

Resynchronization Properties of Arithmetic Coding

Peter W. Moo and Xiaolin Wu*

Arithmetic coding is a popular and efficient lossless compression technique that maps a sequence of source symbols to an interval of numbers between zero and one.¹ In arithmetic coding, an entire source sequence is mapped to a single code stream. Therefore, a single error in an arithmetic code stream often causes error avalanches at the decoder, rendering the decoded code stream useless. In this abstract, we consider the important problem of decoding an arithmetic code stream when an initial segment of that code stream is unknown. We call decoding under these conditions *resynchronizing* an arithmetic code.

This problem has importance in both error resilience and cryptology. If an initial segment of the code stream is corrupted by channel noise, then the decoder must attempt to determine the original source sequence without full knowledge of the code stream. In this case, the ability to resynchronize helps the decoder to recover from the channel errors. But in the situation of encryption one would like to have very high time complexity for resynchronization. We have recently proposed a technique for real-time image/video encryption.² The basic idea is to encrypt only a tiny beginning part of a very long arithmetic code stream of image or video, and leave the large remainder of the compressed file unencrypted. This can be considered as a joint compression and encryption scheme with negligible extra cost for encryption. This strategy only works if the time complexity for resynchronizing an arithmetic coder is prohibitively high. Thus our results on resynchronization have dual implications.

In this paper we consider the problem of resynchronizing simple arithmetic codes. This research lays the groundwork for future analysis of arithmetic codes with high-order context models. In order for the decoder to achieve full resynchronization, the unknown, initial b bits of the code stream must be determined exactly. When the source is approximately i.i.d., the search complexity associated with choosing the correct sequence is at least $O(2^{b/2})$. Therefore, when b is 100 or more, the time complexity required to achieve full resynchronization is prohibitively high.

To partially resynchronize, the decoder must determine the coding interval after b bits have been output by the encoder. For a stationary source and a finite-precision static binary arithmetic coder, the complexity of determining the code interval is $O(2^{2s})$, where the precision is s bits.

*Dept. of Computer Science, Univ. of Western Ontario, London, ON, N6A 5B7. Email: {moo,wu}@csd.uwo.ca

¹T.C Bell, J.G. Cleary and I.H. Witten, *Text Compression*. Prentice Hall, 1990.

²X. Wu, P.W. Moo, and P. Yu, "Joint image/video compression and encryption via high-order conditional entropy coding of wavelet coefficients and chaotic systems," submitted to *1999 IEEE International Conference on Multimedia Computing and Systems*.