Terms

Ray Bernard wrote in his article that today, "frame" is even more accurate than "image" because the image most accurately refers not to what the camera sensor saw, or to the full or partially compressed video image transmitted, but to what the end user sees when viewing a video clip or still image.

I-frames

https://www.securityinfowatch.com/video-surveillance/article/21124160/real-words-or-buzzwords-h264-and-iframes-pframes-and-bframes-part-2article I-frame is short for intra-coded frame, meaning that the compression is done using only the information contained within that frame, the way a JPEG image is compressed. Intra is Latin for within. I-frames are also called keyframes, because each one contains the full image information. This is spatial compression.

P-Frames

P-frame is short for predicted frame and holds only the changes in the image from the previous frame. This is temporal compression. Except for video with high amounts of scene change, the approach of combining I-frames and P-frames can result in compression levels between 50

B-Frames

B-frame is short for bidirectional predicted frame, because it uses differences between the current frame and both the preceding and following frames to determine its content. Figure 1 from Wikipedia shows the relationships between the frame types.

Group of Pictures (GOP) / Group of VidePictures (GOV)

These two terms come from the MPEG video compression standards. A Group of Pictures begins with an I-frame, followed by some number of P-frames and B-frames. Most security video documentation uses GOP Length and GOV Length interchangeably, which only causes confusion if you don't know about it and the camera settings say "GOV Length" while the VMS settings say "GOP Length." So, a shorter GOP length results in more I-frames, a longer GOP length results in fewer I-frames and greater compression.

Difference

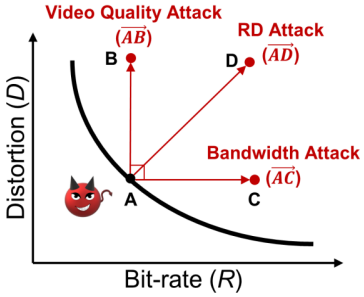
I-frames are the least compressible but don't require other video frames to decode. P-frames can use data from previous frames to decompress and are more compressible than I-frames.

B-frames can use both previous and forward frames for data reference to get the highest amount of data compression.

RoVISQ attacks:

Attack Scenarios `https://arxiv.org/pdf/2203.10183.pdf`

RoVISQ attacks are targeted towards video data generated by front-end sources (e.g., smartphones and surveillance cameras) that is sent to back-end user(s) through a video compression pipeline.



Video Quality Attack increases the distortion D at a given bit-rate, therefore, adding unwanted noise to the video content and reducing the visual quality for viewers. This attack is particularly advantageous when the media server administrator is monitoring the network bandwidth in real time. In this scenario, the service provider can detect anomalies in the bitrate, but the proposed distortion attack remains stealthy.

Bandwidth Attack is formulated to increase bit rate R at a given distortion level. As a result, the compression rate of the video encoder degrades and the amount of data transfer on the underlying network channel increases. This prevents legitimate users from successful communication with the streaming server and induces a high latency.

RD Attack combines the capabilities of the above two attacks by simultaneously targeting R and D to cause a high latency and video distortion. As a result, the back-end users suffer from low-quality or denial-of-service. If the media server lowers the video resolution to reduce network traffic, the RD attack is further exacerbated.

Compression-Robust Classifier Attack Our final attack manipulates the classification result of the decoded video in scenarios where the downstream task uses a DNN-based video recognition. This attack is particularly challenging as the video coding framework inherently invalidates most adversarial examples using DNN-based temporal coding. As such, we carefully craft an optimization problem that generates perturbations that are robust to video coding. We propose two variants of the classifier attack, i.e., targeted and untargeted. Targeted attacks misguide the video classifier to a particular class while untargeted attacks subvert the models to predict any of the incorrect classes.

In cryptography, a watermarking attack is an attack on disk encryption methods where the presence of a specially crafted piece of data can be detected by an

attacker without knowing the encryption key.

The main idea of the proposed methods is to move part of the watermark data from the inter-predicted block in Bframe to a block in I/P frame it depends on. As the result the watermark is reintroduced to the inter-predicted macro-block in B frame during decoding and is not affected by I/P frame removal or residual zeroing.

The first attack is a selective removal of intra-coded(I) and predicted (P) frames during transcoding. Most frames are bi-directional predicted (B) frames in typical High Pro[U+FB01]le Group Of Pictures (GOP), so if the watermark's data is concentrated in I/P frames it will be removed with minor impact on quality. This is true speci[U+FB01]cally for 60 fps video.So the watermark needs to be equally distributed between all frame types.

The second method: ensures mark propagation from I/P to Bframes by dependency analysis. Watermark data is embedded into I/P frame blocks, which affect only a small amount of blocks in B frame. During decoding I/P frame block is copied to dependent blocks in B frames. So if I/P frames are removed,then mark is still extractable from B frames. Small number of dependencies in each frame limits error ampli[U+FB01]cation. The limit of the modi[U+FB01]ed blocks per frame is set for quality control.