Identification scheme for romanized Indian languages from their plain and ciphered bit stream*

Shri Kant †
Coordinator, Joint Cipher Bureau
Department of Defence R & D
M. G. Road, Metcalfe House
Delhi 110054, India

Veena Sharma
Neelam Verma
Scientific Analysis Group
Defense Research and
Development Organization
Metcalfe House Complex
Delhi 110054, India

B. K. Dass
Department of Mathematics
University of Delhi
Delhi 110007, India

Abstract

Identification of the Indian languages, when they are communicated in their plain bit stream after Romanizing their script has been dealt. An Attempt has also been made to identify them from their enciphered bit stream obtained through standard encryption schemes. In this context plain and cipher bit stream of four Indian languages viz. Hindi, Punjabi, Oriya and Bengali have been studied. A novice method proposed earlier [6] has been extended for extraction of statistical features. Several other feature extraction and features selection technique have been used for experimenting with four classifiers and finally the results are summarized. Maximum Likelihood Classifier (MLC) has performed better than Minimum Distance Classifier (MDC), Linear Statistical Classifier (LSC) and Piecewise Linear Classifier (PLC) in terms of performance accuracy and consistency.

Keywords and phrases : *Romanized Indian languages, enciphered bit stream, feature extraction, binary to real conversion, identification and classifier performance.*

*This paper was presented at '*Pre-ICM International Convention on Mathematical Sciences*' held at Department of Mathematics, University of Delhi, 18-20 December, 2008.

†*E-mail*: shrikant.ojha@gmail.com

Journal of Discrete Mathematical Sciences & Cryptography
Vol. 13 (2010), No. 4, pp. 329–345