Towards a Privacy-Preserving Video Codec: A Selective Encryption Framework for H.264

Guilherme Toledo

[1]Angelicam, Belo horizonte MG

`toledo@guilherme.zip`

Abstract. *We propose a backward-compatible extension to H.264/AVC that embeds a selectable privacy layer inside the bitstream. A new Supplemental Enhancement Information (SEI) payload,* `privacy_metadata`*, carries a compressed, optionally encrypted mask identifying sensitive regions such as faces and license plates. We implement the proposal as an FFmpeg bitstream filter, a decode-side video filter and a small Python pipeline that detects faces and plates and drives the encoder.*

Resumo. *Propomos uma extensão retrocompatível ao H.264/AVC que incorpora uma camada opcional de privacidade ao próprio* bitstream. *Um novo tipo de mensagem SEI,* `privacy_metadata`*, transporta uma máscara comprimida e opcionalmente cifrada, identificando regiões sensíveis como rostos e placas de veículos. Implementamos a proposta como um* bitstream filter *e um* video filter *no FFmpeg, com um pequeno* pipeline *em Python que detecta rostos e placas e alimenta o codificador.*

1. Introduction

Video surveillance is ubiquitous and, with it, the capture of personally identifiable information: faces, gait, license plates. The *Lei Geral de Proteção de Dados* (LGPD) [LGP 2018] and the General Data Protection Regulation (GDPR) [GDP 2016] both mandate *protection by design*, a requirement that does not fit well with either post-hoc blurring (which destroys evidence) or full-stream encryption (which blocks any analytics). Selective encryption of sensitive regions [Dufaux and Ebrahimi 2008, Tew and Wong 2013, Huang et al. 2015] has been studied for more than a decade, but no backward-compatible, codec-level solution has reached the H.264/AVC standard, limiting interoperability in real deployments [Padilla-López et al. 2023, Ribarić et al. 2016].

Modern capture devices already let the user select the acquisition codec. We exploit that flexibility: rather than patch privacy in after the fact, we propose a Privacy Protection Profile (PPP) that lives *inside* H.264 as a new Supplemental Enhancement Information (SEI) payload type. Our contributions are threefold:

- A PPP for H.264/AVC that defines how a binary privacy mask is compressed and carried alongside the primary video, without breaking existing decoders.
- A new SEI message type, `privacy_metadata`, that carries the mask together with encryption metadata and a key identifier.
- A reference implementation in FFmpeg: a bitstream filter (BSF) that injects the SEI, a video filter that applies obfuscation or reconstruction, and a detection pipeline released as an **open source fork**.

2. Related Work

Selective encryption of H.264. Dufaux and Ebrahimi [Dufaux and Ebrahimi 2008] used Flexible Macroblock Ordering to scramble face regions. Tew and Wong [Tew and Wong 2013] encrypted intra prediction modes of I-frames to avoid drift. Huang et al. [Huang et al. 2015] restricted intra prediction to eliminate drift errors, at the cost of coding efficiency. These methods deliver region-of-interest protection but remain outside the standard: every decoder in the field is a bespoke build, and interoperability with unmodified decoders is lost.

Encrypted-domain analytics. Tian et al. [Tian et al. 2021] perform motion detection and object tracking directly on encrypted bitstreams, showing that privacy-preserving analytics is feasible without full decryption. Our work is complementary: we standardize the carrier, not the analytics.

Visual privacy at the pixel level. Face-anonymization surveys catalogue blurring, pixelation, avatar replacement and face-swapping techniques [Ribarić et al. 2016, Padilla-López et al. 2023]. Recent generative methods swap identities with good temporal consistency [Dogan and Yüksel 2025]. All of these act on decoded pixels and, importantly, are irreversible: once the video is anonymized, the original content is gone.

Regulatory context. GDPR articles 5 and 25 require data protection by design and by default. The LGPD articles 6 and 46 mirror these duties in Brazil. Work at the Brazilian Symposium on Cybersecurity (SBSeg) has already addressed privacy in connected devices, for instance by pseudonymizing traffic in IoT networks [Pinheiro et al. 2018]; however, the codec itself has so far been left untouched.

3. Proposed Architecture

3.1. Overview

Figure 1 shows the encoding pipeline. Input video is compressed by an off-the-shelf H.264 encoder (libx264 in our prototype). In parallel, a detector produces a binary mask at macroblock granularity (16×16 pixels), aligned with H.264's coding structure. A bitstream filter, `h264_privacy`, takes the encoded stream and the raw mask, compresses the mask with run-length encoding, optionally for this initial proposal encrypts it with AES in CBC mode, and injects a `privacy_metadata` Supplemental Enhancement Information Network Abstraction Layer (SEI NAL) unit in front of each coded picture. Those units are additional information transmitted alongside the video designed to enable capable decoders to provide an improved experience to the user while still being easily ignored by legacy decoders. Three decoder modes coexist:

- **Legacy**: decoders without support for SEI payload type 215 skip the payload per H.264 Annex D and output the original video.
- **Privacy**: decoders with support but no key apply obfuscation to the masked regions (pixelation, blur or black).
- **Authorized**: decoders with the correct key bypass obfuscation and output the original video.

Figure 3 summarizes the decoder side.

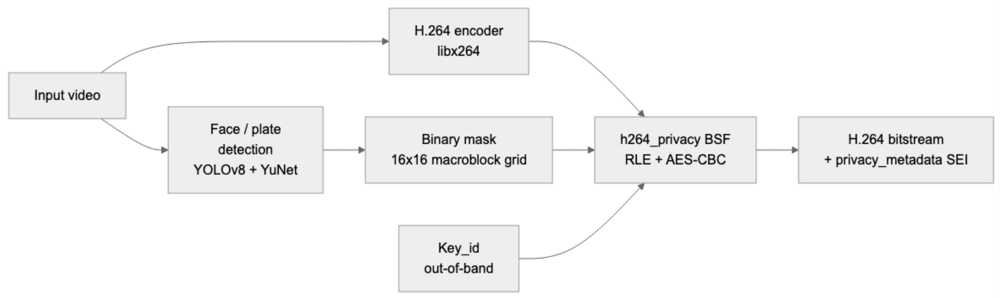

Figure 1. Encoding pipeline. Detection produces a binary mask at macroblock granularity; the `h264_privacy` BSF compresses and optionally encrypts it, then injects it as a `privacy_metadata` SEI NAL unit.

3.2. Detection and mask generation

The reference pipeline uses YOLOv8 [Jocher et al. 2023] for general object detection and vehicle localization, and YuNet [Wu et al. 2023] for face detection, with a Viola–Jones Haar cascade [Viola and Jones 2001] fallback when neither neural model is available. License plates are localized by detecting vehicles (cars, buses, trucks, motorcycles) and estimating a plate region in the lower portion of each vehicle bounding box. Detection output is binarized, downsampled to the 16×16 grid and written as raw grayscale frames, which the BSF consumes directly. Temporal tracking is not yet integrated and is left as future work; the FFmpeg project [ffm 2024] provides a natural integration point.

3.3. SEI message definition

Following H.264 Annex D conventions [ITU-T 2024], we define the SEI payload shown in Figure 2 as follows: ue(v) is an Unsigned Exponential-Golomb with variable length v, u(n) are unsigned integers of size n and b(n) are n byte aligned bits. The payload type value 215 is taken from the application-reserved range (100–254) and is currently used by no other profile in the standard.

```
privacy_metadata( payloadSize ) {
  privacy_metadata_id                      ue(v)
  privacy_metadata_cancel_flag             u(1)
  if( !privacy_metadata_cancel_flag ) {
    mask_encryption_algorithm              u(8)
    key_id                                 u(128)
    obfuscation_type                       u(8)
    encrypted_mask_data                    b(8)
  }
}
```

Figure 2. Syntax of the proposed `privacy_metadata` SEI payload.

The `mask_encryption_algorithm` byte selects the primitive: 0 for none, 1 for AES-128-CBC, 2 for AES-256-CBC. The 128-bit `key_id` identifies the key (not the key itself) so that keys can be distributed out of band, for instance via an existing PKI or key-management service. The `obfuscation_type` signals the decoder's preferred action for non-authorized playback: pixelate, blur, black fill and is left easily extensible.

The `encrypted_mask_data` block is the RLE-compressed mask, optionally encrypted under the algorithm indicated. Authenticated encryption, e.g. AES-GCM as specified in [Dworkin 2007], is a natural extension we have left for future work.

3.4. Compression and encryption

Because the mask is a sparse binary signal over a macroblock grid, RLE typically compresses it to around an order of magnitude smaller than the raw grid; the exact ratio depends on scene content. The compressed mask is then encrypted, if an algorithm is selected. Keys never travel in the stream: only the `key_id` does, decoupling the access control plane from the media plane.

3.5. Decoding

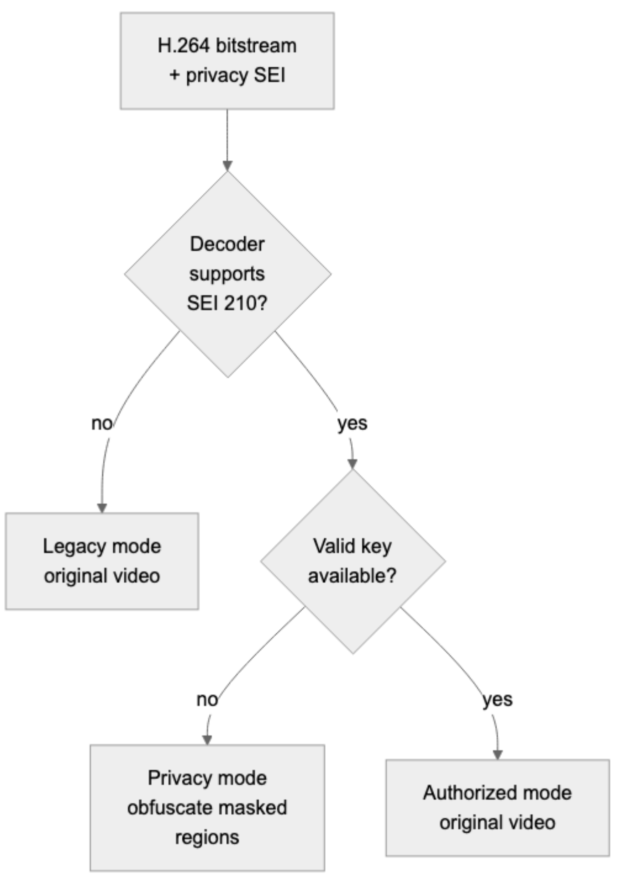

Figure 3. Decoder flow. Unknown SEI payload types are skipped by legacy decoders. Privacy-aware decoders branch on key availability.

A privacy-aware decoder runs the standard H.264 pipeline and, after picture reconstruction, inspects the `privacy_metadata` SEI. If the algorithm is zero, the mask is read directly; otherwise the decoder queries its key store with `key_id` and, on success, decrypts the mask. The masked regions are then either passed through (authorized mode) or obfuscated according to `obfuscation_type` (privacy mode). The modification is strictly local and post-decode, so neither the bitstream nor the decoded picture is altered for downstream tools such as motion estimation or analytics.

4. Implementation and Preliminary Evaluation

We implemented the PPP as an FFmpeg fork. The changes are small and concentrated:

- `libavcodec/sei.h`: a new payload type `SEI_TYPE_PRIVACY_METADATA = 215`.
- `libavcodec/h264_sei.{h,c}`: a new `H264SEIPrivacyMetadata` struct and its parser, plumbed into the existing `ff_h264_sei_decode` dispatcher.
- `libavcodec/bsf/h264_privacy.c`: a new bitstream filter(BSF) that reads raw grayscale mask frames, downsamples them to the 16×16 grid, RLE-compresses them, and injects the SEI NAL.
- `libavfilter/vf_privacy_mask.c`: a new video filter that applies obfuscation (or, given a matching passkey, no obfuscation) to regions identified by the mask.
- `scripts/`: a Python pipeline for auto-detection and for driving the encode/decode loop, together with a demo that produces a side-by-side comparison video.

Table 1 reports preliminary per-frame measurements for one 720p sequence. These figures are indicative: they come from a single scene on a developer workstation (Intel i7, NVIDIA RTX 4060) rather than from a benchmark suite, and they are intended to show that the pipeline is feasible in real time in some situations on commodity hardware, not to certify performance. The mask-compression and encryption stages are negligible; detection dominates, which is expected.

Table 1. Indicative per-frame cost on a 720p sequence. Single-scene measurements; not a benchmark.

<table>
<tr><th>Stage</th><th>GPU (ms)</th><th>CPU-only (ms)</th></tr>
<tr><td>Detection (YOLOv8 + YuNet)</td><td>15.1</td><td>205.9</td></tr>
<tr><td>Mask RLE and AES-CBC</td><td colspan="2"><1.0</td></tr>
<tr><td>SEI injection (BSF)</td><td colspan="2">≈ 0.5</td></tr>
<tr><td>Total overhead per frame</td><td>16.9</td><td>220.8</td></tr>
</table>

On the same clip, the SEI adds between 2% and 5% to the encoded bitrate, consistent with RLE behavior on masks whose coverage is a few percent of the frame. The overhead grows with coverage and scene complexity; a worst-case clip with crowded faces remains to be measured.

4.1. Regulatory fit

In Privacy mode, any decoder without a key produces a visually anonymized stream, which is the behavior the LGPD and GDPR ask of video-handling systems by default. Authorized analytics or law-enforcement review happens through the same stream with the key, without a separate archive. Legacy decoders, finally, keep working; a surveillance operator does not have to replace existing infrastructure to start tagging new recordings.

4.2. Limitations

Three limitations matter for adoption. First, detection accuracy is the floor: anything the detector misses is not protected. Second, frame-to-frame consistency of masks is not yet enforced, so brief flickers can occur at scene boundaries; a tracker [Redmon et al. 2016] is a natural addition. Third, the current encryption is CBC; authenticated encryption such as AES-GCM [Dworkin 2007] is simple to add and would close a malleability gap we do not claim to address here.

5. Conclusion and Future Work

We have proposed a Privacy Protection Profile for H.264/AVC that embeds an optional, revocable privacy layer inside the bitstream, carried by a new SEI message type. The prototype sits atop FFmpeg with small, self-contained changes and keeps legacy decoders working. The measured overhead is modest and the implementation path is straightforward, making the proposal a practical candidate for LGPD- and GDPR-aware video surveillance.

Open fronts include: extending the profile to H.265/HEVC and H.266/VVC where mask compression is likely to improve; integrating a temporal tracker to stabilize masks across frames; upgrading the primitive to authenticated encryption; and a systematic benchmark across diverse surveillance scenes, including adversarial cases, before any claim beyond feasibility is made.

Furthermore, addition of more SEI messages can further the capabilities of the current proposed approach. For example, inverting the desing and maintaining the original video stream as the censored counterpart and providing an encrypted bitstream with the original data that the decoder will use to replace the mask with.

Acknowledgments

The author of this paper would like to thank Angelicam for their ongoing support on the development of this paper and acknowledge the use of GitHub Copilot during the implementation of the proof-of-concept defined above.

References

(2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation). Official Journal of the European Union, L 119.

(2018). Lei n. 13.709, de 14 de agosto de 2018 – Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brazil.

(2024). FFmpeg: A complete, cross-platform solution for recording, converting and streaming audio and video. Project documentation.

Dogan, S. and Yüksel, A. F. (2025). Assessing the use of face swapping methods as face anonymizers in videos. arXiv:2505.20985.

Dufaux, F. and Ebrahimi, T. (2008). Scrambling for privacy protection in video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1168–1174.

Dworkin, M. (2007). Recommendation for block cipher modes of operation: Galois/Counter Mode (GCM) and GMAC. Technical Report NIST SP 800-38D, National Institute of Standards and Technology.

Huang, H., Wang, H., and Kwong, S. (2015). Restricted H.264/AVC video coding for privacy protected video scrambling. *Journal of Visual Communication and Image Representation*, 31:121–128.

ITU-T (2024). Recommendation H.264: Advanced video coding for generic audiovisual services. Technical report, International Telecommunication Union. Edition 15.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). YOLO by Ultralytics (yolov8). Ultralytics open-source release.

Padilla-López, J. R., Chaaraoui, A. A., and Flórez-Revuelta, F. (2023). A review on visual privacy preservation techniques for active and assisted living. *Multimedia Tools and Applications*, 82:31–61.

Pinheiro, A. J., Burgardt, C. A. P., and Campelo, D. R. (2018). Preservando a privacidade na Internet das Coisas com pseudônimos usando SDN. In *Anais do XVIII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg)*, pages 457–464, Porto Alegre, RS. SBC.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

Ribarić, S., Ariyaeeinia, A., and Pavešić, N. (2016). De-identification for privacy protection in multimedia content: A survey. *Signal Processing: Image Communication*, 47:131–151.

Tew, Y. and Wong, K. (2013). Privacy region protection for H.264/AVC by encrypting the intra prediction modes without drift error in I frames. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2912–2916.

Tian, X., Zheng, P., and Huang, J. (2021). Robust privacy-preserving motion detection and object tracking in encrypted streaming video. *arXiv preprint arXiv:2108.13141*.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 511–518.

Wu, W., Peng, H., and Yu, S. (2023). YuNet: A tiny millisecond-level face detector. *Machine Intelligence Research*, 20(5):656–665.