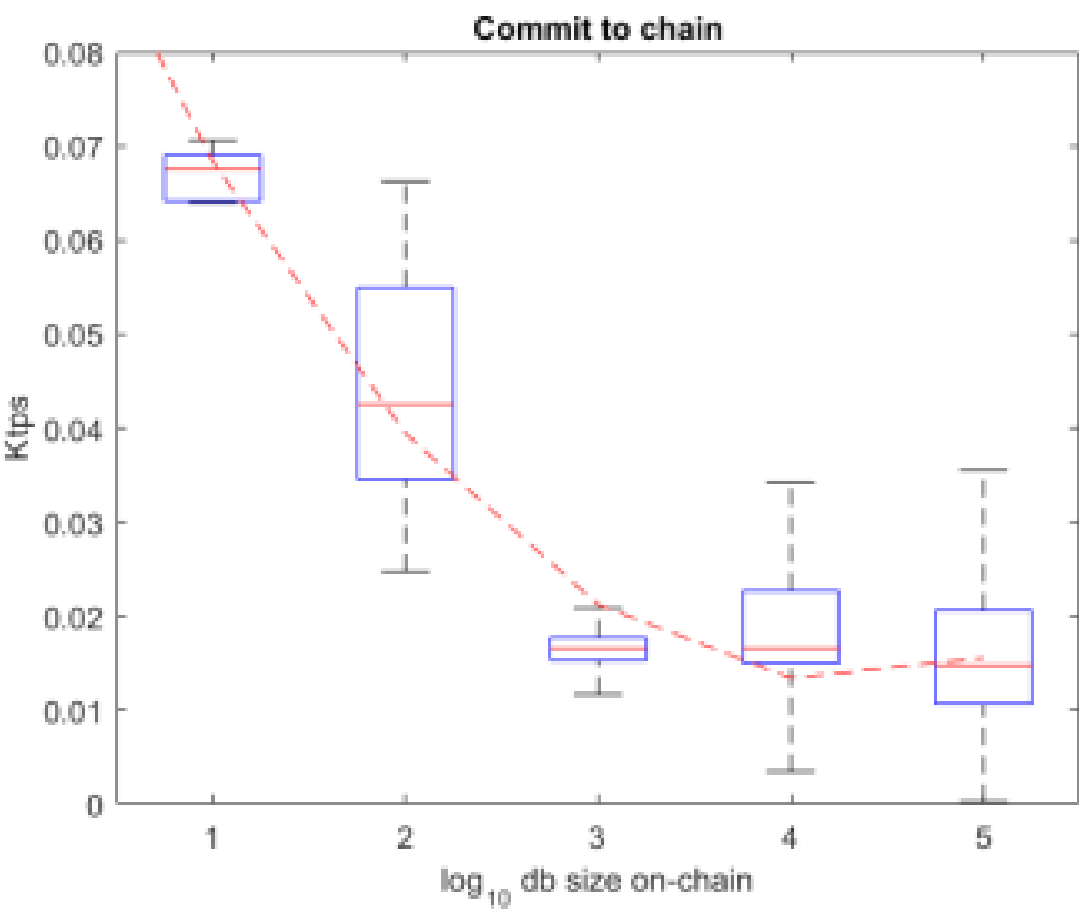
Detection accuracy vs. video length

Datasets	Precision	Recall	F1
ASSAVID	0.981	0.756	0.854
OLYMPICS	0.944	0.823	0.879
TNA	0.919	0.925	0.922

Precision, Recall and F1-score

ARCHANGEL – Evaluation (Scalability)

ARCHANGEL Blockchain read/write throughput statistics (Ethereum, Proof of Authority)



Ktps: thousands of transactions per second

ARCHANGEL – Evaluation (User Trial)

Deployed for trial across National Archives of 5 countries (UK, USA, Australia, Norway, Estonia)

“On Blockchain”

"... The key thing for us is the blockchain environment really. I think there's a lot of exciting possibilities around blockchain. Where I work in XXXX there's a 20-year closed period, so a record does not become publicly accessible 20 years after creation, and that's a long time for files and metadata to be sitting around. If we can prove authenticity and integrity through a tried and tested technology, which hopefully distributed ledger will become, then that's going to help us a lot just through these multiple generations of technology."

"... I have done some private blockchain investigation and it looks like the technology is sound. How you share, store information across several nodes is a very good idea, and your implementation is quite good actually."

"... The idea is really good and never thought that blockchain could be used like this and this would be a great alternative from other cryptocurrencies and everything. Finally something useful!"

Anonymised excerpts of interview transcripts gathered during trial with these 5 archive and memory institutions.

ARCHANGEL – Evaluation (User Trial)

Deployed for trial across National Archives of 5 countries (UK, USA, Australia, Norway, Estonia)

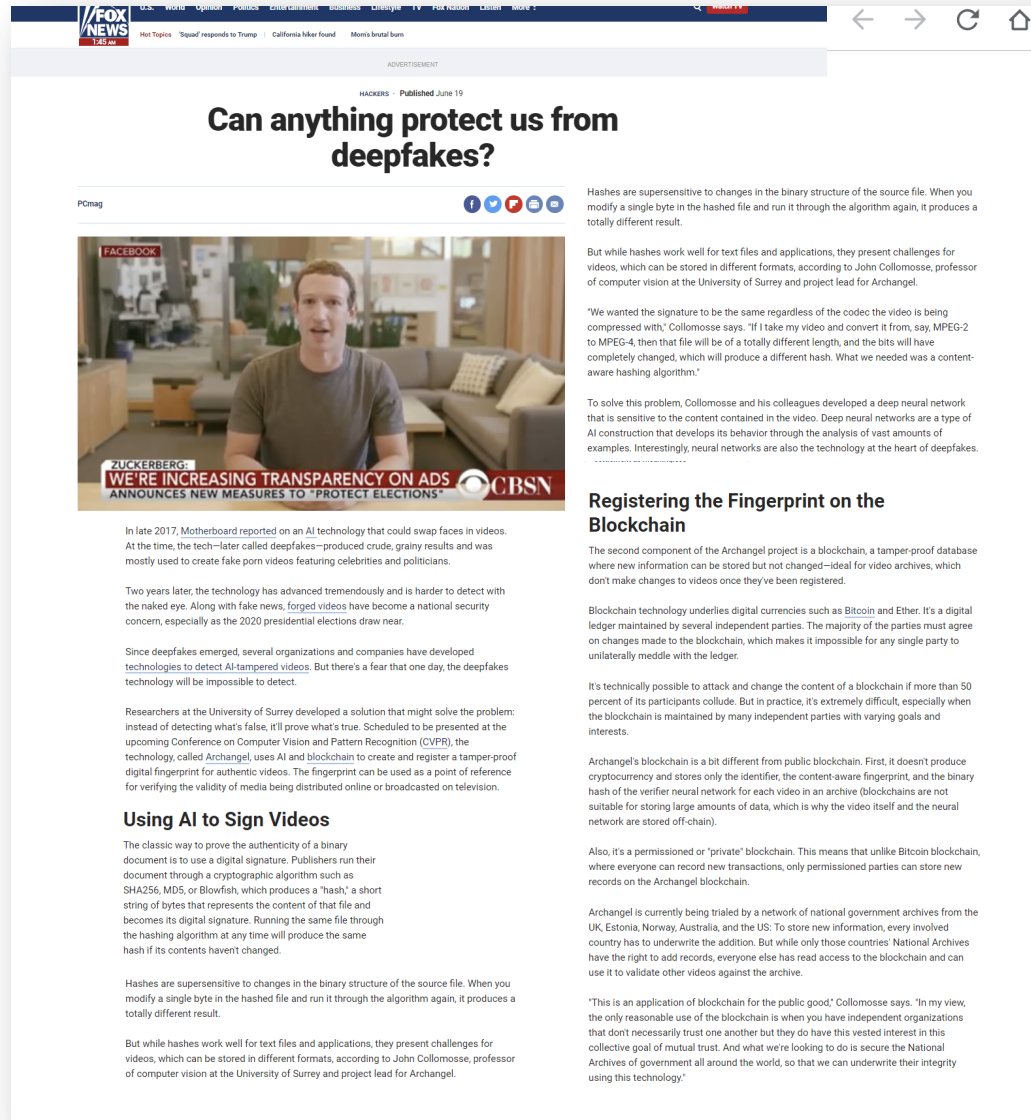
“On novelty of ARCHANGEL”

"... I really like that you guys have expanded it out to a really wide audience of archival professionals around the world; that's really impressive that you guys accomplished that. "

"... I feel like you're really spearheading the use of blockchain in our whole community and something that I think is going to become the power for the course and it's interesting - I can't wait to see the final reports and outputs of this project. It's really probably going to help inform what we do as an organisation for sure. "



ARCHANGEL - Applications beyond National Archives



Codec invariant but tamper sensitive video hashes

Potential applications to:

Social media

Fake news / deep faked videos

Journalism

Any context where video provenance holds value but where the content is likely to be transcoded e.g. cross-platform distribution

<https://www.foxnews.com/tech/can-anything-protect-us-from-deepfakes>

Conclusion

ARCHANGEL combines Blockchain and AI to protect the archival record

Why Blockchain?

- ✓ Multiple independent organizations need a shared, de-centralized point of trust
- ✓ A move from institutional underscoring of trust to a technological underscoring

Why AI?

- ✓ Current best way to compute content aware hash
- ✓ Protect audio-visual streams hours in length against tampers ~3s or more

Value proposition verified in live trial with the national archives of five nations

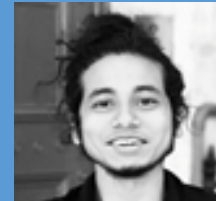
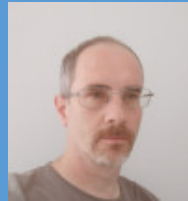
Further work

- ✓ Standardization
- ✓ Further international rollout

ARCHANGEL: Trusted Archives of Digital Public Records

EPSRC

John Collomosse | Centre for Vision Speech and Signal Processing, University of Surrey



www.archangel.ac.uk

J.Collomosse@surrey.ac.uk