Media Protection Mechanisms in the SYNAT Digital Library

A. Dziech, A. Głowacz, J. Białas, and P. Korus

ACK Cyfronet AGH

March 1, 2013

Outline

SYNAT Project

The main goal of the SYNAT project is to create an advanced and open distributed data repository intended for development in digital libraries.

The project started in 2010, the consortium consists of:

- ICM, University of Warsaw,
- Warsaw University of Technology,
- Academic Computer Center CYFRONET AGH,
- National Library of Poland,
- Institute of Bioorganic Chemistry, PAS - PCSS,
- National Institute of Telecommunications,
- Institute of Computer Science, PAN
- MIM, University of Warsaw
- NASK
- Polish-Japanese Institute of Information Technology
- Wrocław University of Technology
- Łazarski University
- Jagiellonian University in Cracow
- Cardinal Stefan Wyszyński University in Warsaw
- Military University of Technology
- WETI, Gdańsk University of Technology

SYNAT Project

The main goal of the SYNAT project is to create an advanced and open distributed data repository intended for development in digital libraries.

The project started in 2010, the consortium consists of:

- ICM, University of Warsaw,
- Warsaw University of Technology,
- Academic Computer Center CYFRONET AGH,
- National Library of Poland,
- Institute of Bioorganic Chemistry, PAS - PCSS,
- National Institute of Telecommunications,
- Institute of Computer Science, PAN
- MIM, University of Warsaw
- NASK
- Polish-Japanese Institute of Information Technology
- Wrocław University of Technology
- Łazarski University
- Jagiellonian University in Cracow
- Cardinal Stefan Wyszyński University in Warsaw
- Military University of Technology
- WETI, Gdańsk University of Technology

Contributions of ACK Cyfronet

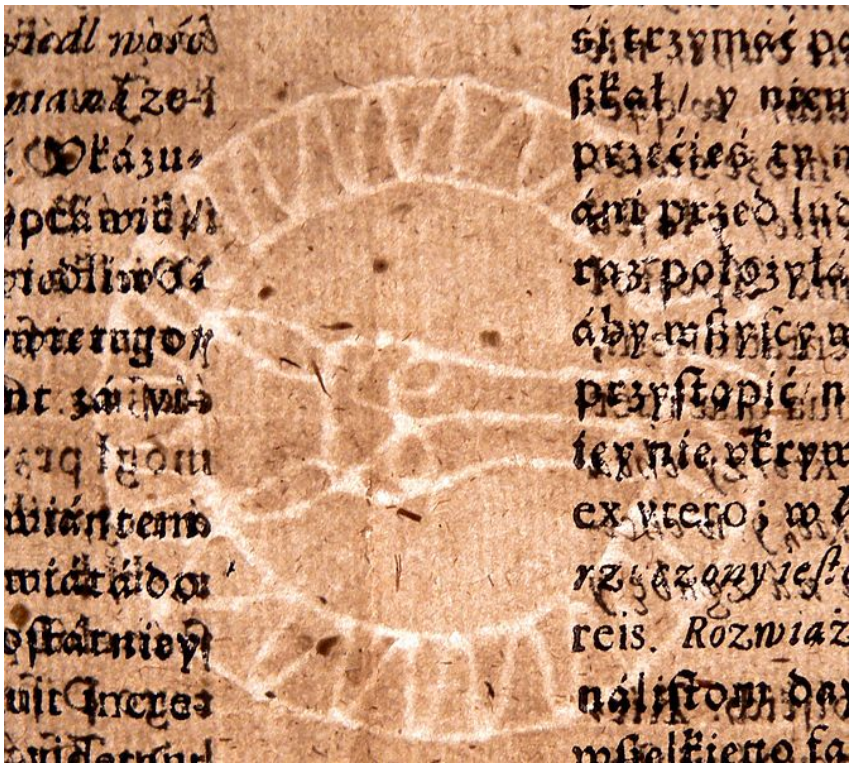
- Multimedia data indexing for purpose of content searching,
- Hardware acceleration of repository information processing,
- Creating distributed data repositories for multimedia.

Outline

Requirements

- Scalability - capable to store and maintain vast amount of data,
- Format independent - supports any data exchange format (not only those most popular),
- Security - provides access control, detects sources of illegal distribution.

Digital Watermarking



- A way of embedding information into a digital cover work in a permanent manner,

Applications of Digital Watermarking (continues)

- Content authentication:
  - Strict and robust authentication,
  - Tampering localization.
- Content reconstruction:
  - Restoration of original image content in case of tampering.
- Reversible privacy protection:
  - Blurring of selected regions of interest,
  - High-quality reconstruction for authorized users.
- Annotation watermarking:
  - Embedding meta-data or textual descriptions: e.g., subtitles, patient's medical history, doctor ID.

Applications of Digital Watermarking (continued)

- Copyright protection:
  - Embedding information about the author, owner, etc.
- Broadcast monitoring.
- Embedding low-quality soundtrack for audio-synchronization.
- Audio watermarking - e.g., echo cancellation
- Hidden communication - *Steganography*
  - File systems.
  - Executable files.
  - Web 2.0 applications - e.g., the avatars carry the messages while the text is simple smalltalk
  - Network protocols

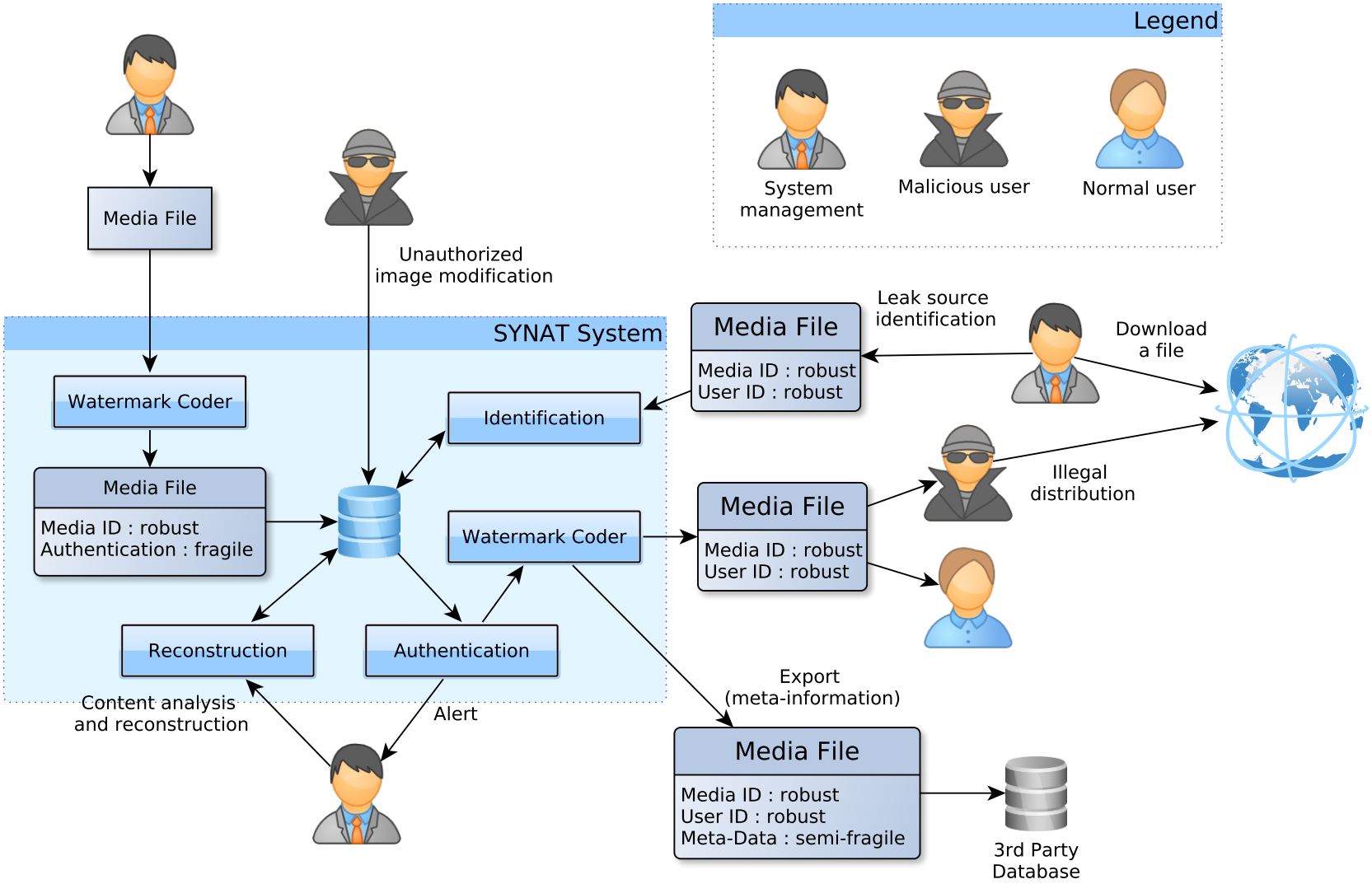
Outline

Multi-watermarking scheme

The system uses many independent co-existing watermarks:

- Robust resource identifier - identifies media files within repository,
- Digital fingerprint - identifies individual copy of a digital file,
- Authentication/reconstruction - detects unauthorized modifications/recovers original content,
- Meta-data - embeds additional information.

Use cases


Legend
System management
Malicious user
Normal user
Media File
Unauthorized image modification
SYNAT System
Watermark Coder
Media File
Media ID : robust
Authentication : fragile
Identification
Watermark Coder
Reconstruction
Authentication
Content analysis and reconstruction
Alert
Media File
Media ID : robust
User ID : robust
Leak source identification
Download a file
Media File
Media ID : robust
User ID : robust
Illegal distribution
Export (meta-information)
Media File
Media ID : robust
User ID : robust
Meta-Data : semi-fragile
3rd Party Database

Robust resource identifier

- Synchronization watermark:
  - Randomly generated pattern is embeded using Spread Spectrum technique in spatial domain,
  - SPOMF detector (Symmetric Phase-Only Matched Filter).
- Information watermark:
  - The information consists of system signature and media identifier,
  - DSSS (Direct Sequence Spread-Spectrum) method is used for watermark generation,
  - Embedding in LH and HL subbands of 2^{nd} level of DWT CDF 9/7 using Quantization Index Modulation.

Robust resource identifier Synchronization Example

Digital fingerprint (robust)

- User-specific identification signature - Randomly generated from normal distribution,
- Embedded in HH subband of 2^{nd} level of DWT CDF 9/7.
- Informed detector (higher accuracy, better performance),
- Capable of detect identity in case of colluding attack,
- Tree-based detection algorithm (slightly decreases the accuracy, but greatly increases the performance).

Robustness

- Cropping,
- Resizing,
- Lossy compression,
- Gamma correction,
- Collage attacks.

Robustness

- Cropping,
- Resizing,
- Lossy compression,
- Gamma correction,
- Collage attacks.

Robustness

- Cropping,
- Resizing,
- Lossy compression,
- Gamma correction,
- Collage attacks.

Robustness

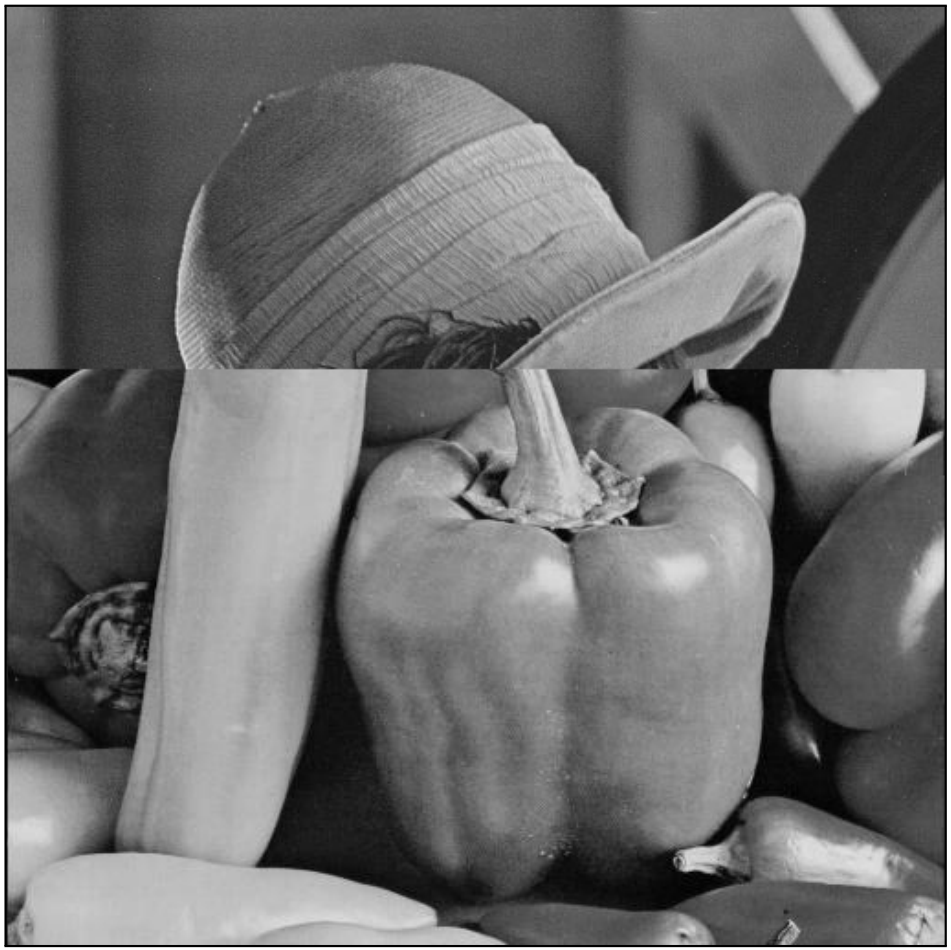
- Cropping,
- Resizing,
- Lossy compression,
- Gamma correction,
- Collage attacks.

Robustness

- Cropping,
- Resizing,
- Lossy compression,
- Gamma correction,
- Collage attacks.

Robustness

- Cropping,
- Resizing,
- Lossy compression,
- Gamma correction,
- Collage attacks.



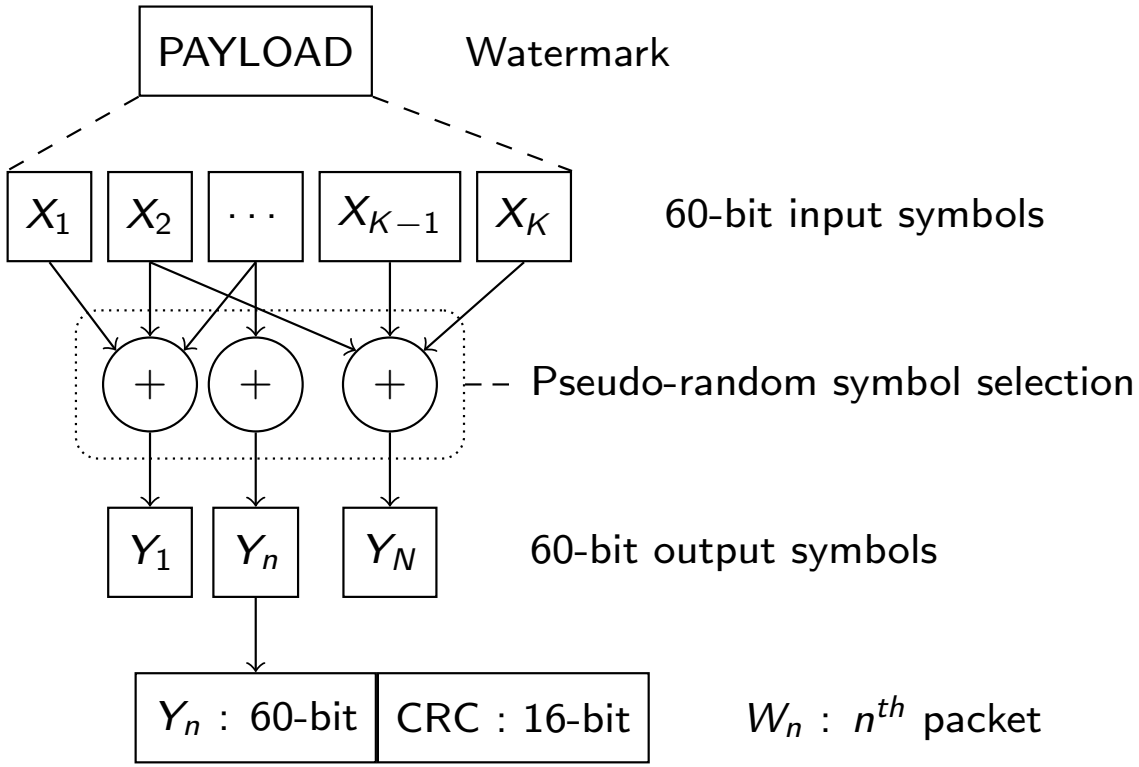
Authentication/Reconstruction

- Signature is generated by using a block content and a secret key,
- Watermark is embedded in low-frequency coefficients of DCT,
- Reconstruction reference quality depends on tampering rate,
- Robust against lossy compression,
- Watermark is generated by using Fountain Codes.

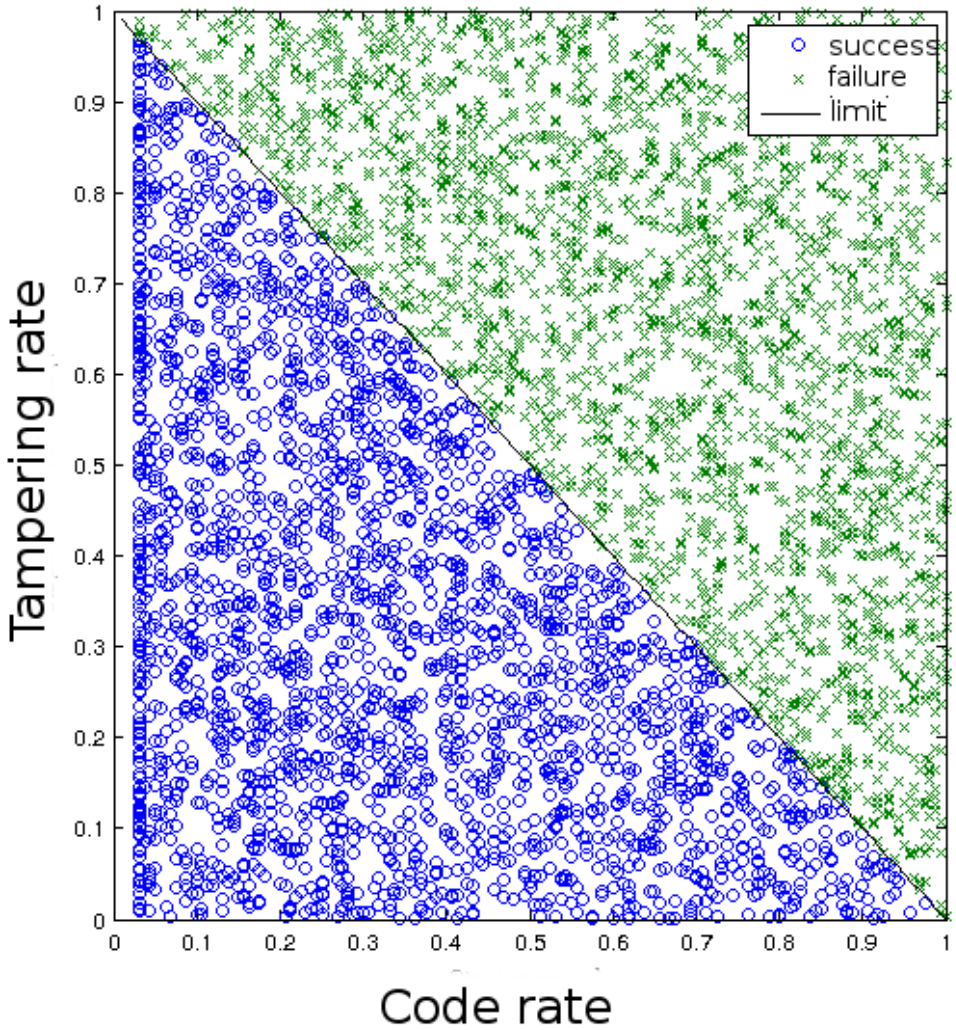
Fountain Codes

- Class of Rateless, erasure codes,
- In theory, a limitless stream of symbols can be generated from a given set of source symbols,
- Original set can be idealy recovered from any subset of symbols of size equal or slightly larger than a the number of encoding symbols.

Fountain Codes

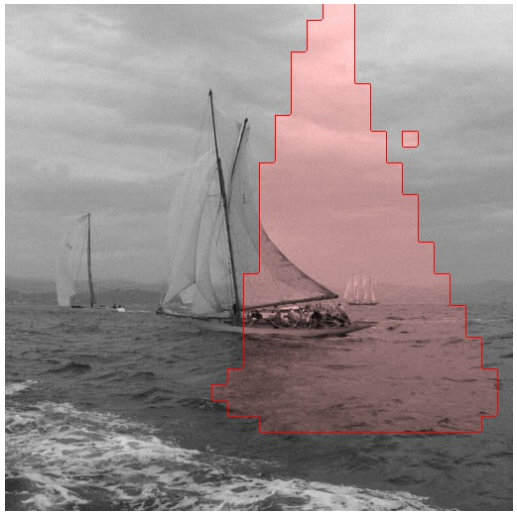


Decoding success bounds



Authentication/Reconstruction Example





Meta-data

- The idea is to embed additional meta-data related to an image directly in the image itself.
- Facilitates easier information sharing, as it suffices to share the watermarked images.
- The additional information is stored in a secure manner.
- A dedicated agent might be implemented to keep the embedded information up to date with a database.

Outline

Conclusions

- Digital watermarking is a mechanism which can be used to protect digital multimedia,
- Co-existance of many separete watermarks allows for handling various misuse scenarios.

Acknowledgements

Work financed by The National Centre for Research and Development (NCBiR) within SYNAT project no. SP/I/1/77065/10.

Q&A

- Thank you for your attention
- Questions?