Performance Optimization on Model Synchronization in Parallel Stochastic Gradient Descent Based SVM - Results

Vibhatha Abeykoon, *vlabeyko@iu.edu*, *Indiana University Bloomington*
Geoffrey Fox, *gcf@iu.edu*, *Indiana University Bloomington*
Minje Kim, *minje@iu.edu*, *Indiana University Bloomington*

I. Experiments

Term MSF : Model Synchronizing Frequency, we have included both MSF and Block Size, MSF = 1 resembles to the experiment done corresponding to block size 1. This is clearly mentioned with detail in the paper. In the paper we have only used the block size term in the graphs.

II. Summary of Cross Validation Accuracy Variation with MSF

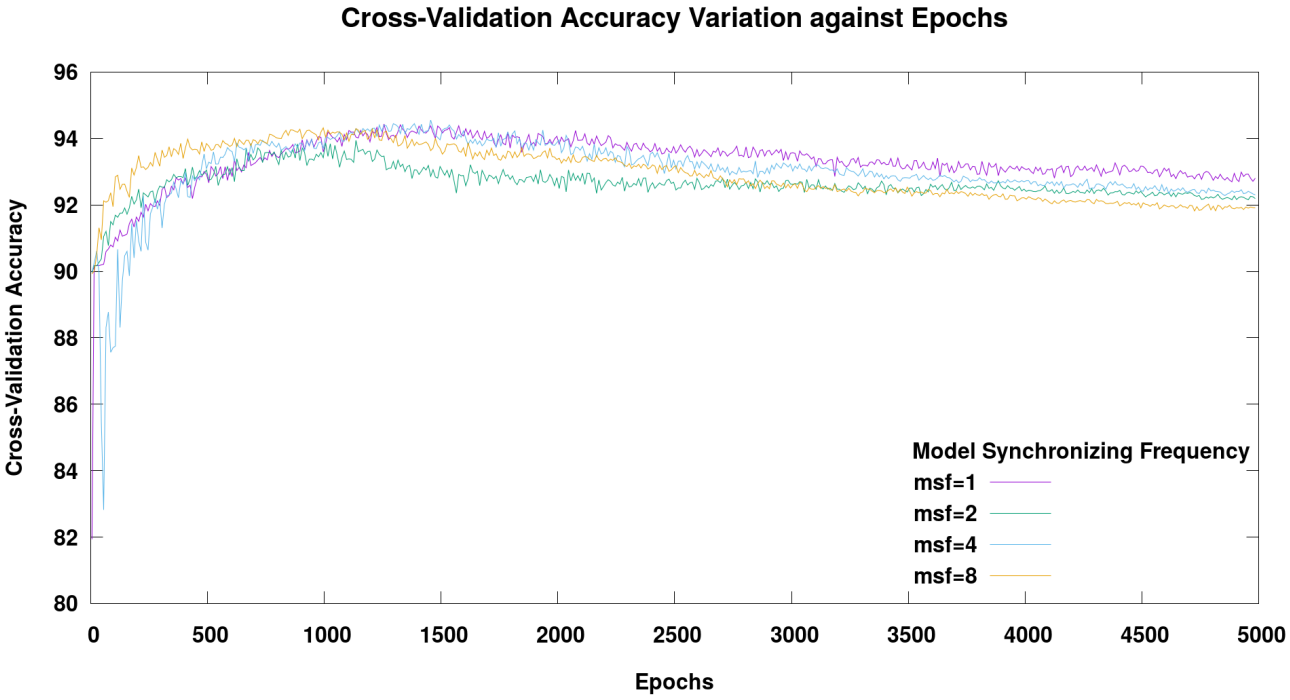

Fig. 1. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for MSF = [1,2,4,8]

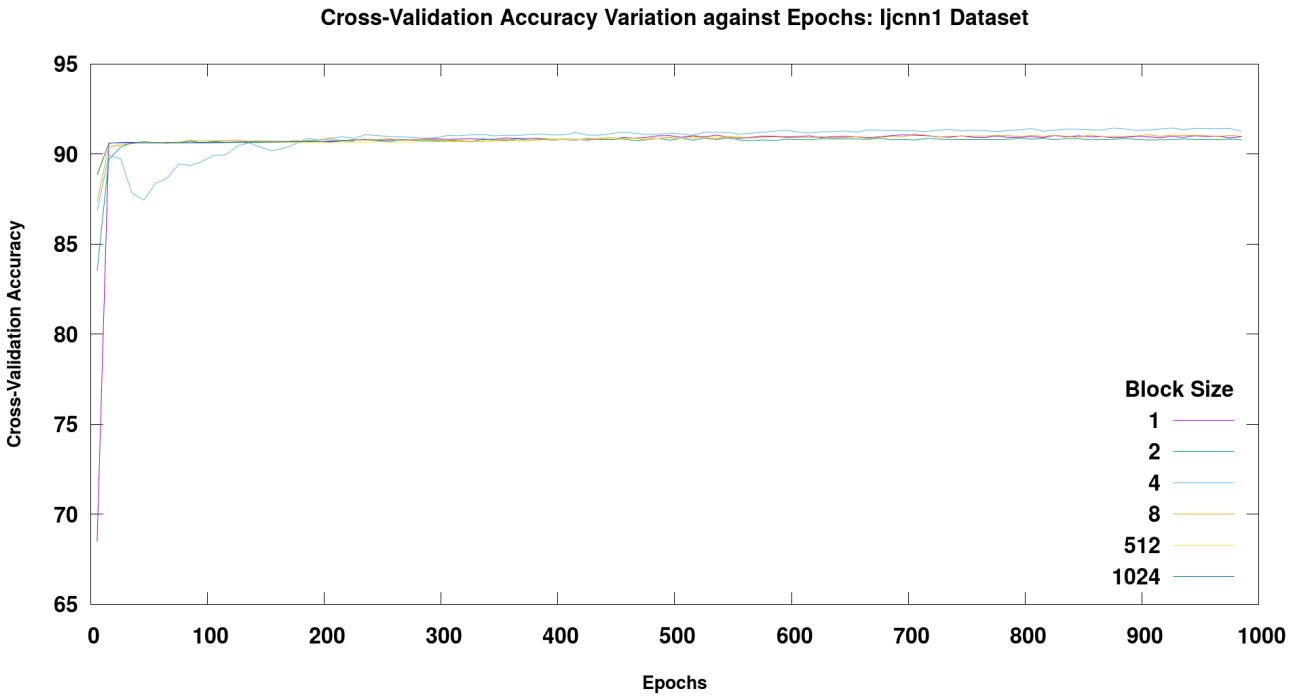

Fig. 2. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for MSF = [1,2,4,4096]

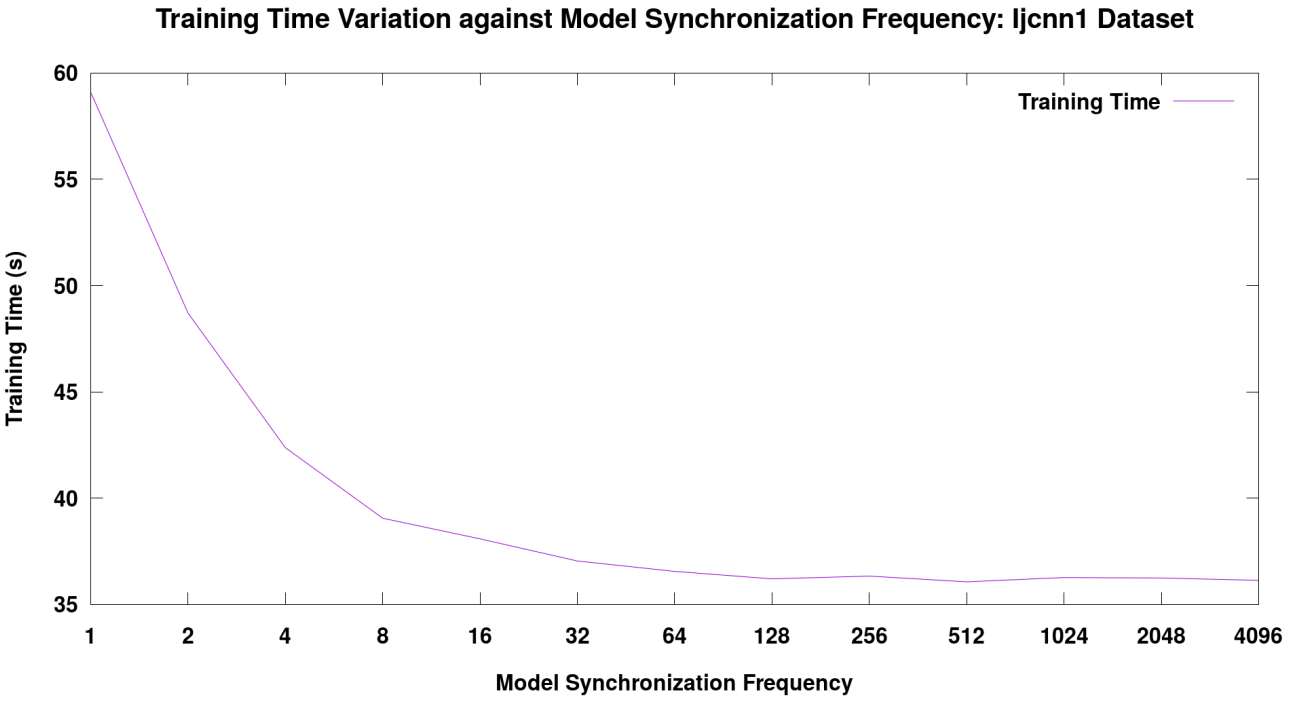

Fig. 3. Training Time Variation with Model Synchronization Frequency of Ijcnn1 Dataset with Variable MSFs

Fig. 4. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for MSF = [1,2,4,8]

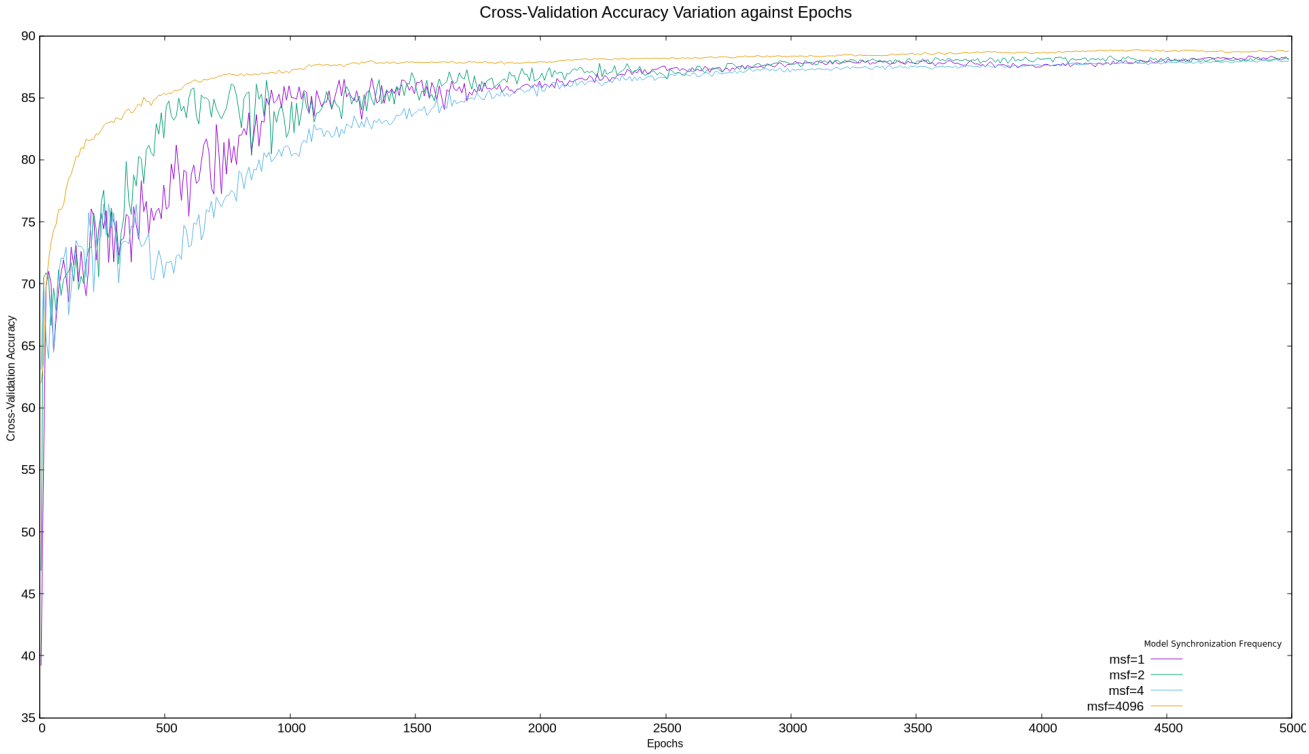

Fig. 5. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for MSF = [1,2,4,4096]

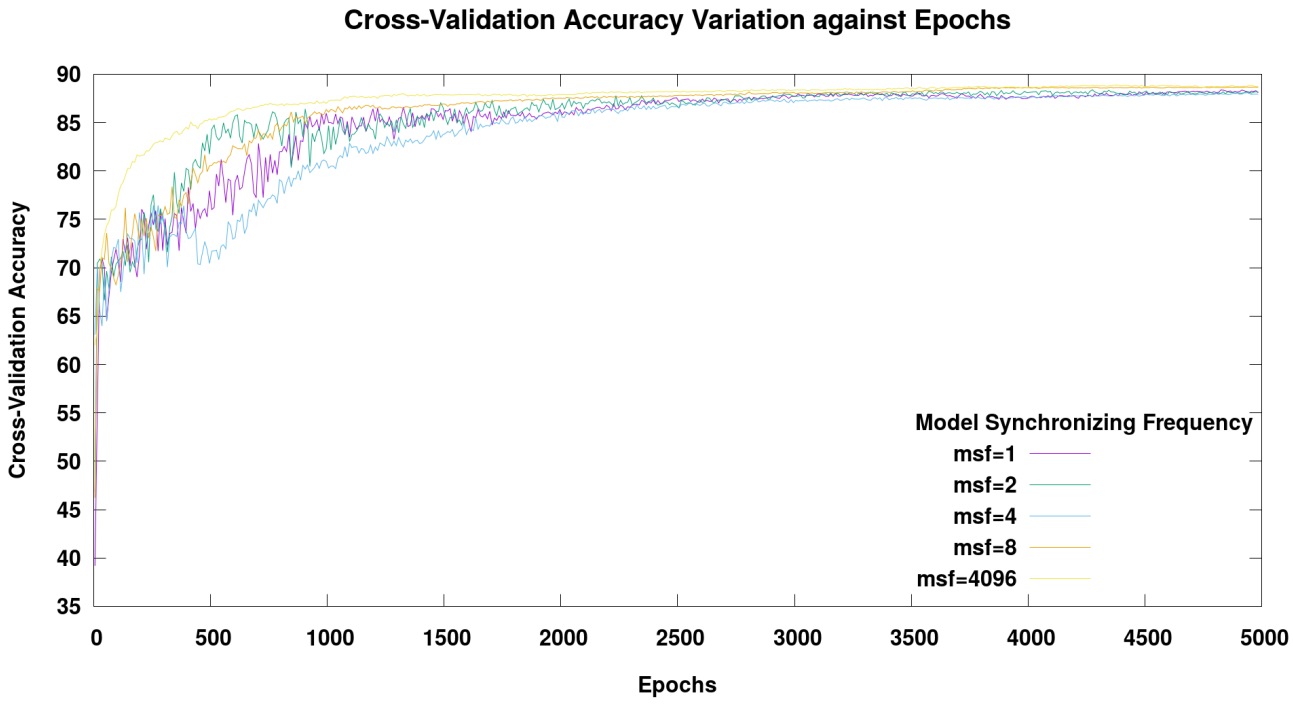

Fig. 6. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for MSF = [1,2,4,8,4096]

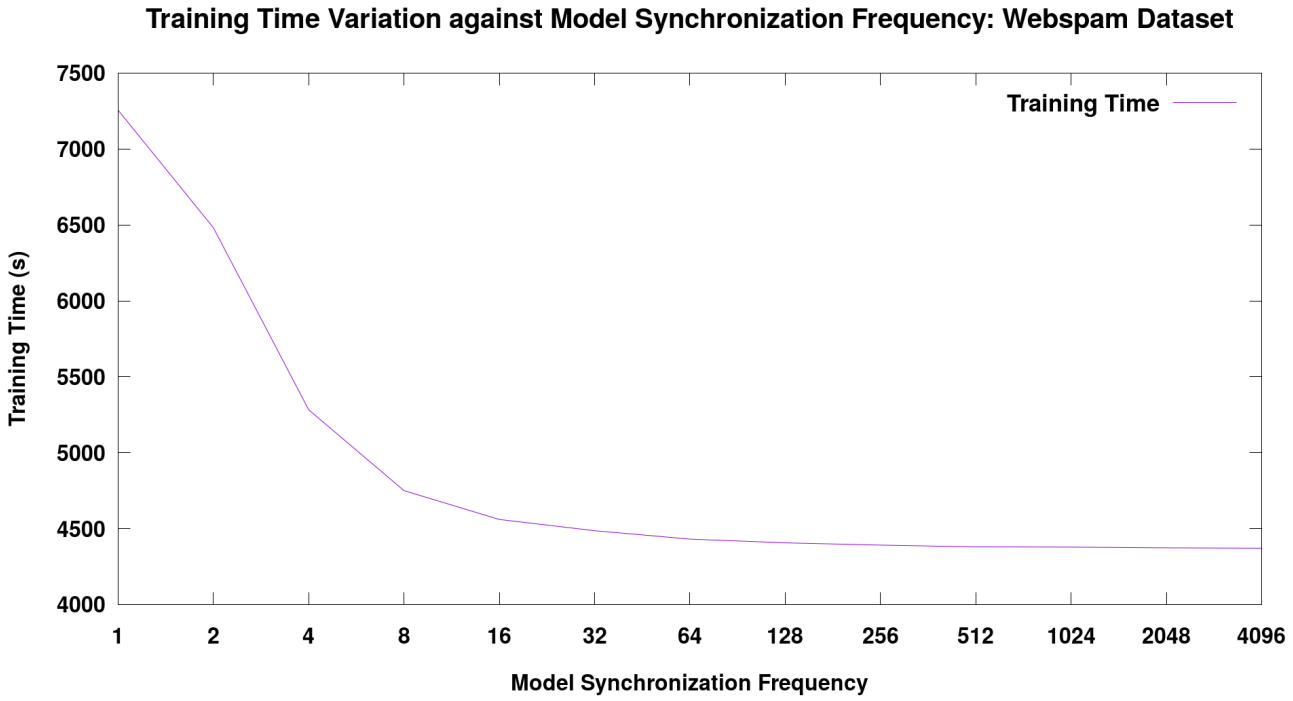

Fig. 7. Training Time Variation with Model Synchronization Frequency of Webspam Dataset with Variable MSFs

III. All Experiments on Ijcnn1, Webspam and Epsilon Datasets on All MSFs

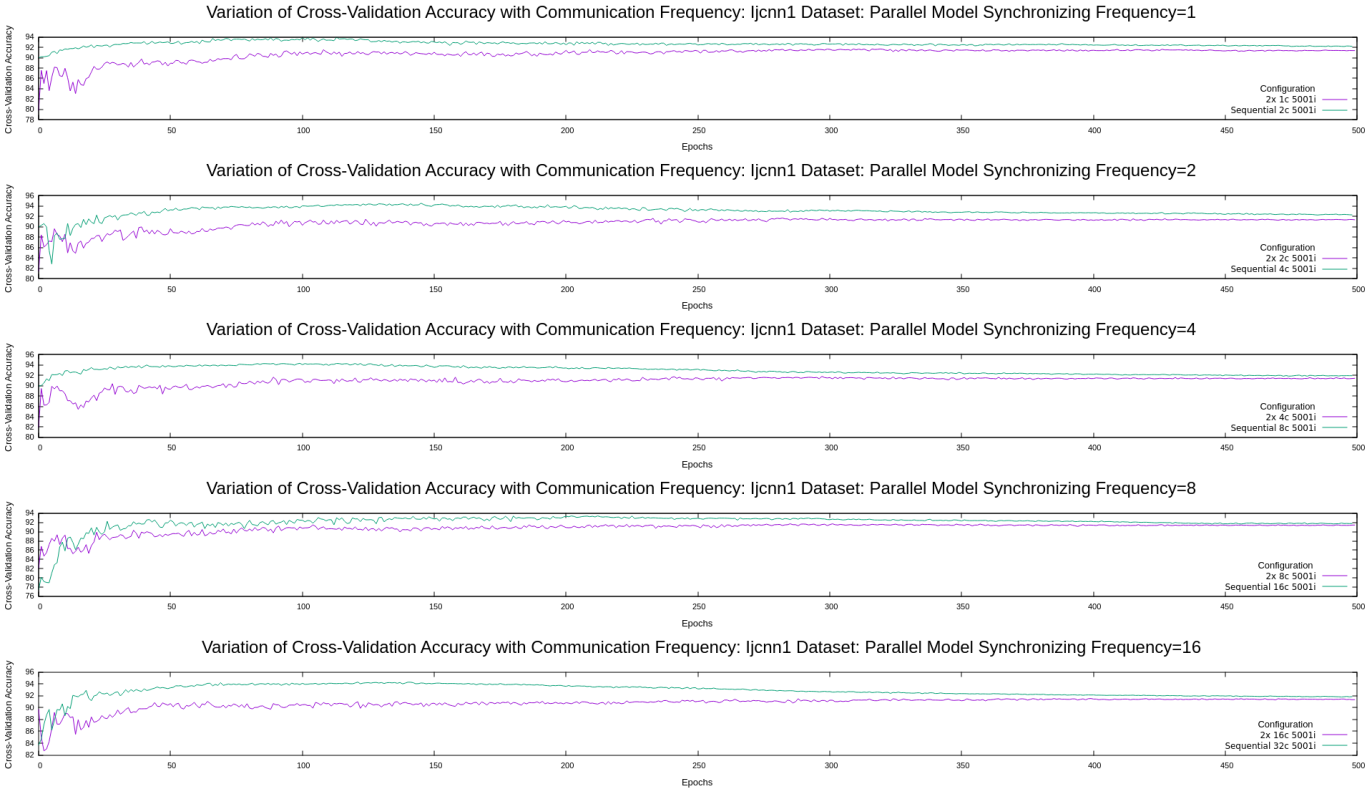

Fig. 8. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 2 MSF = [1,2,4,8,16]

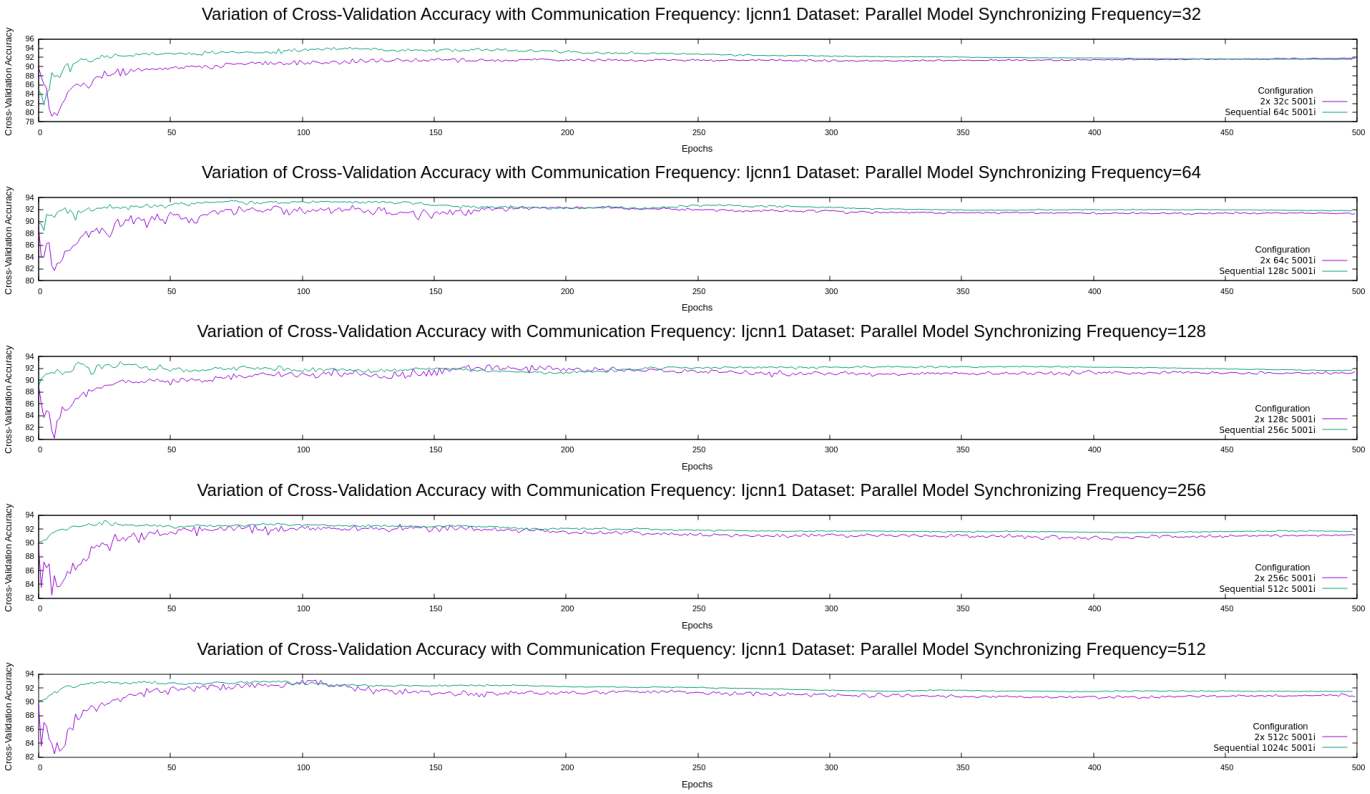

Fig. 9. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 2, MSF = [32,64,128,256,512]

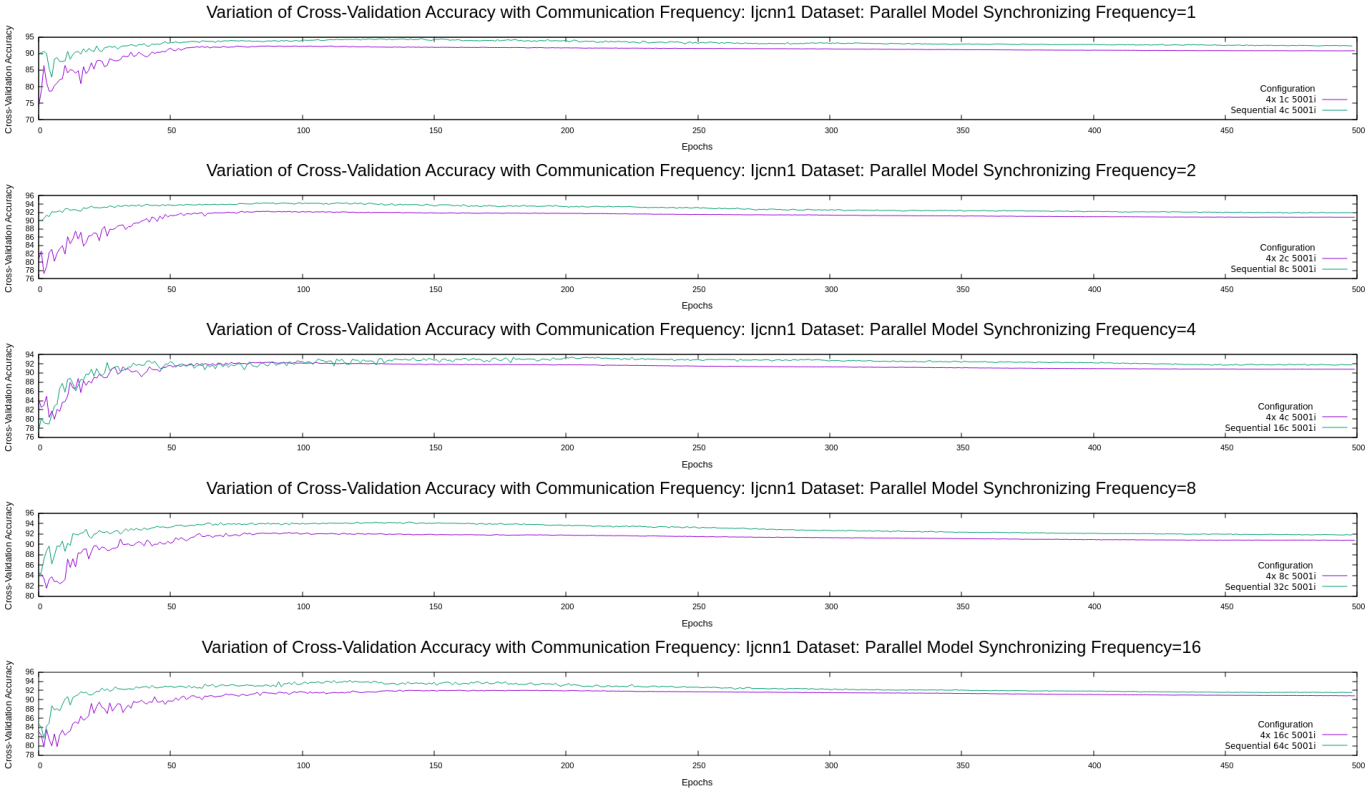

Fig. 10. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 4 MSF = [1,2,4,8,16]

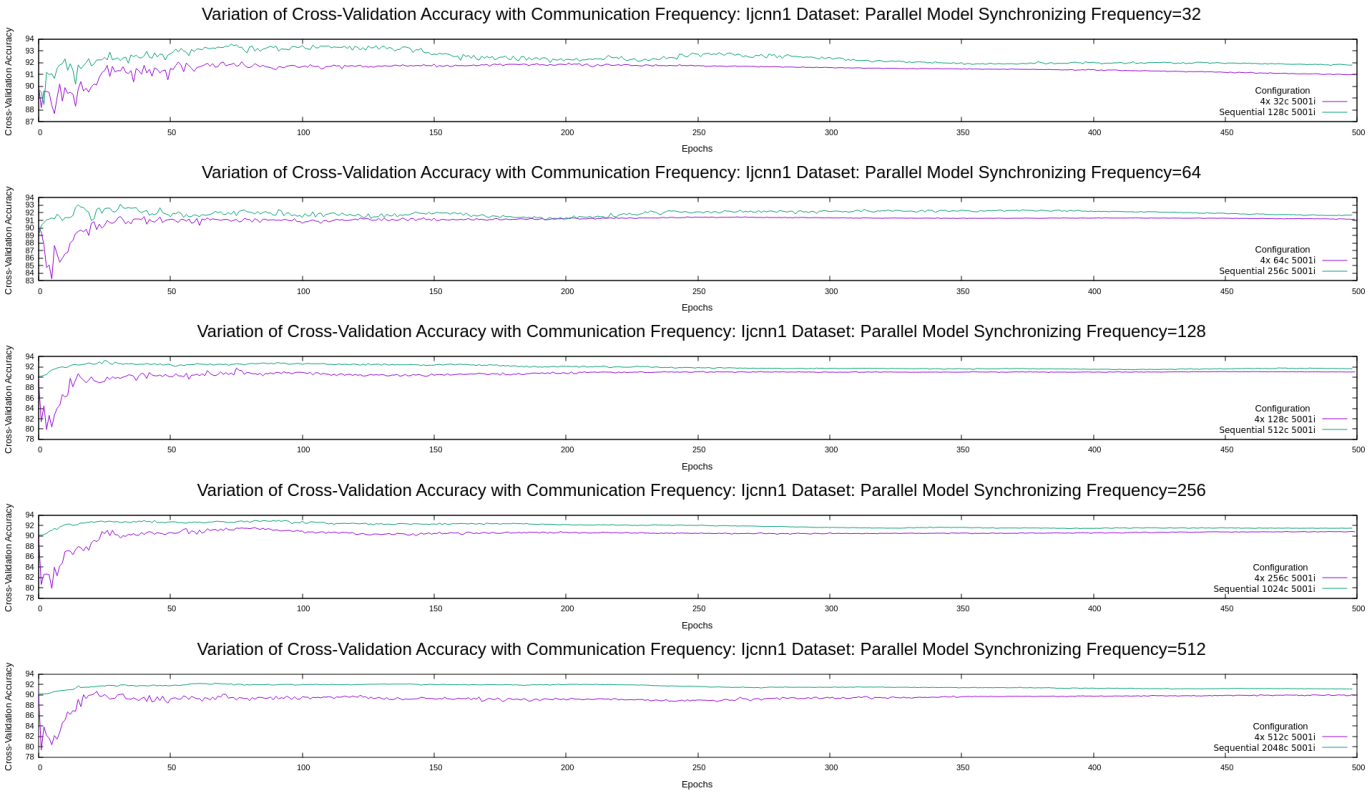

Fig. 11. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 4, MSF = [32,64,128,256,512]

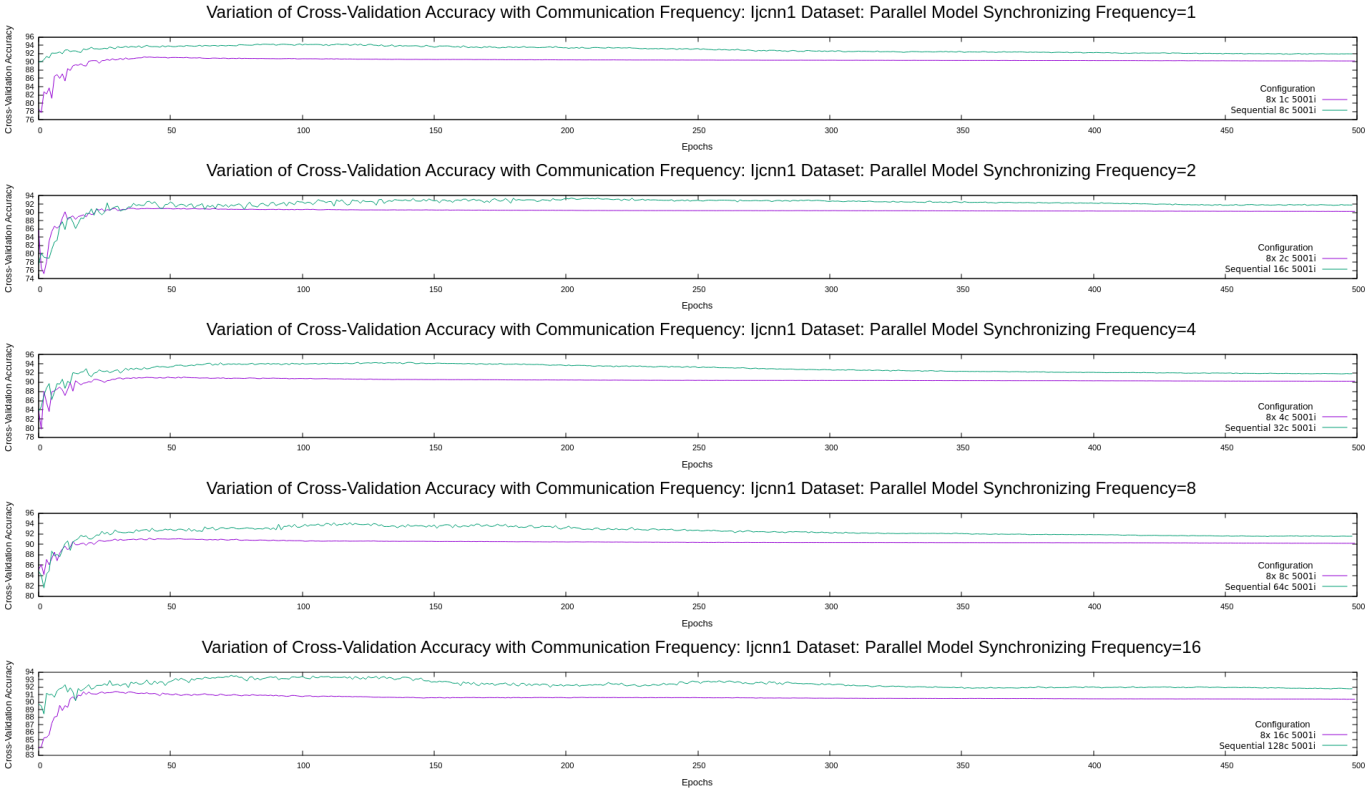

Fig. 12. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 8 MSF = [1,2,4,8,16]

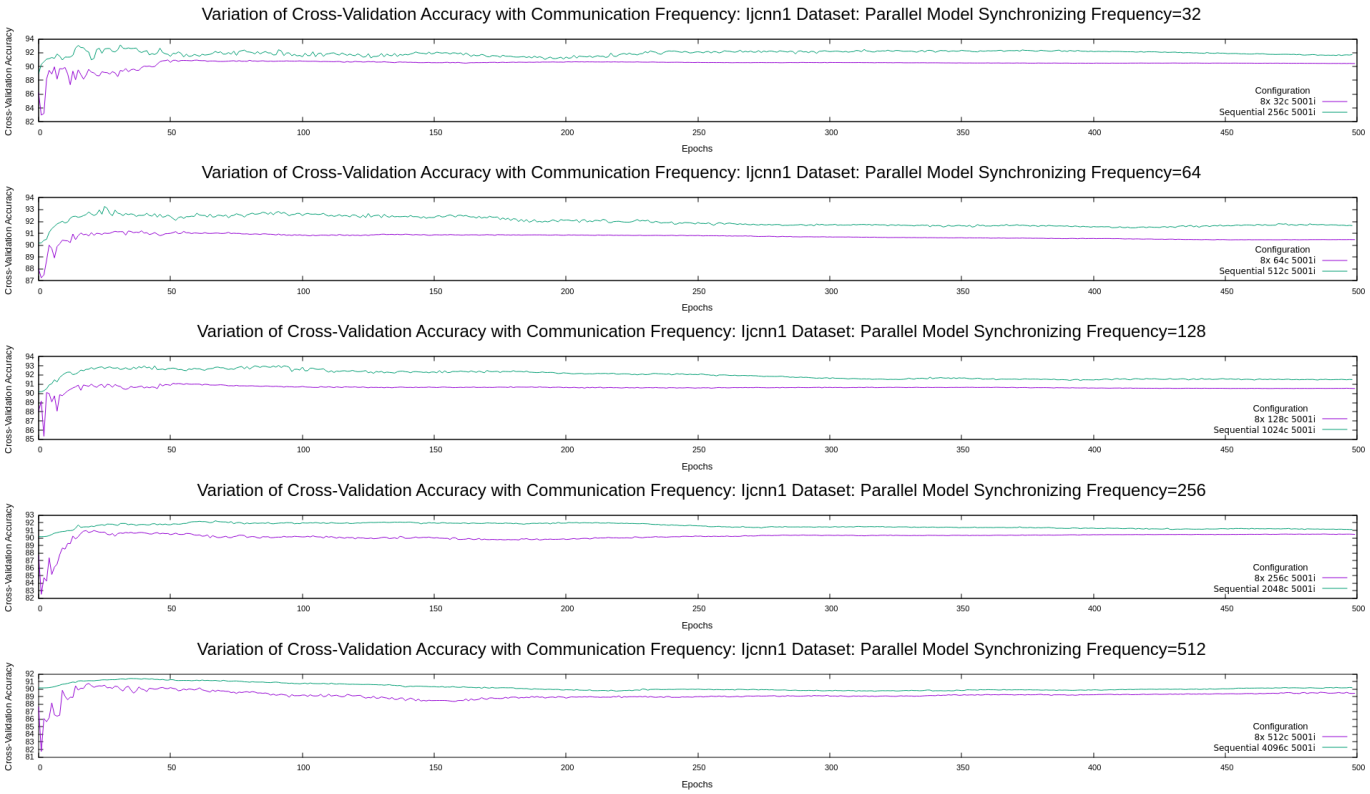

Fig. 13. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 8, MSF = [32,64,128,256,512]

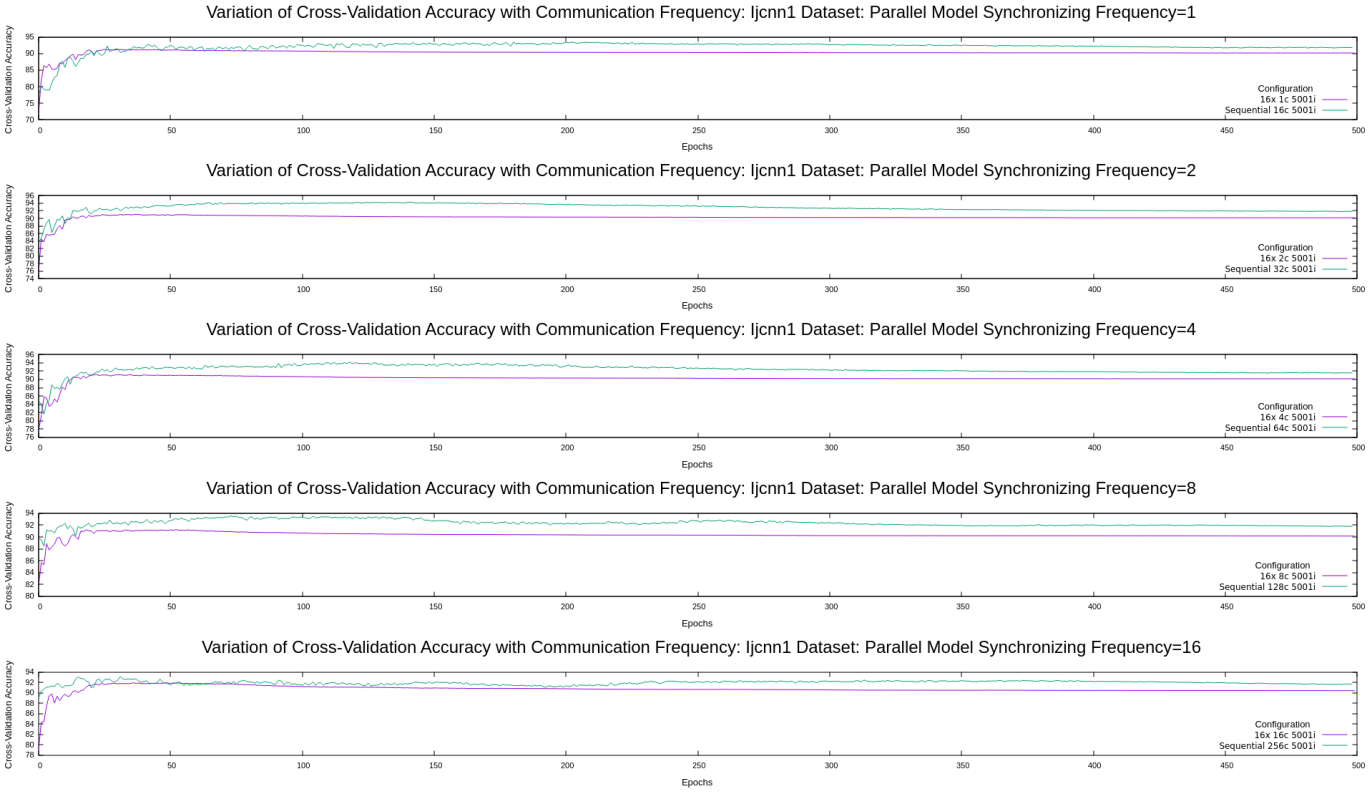

Fig. 14. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 16 MSF = [1,2,4,8,16]

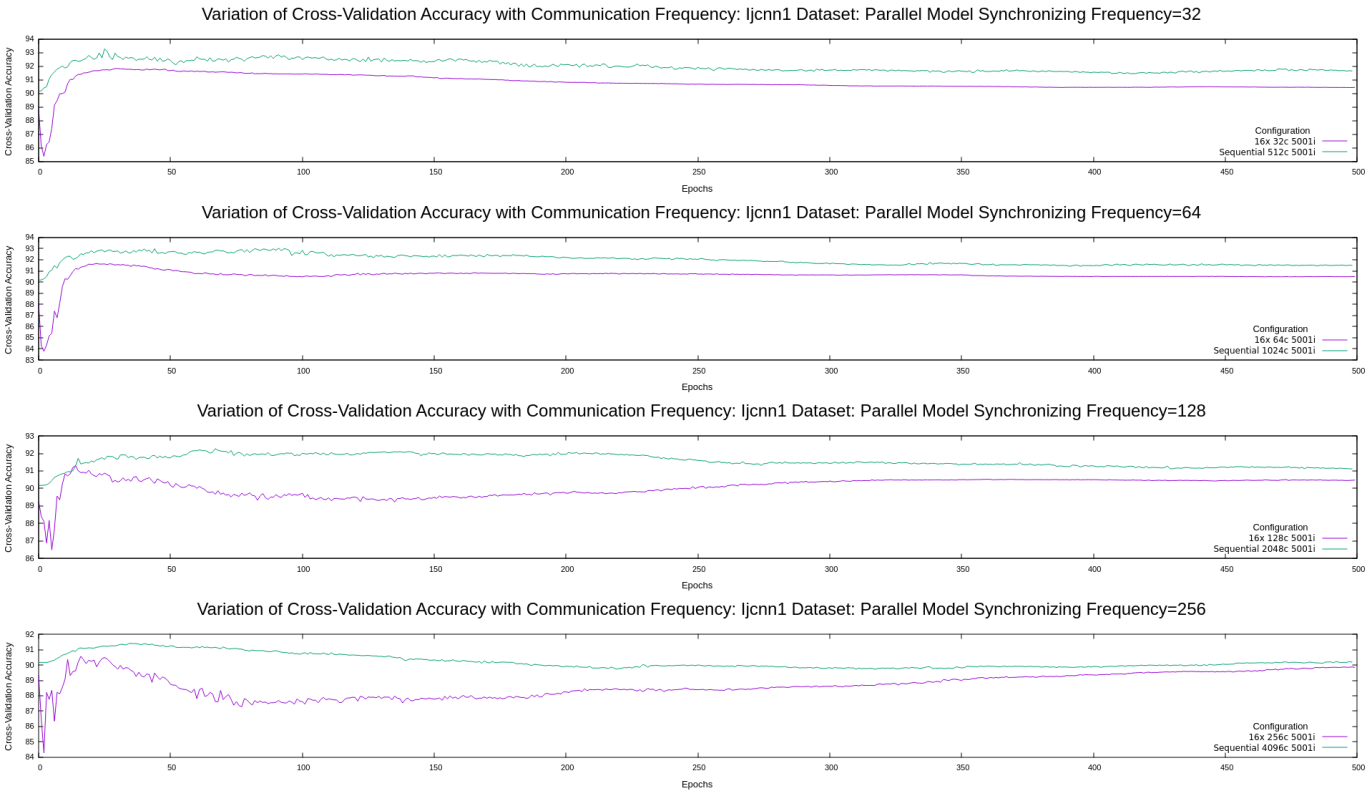

Fig. 15. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 16, MSF = [32,64,128,256]

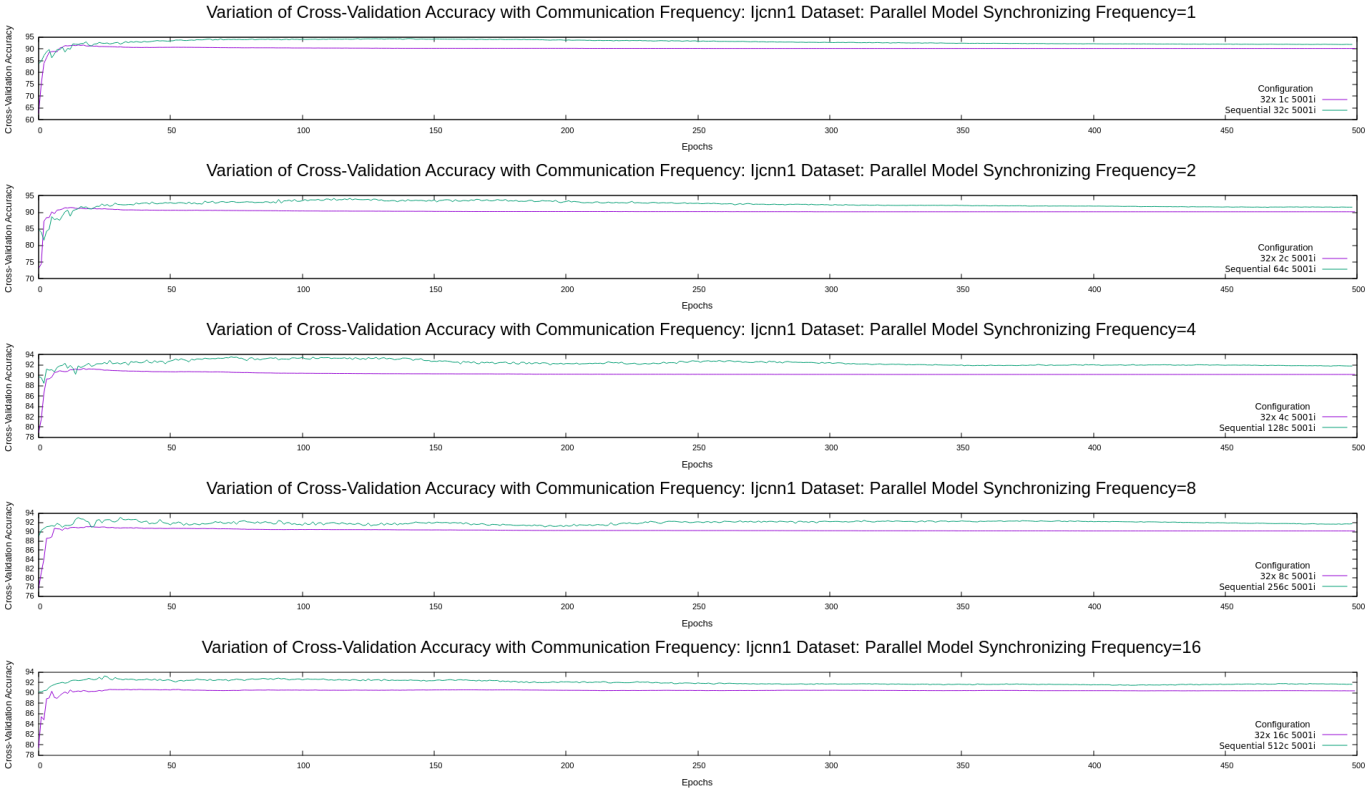

Fig. 16. Cross Validation Accuracy Variation with Model Synchronization Frequency of Ijcnn1 Dataset for Parallelism 32 MSF = [1,2,4,8,16]

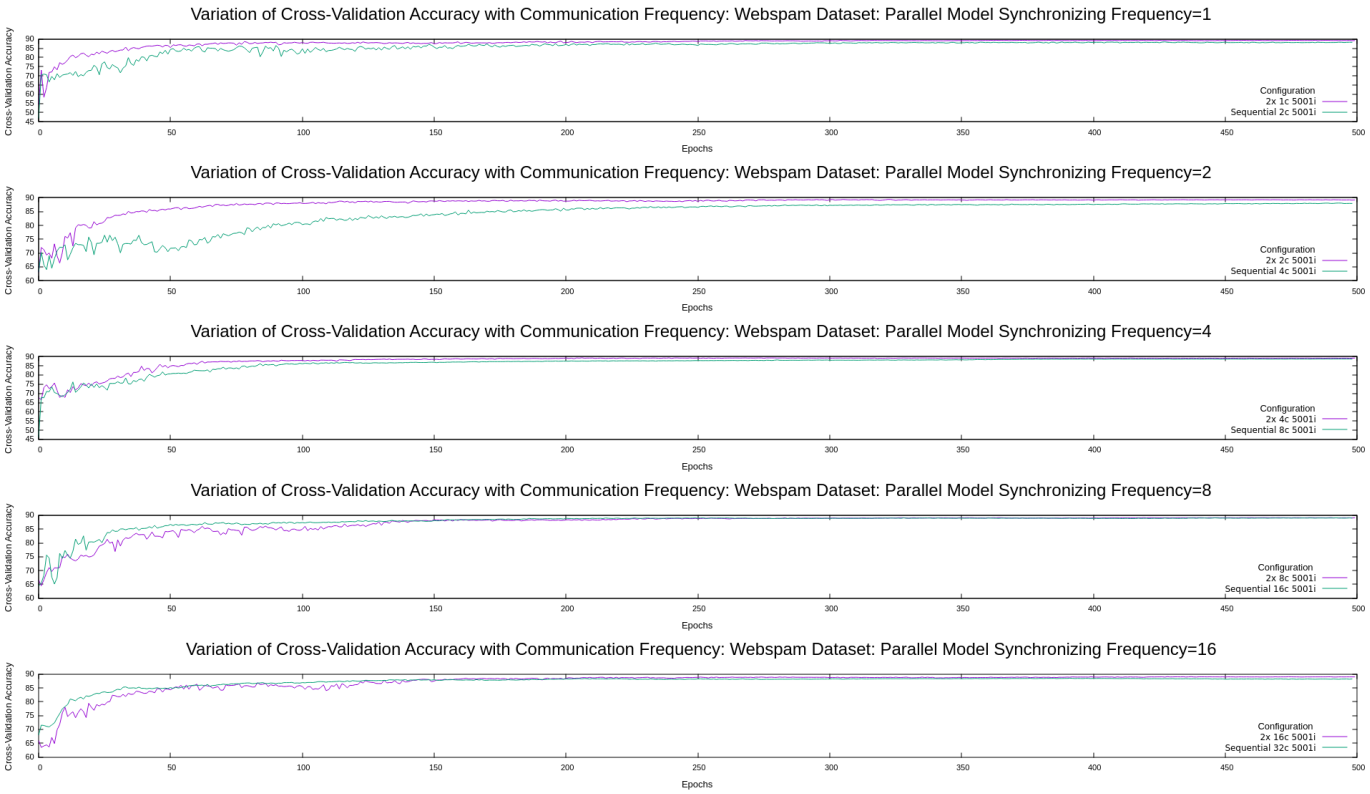

Fig. 17. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 2 MSF = [1,2,4,8,16]

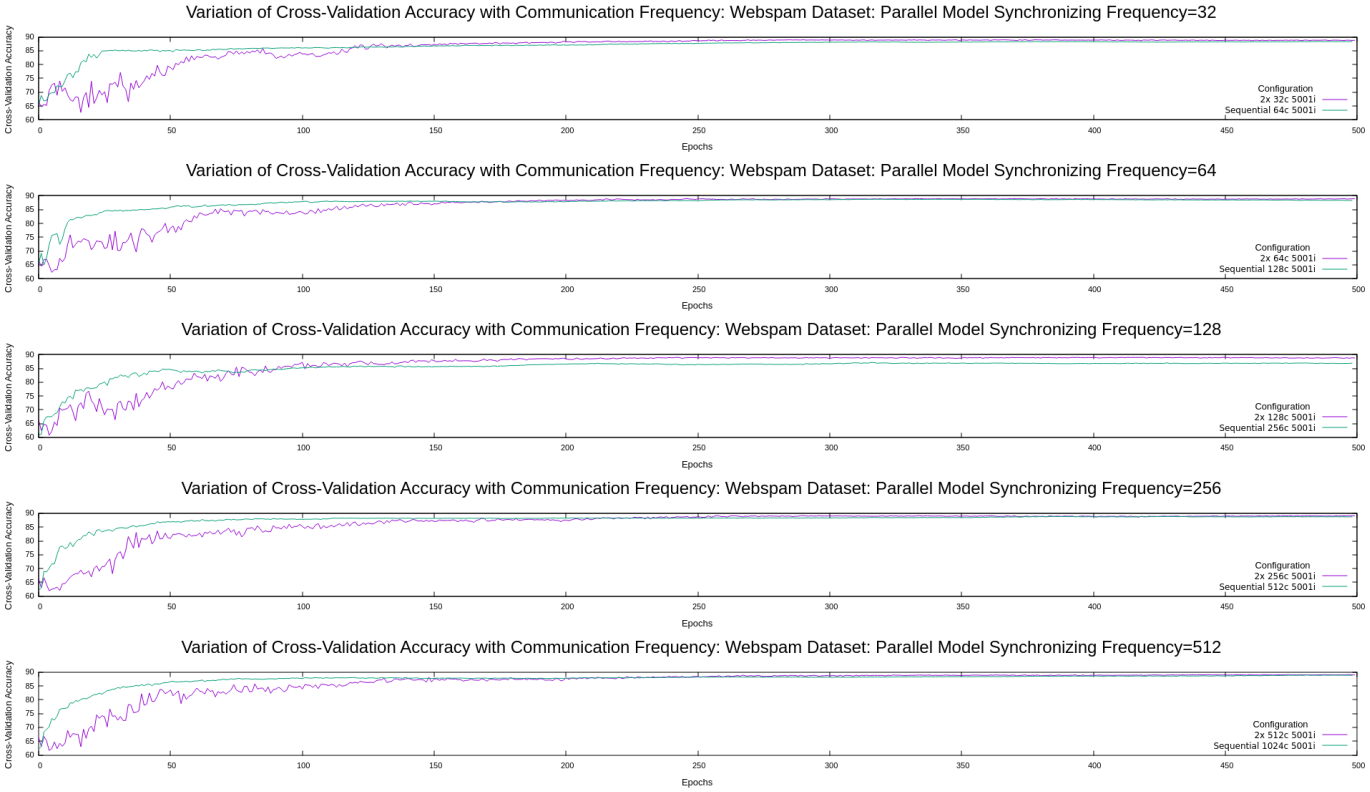

Fig. 18. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 2, MSF = [32,64,128,256,512]

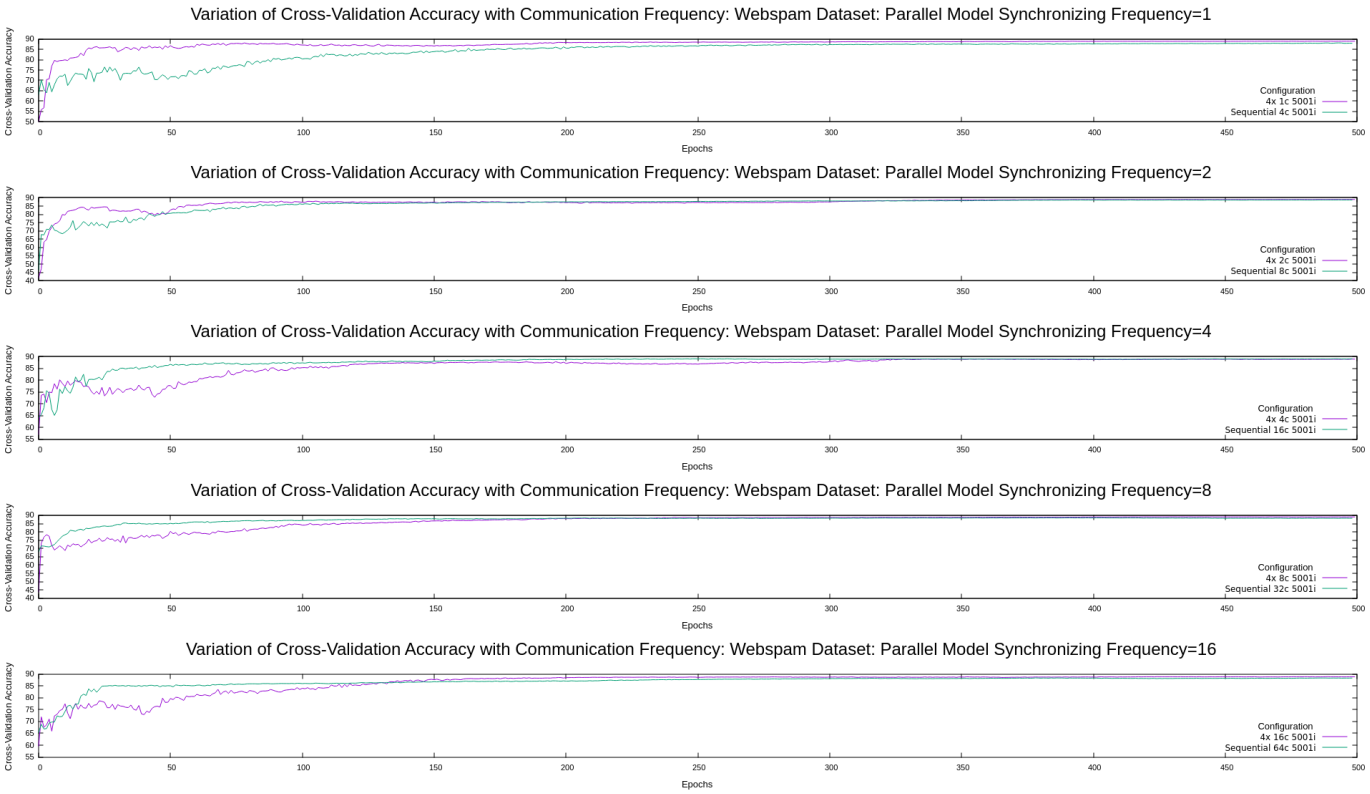

Fig. 19. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 4 MSF = [1,2,4,8,16]

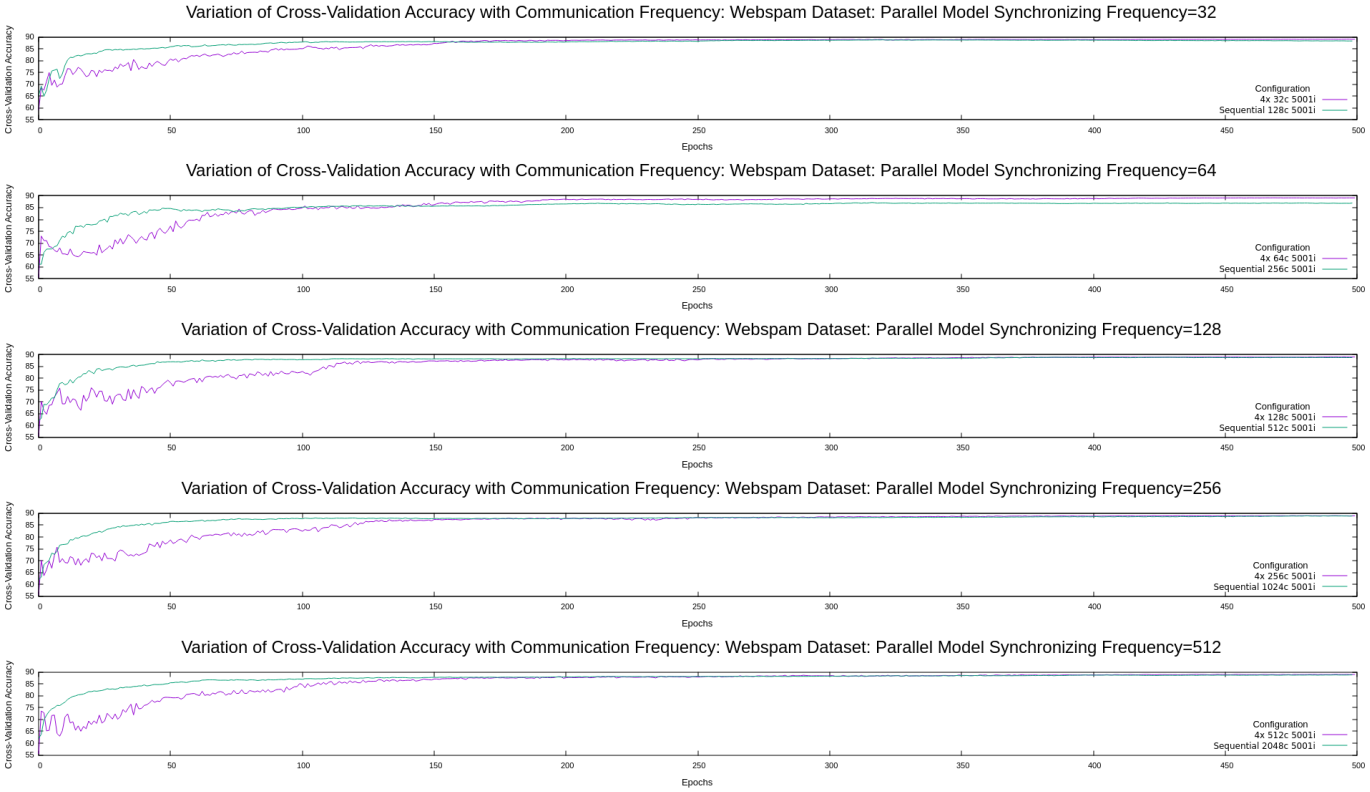

Fig. 20. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 4, MSF = [32,64,128,256,512]

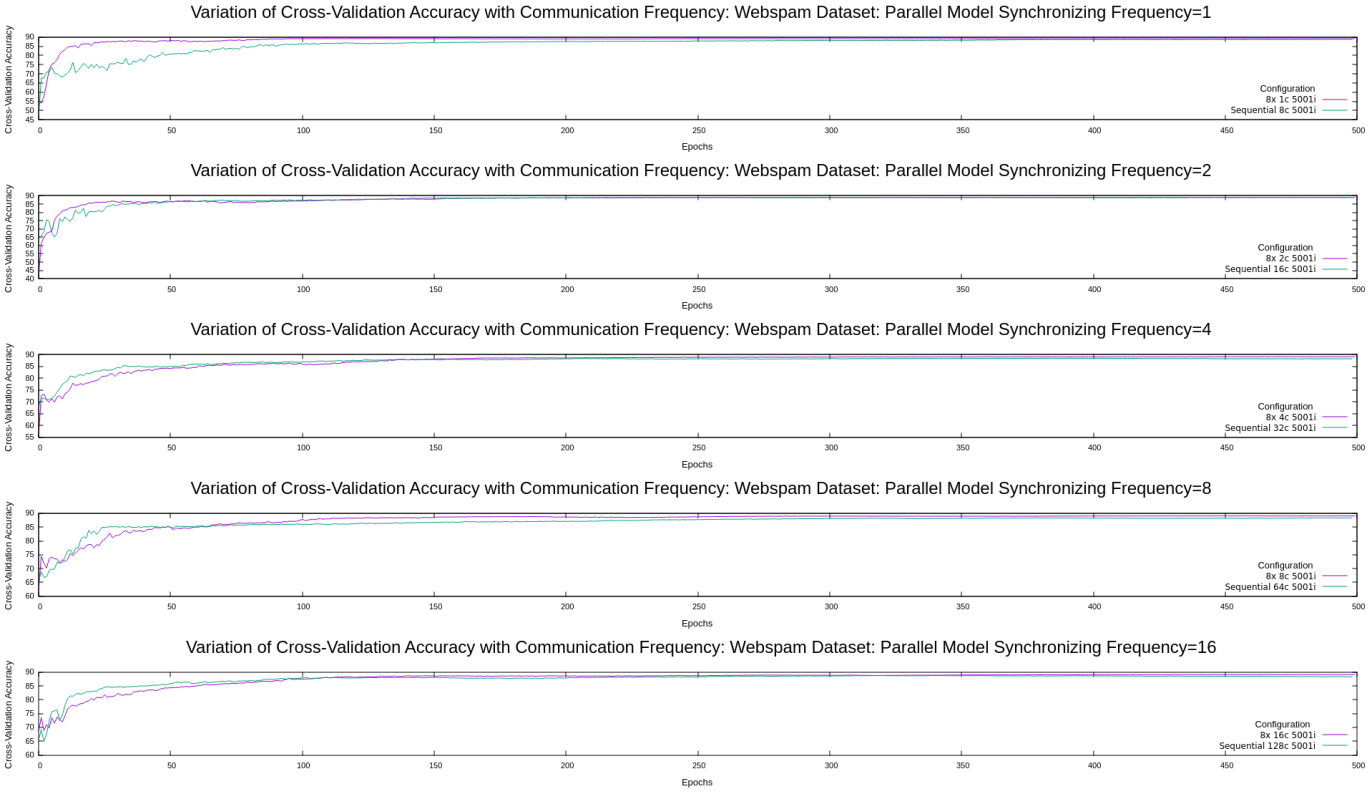

Fig. 21. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 8 MSF = [1,2,4,8,16]

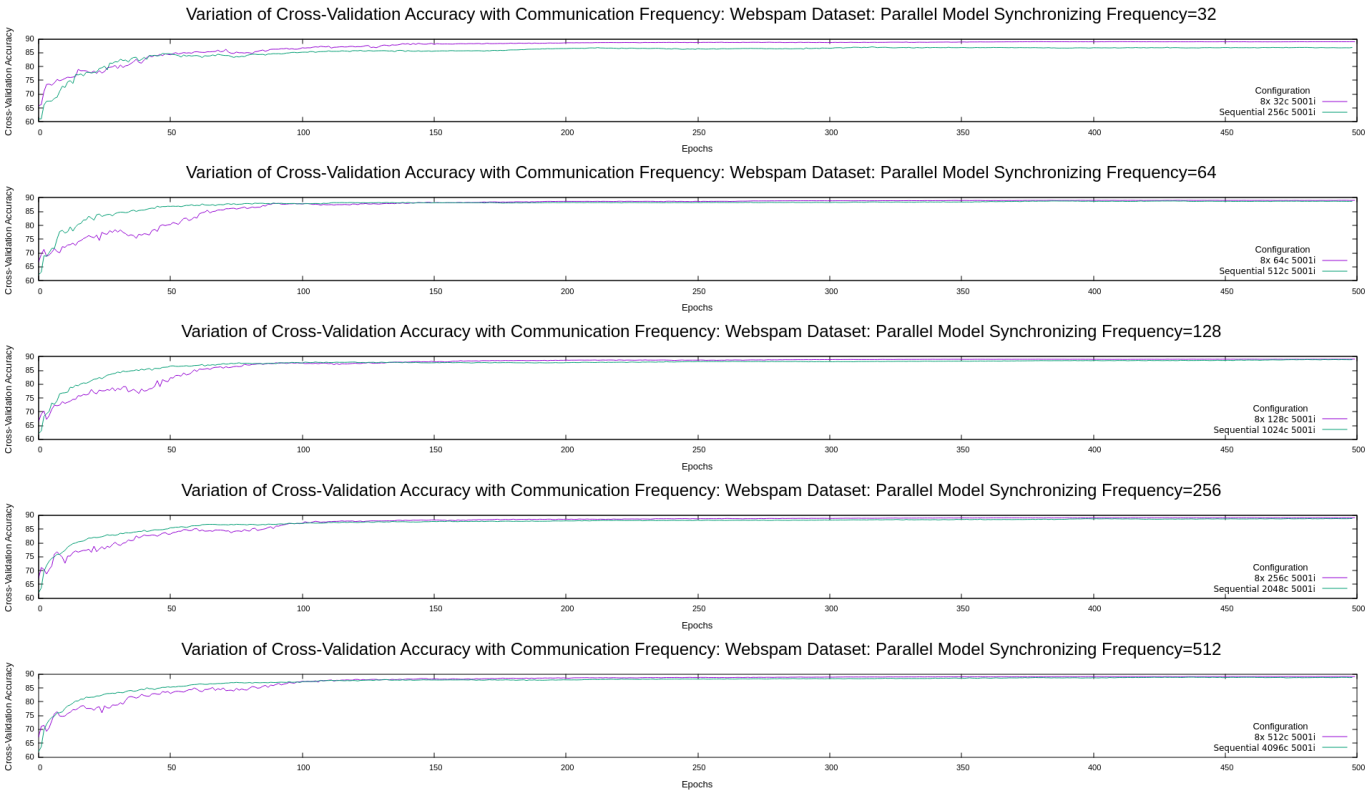

Fig. 22. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 8, MSF = [32,64,128,256,512]

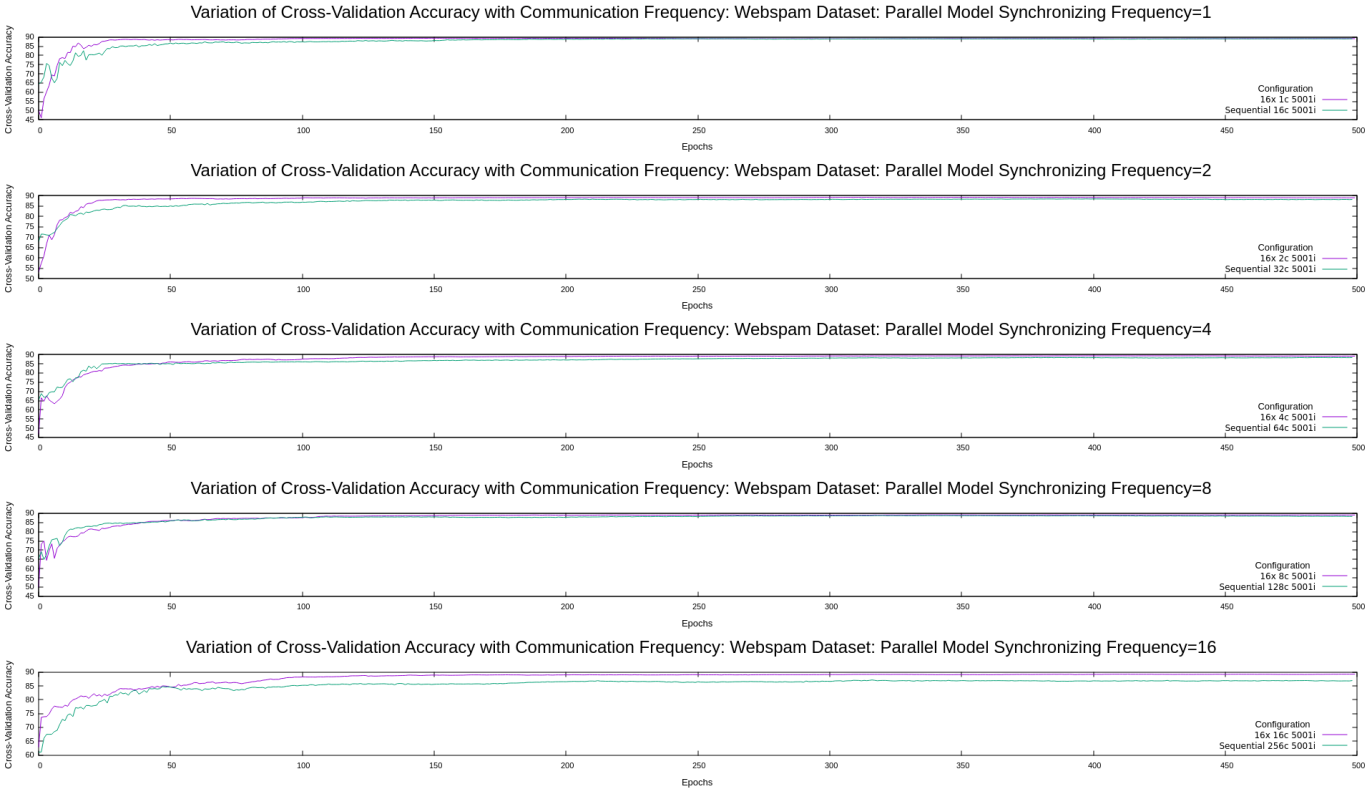

Fig. 23. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 16 MSF = [1,2,4,8,16]

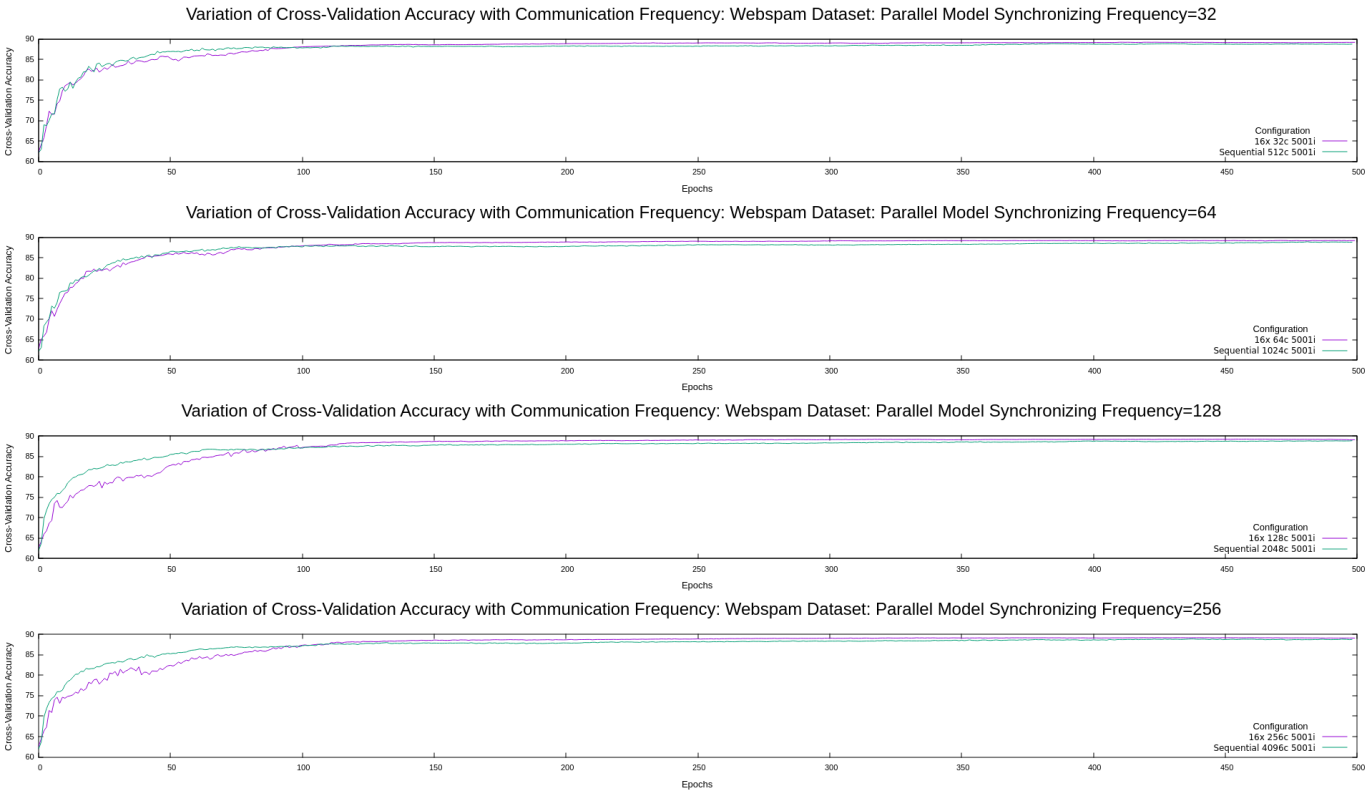

Fig. 24. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 16, MSF = [32,64,128,256]

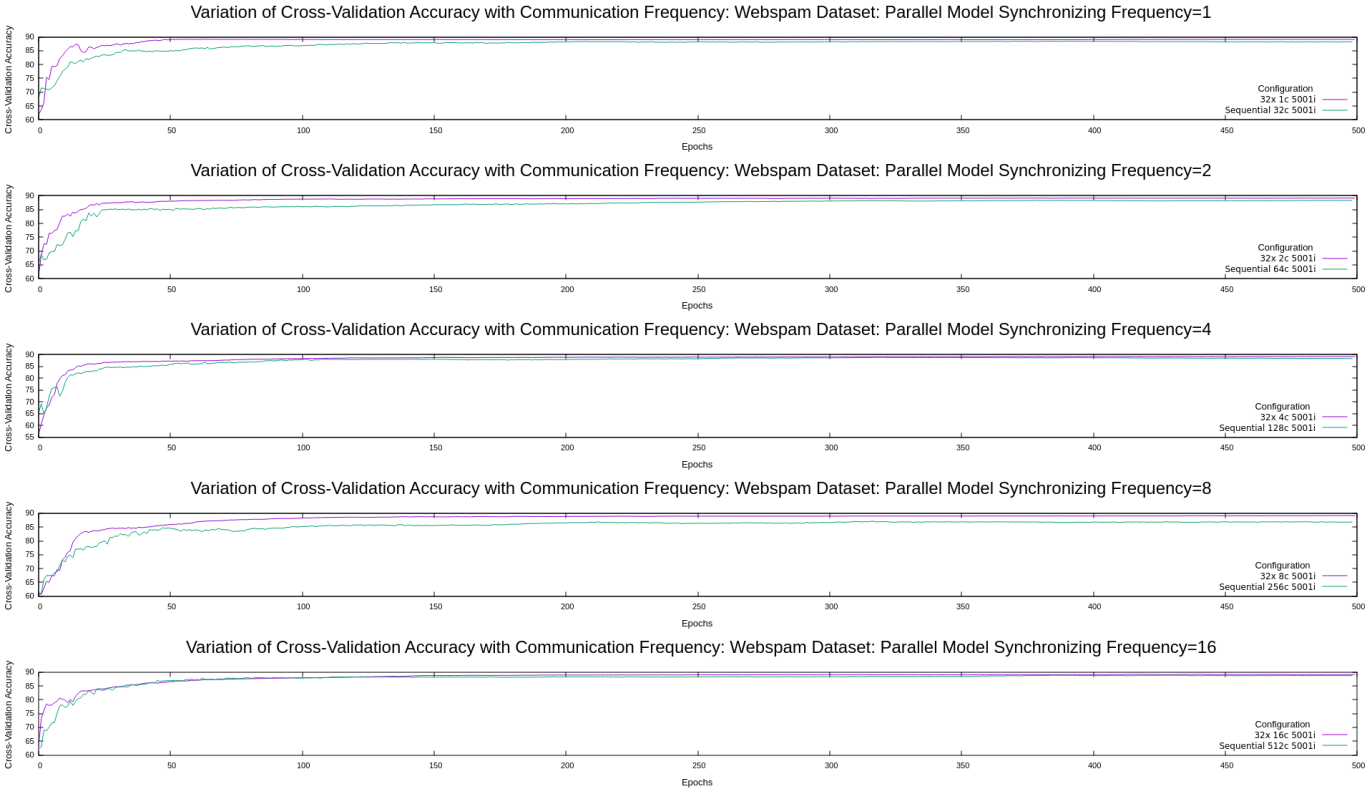

Fig. 25. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 32 MSF = [1,2,4,8,16]

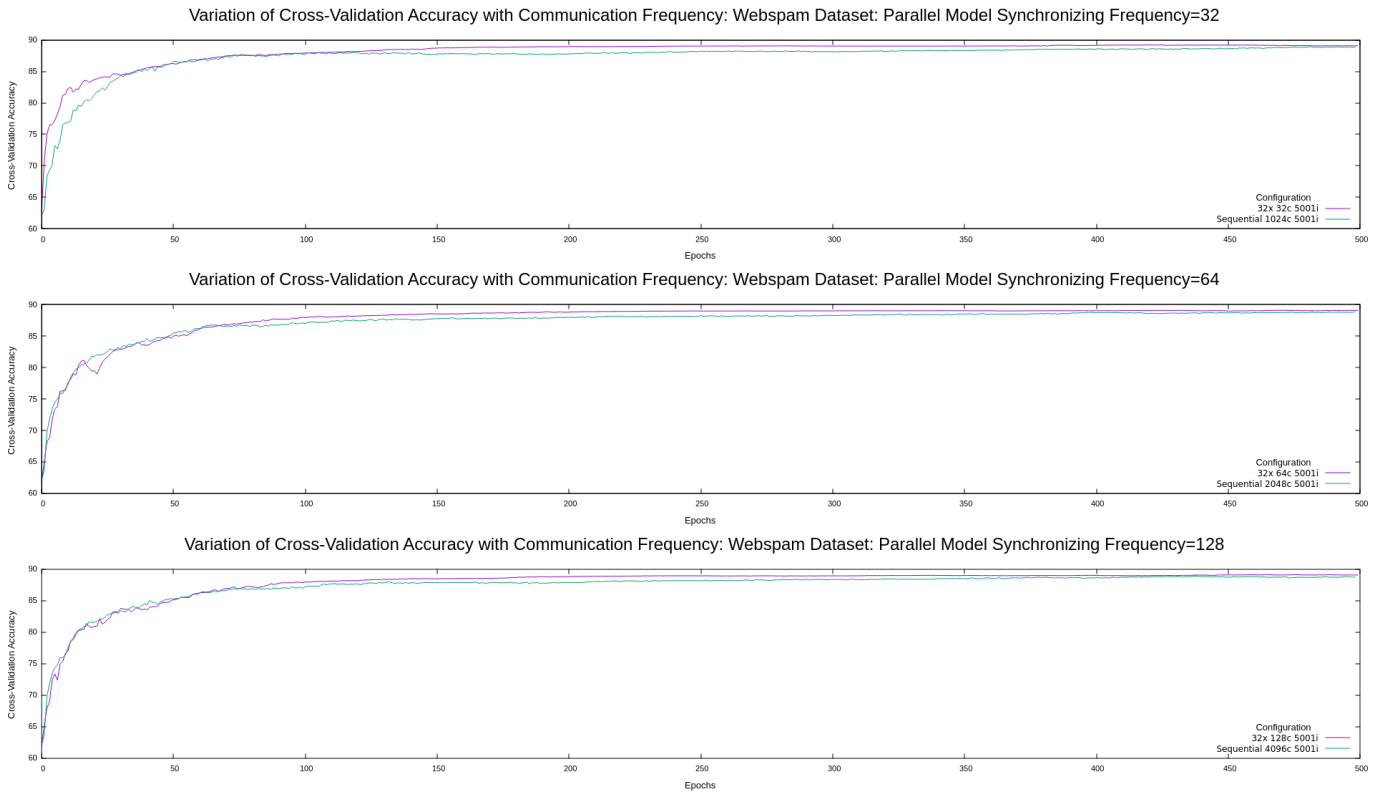

Fig. 26. Cross Validation Accuracy Variation with Model Synchronization Frequency of Webspam Dataset for Parallelism 32, MSF = [32,64,128]

IV. Distributed MSF Experiments

A. *Ijcnn1 All Cross Validation Accuracy and Objective Function Value Variation against MSFs*

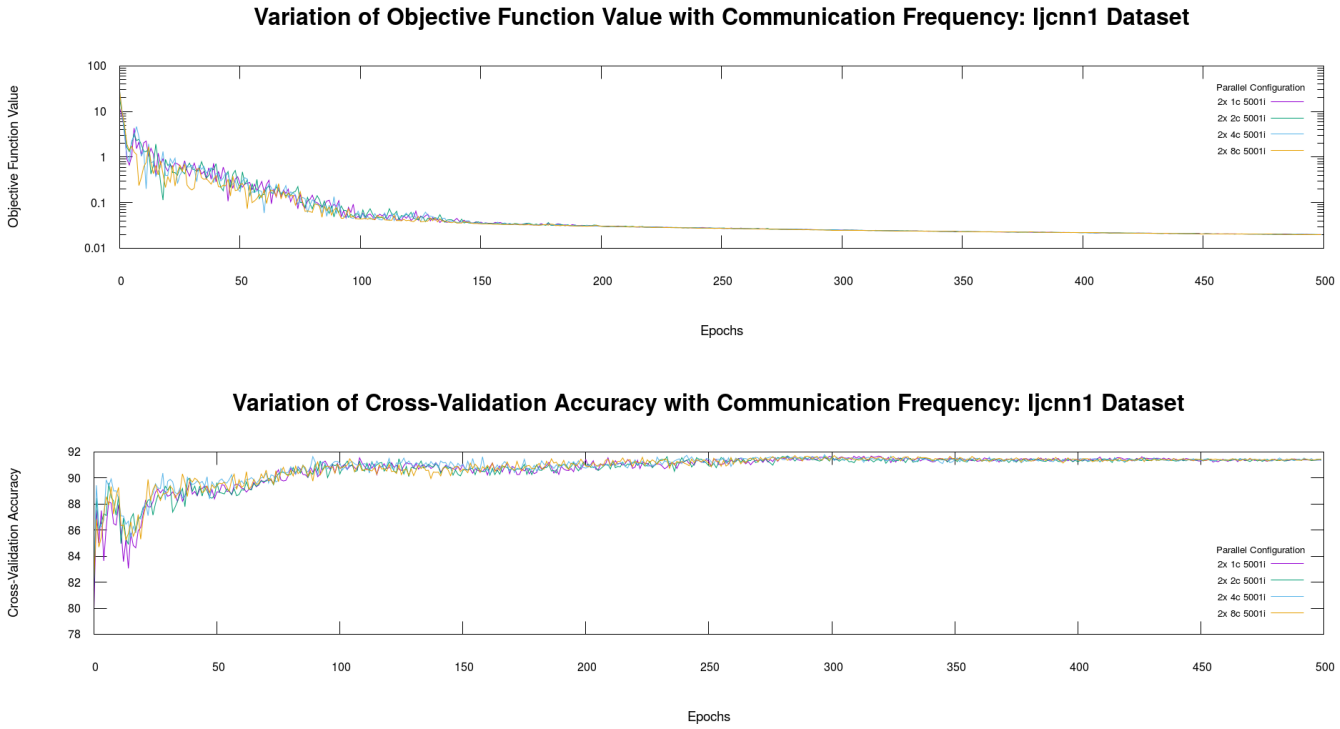

Fig. 27. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,], Parallelism = 2

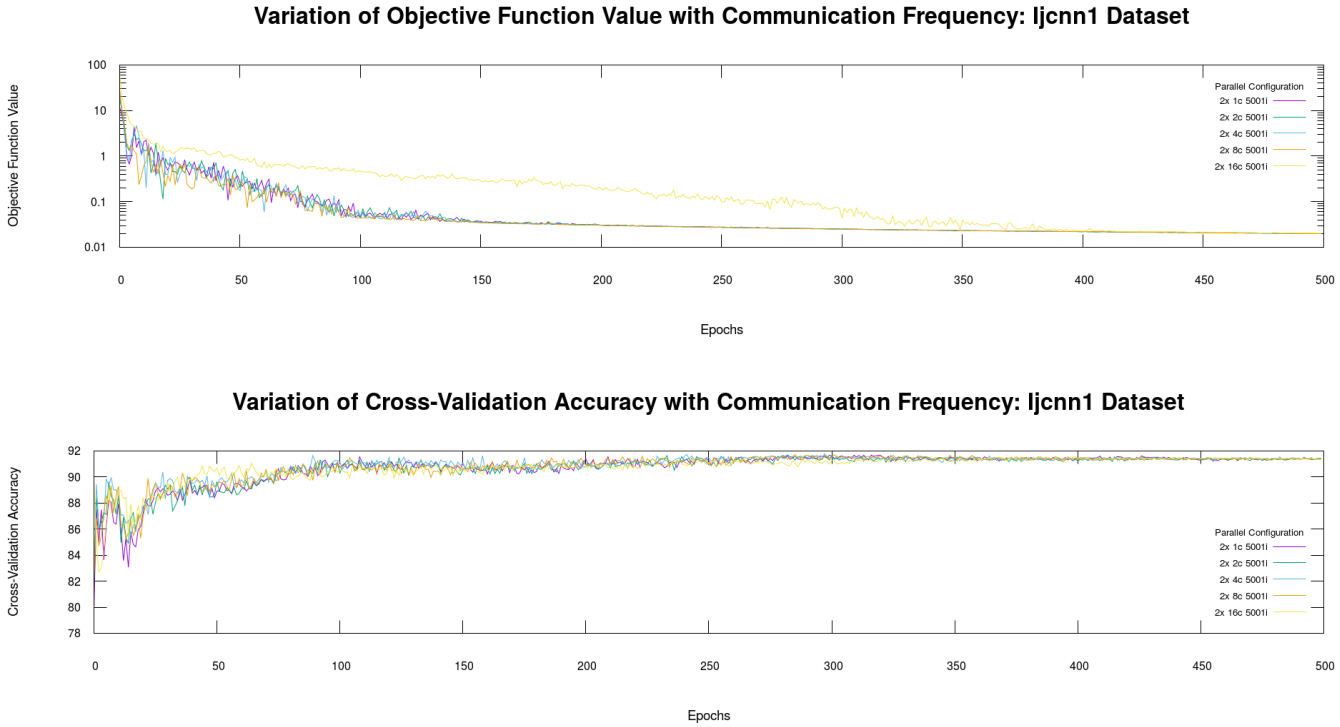

Fig. 28. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,16,], Parallelism = 2

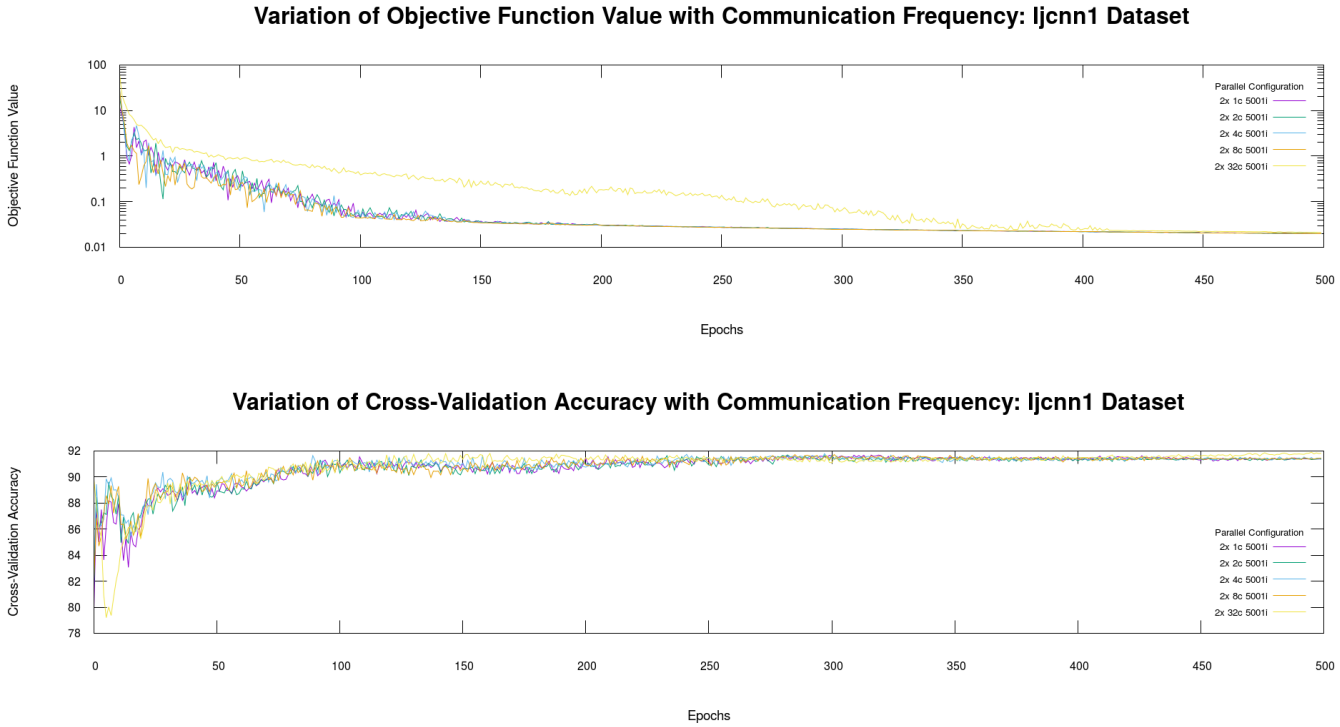

Fig. 29. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,32,], Parallelism = 2

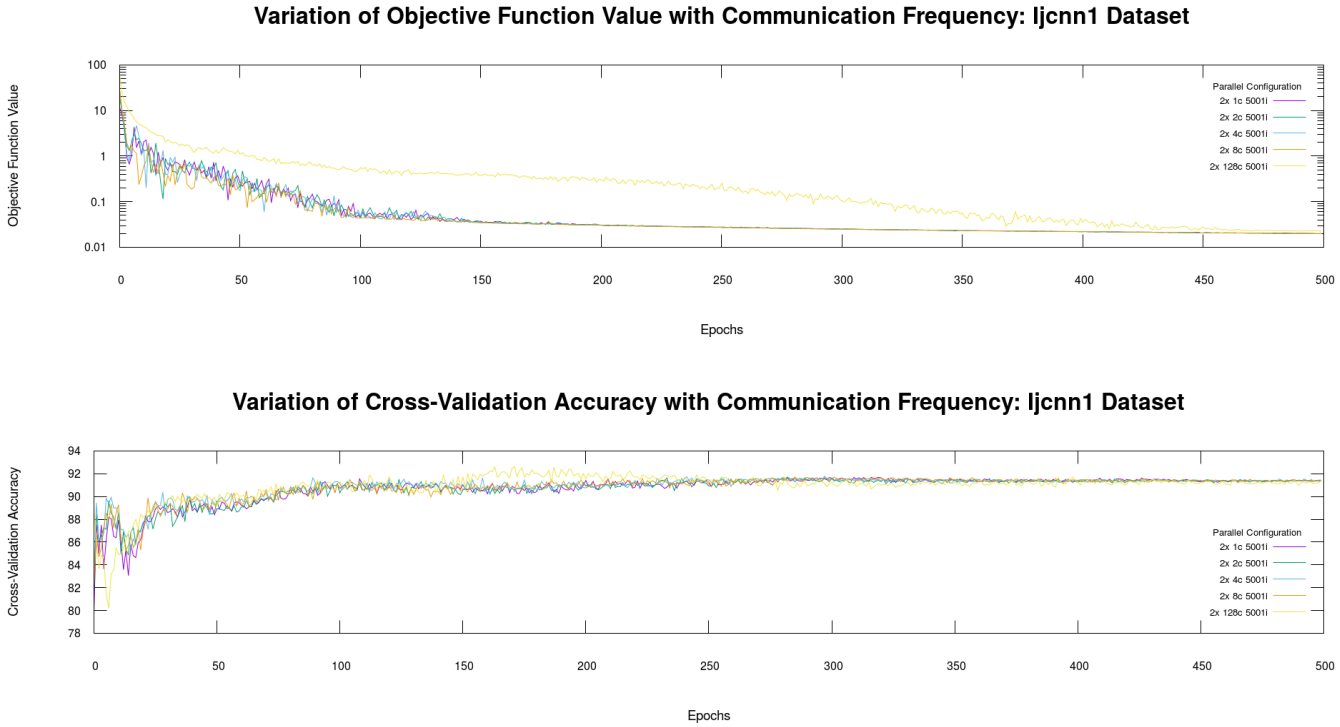

Fig. 30. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,128,], Parallelism = 2

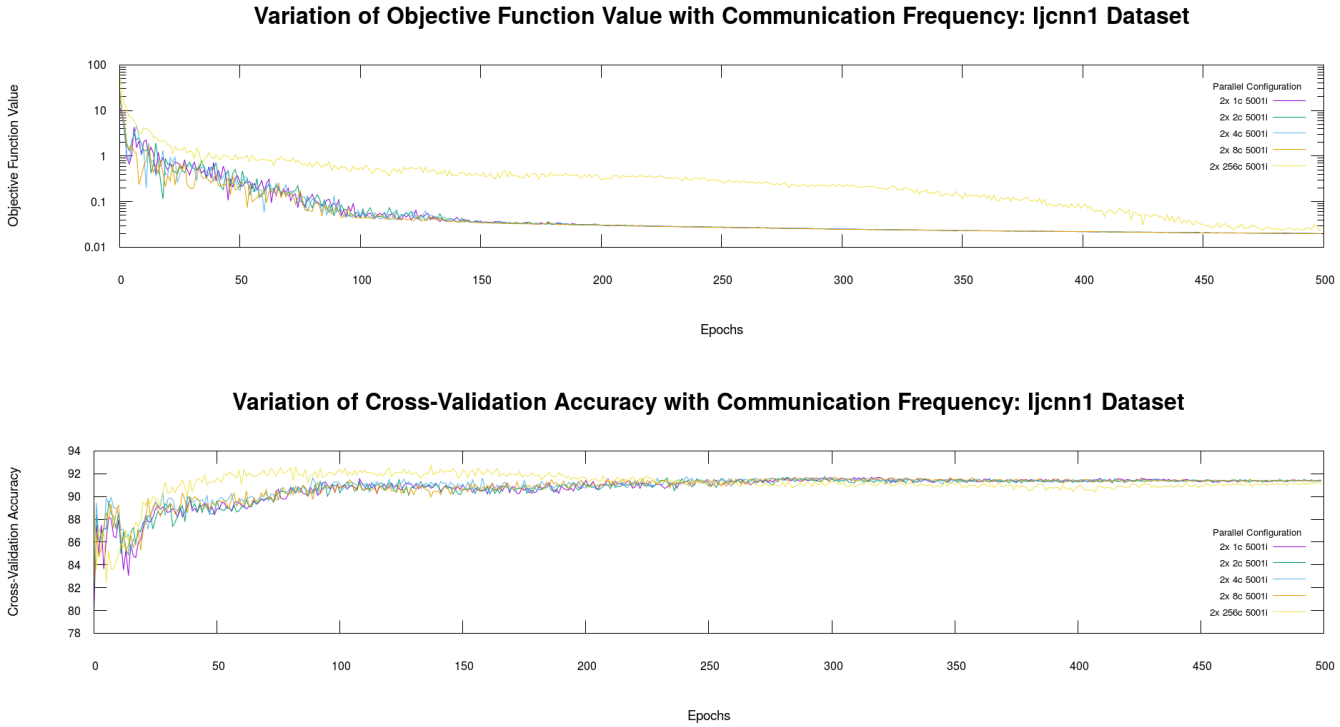

Fig. 31. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,256,], Parallelism = 2

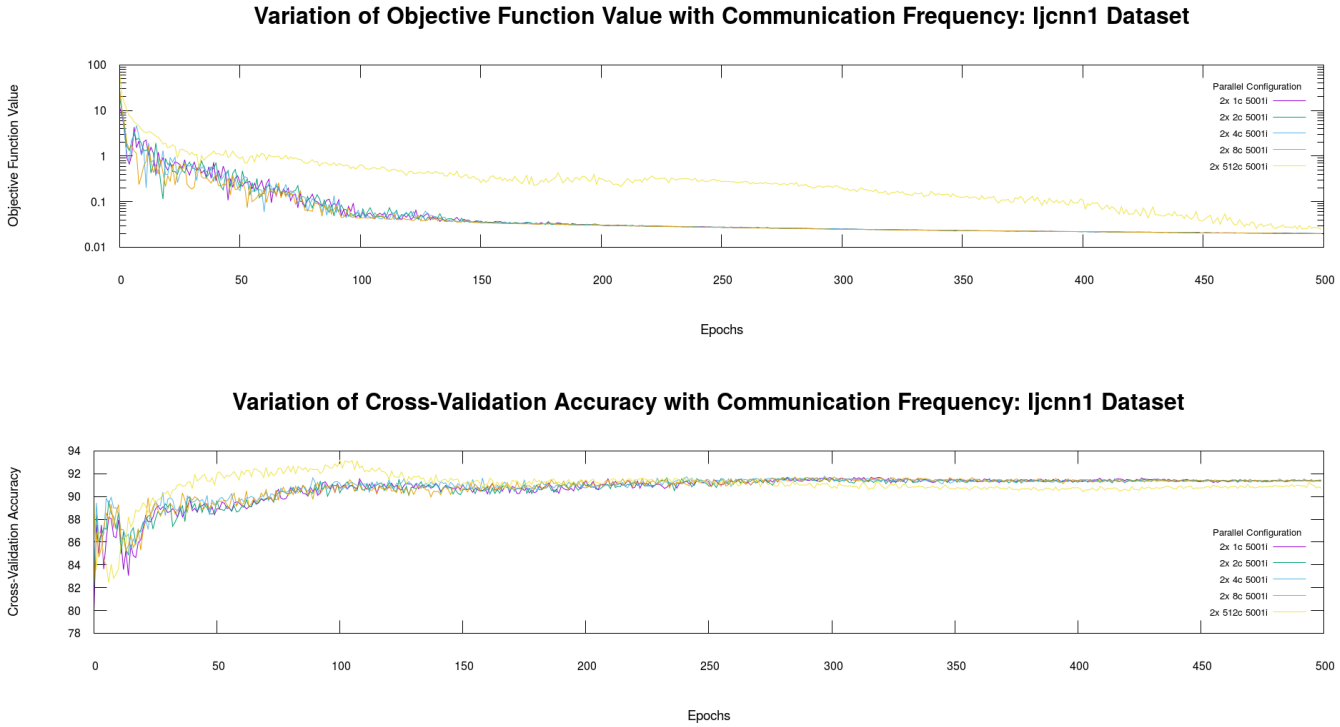

Fig. 32. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,512,], Parallelism = 2

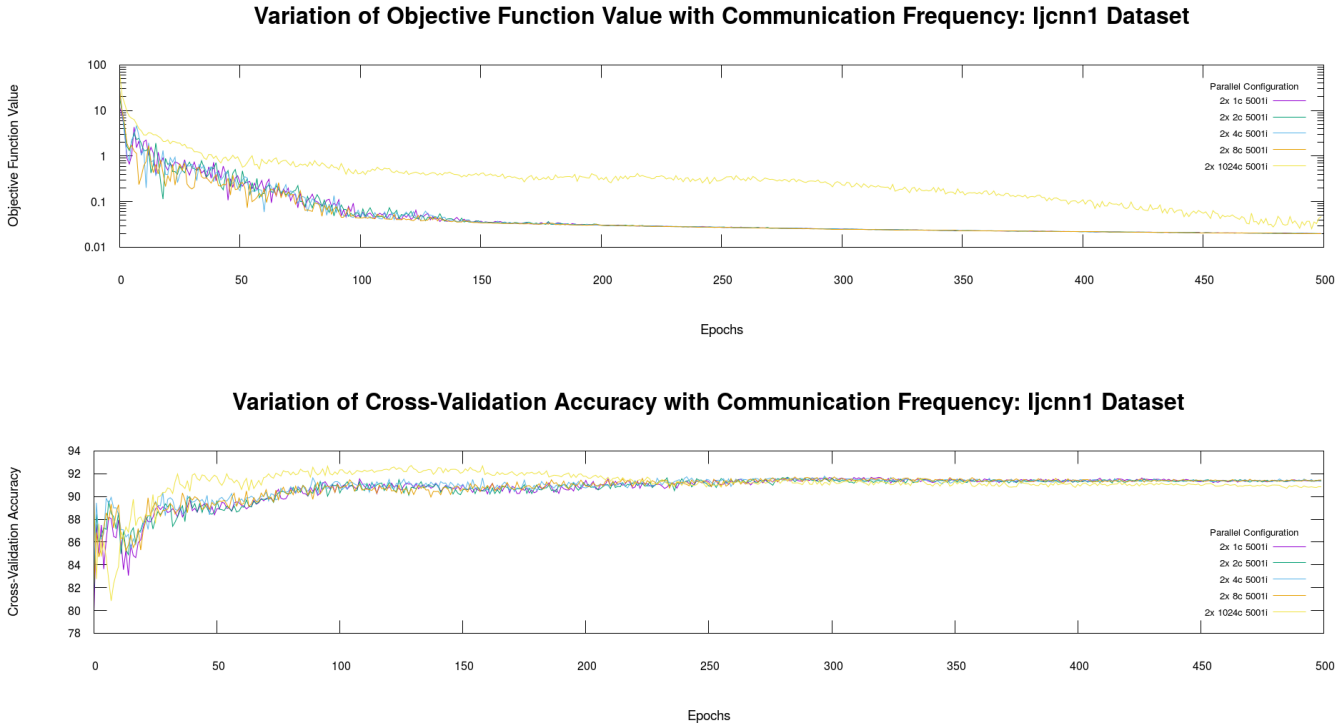

Fig. 33. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 2

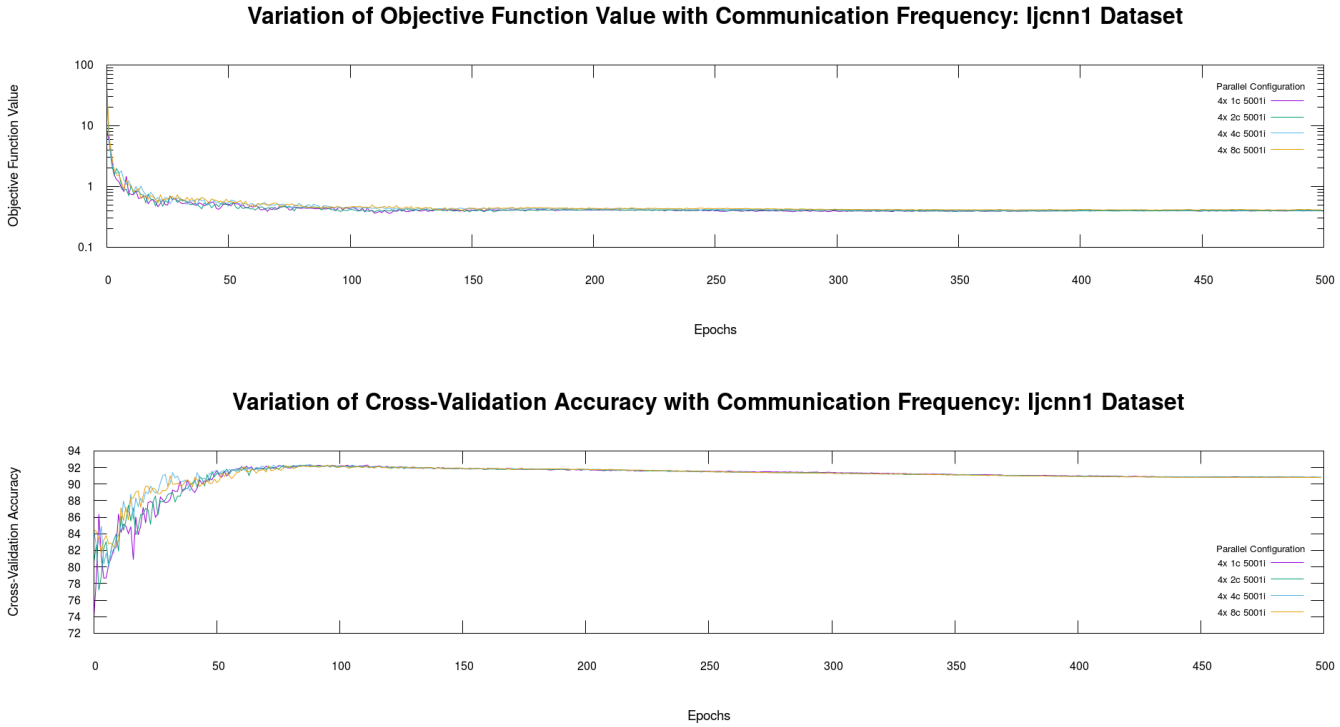

Fig. 34. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,], Parallelism = 4

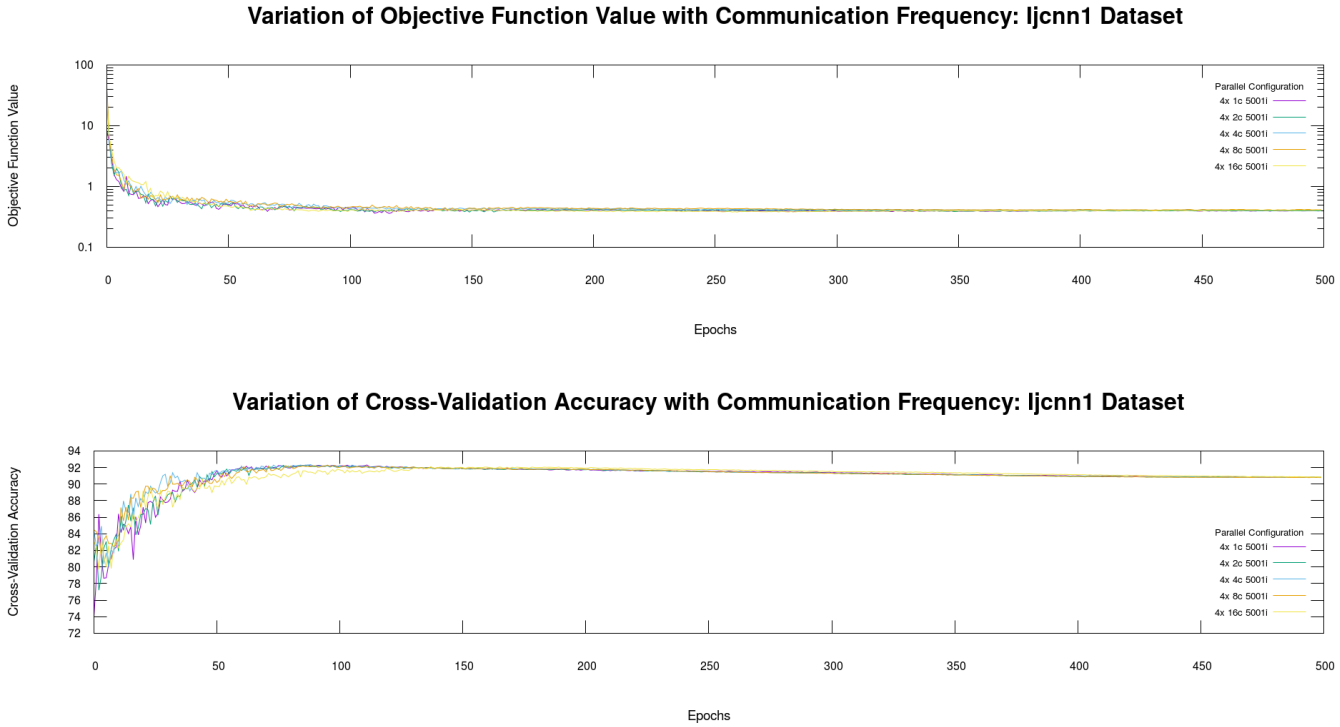

Fig. 35. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,16,], Parallelism = 4

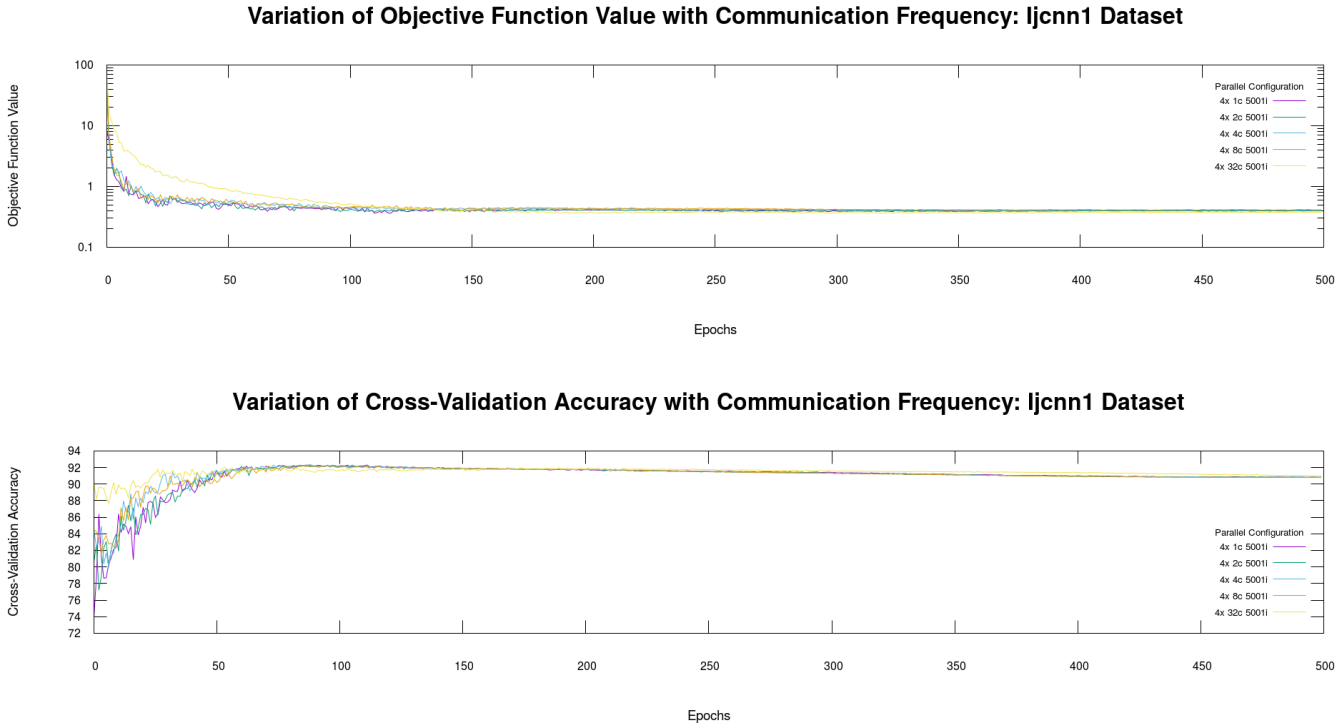

Fig. 36. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,32,], Parallelism = 4

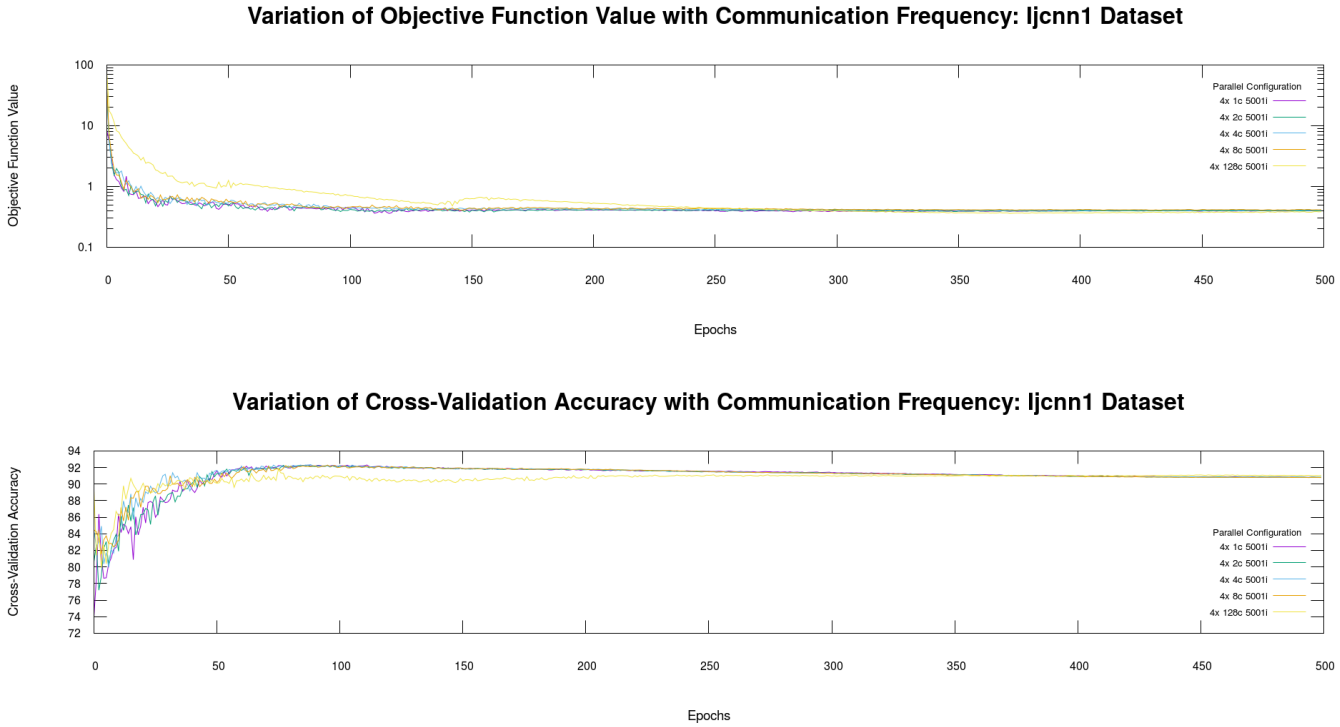

Fig. 37. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,128,], Parallelism = 4

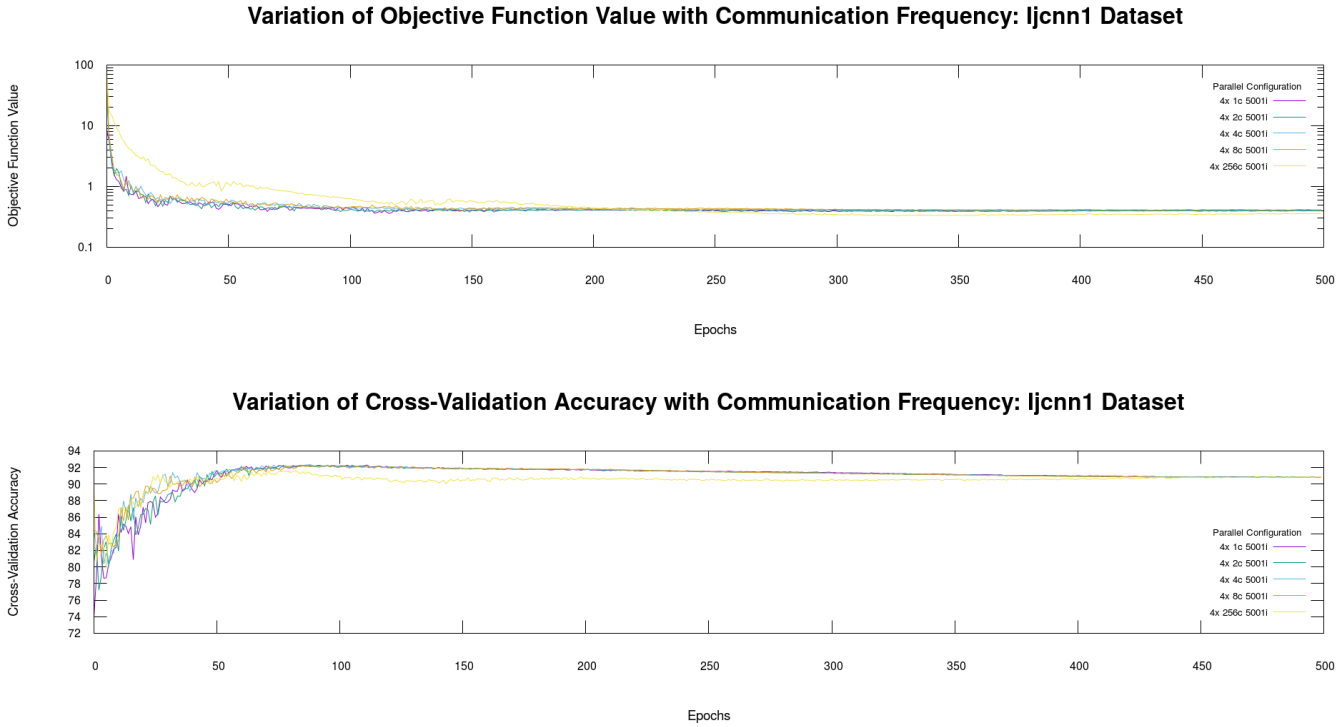

Fig. 38. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,256,], Parallelism = 4

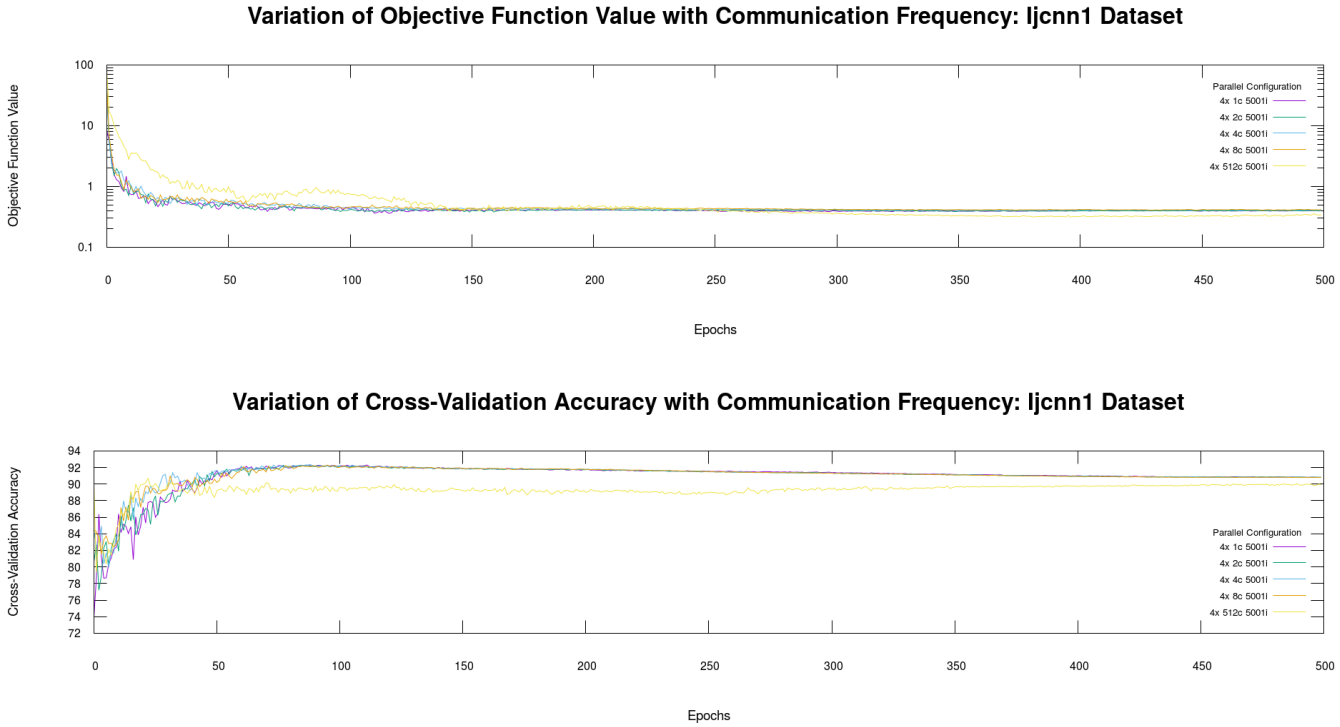

Fig. 39. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,512,], Parallelism = 4

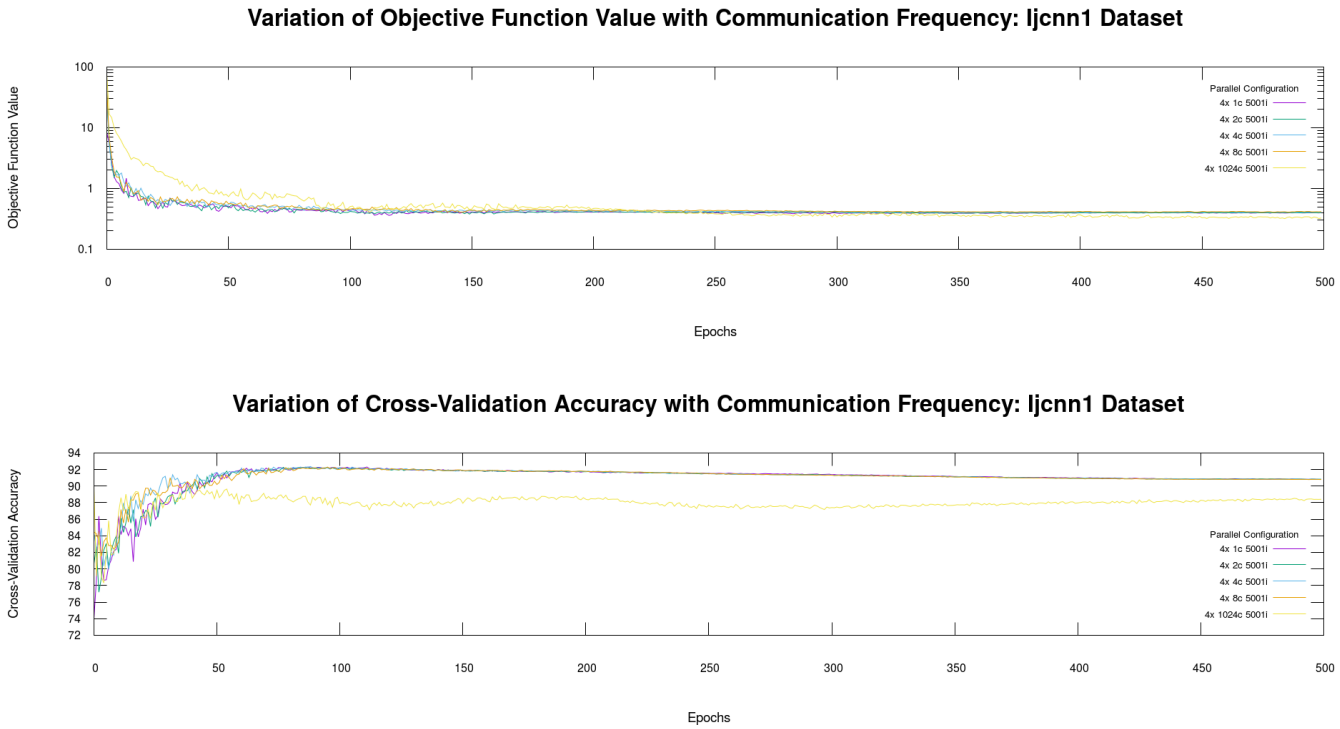

Fig. 40. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 4

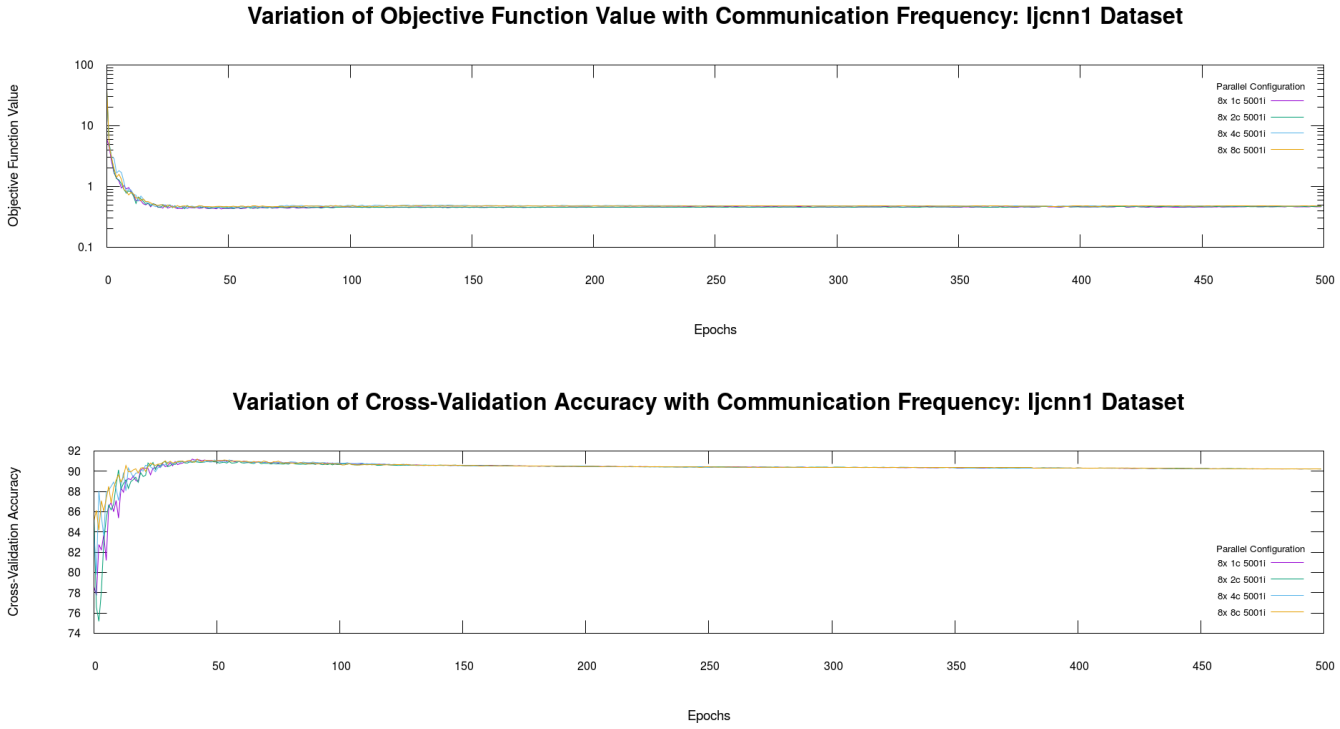

Fig. 41. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,], Parallelism = 8

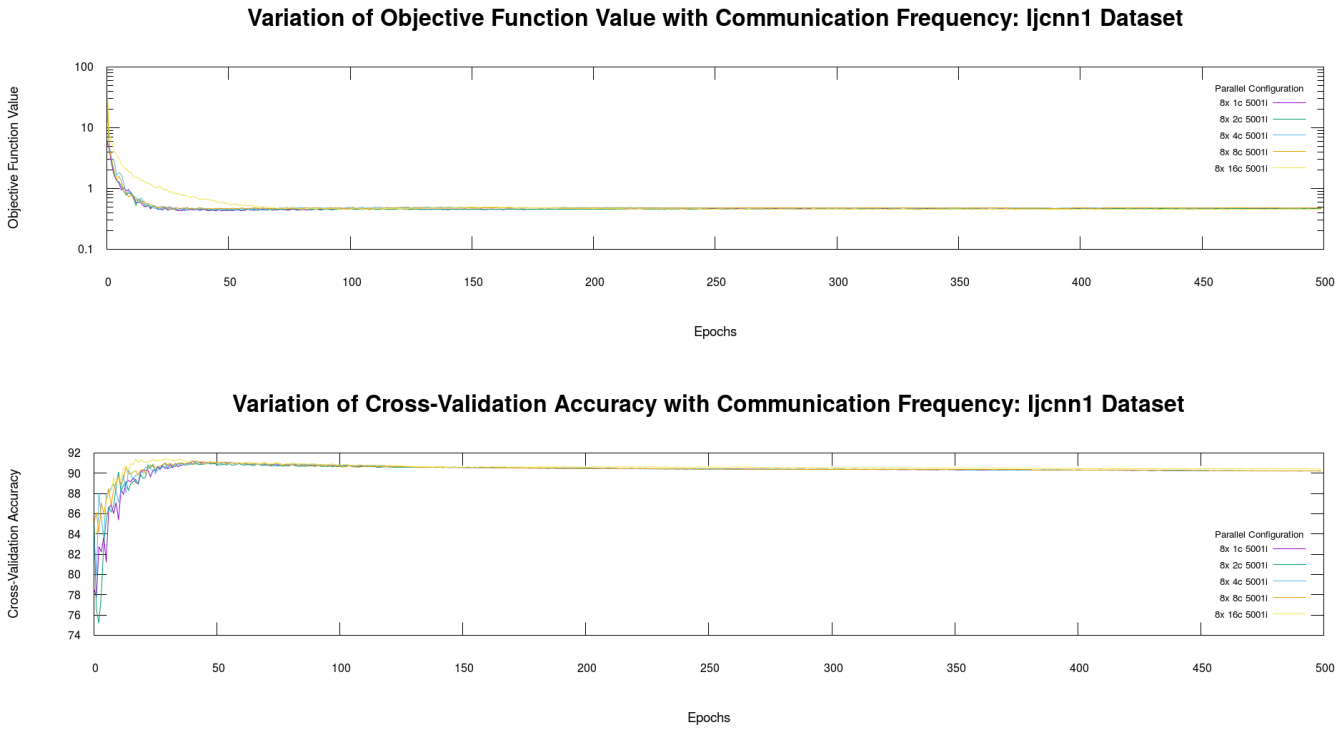

Fig. 42. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,16,], Parallelism = 8

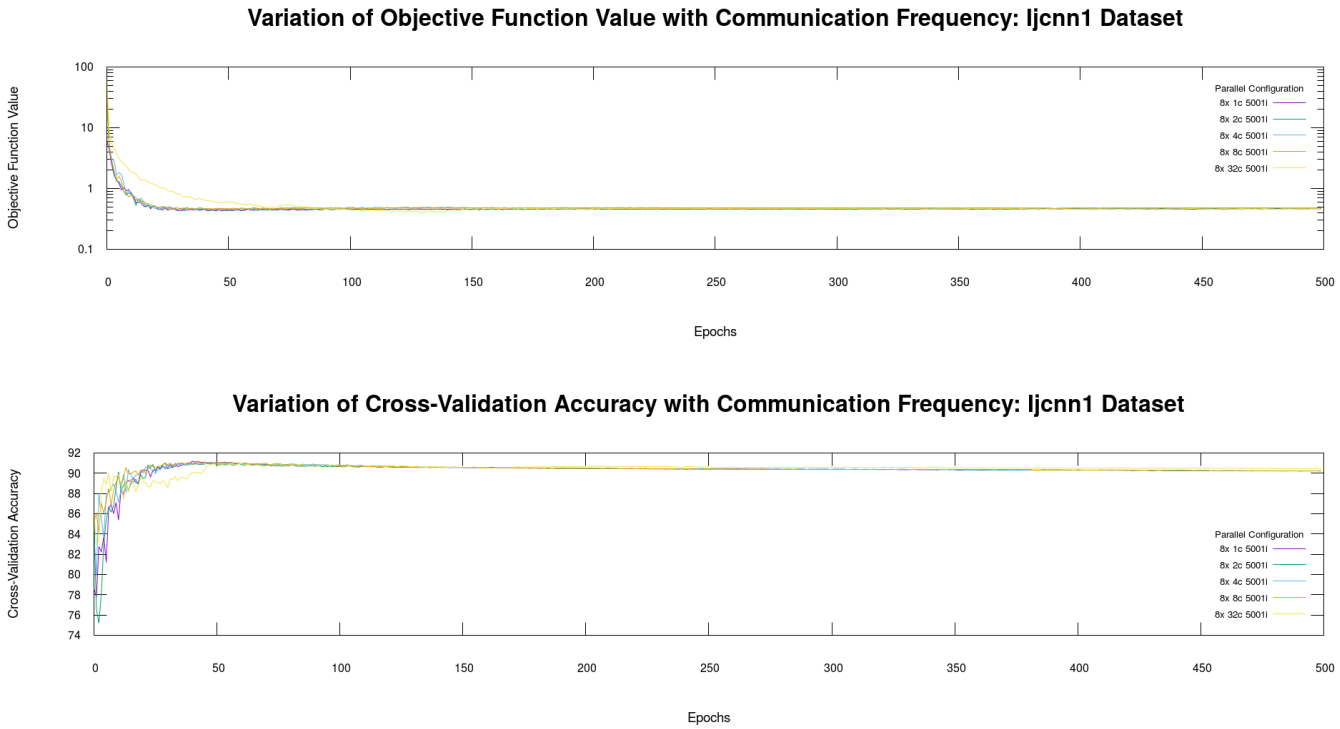

Fig. 43. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,32,], Parallelism = 8

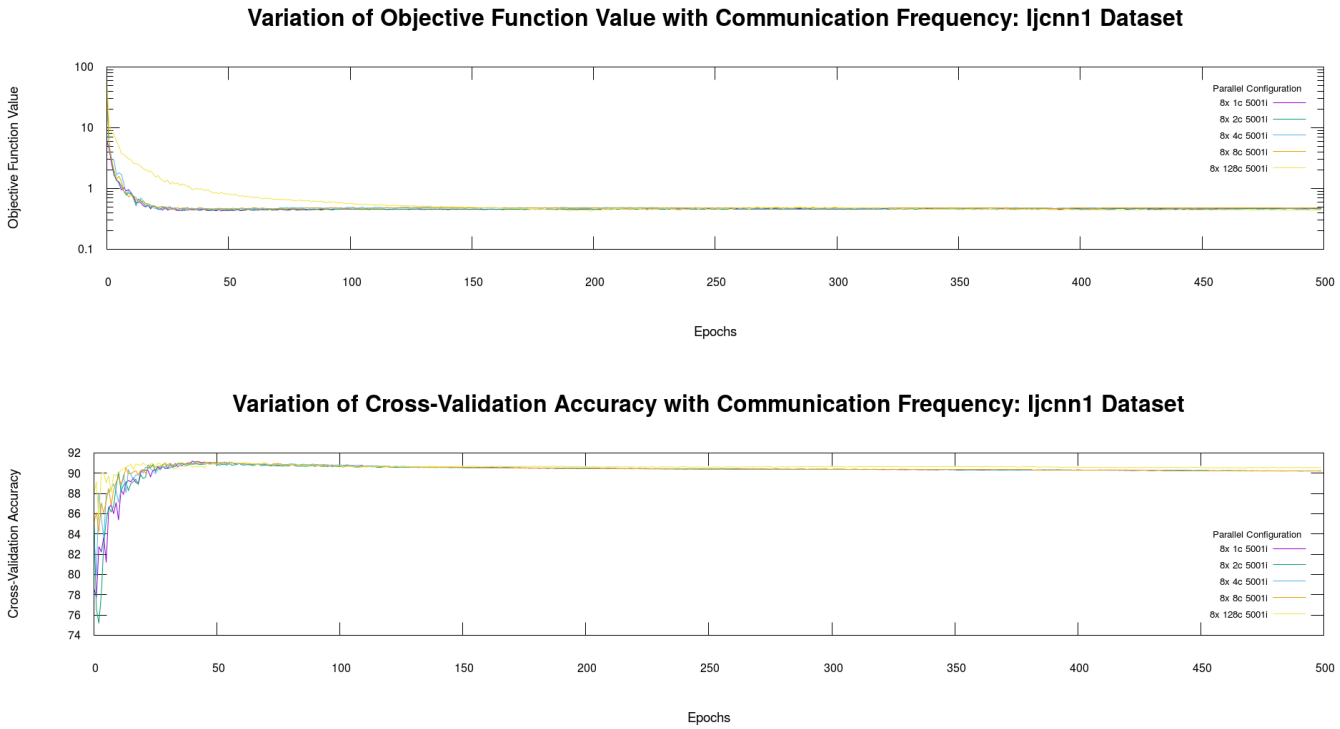

Fig. 44. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,128,], Parallelism = 8

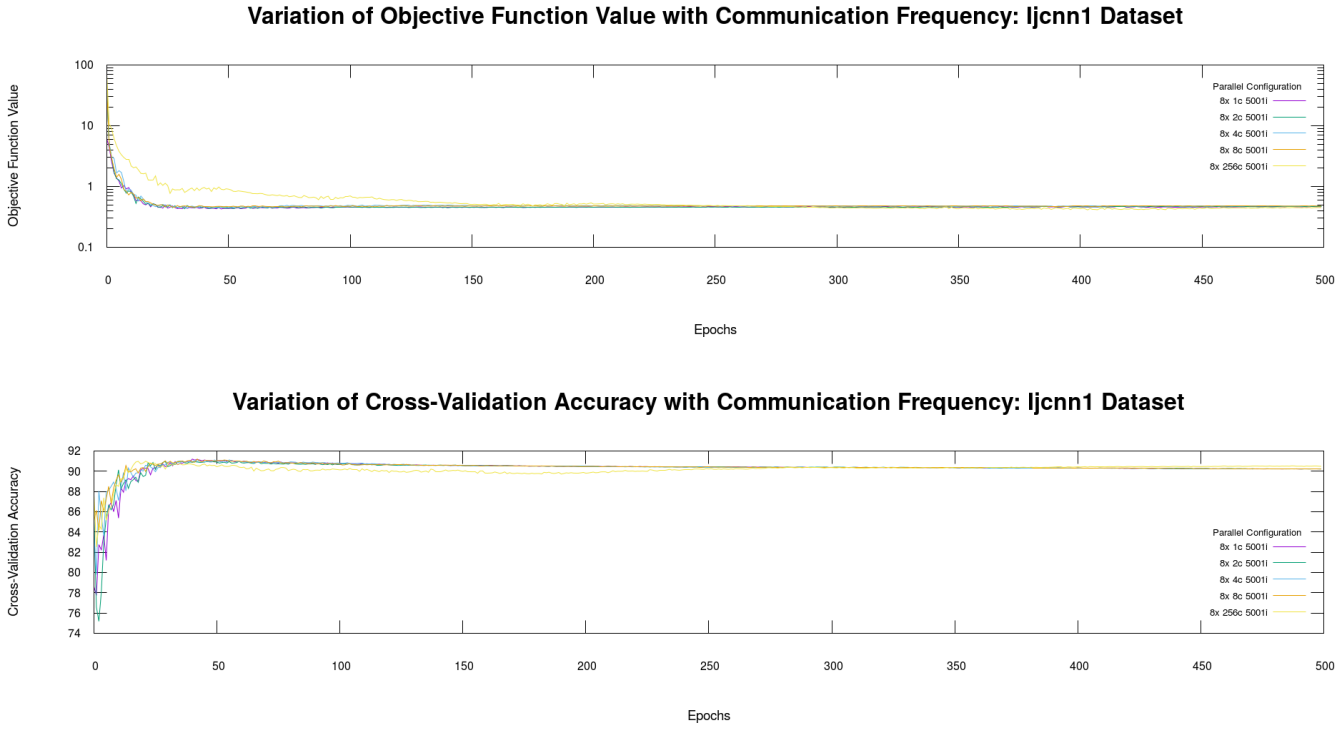

Fig. 45. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,256,], Parallelism = 8

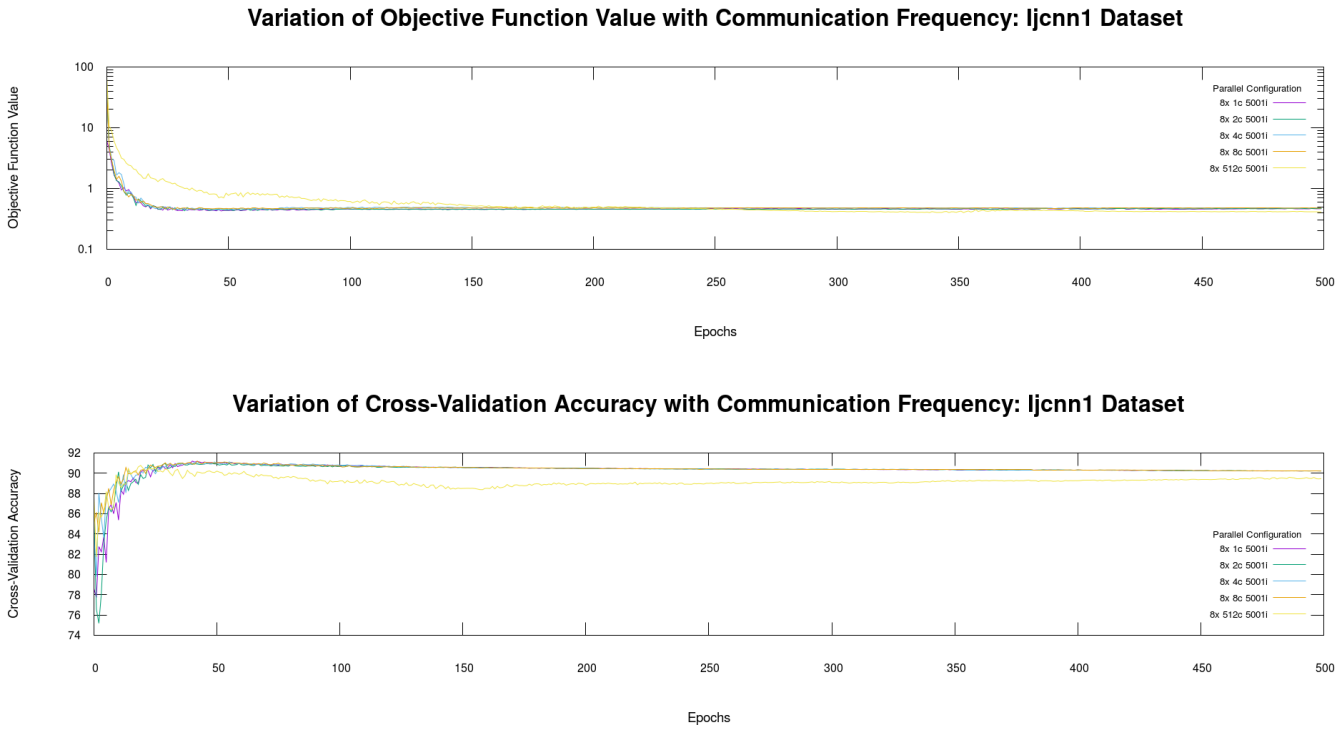

Fig. 46. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,512,], Parallelism = 8

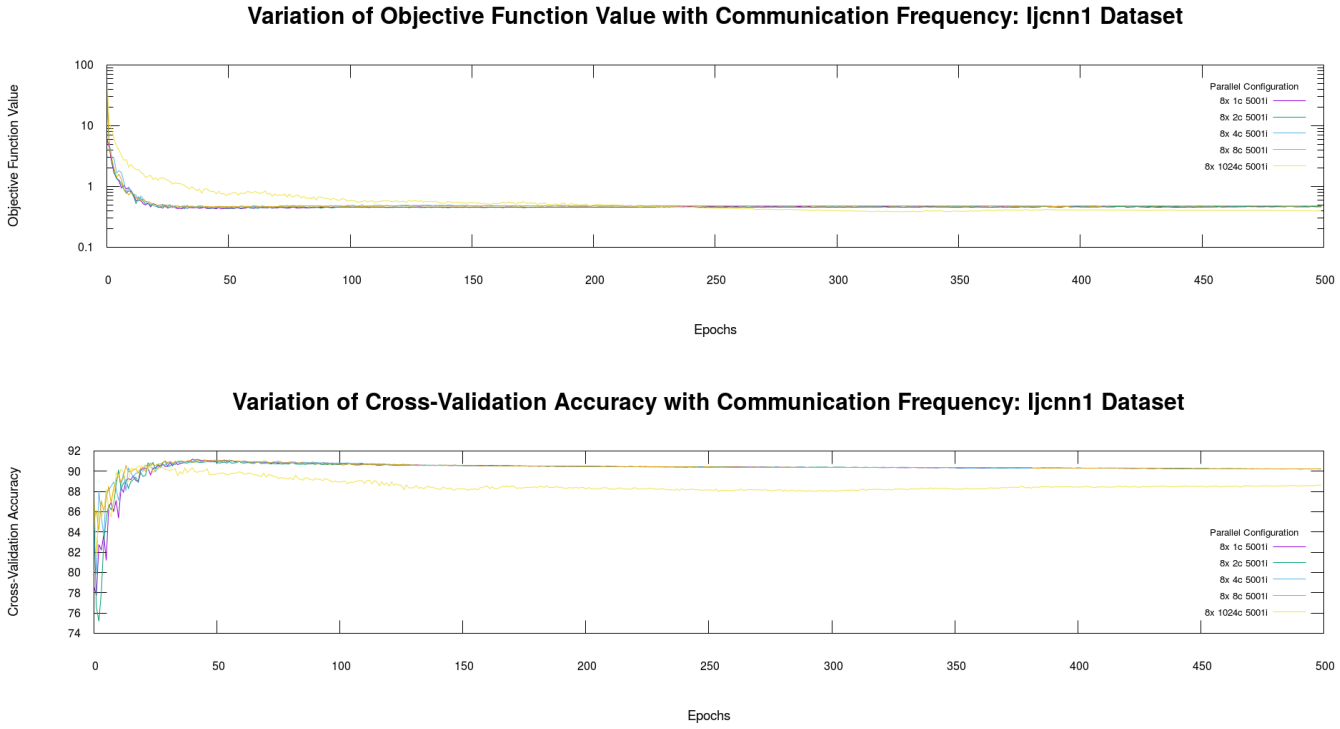

Fig. 47. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 8

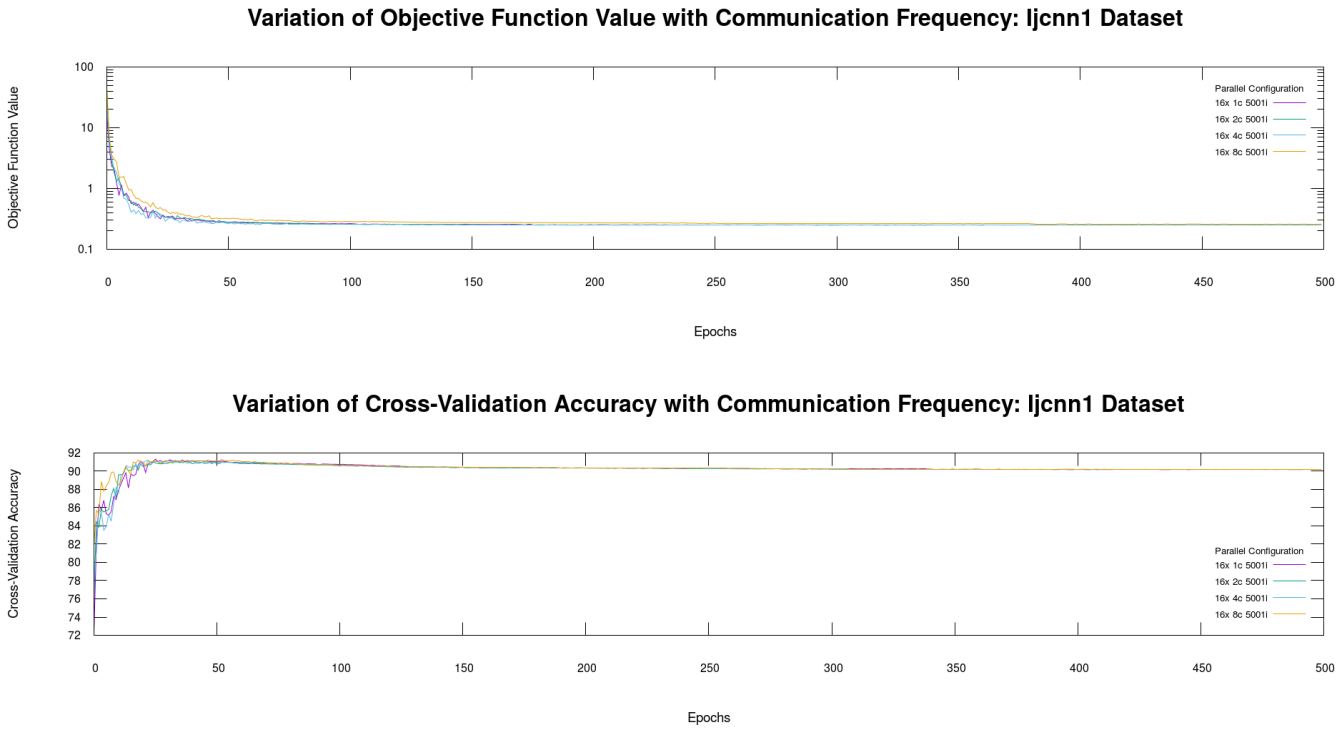

Fig. 48. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,], Parallelism = 16

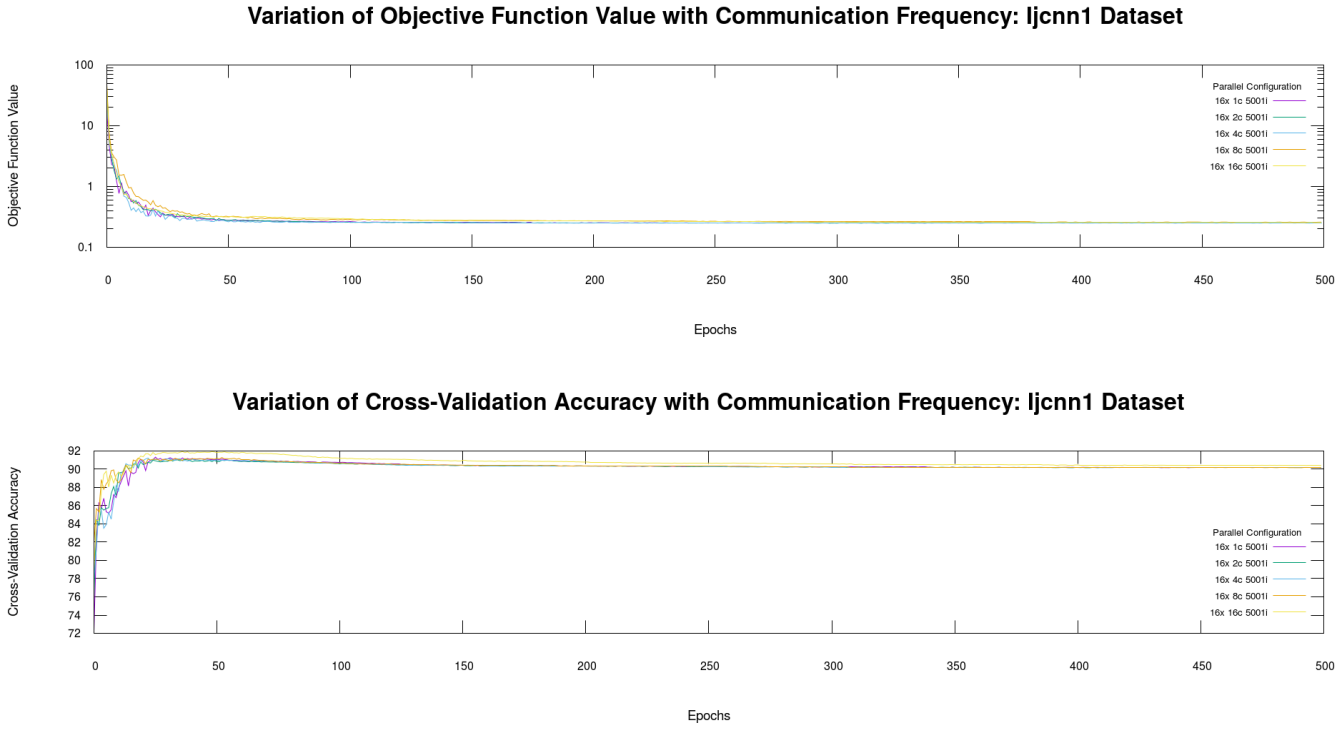

Fig. 49. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,16,], Parallelism = 16

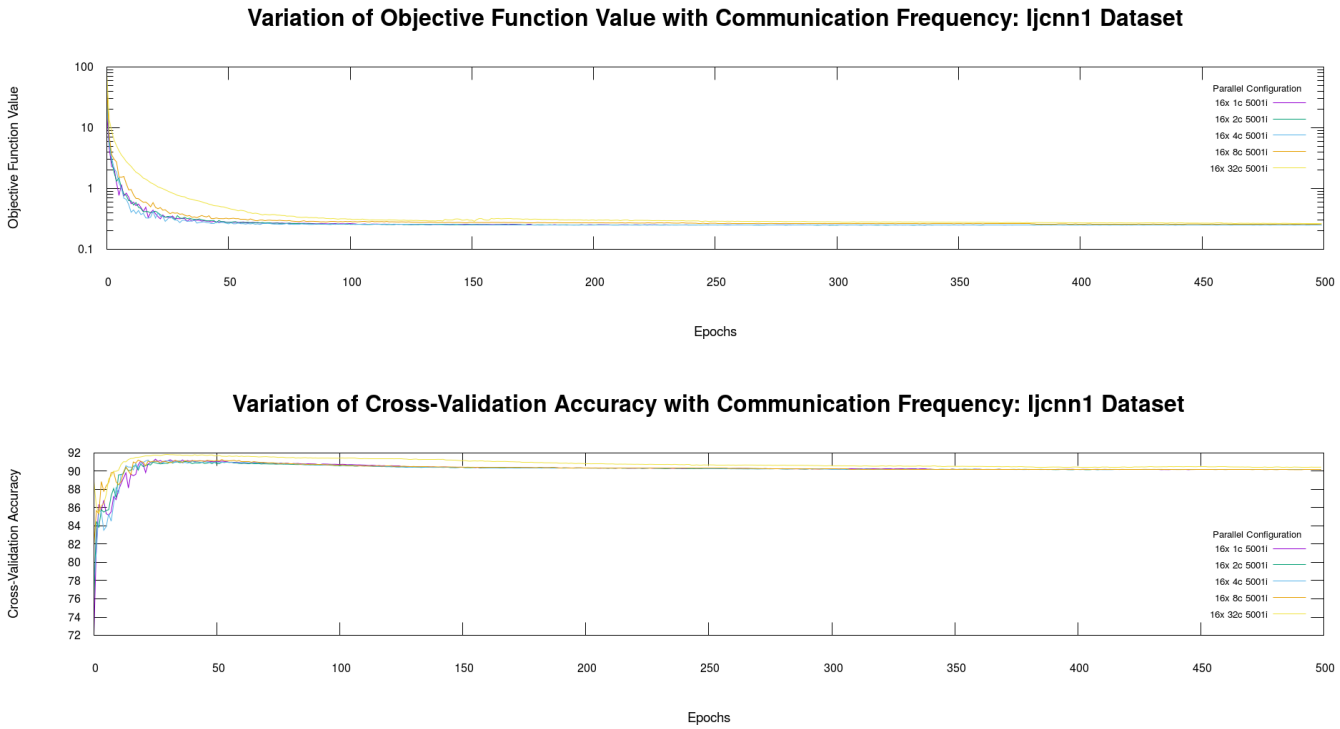

Fig. 50. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,32,], Parallelism = 16

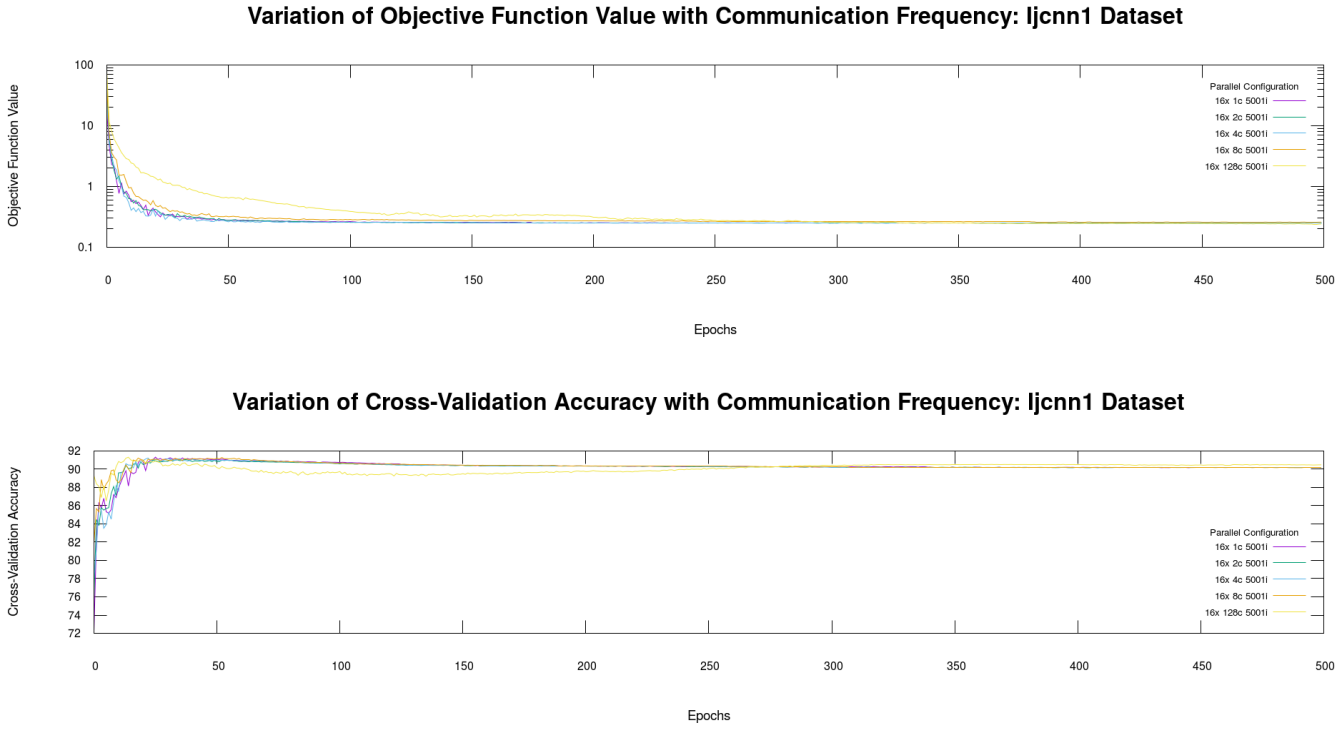

Fig. 51. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,128,], Parallelism = 16

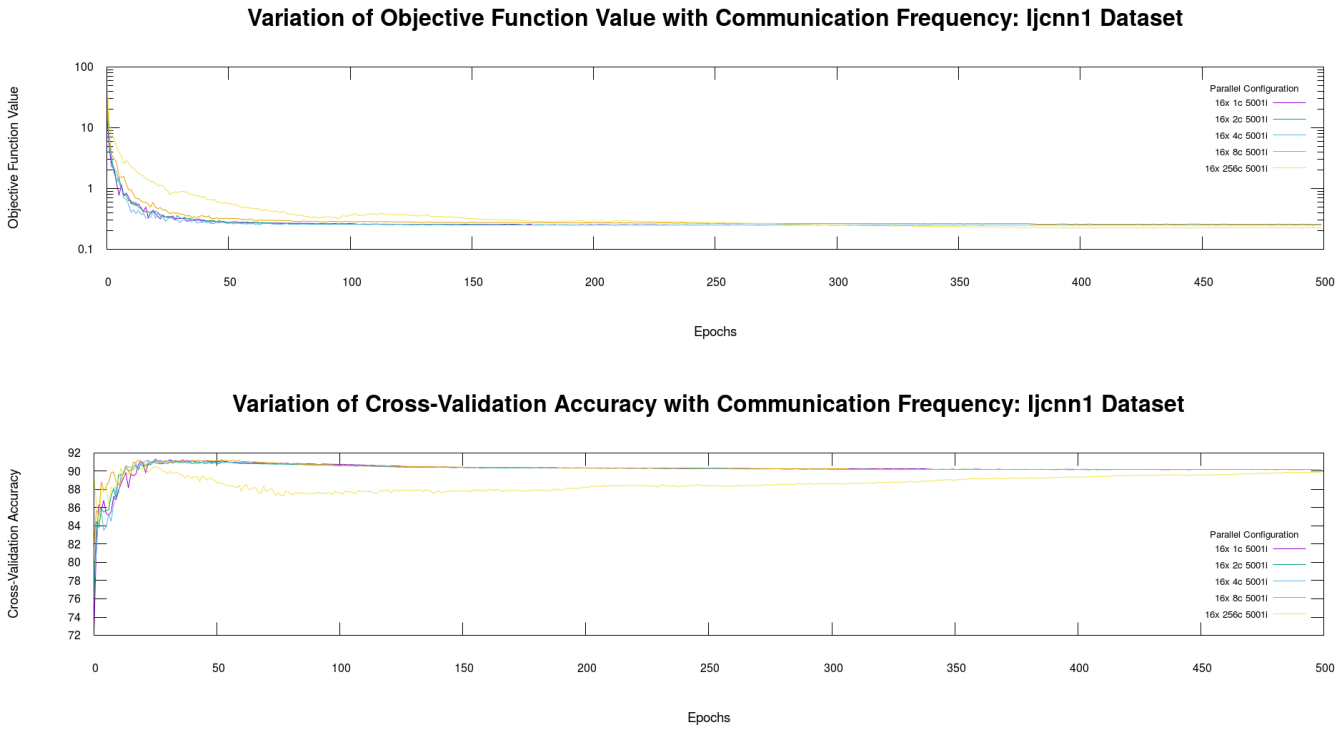

Fig. 52. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,256,], Parallelism = 16

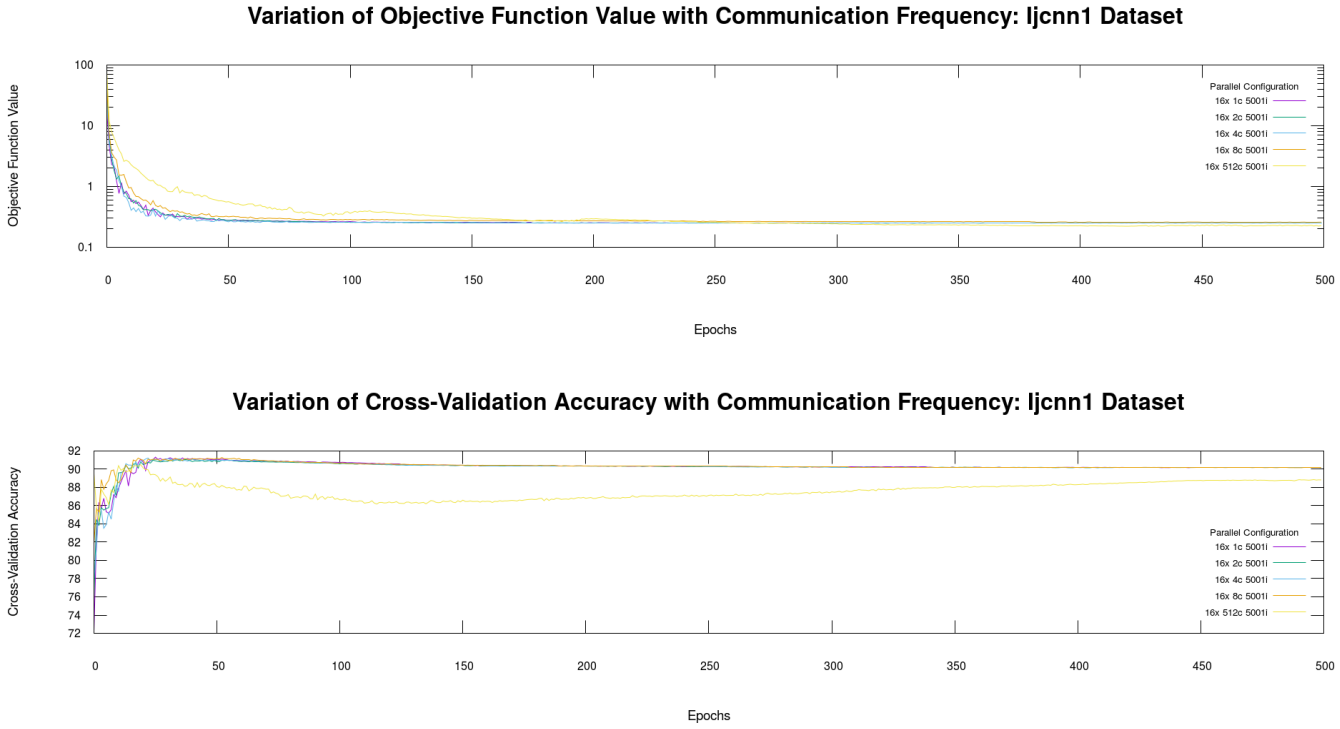

Fig. 53. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,512,], Parallelism = 16

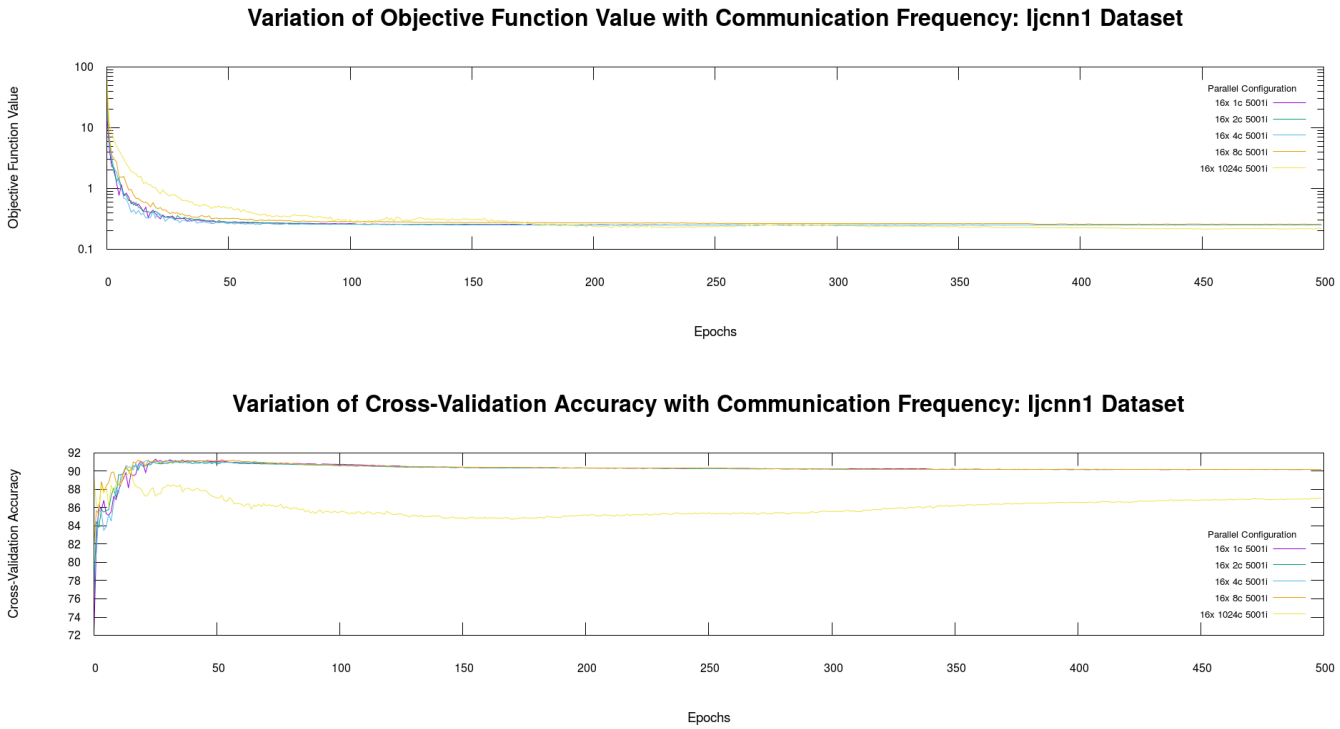

Fig. 54. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 16

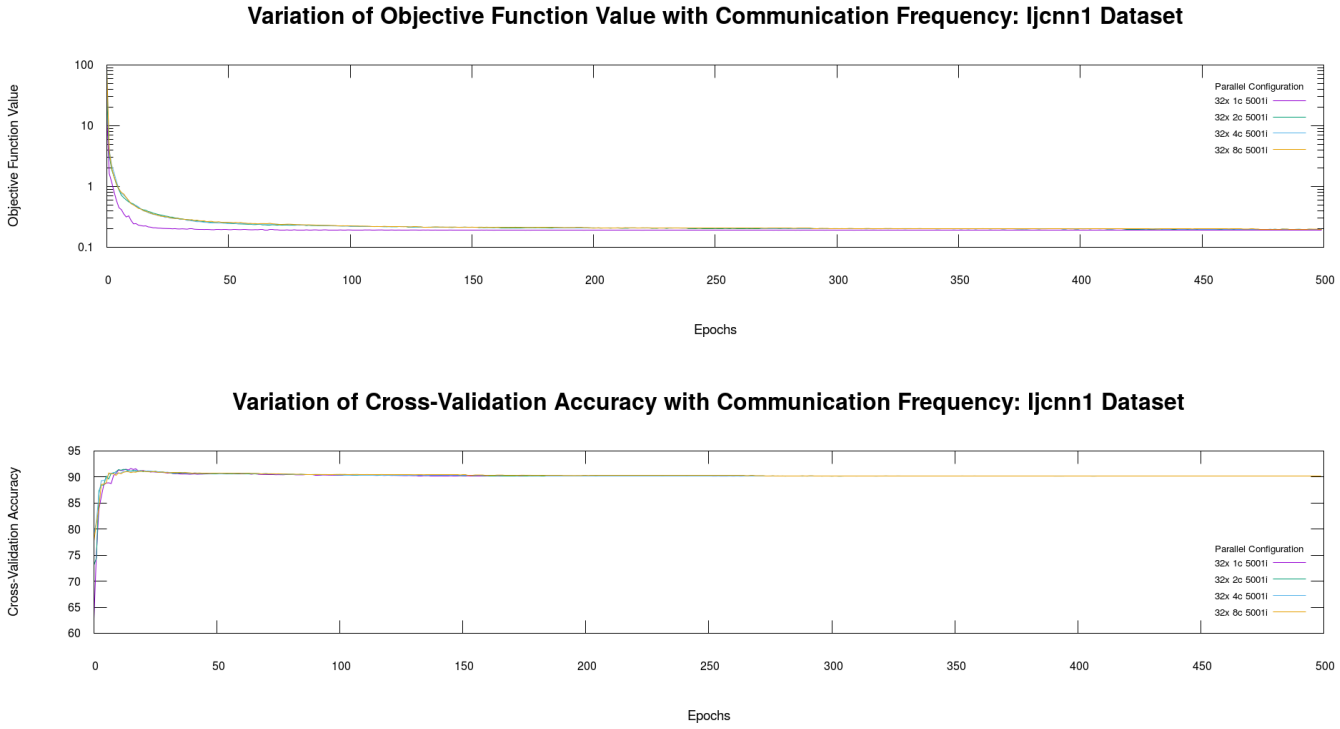

Fig. 55. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,], Parallelism = 32

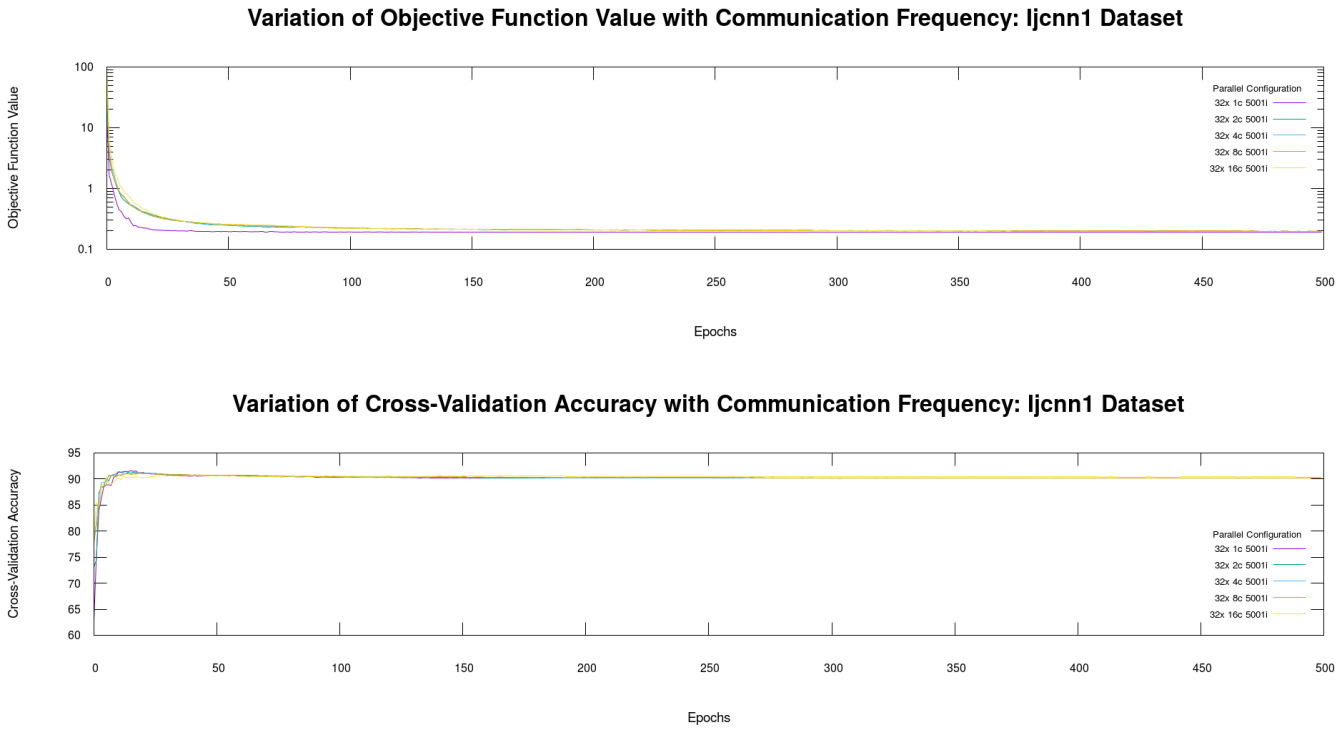

Fig. 56. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,16,], Parallelism = 32

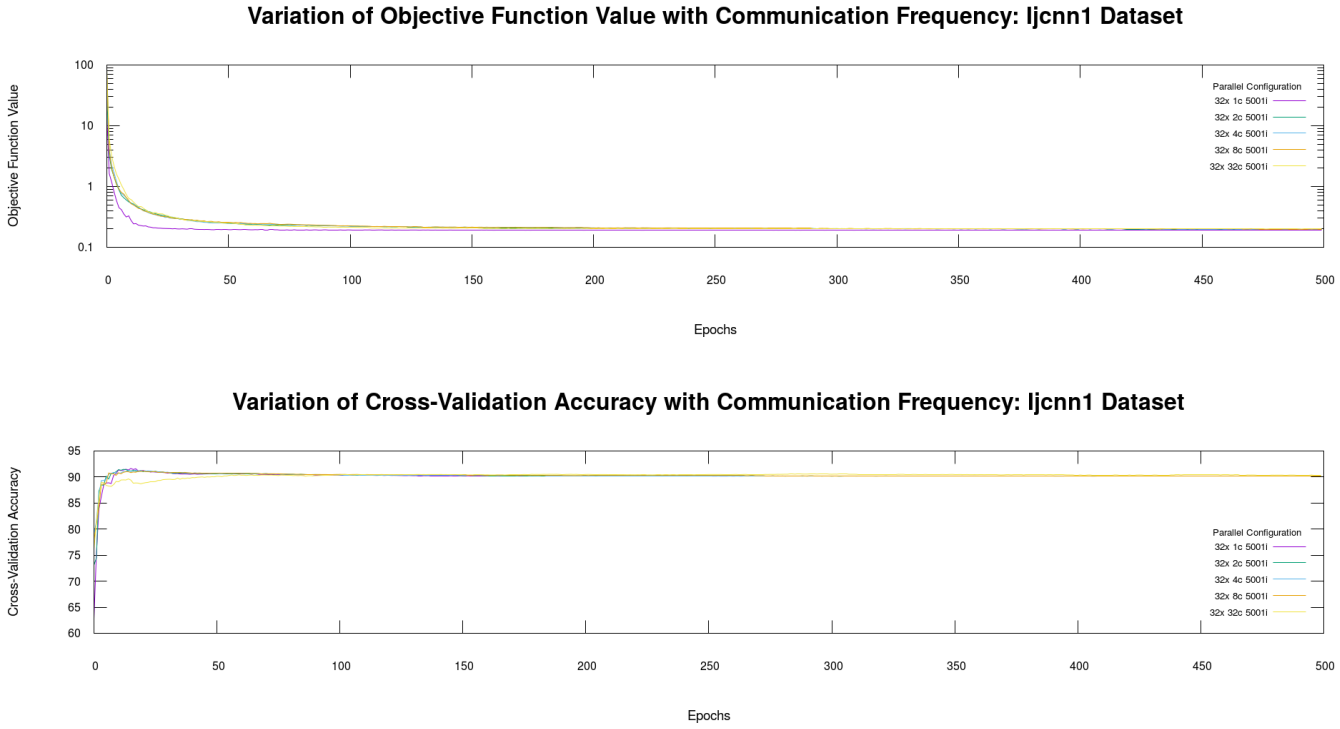

Fig. 57. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,32,], Parallelism = 32

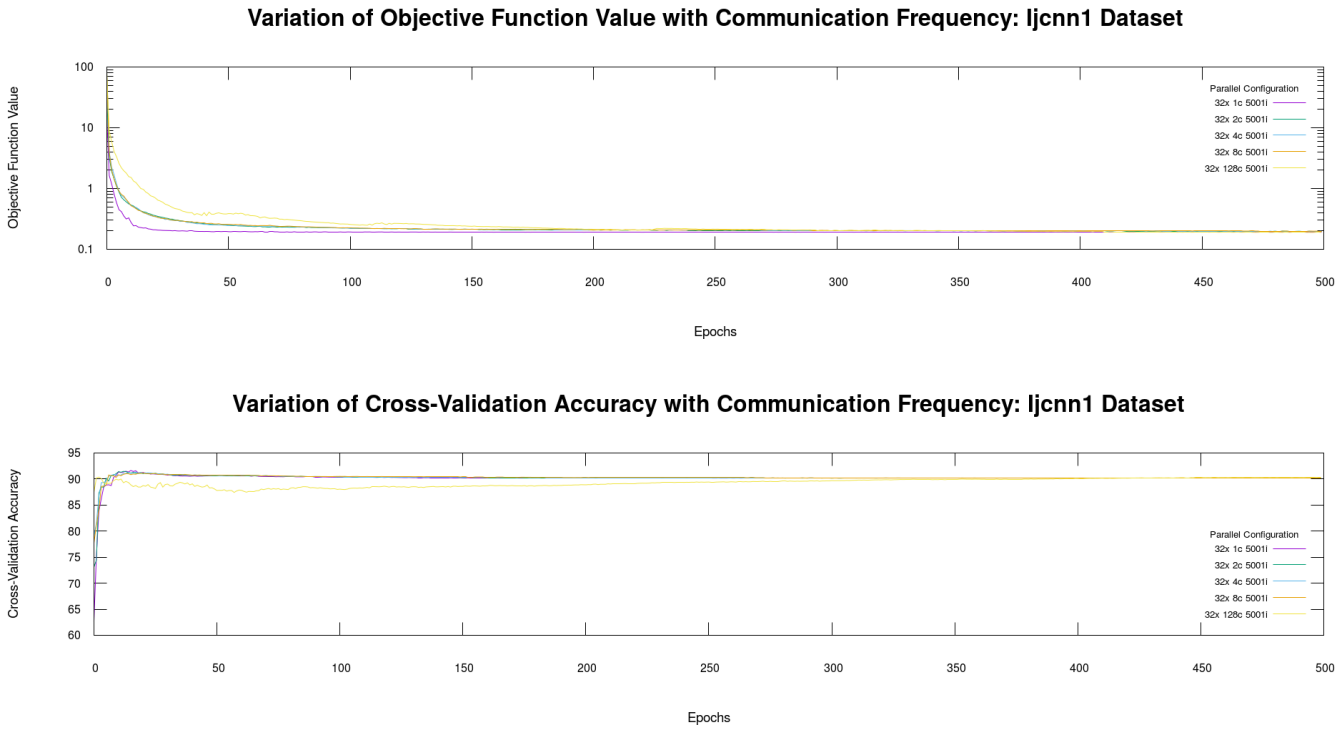

Fig. 58. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,128,], Parallelism = 32

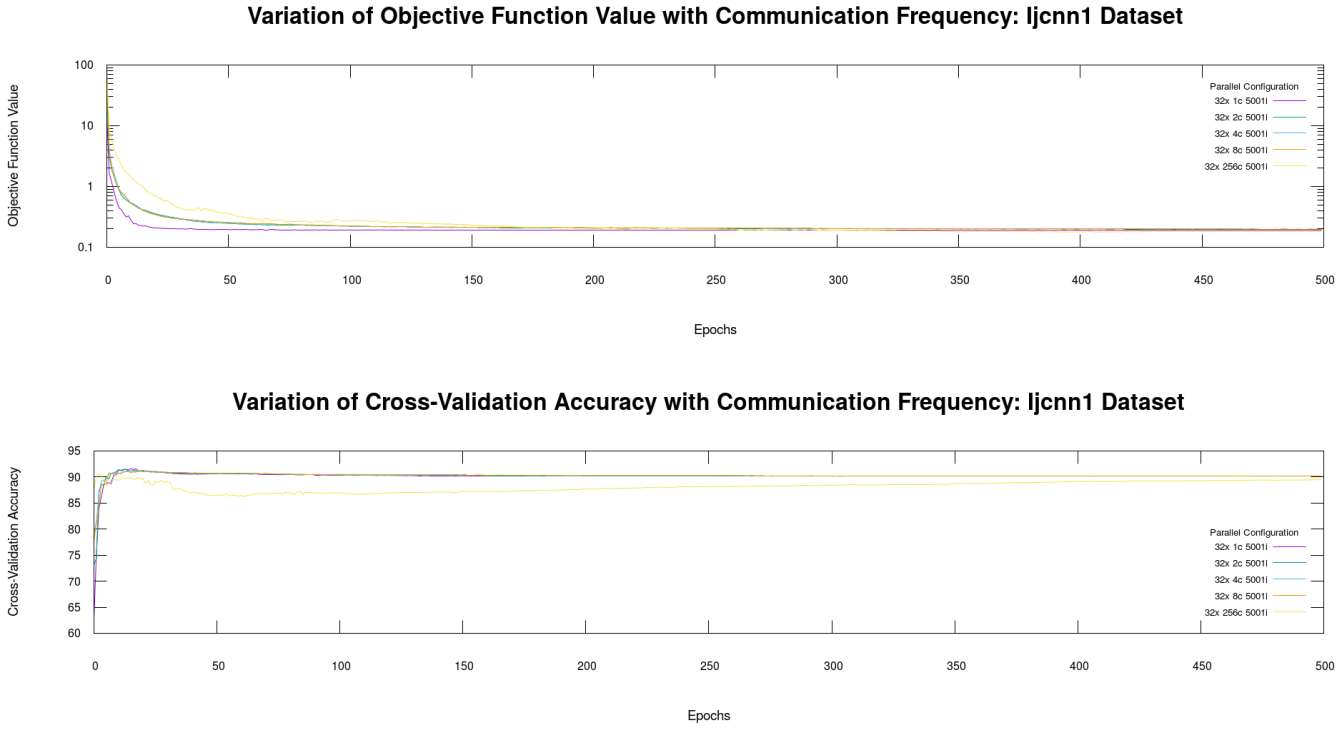

Fig. 59. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,256,], Parallelism = 32

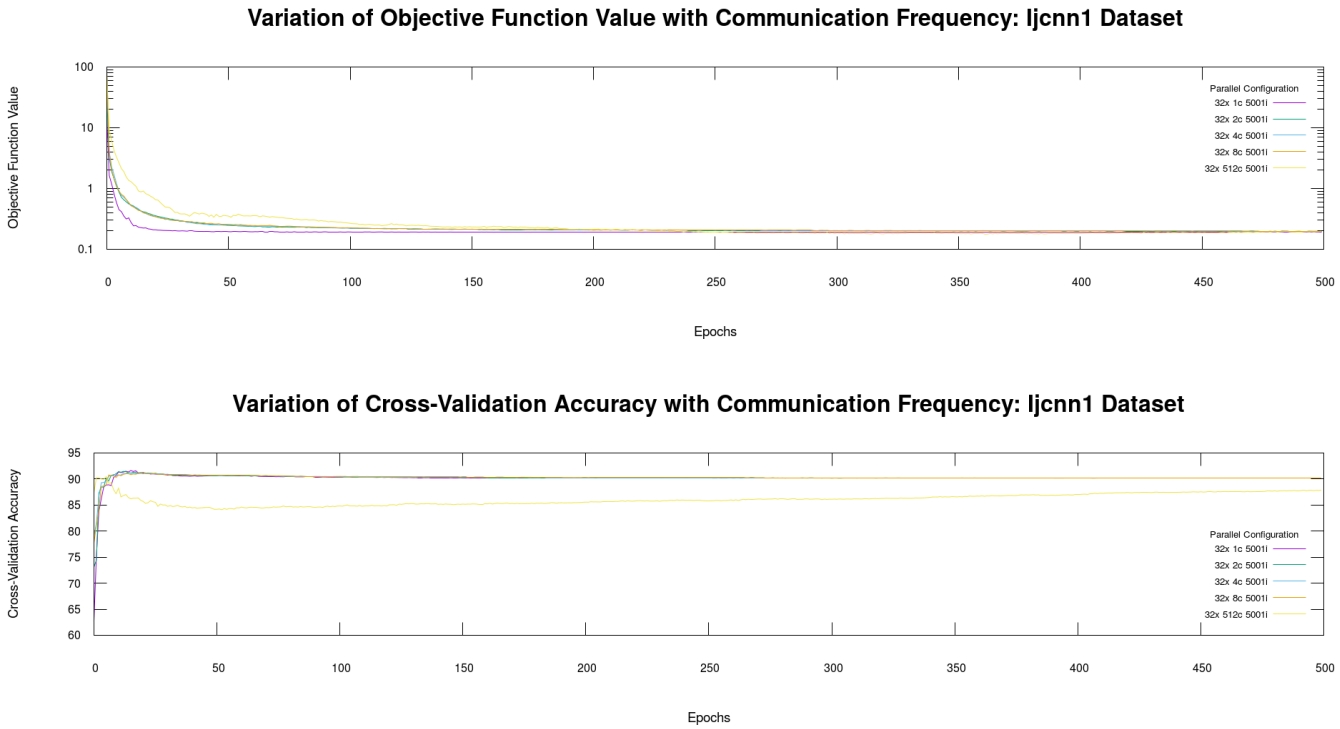

Fig. 60. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,512,], Parallelism = 32

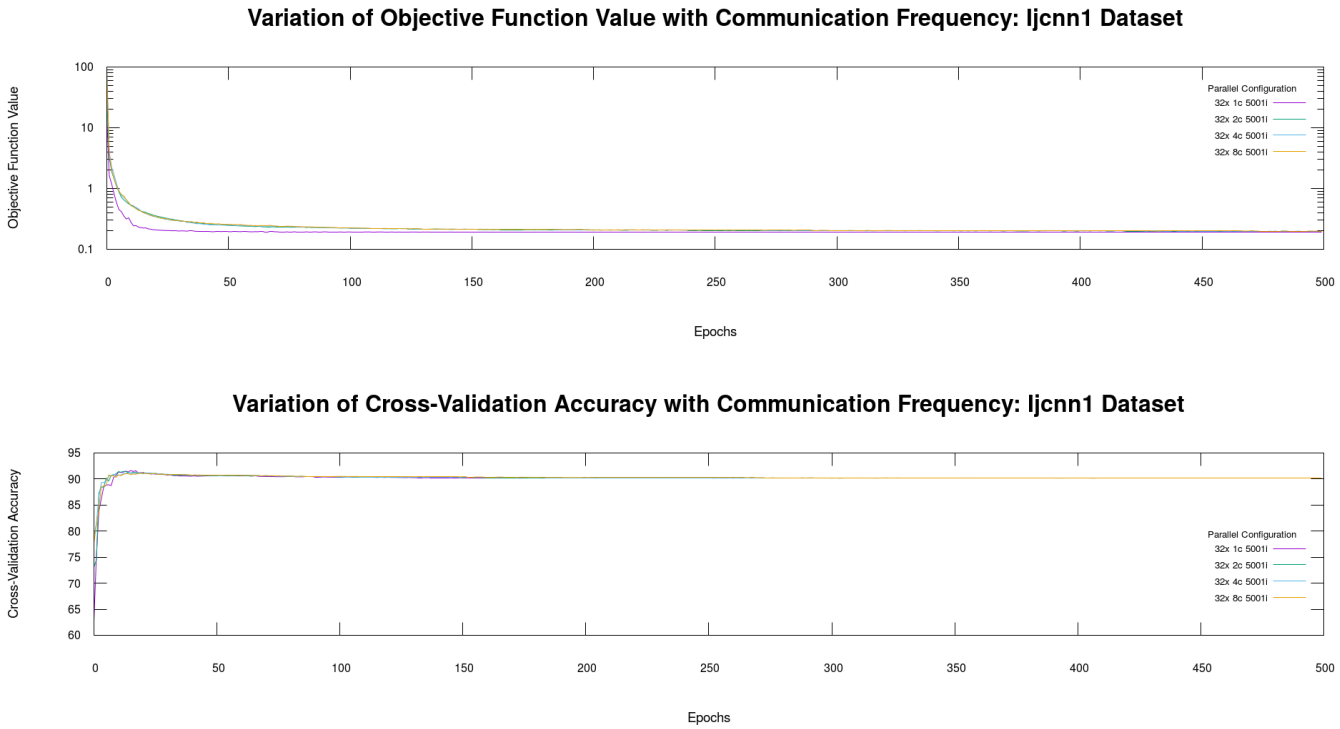

Fig. 61. Distributed Training Time : Dataset Ijcnn1 , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 32

B. Epsilon All Cross Validation Accuracy and Objective Function Value Variation against MSFs

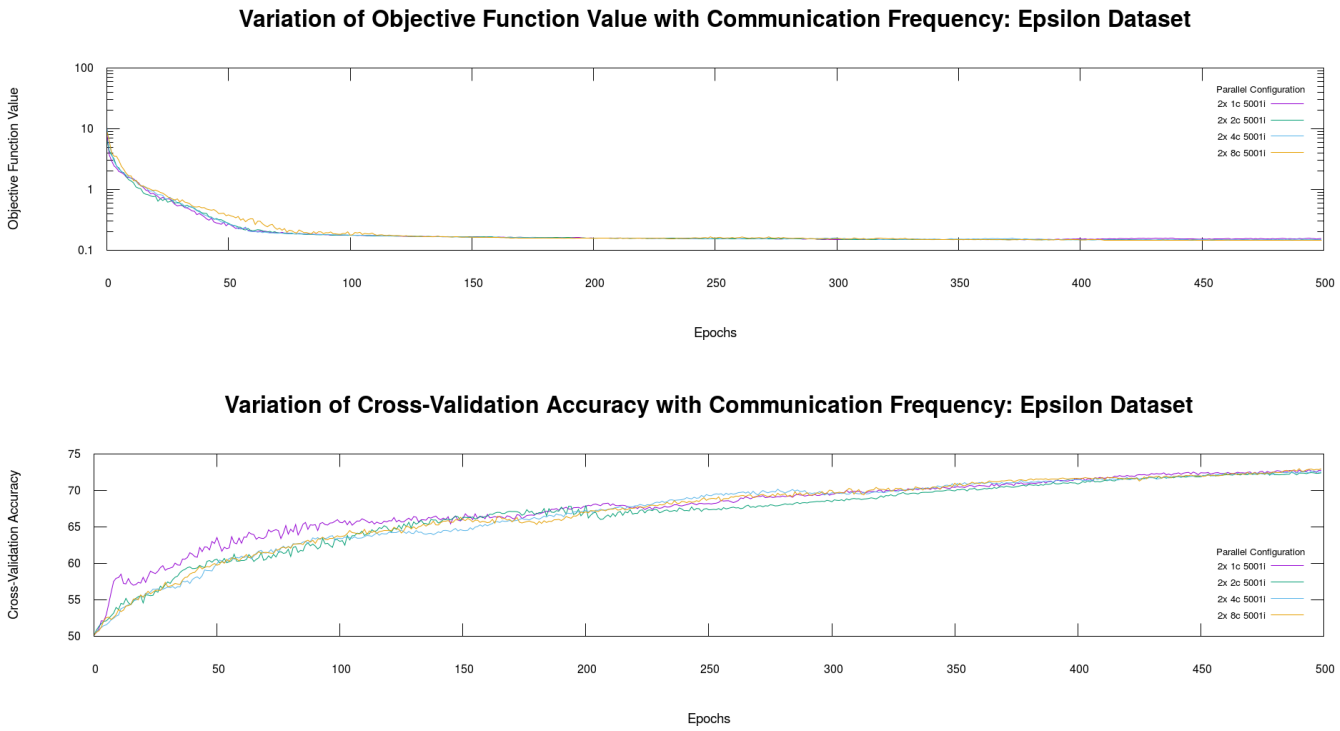

Fig. 62. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,], Parallelism = 2

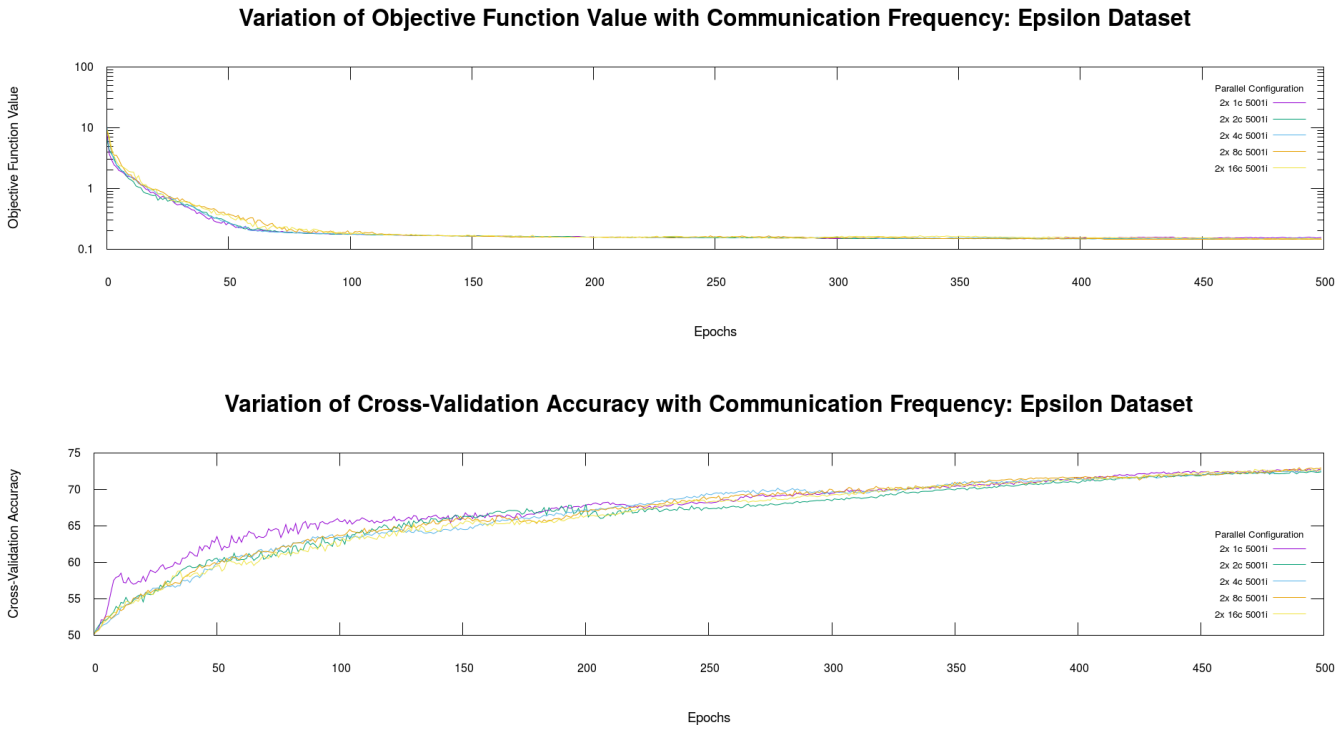

Fig. 63. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,16,], Parallelism = 2

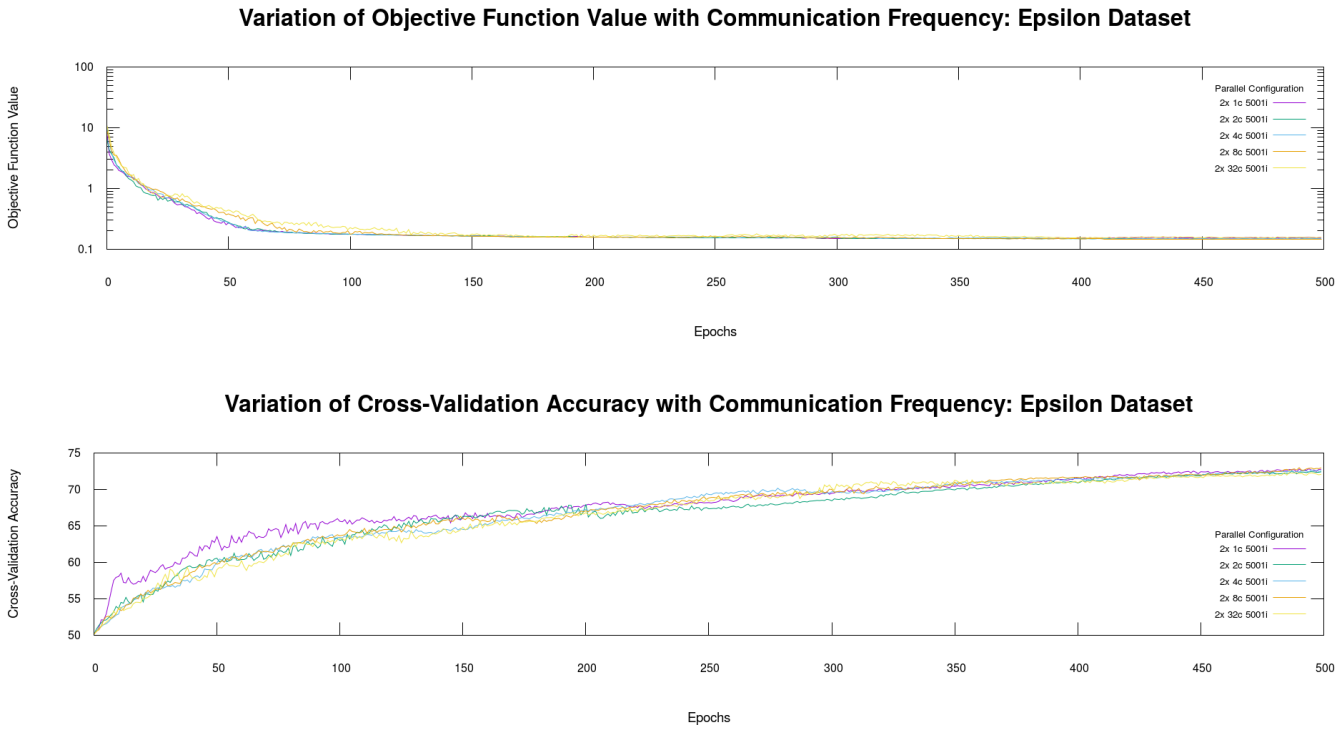

Fig. 64. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,32,], Parallelism = 2

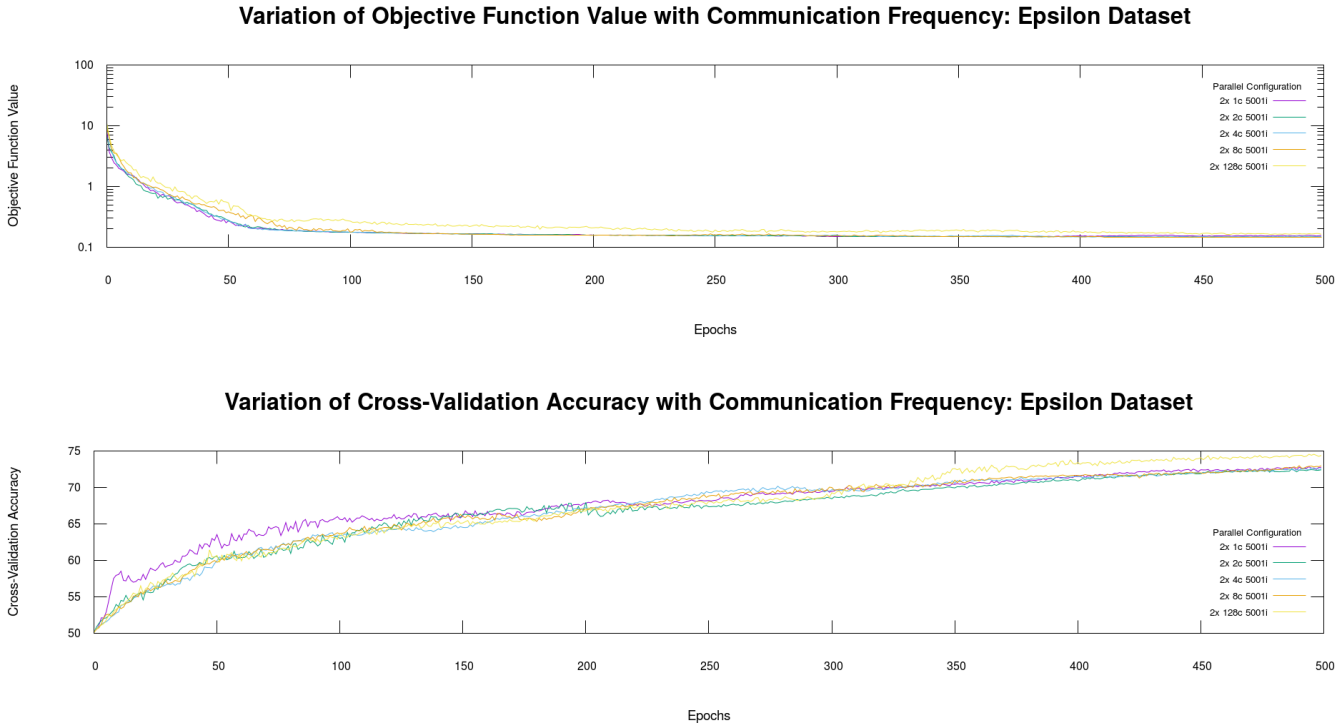

Fig. 65. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,128,], Parallelism = 2

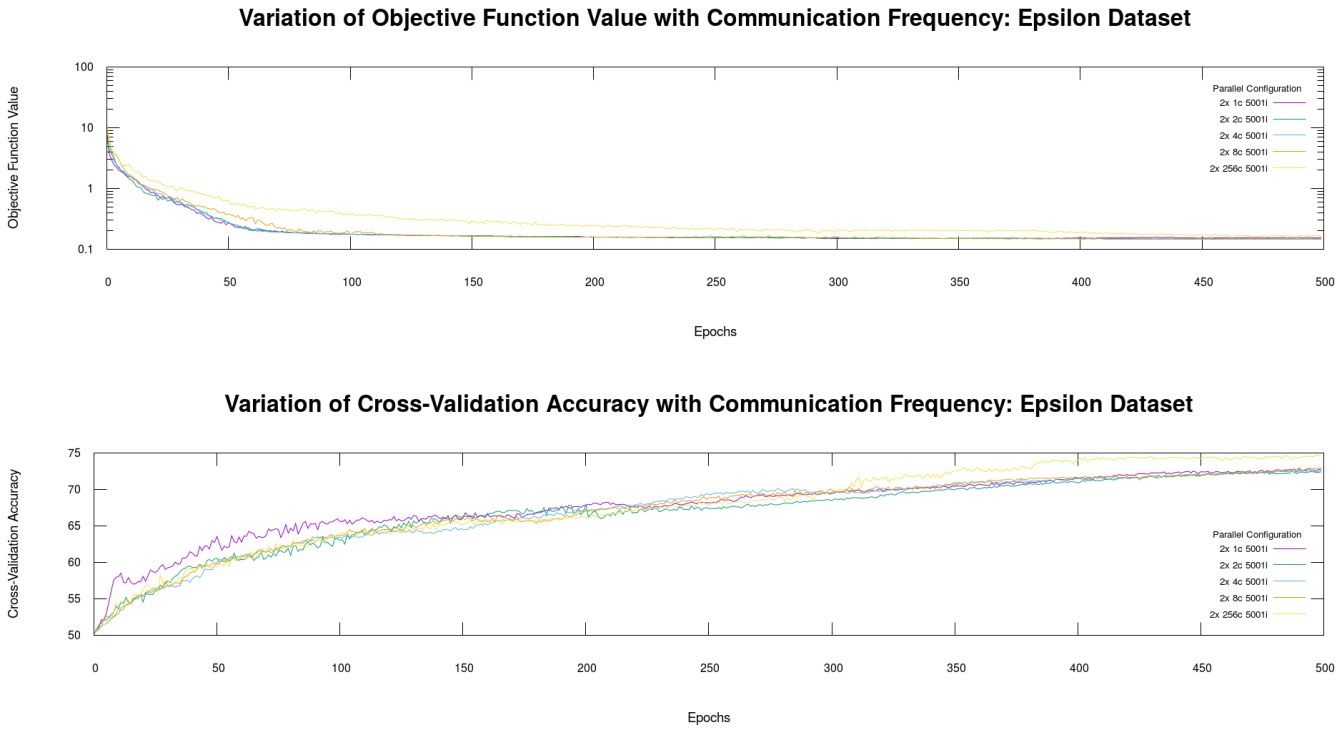

Fig. 66. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,256,], Parallelism = 2

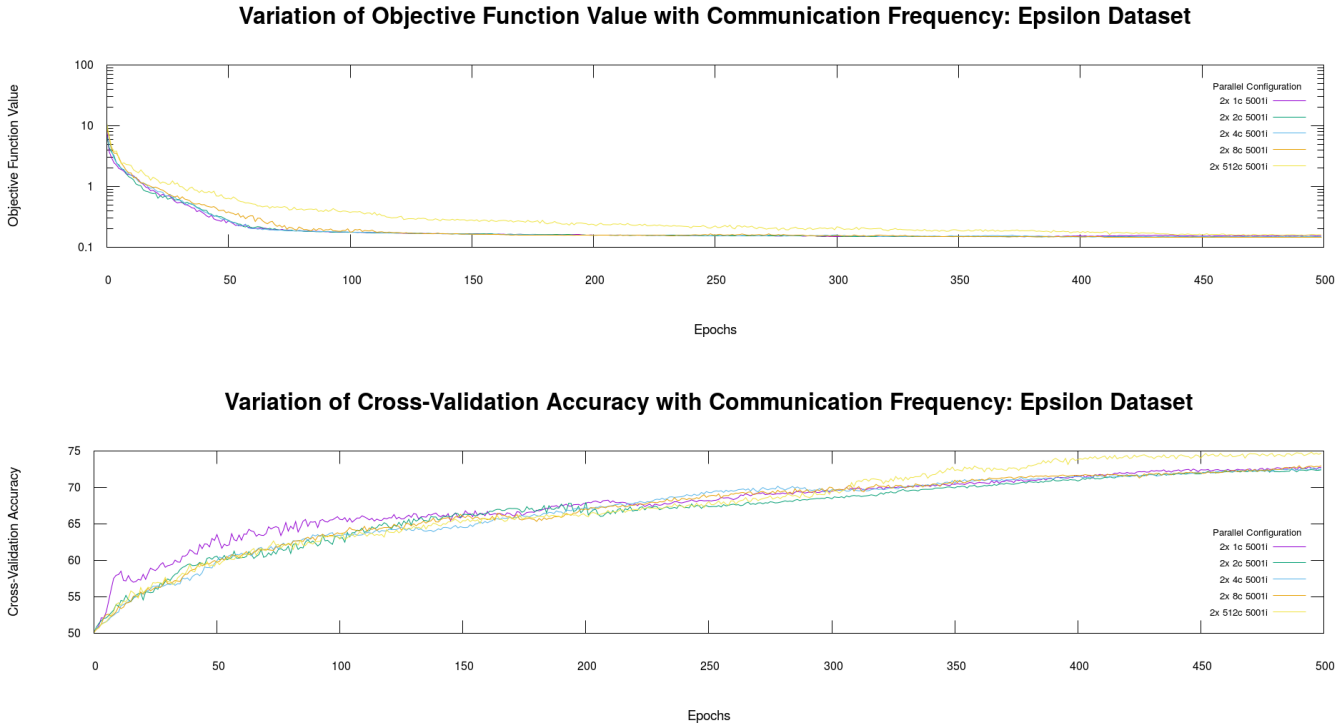

Fig. 67. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,512,], Parallelism = 2

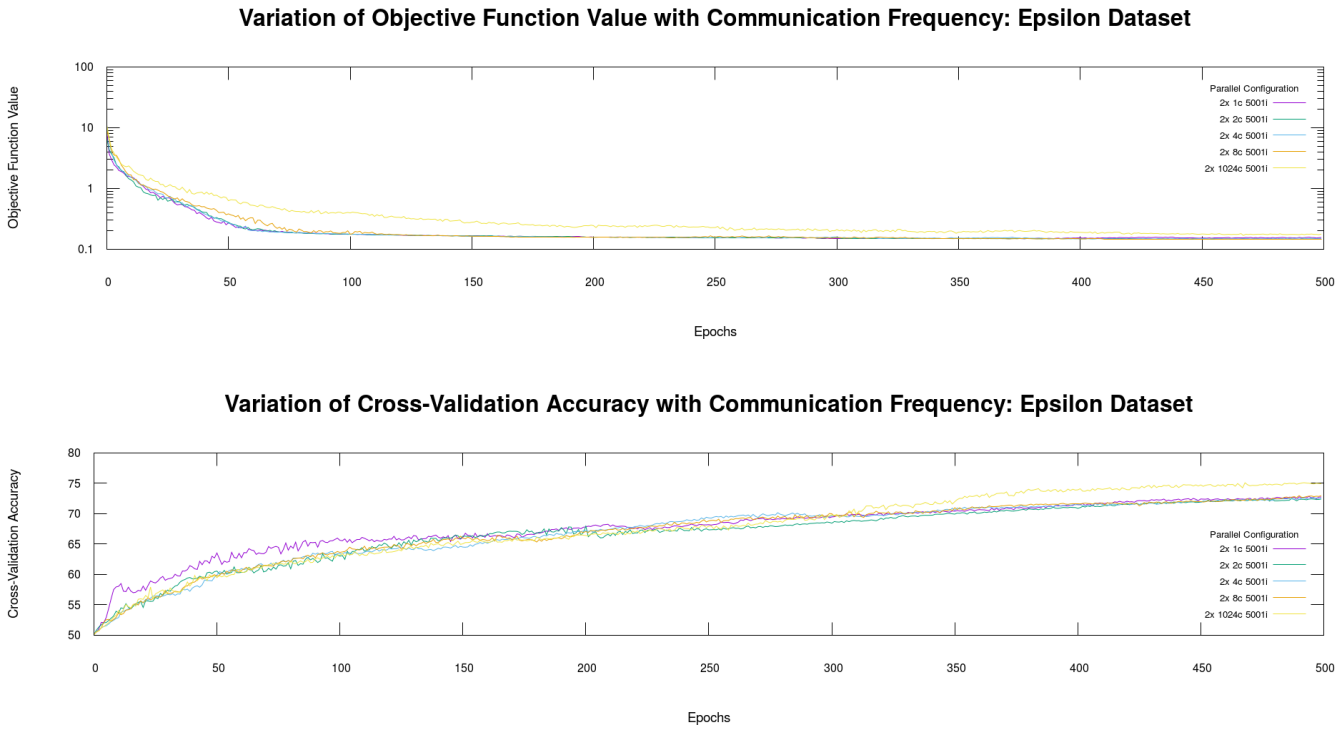

Fig. 68. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 2

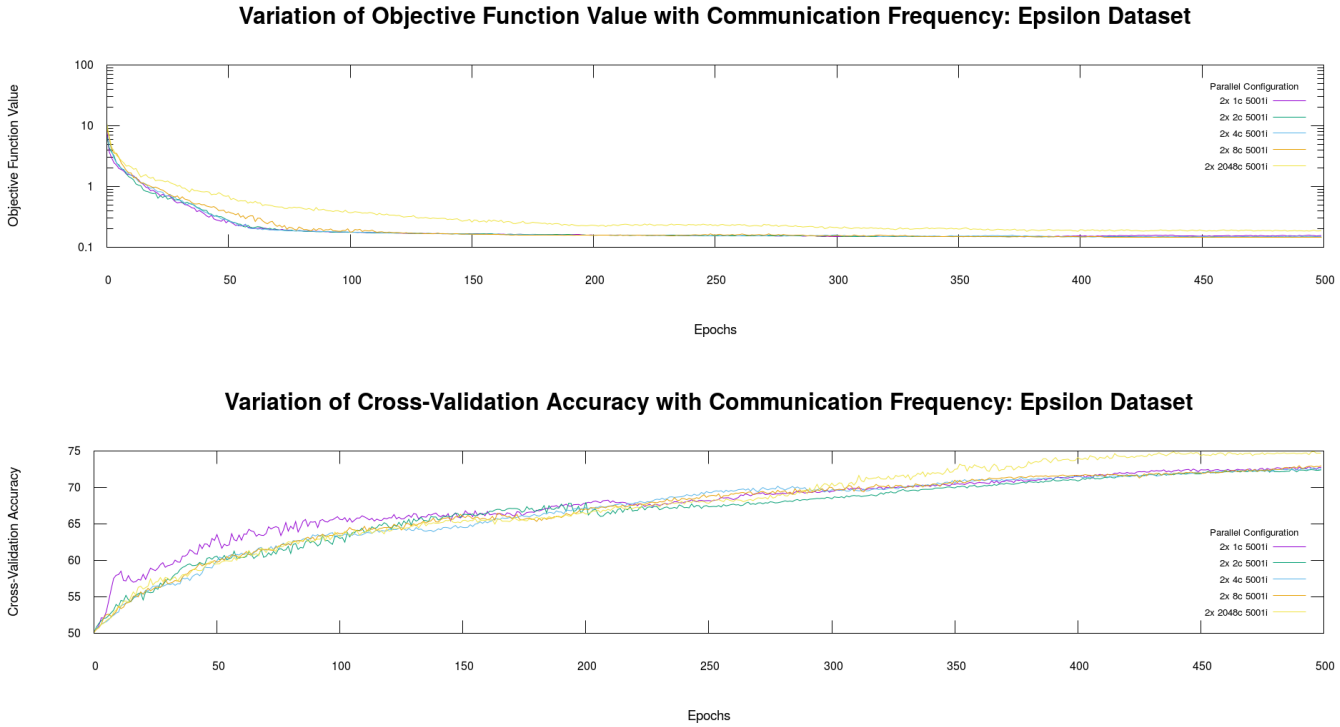

Fig. 69. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,2048,], Parallelism = 2

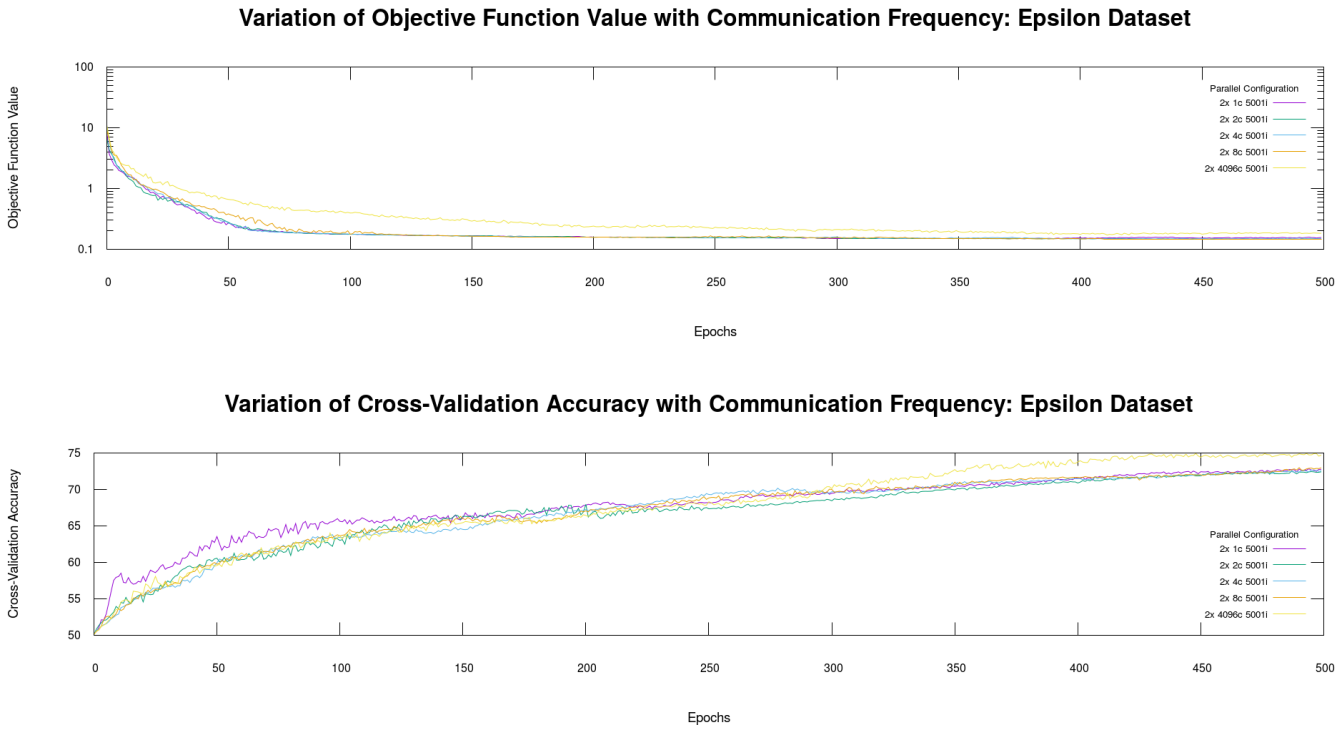

Fig. 70. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,4096,], Parallelism = 2

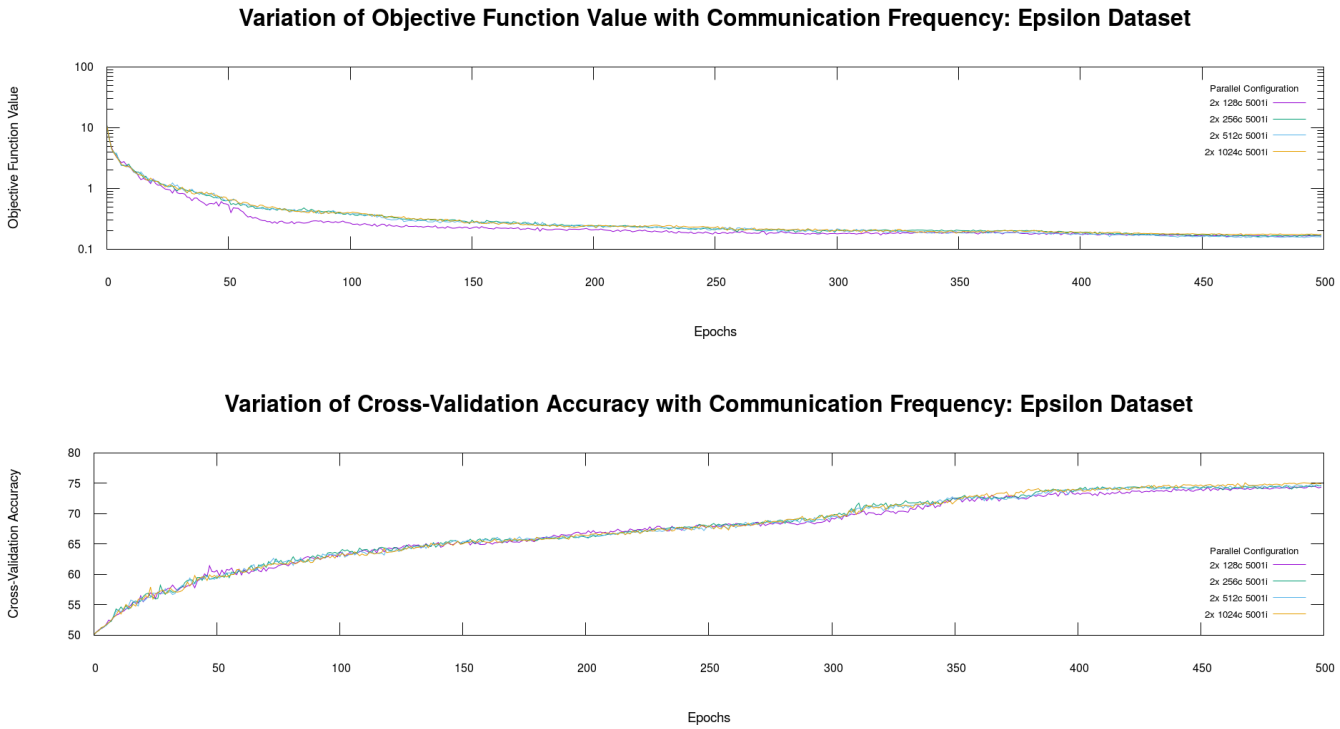

Fig. 71. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [128,256,512,1024,], Parallelism = 2

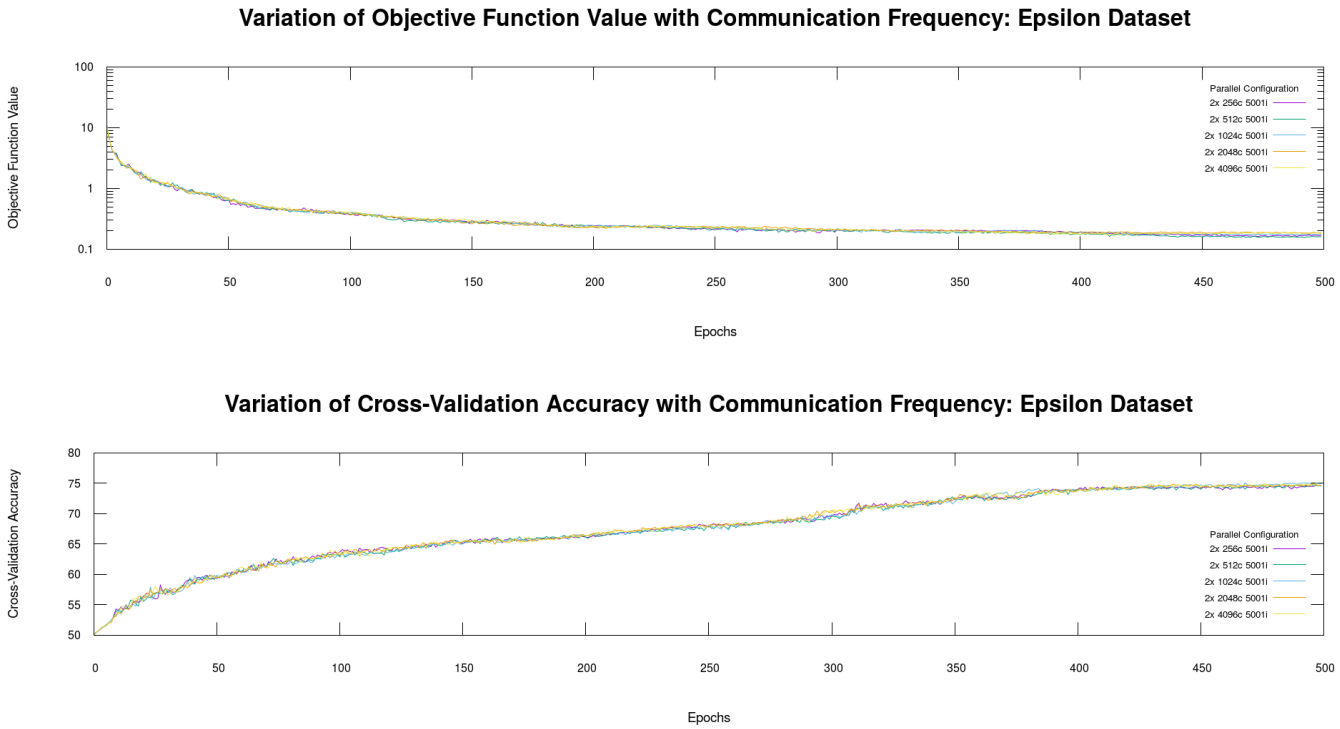

Fig. 72. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [256,512,1024,2048,4096,], Parallelism = 2

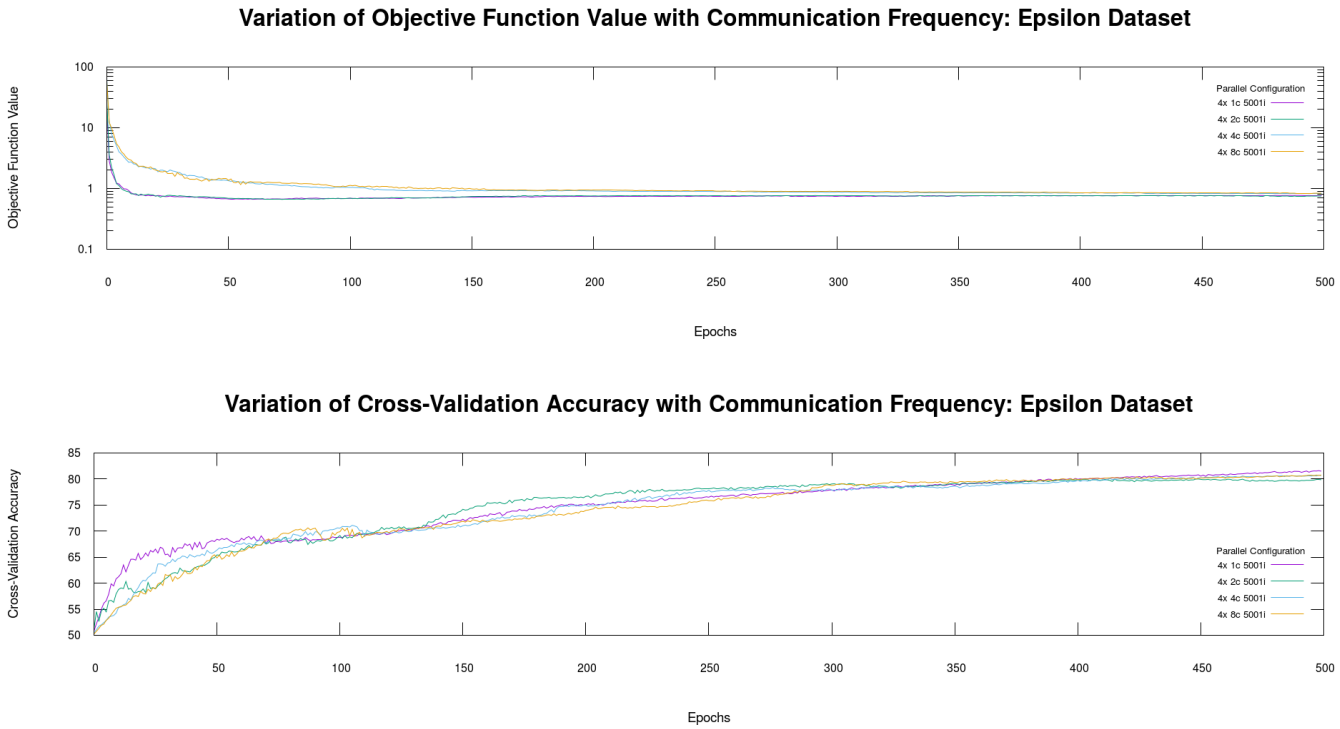

Fig. 73. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,], Parallelism = 4

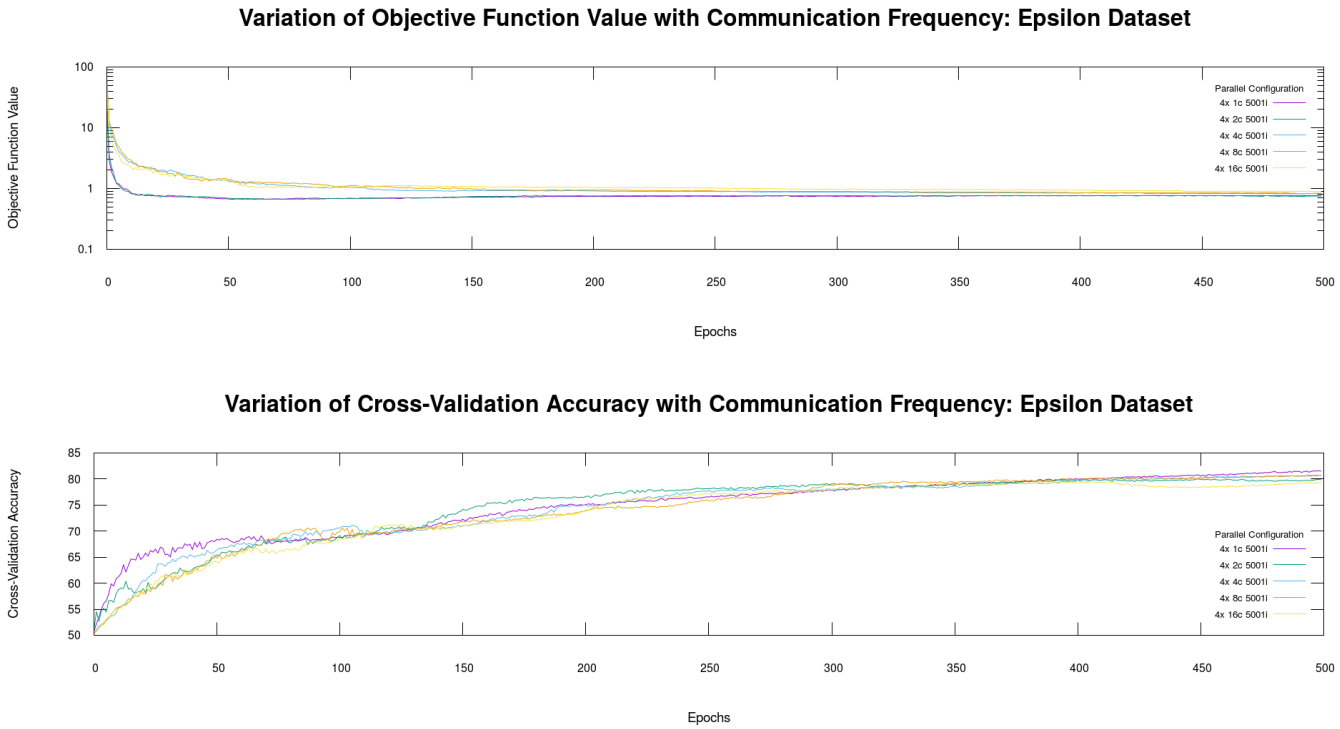

Fig. 74. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,16,], Parallelism = 4

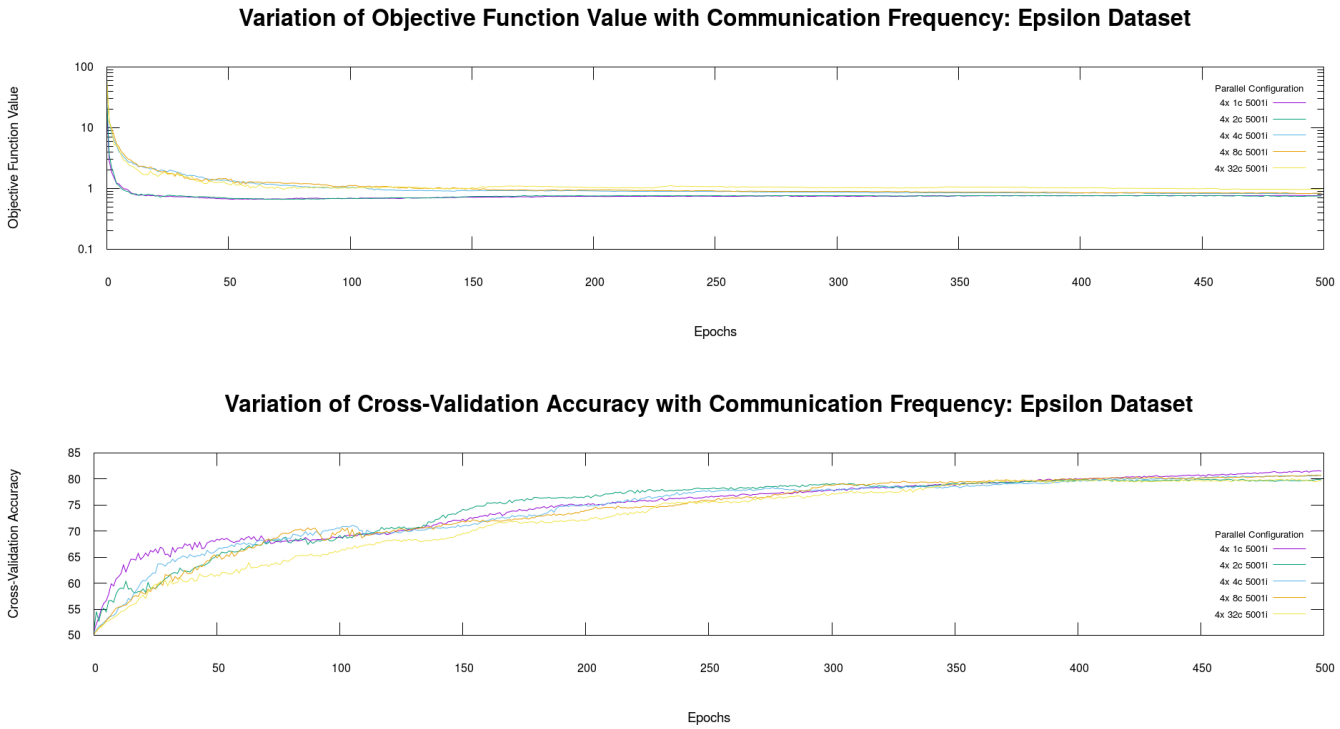

Fig. 75. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,32,], Parallelism = 4

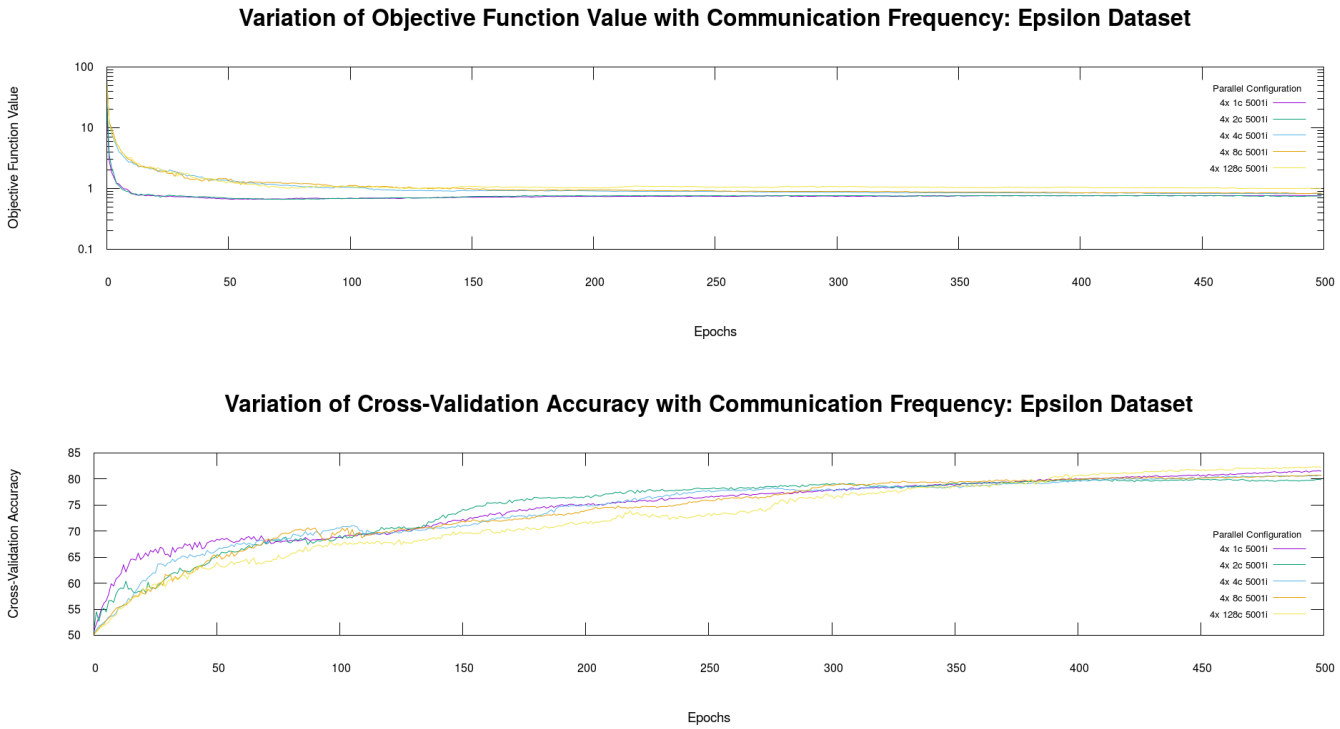

Fig. 76. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,128,], Parallelism = 4

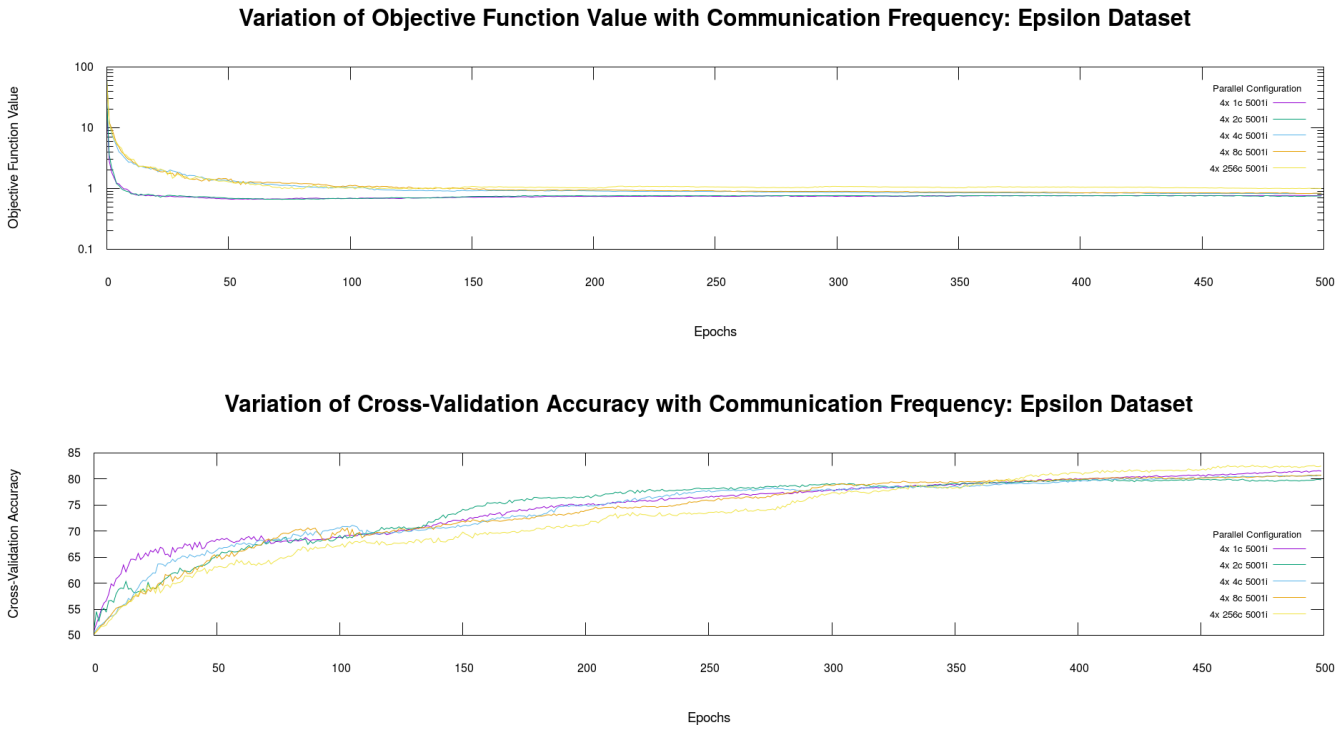

Fig. 77. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,256,], Parallelism = 4

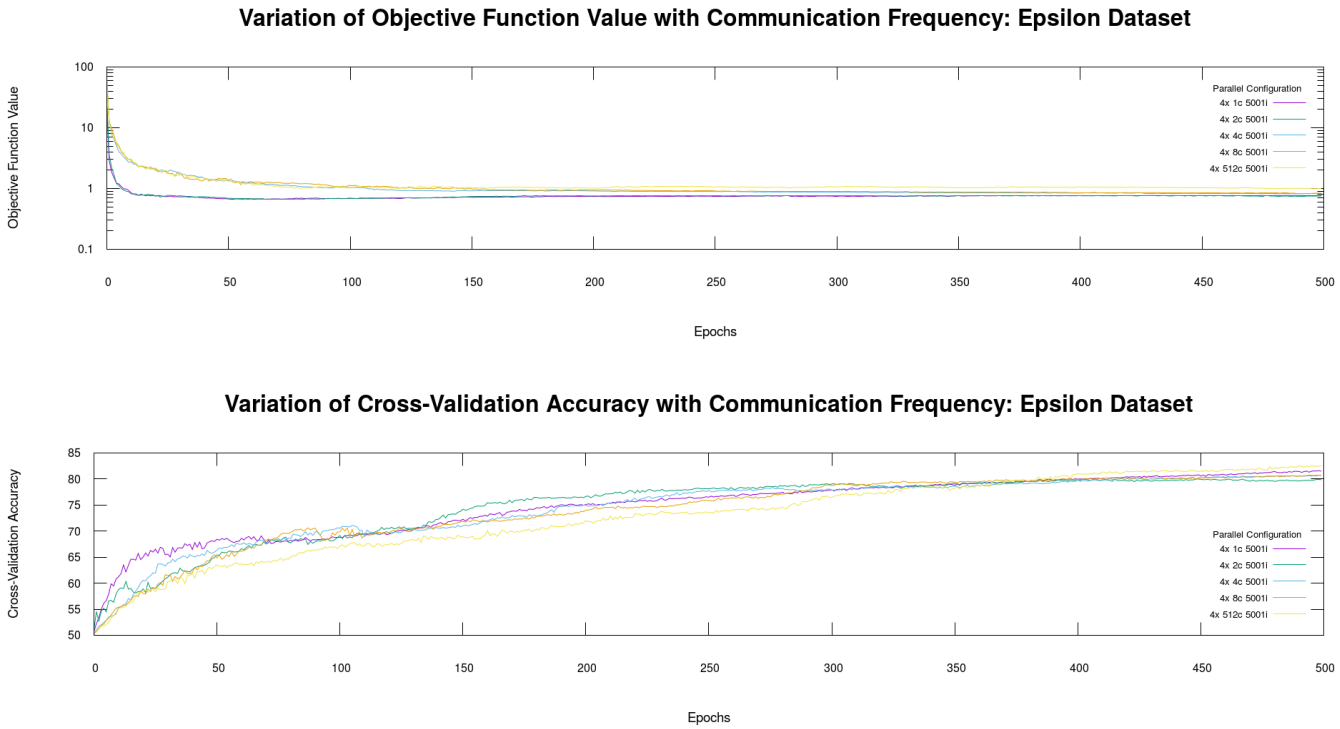

Fig. 78. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,512,], Parallelism = 4

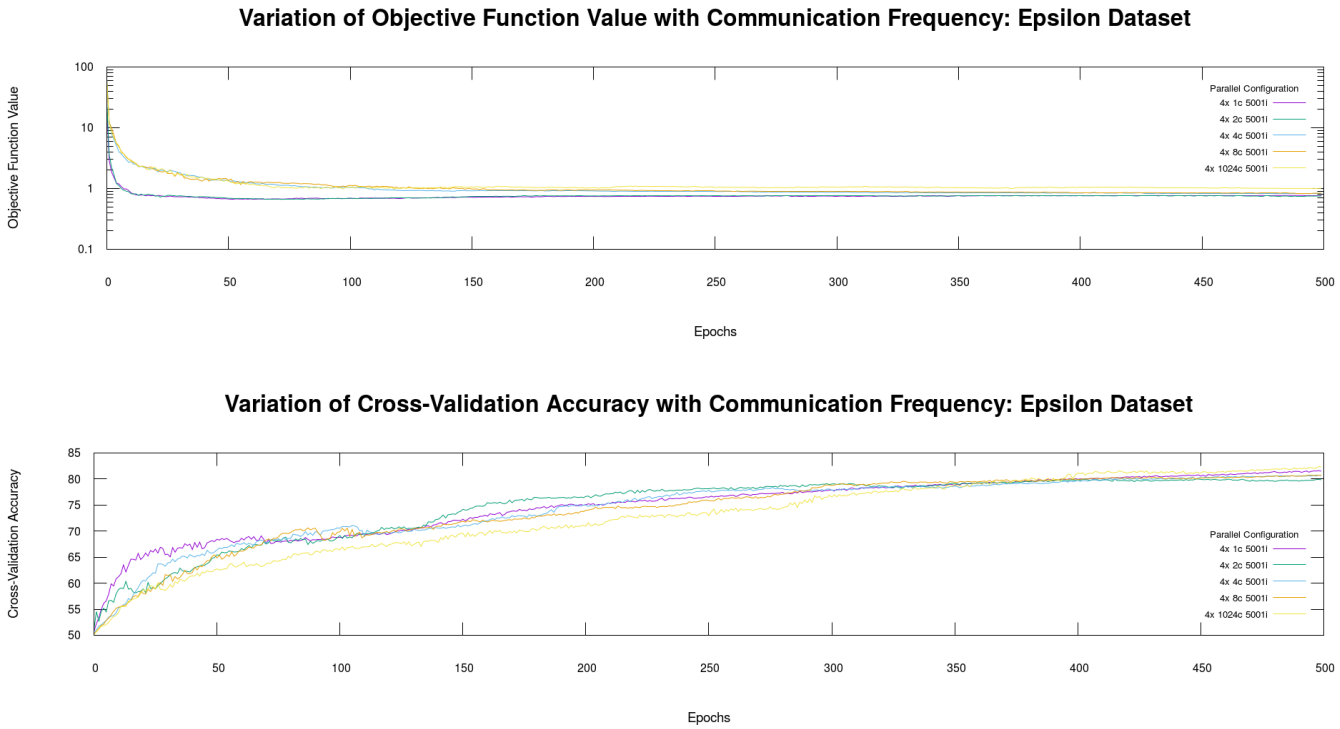

Fig. 79. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 4

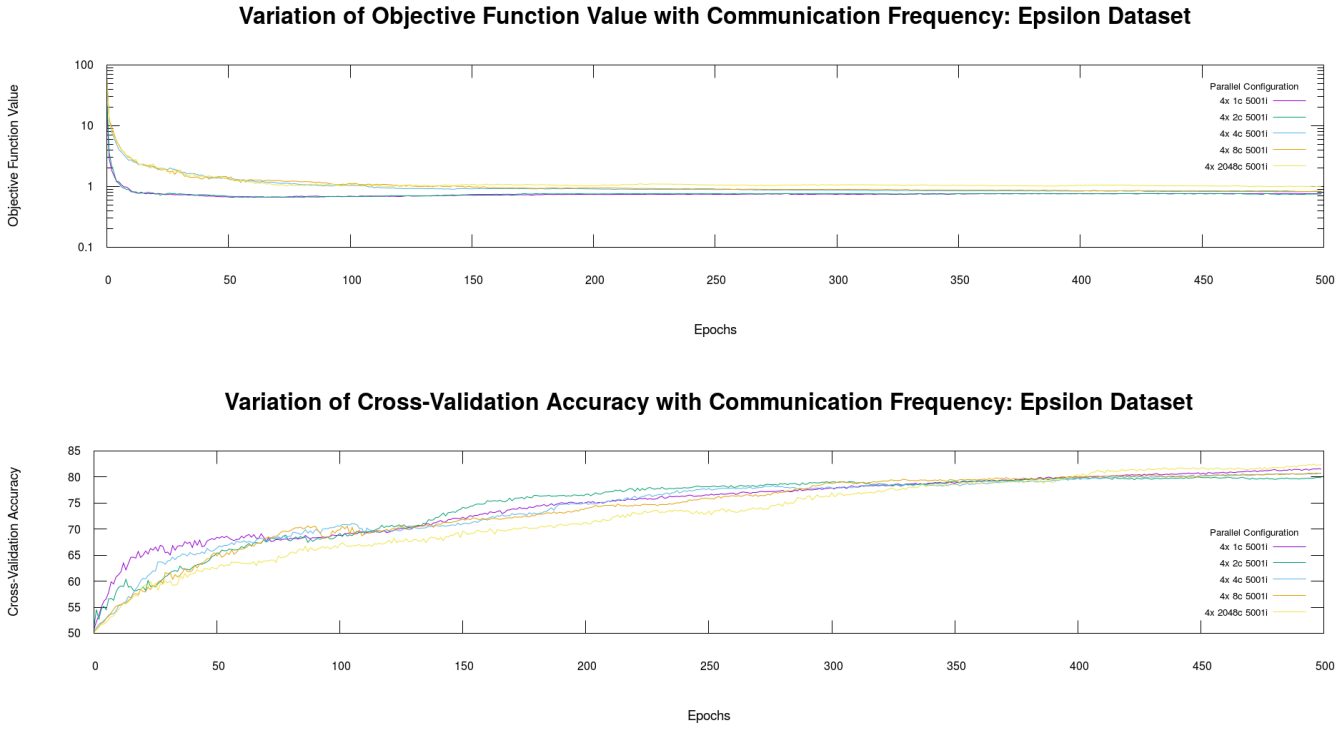

Fig. 80. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,2048,], Parallelism = 4

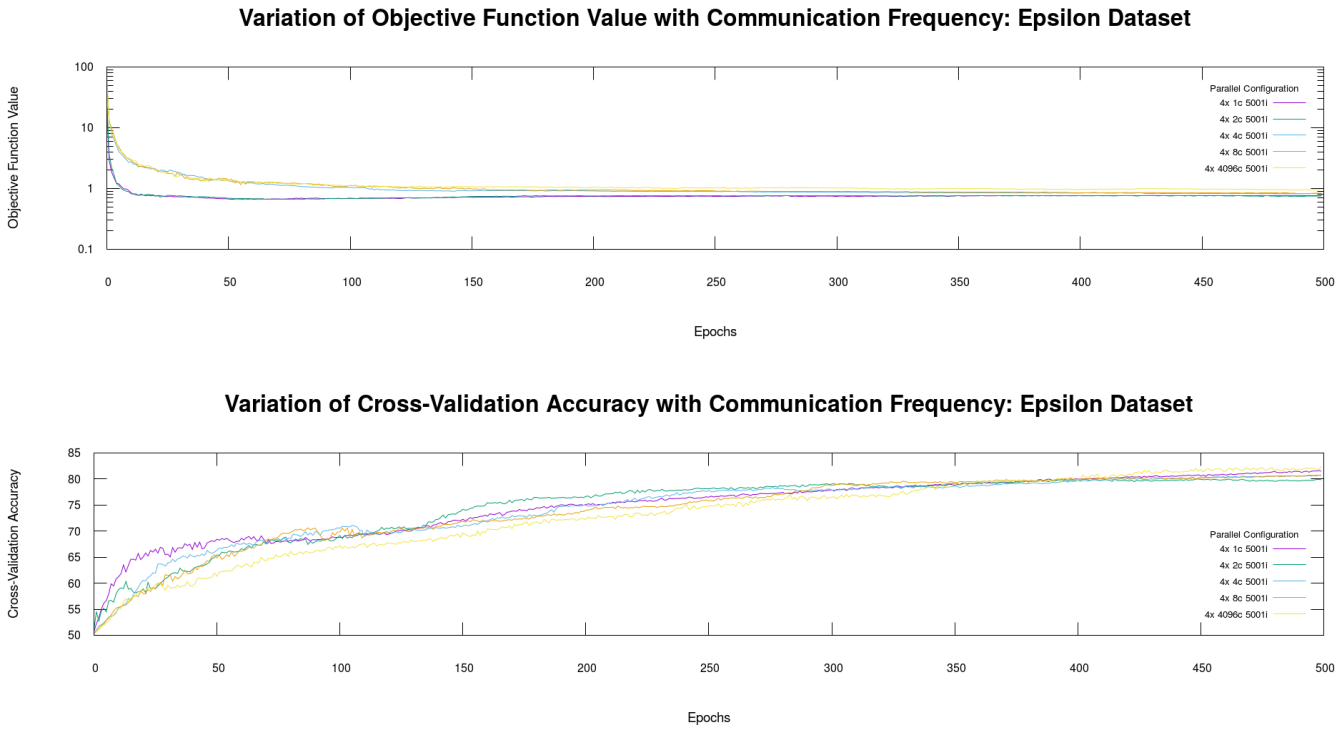

Fig. 81. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,4096,], Parallelism = 4

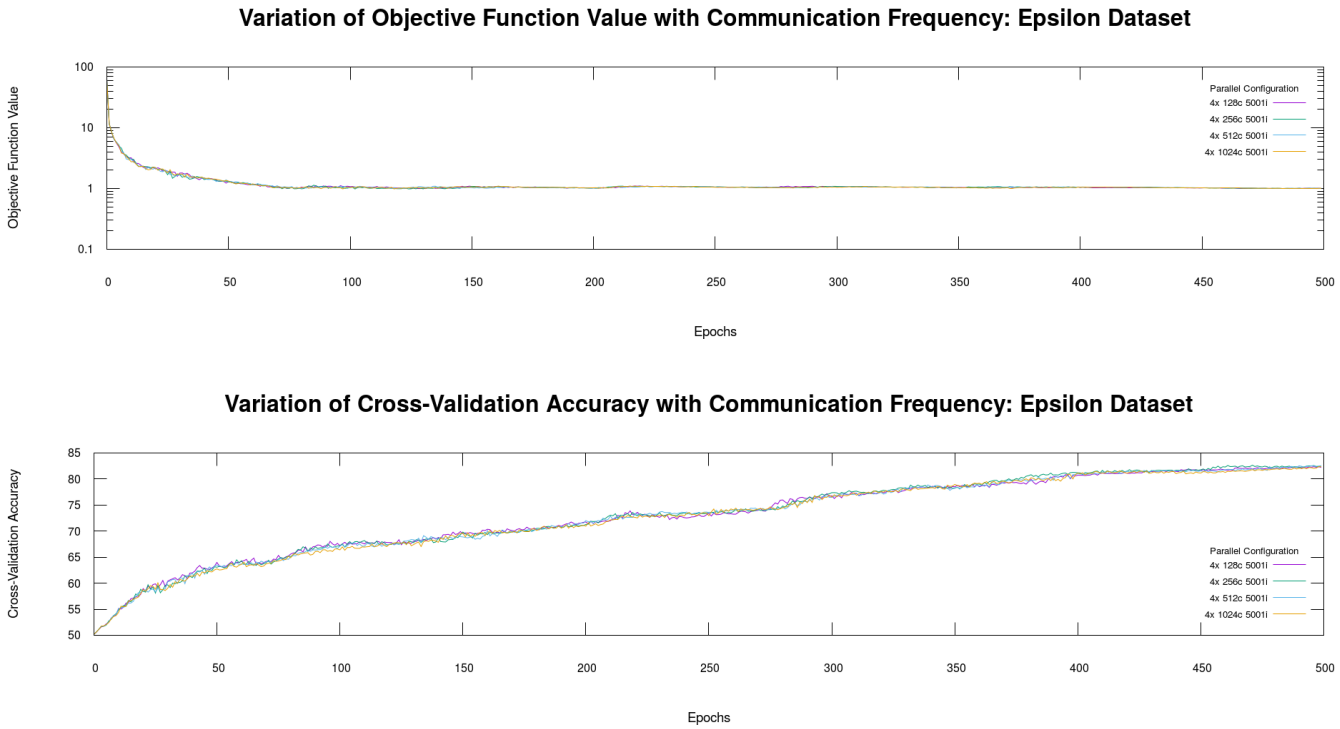

Fig. 82. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [128,256,512,1024,], Parallelism = 4

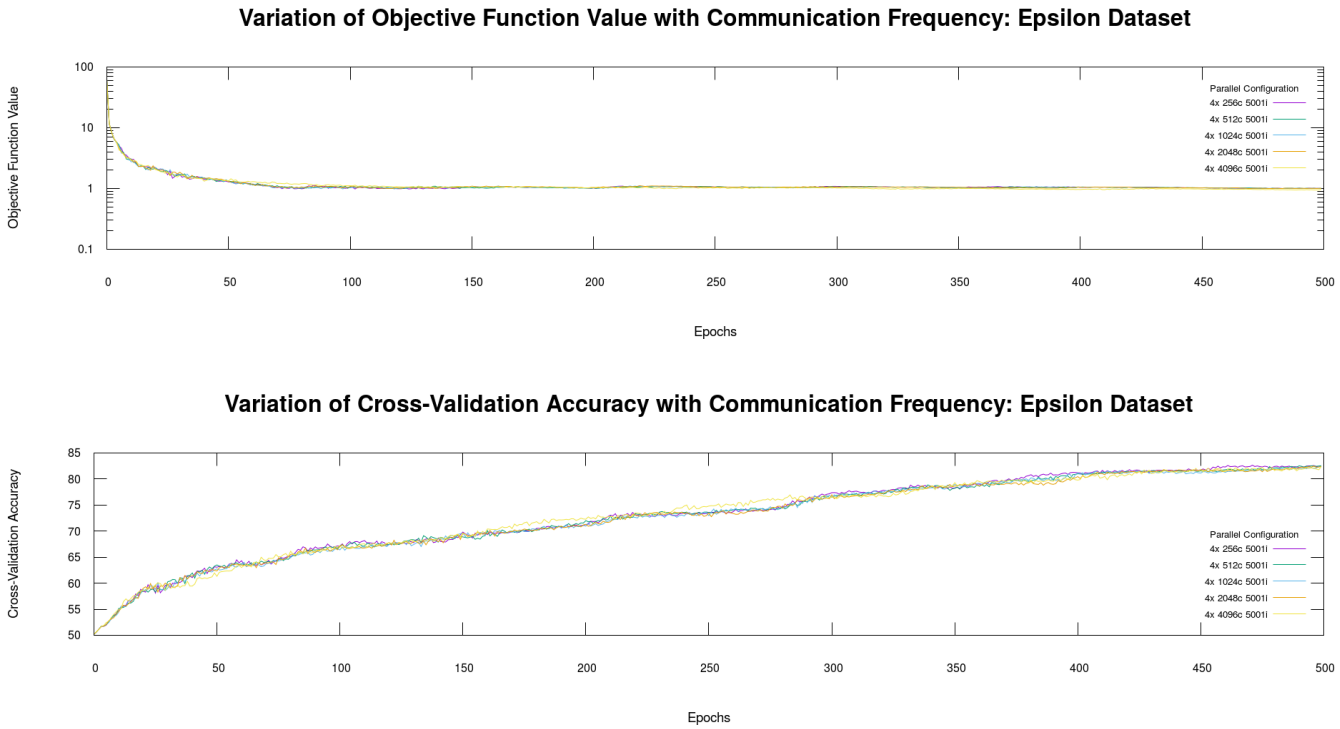

Fig. 83. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [256,512,1024,2048,4096,], Parallelism = 4

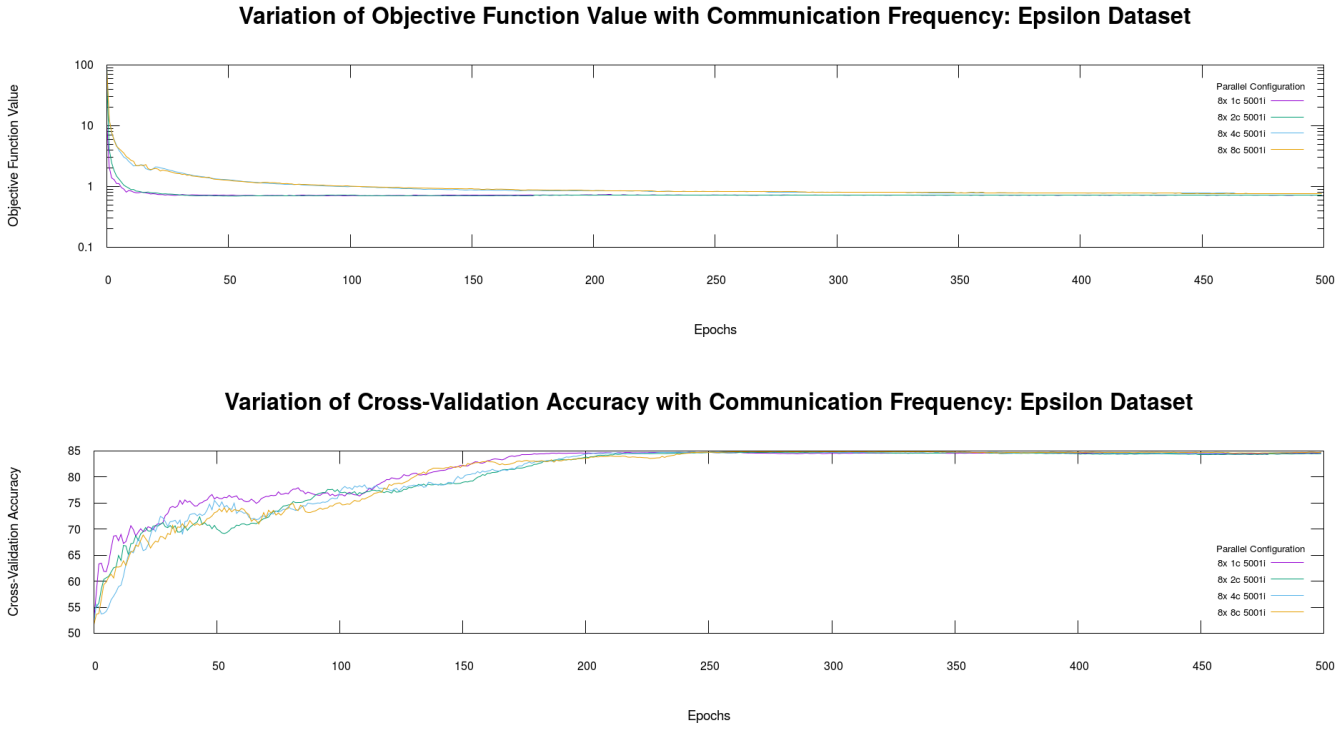

Fig. 84. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,], Parallelism = 8

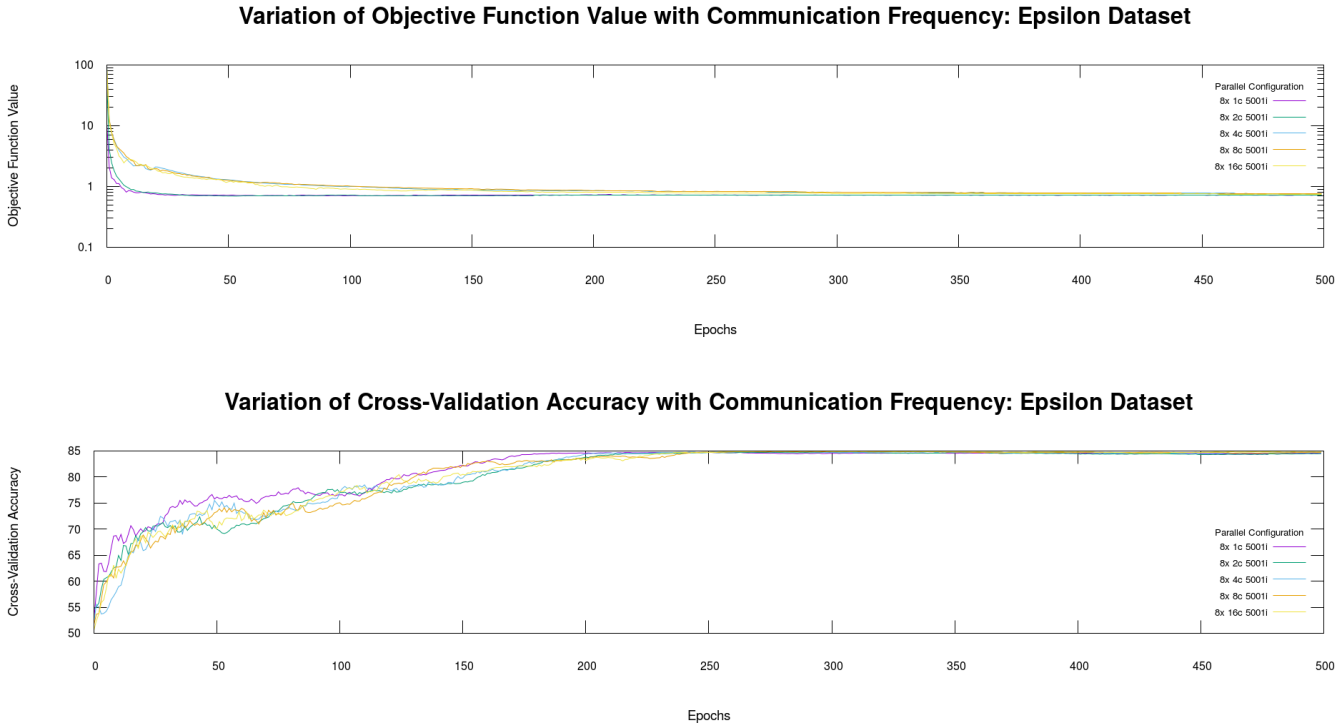

Fig. 85. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,16,], Parallelism = 8

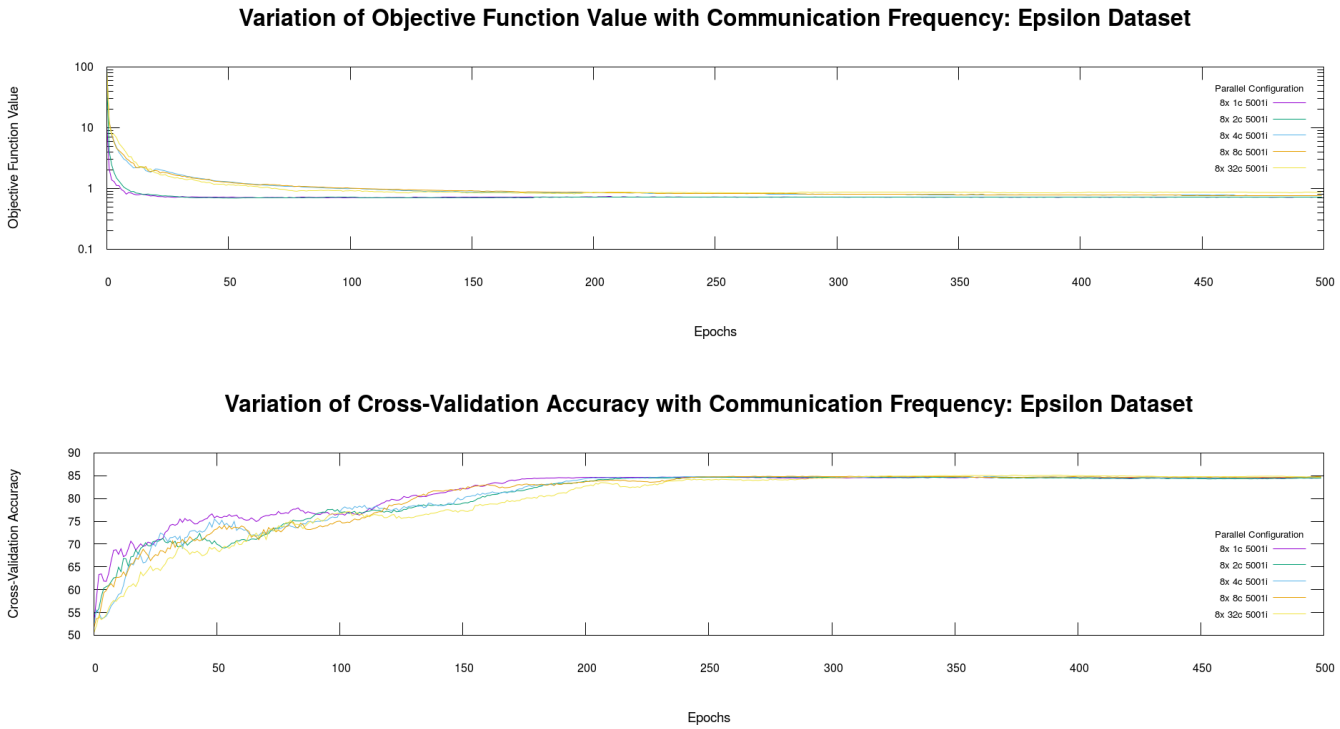

Fig. 86. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,32,], Parallelism = 8

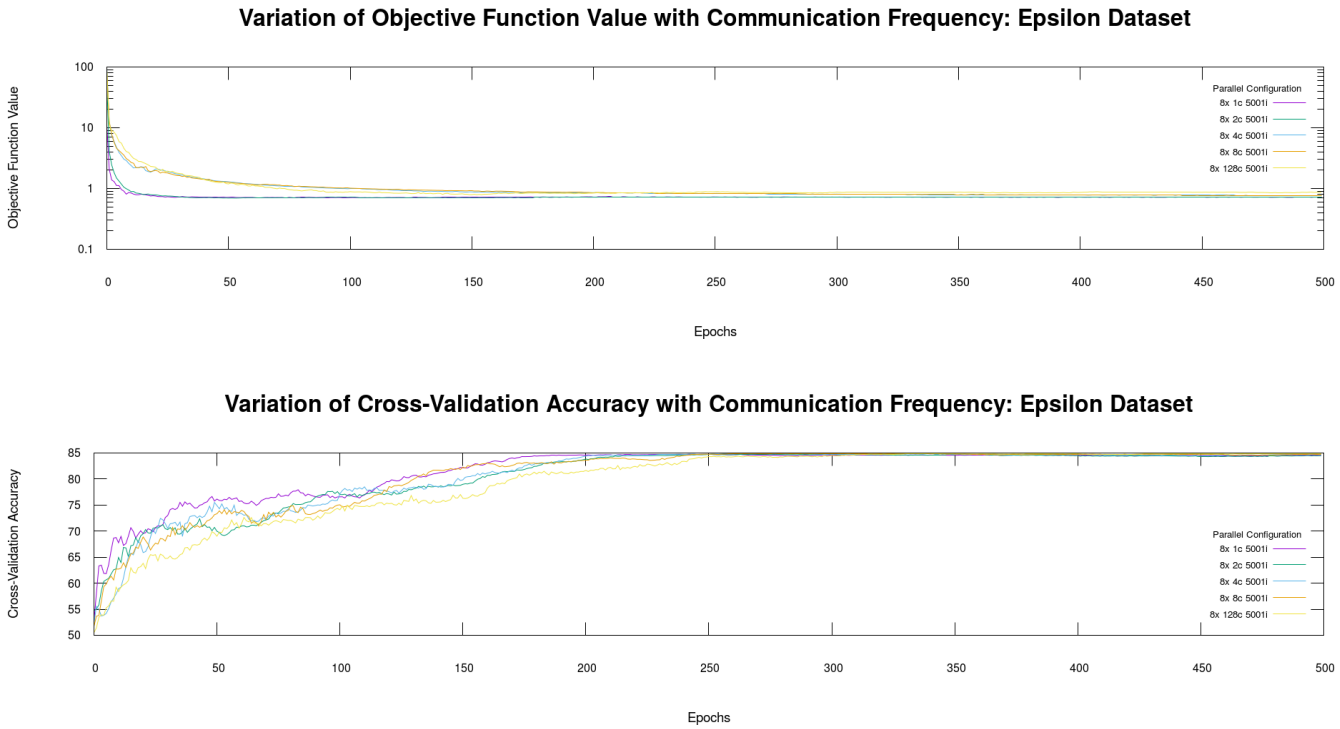

Fig. 87. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,128,], Parallelism = 8

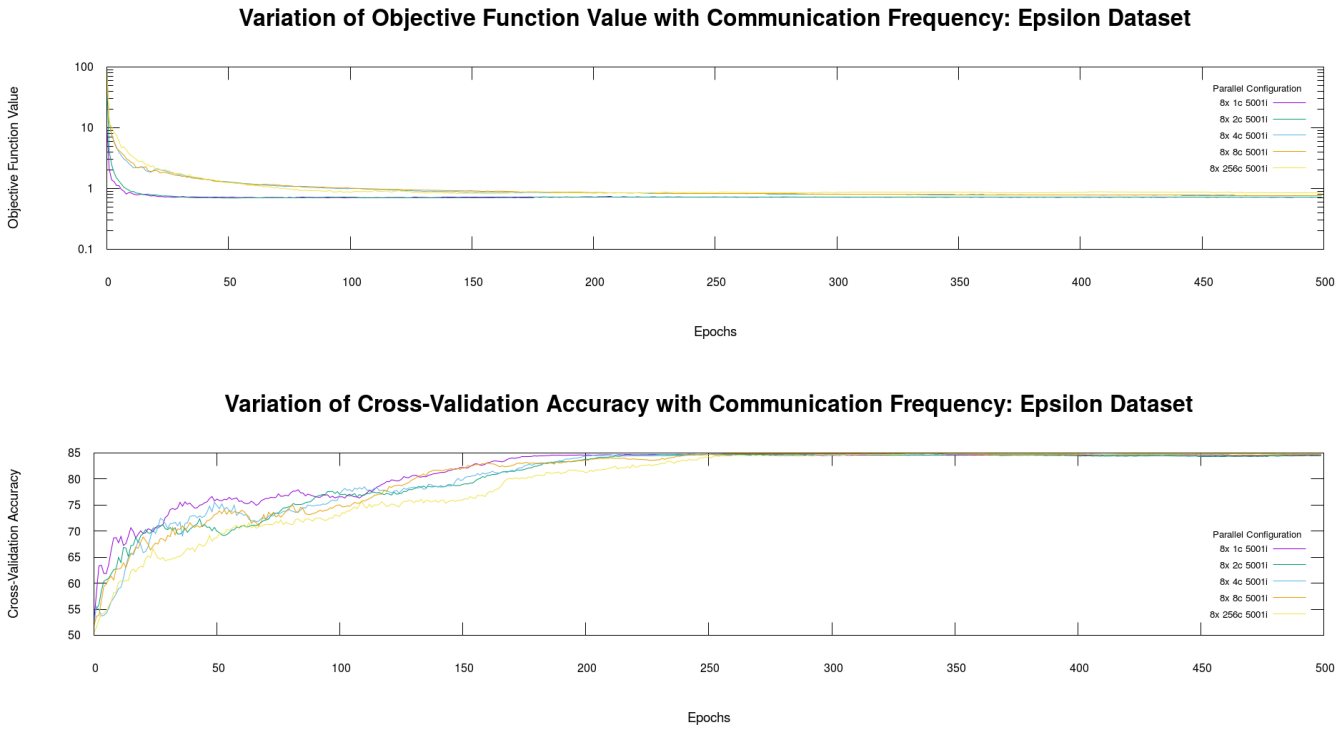

Fig. 88. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,256,], Parallelism = 8

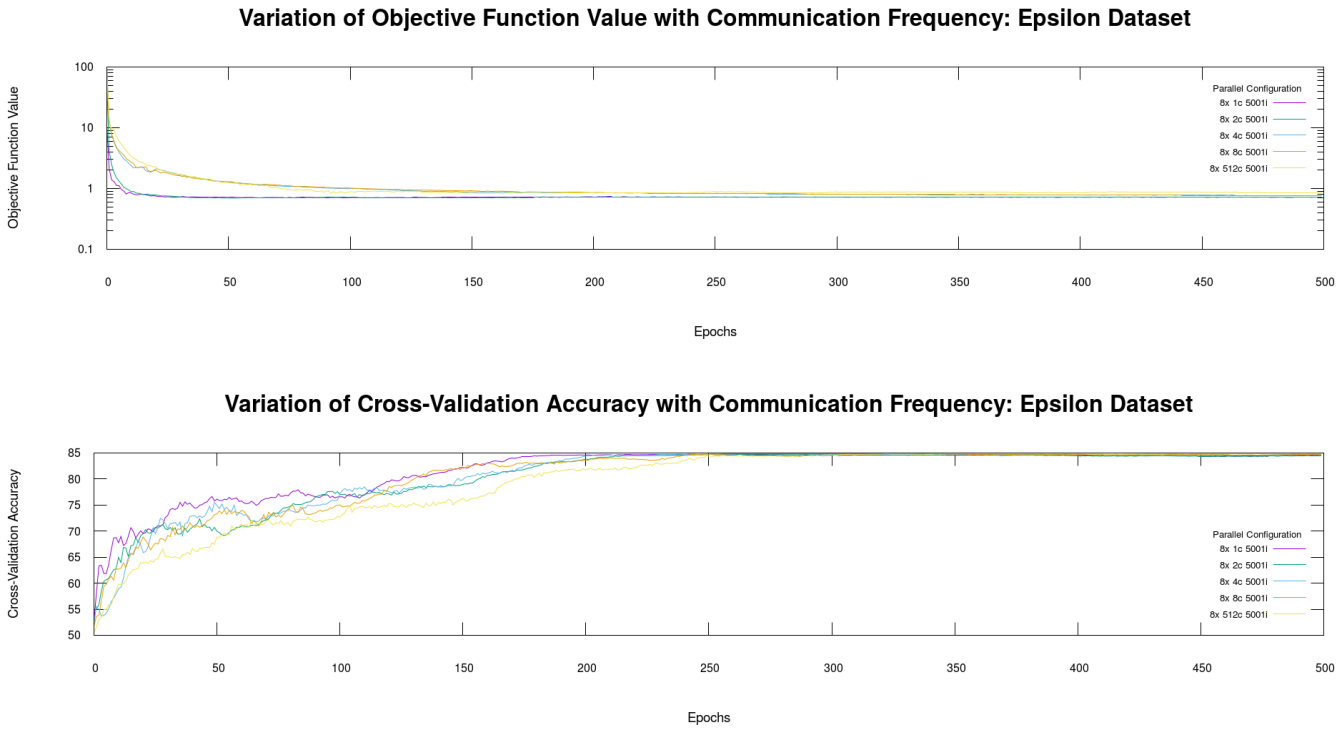

Fig. 89. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,512,], Parallelism = 8

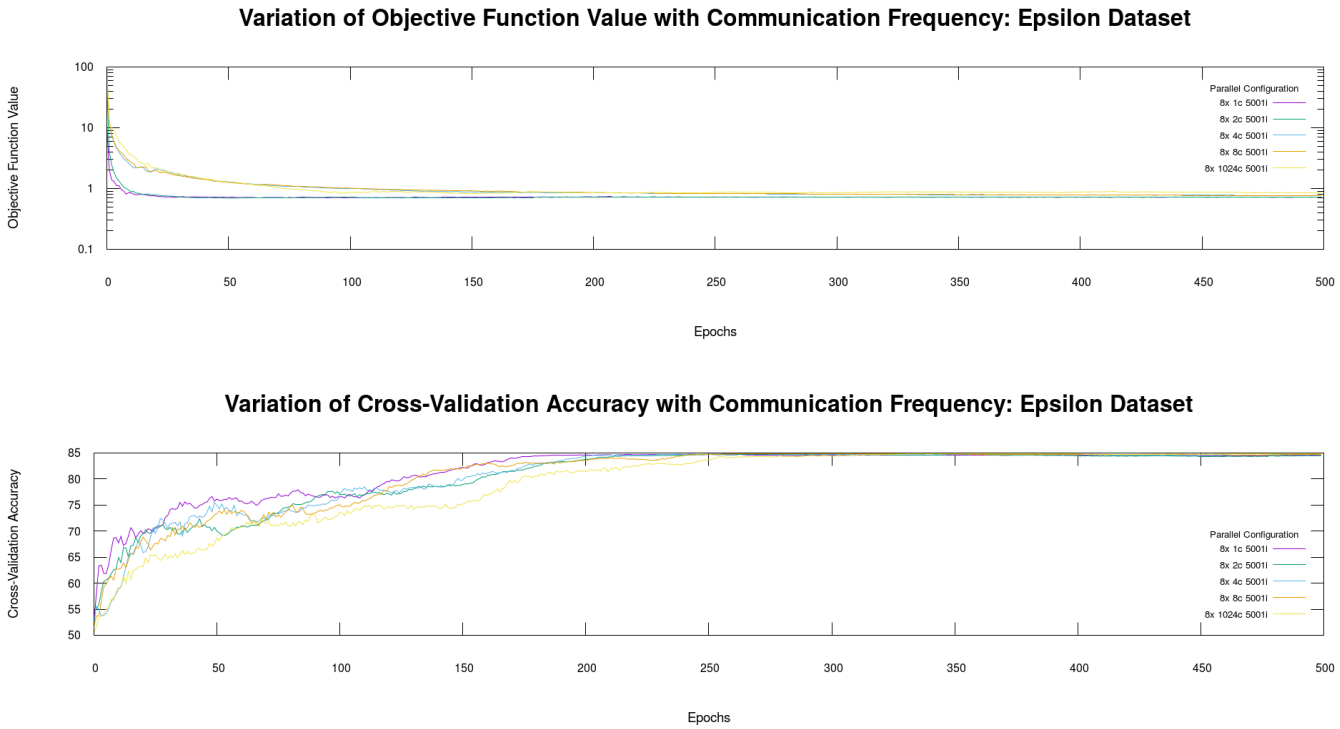

Fig. 90. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 8

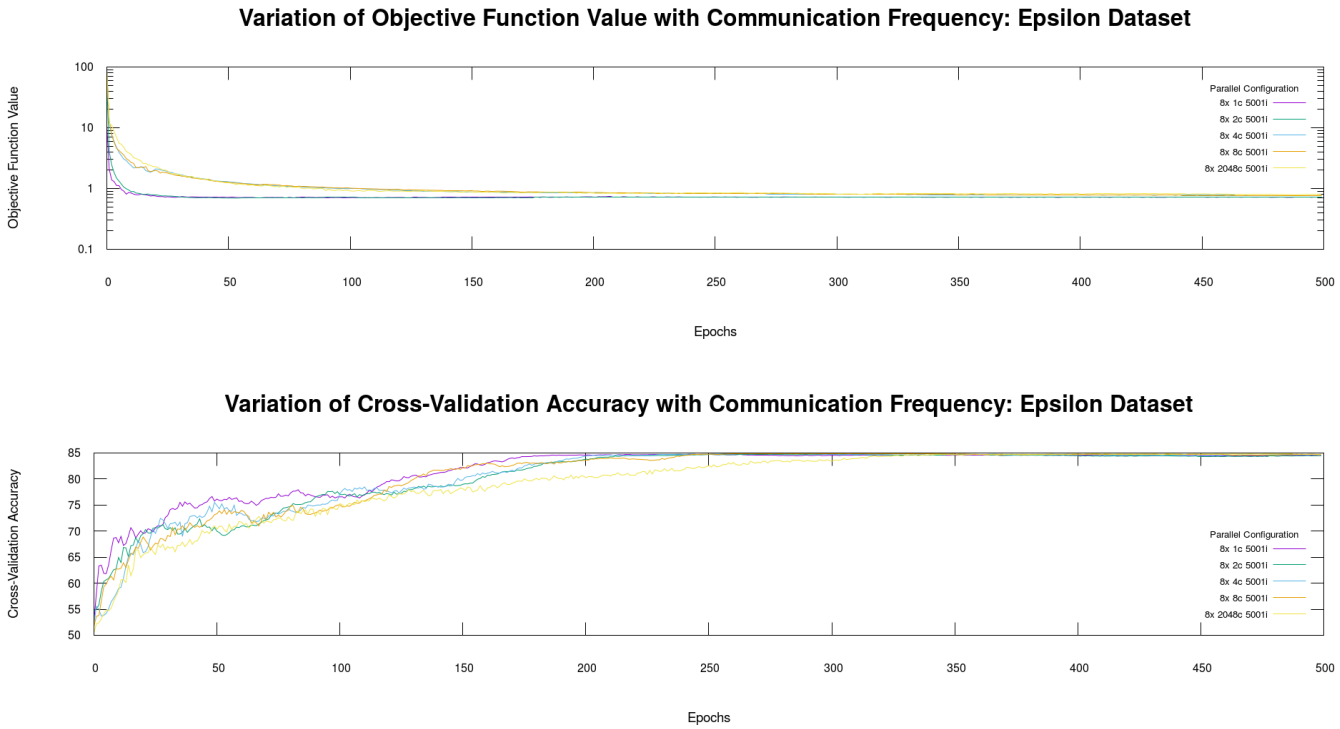

Fig. 91. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,2048,], Parallelism = 8

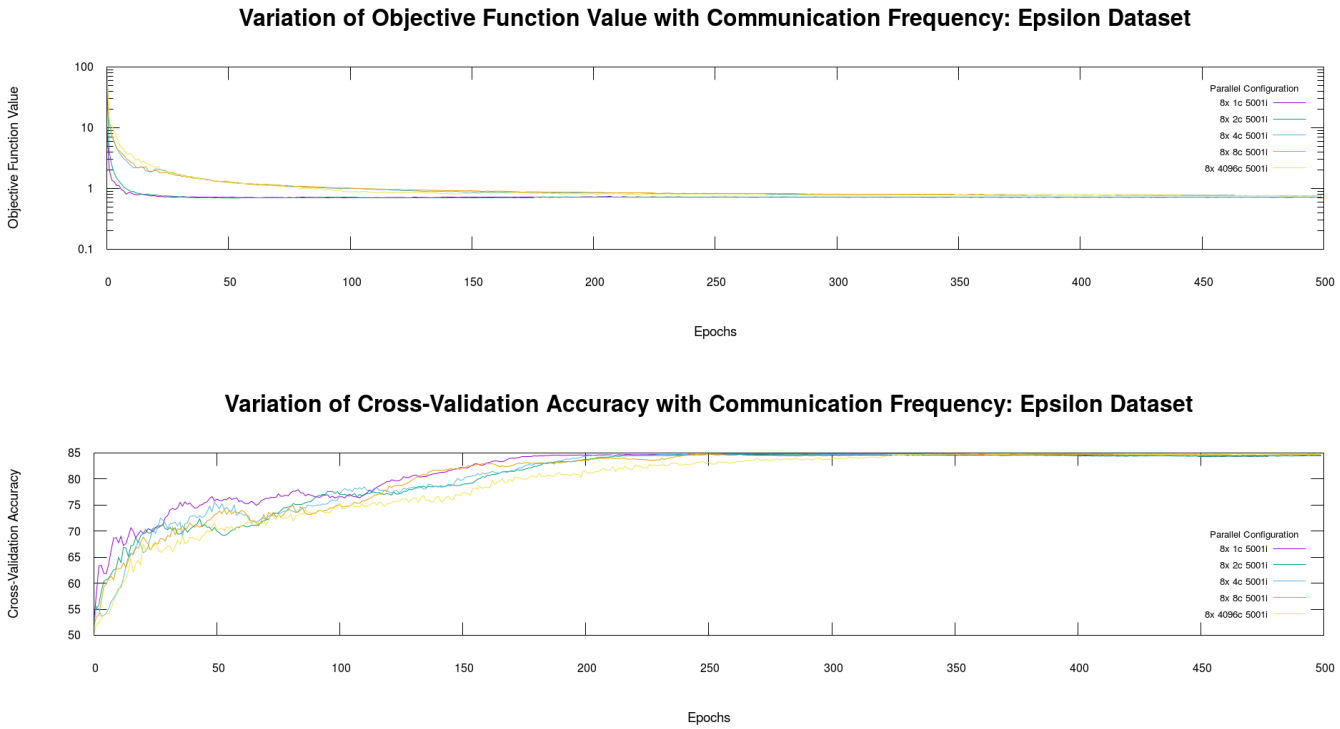

Fig. 92. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,4096,], Parallelism = 8

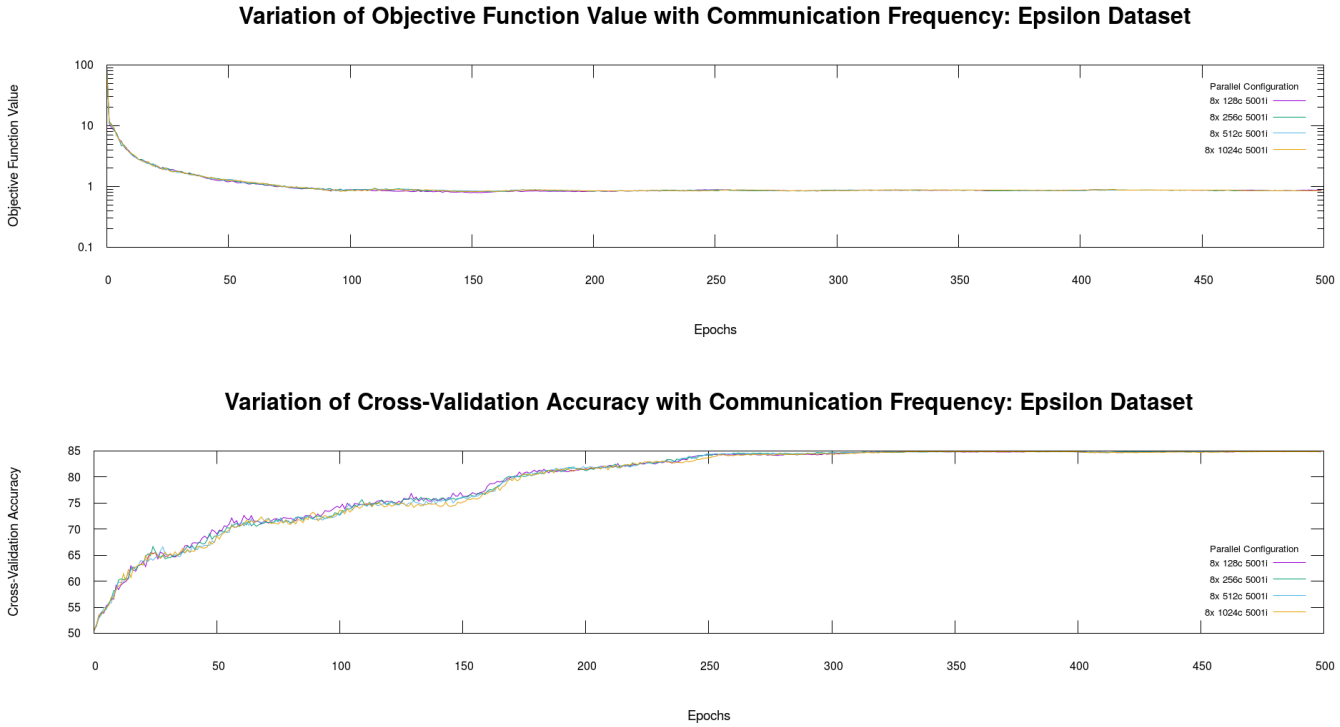

Fig. 93. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [128,256,512,1024,], Parallelism = 8

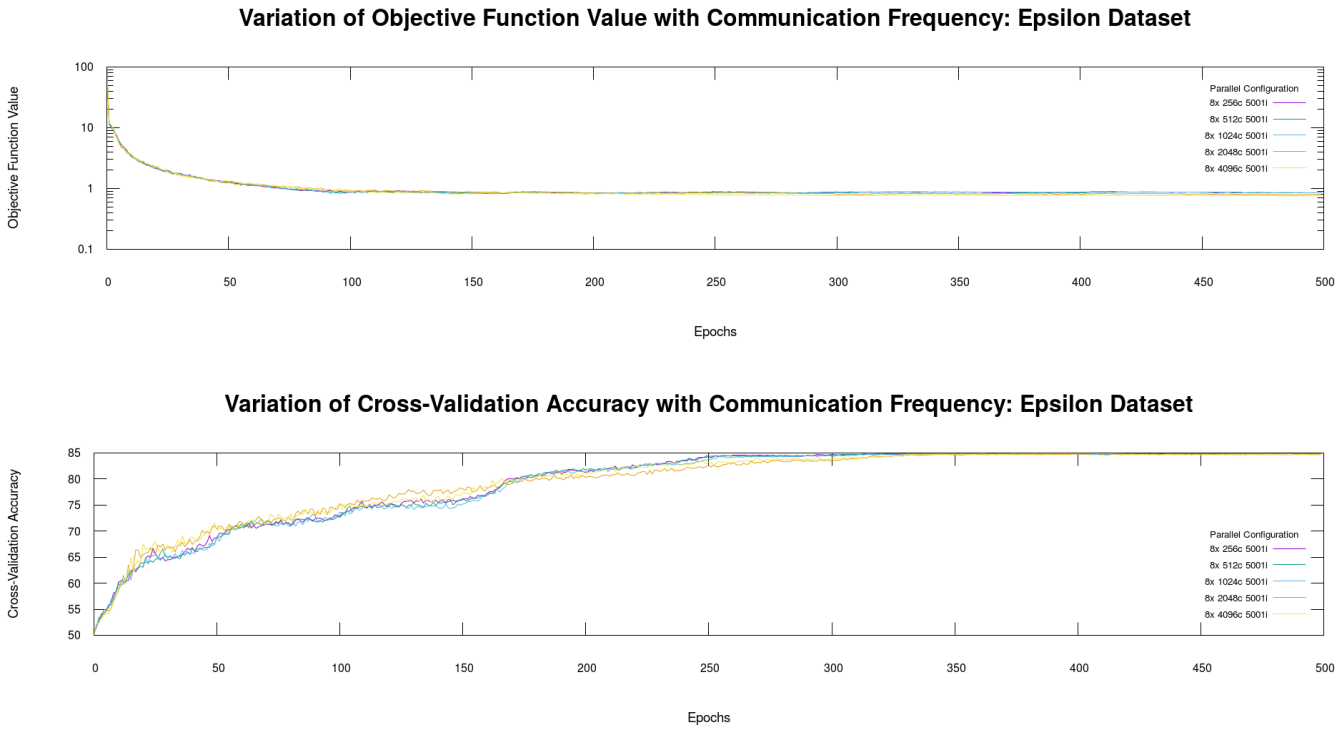

Fig. 94. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [256,512,1024,2048,4096,], Parallelism = 8

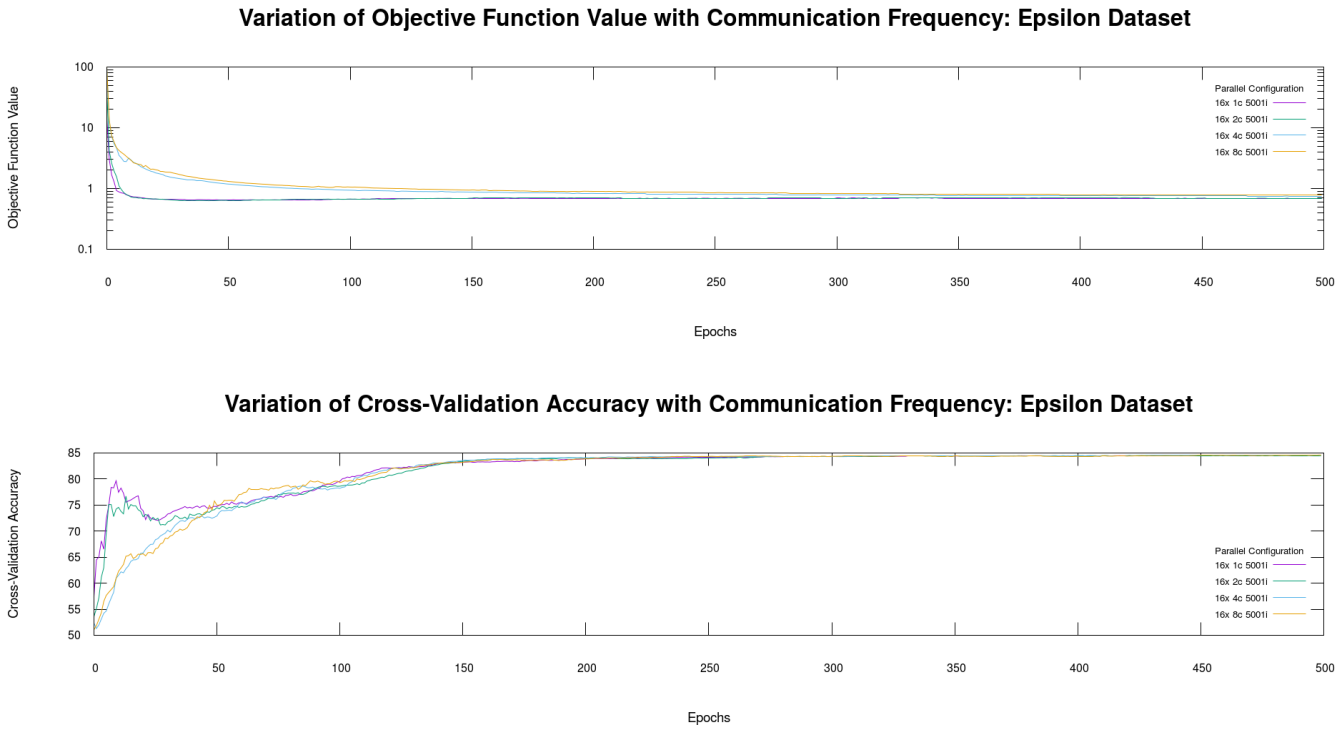

Fig. 95. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,], Parallelism = 16

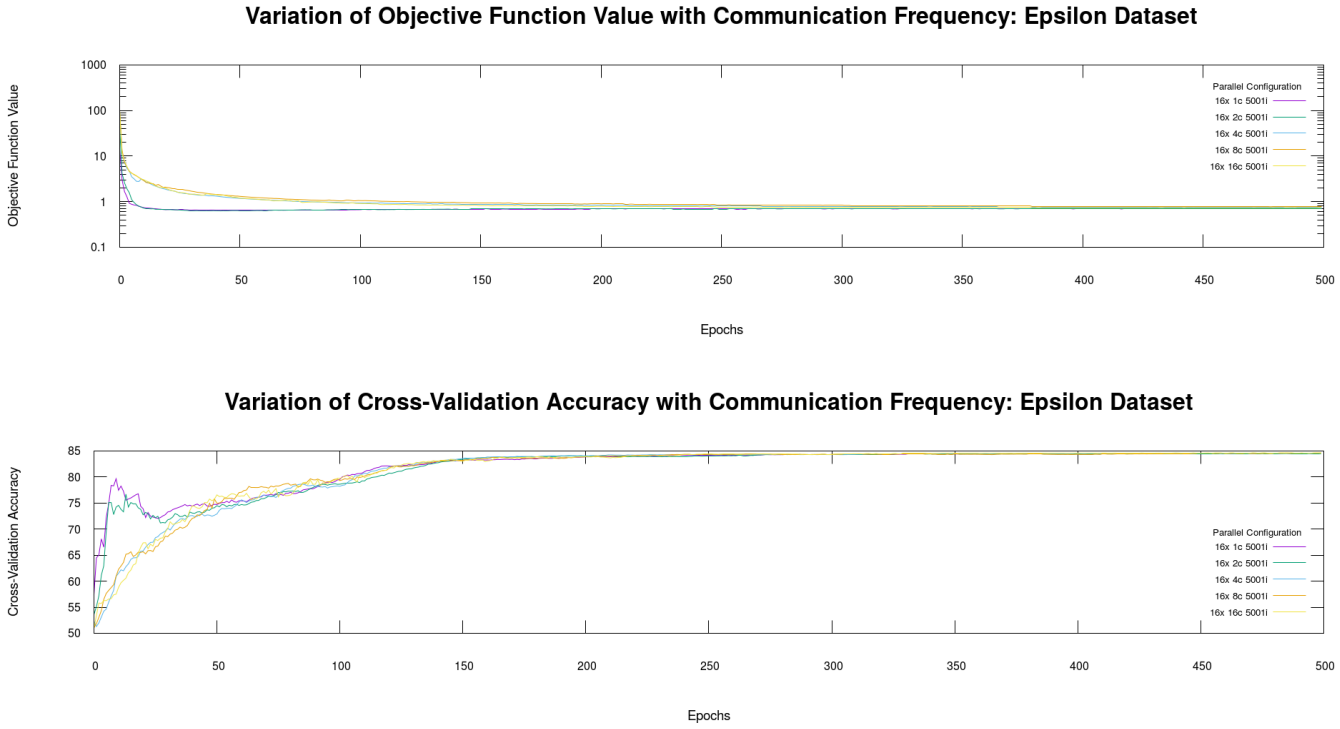

Fig. 96. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,16,], Parallelism = 16

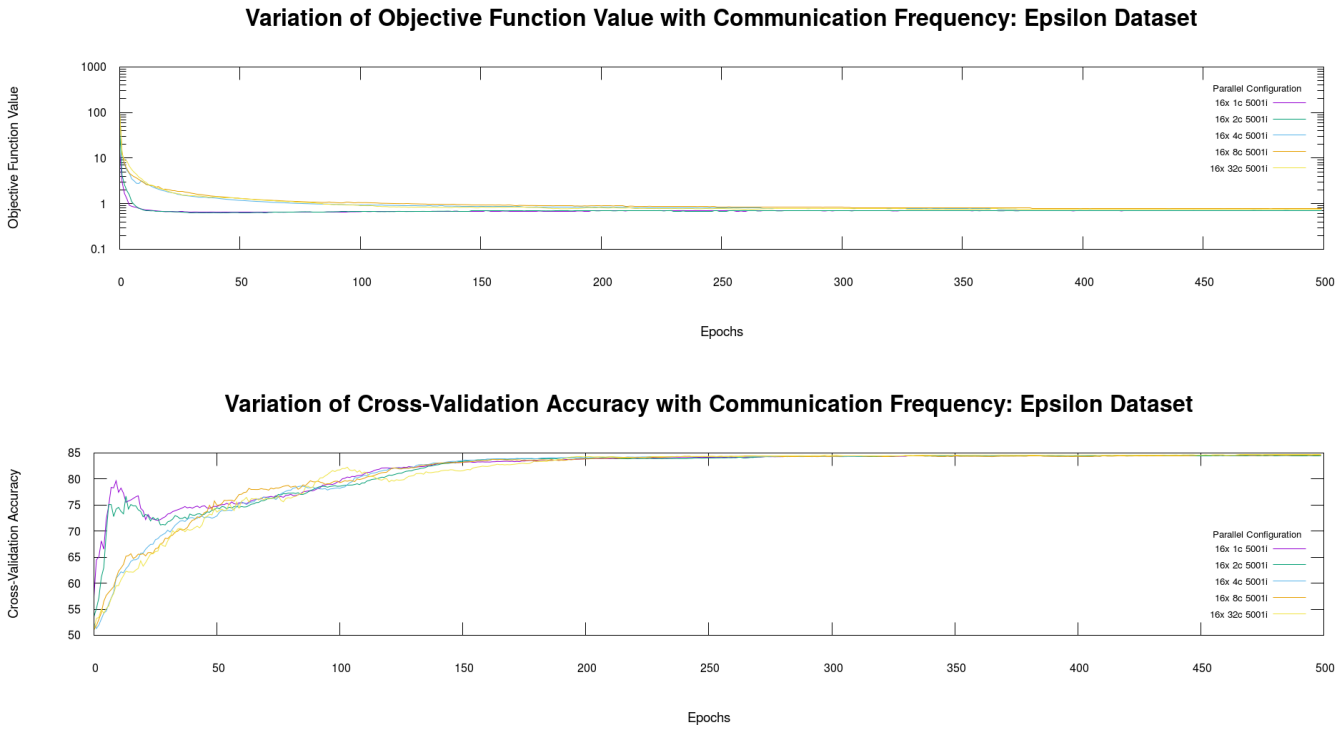

Fig. 97. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,32,], Parallelism = 16

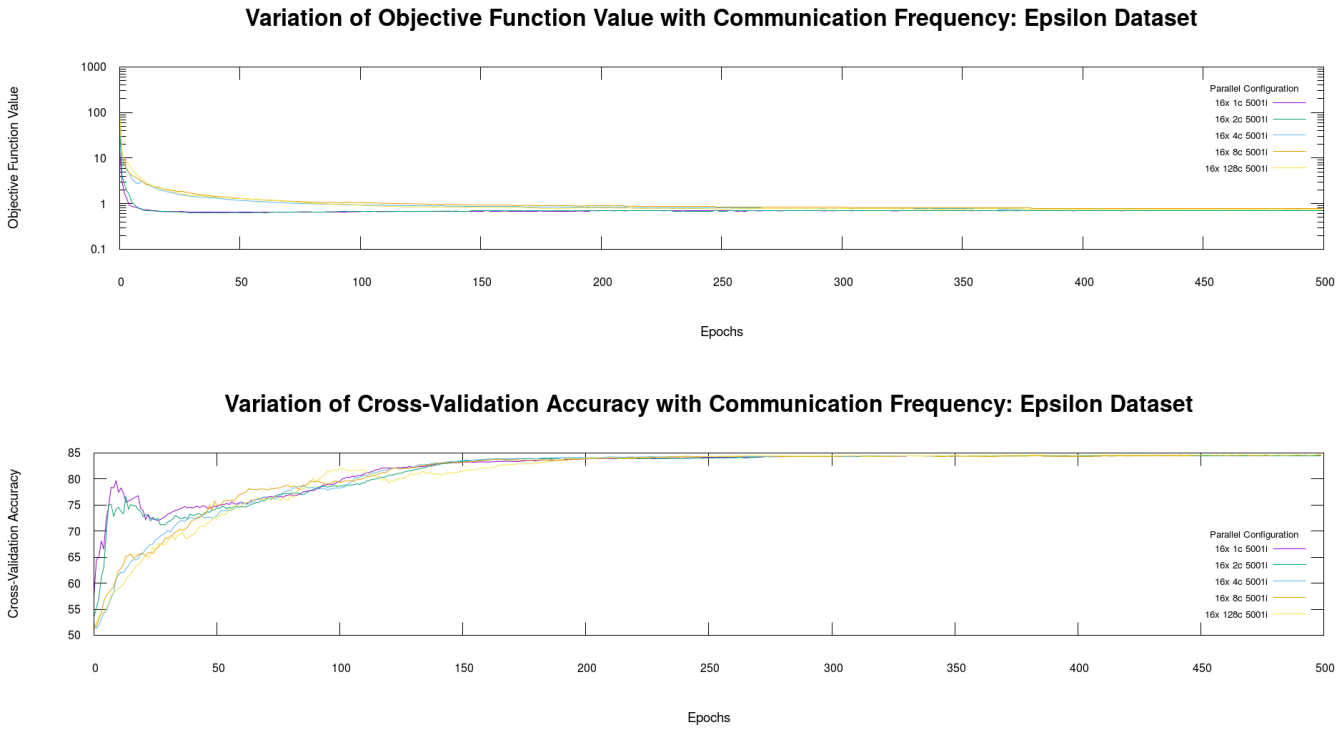

Fig. 98. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,128,], Parallelism = 16

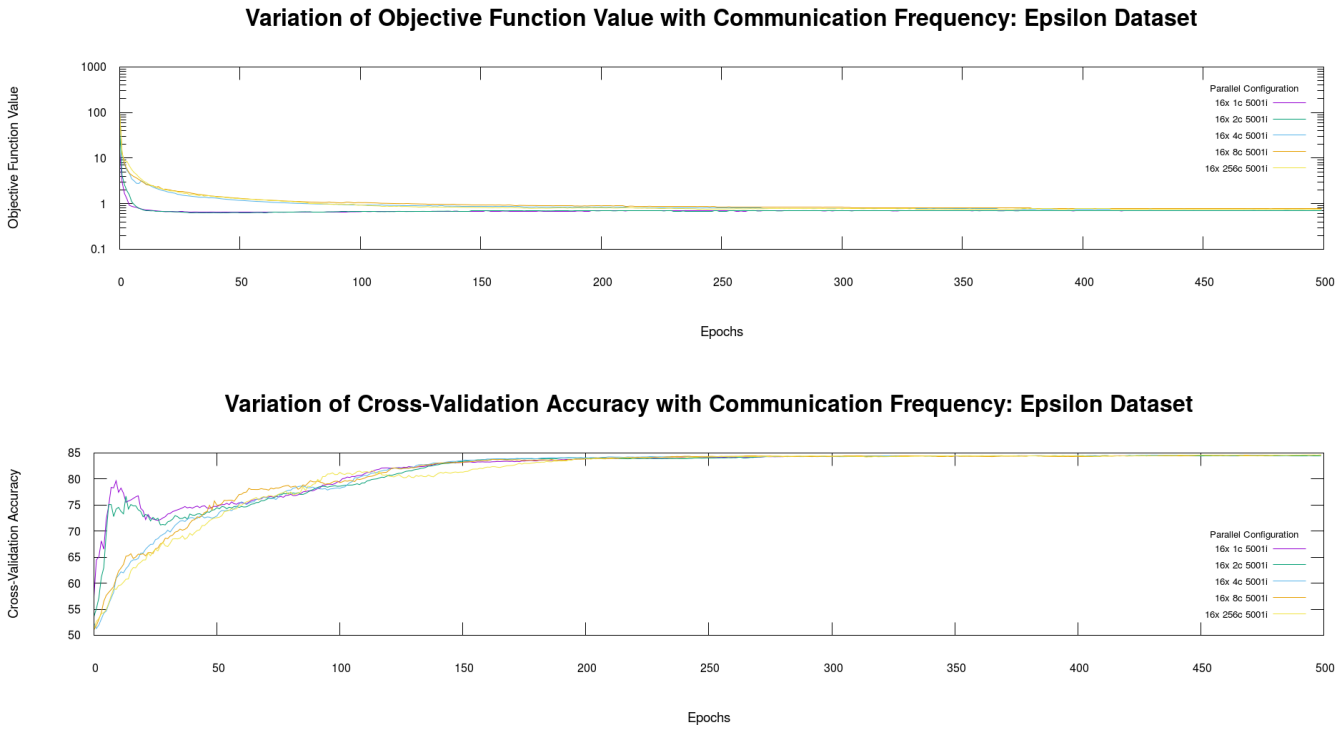

Fig. 99. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,256,], Parallelism = 16

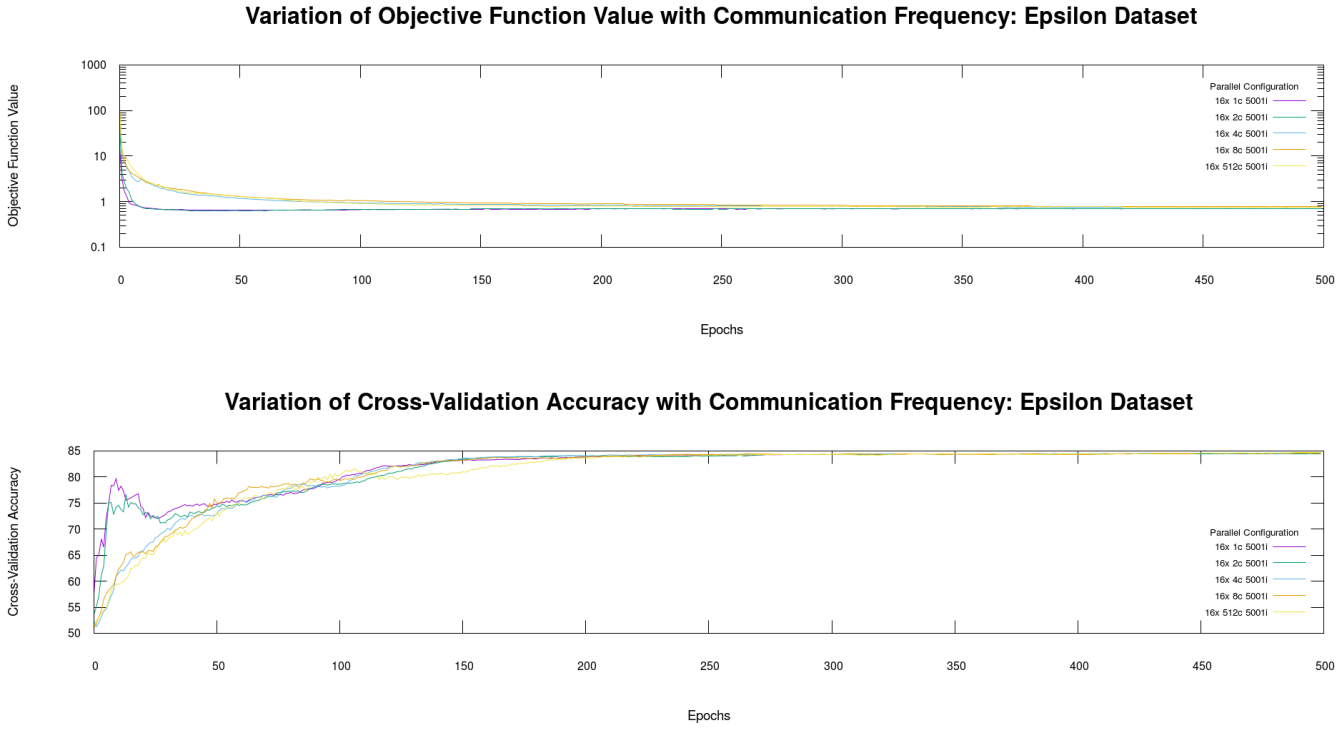

Fig. 100. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,512,], Parallelism = 16

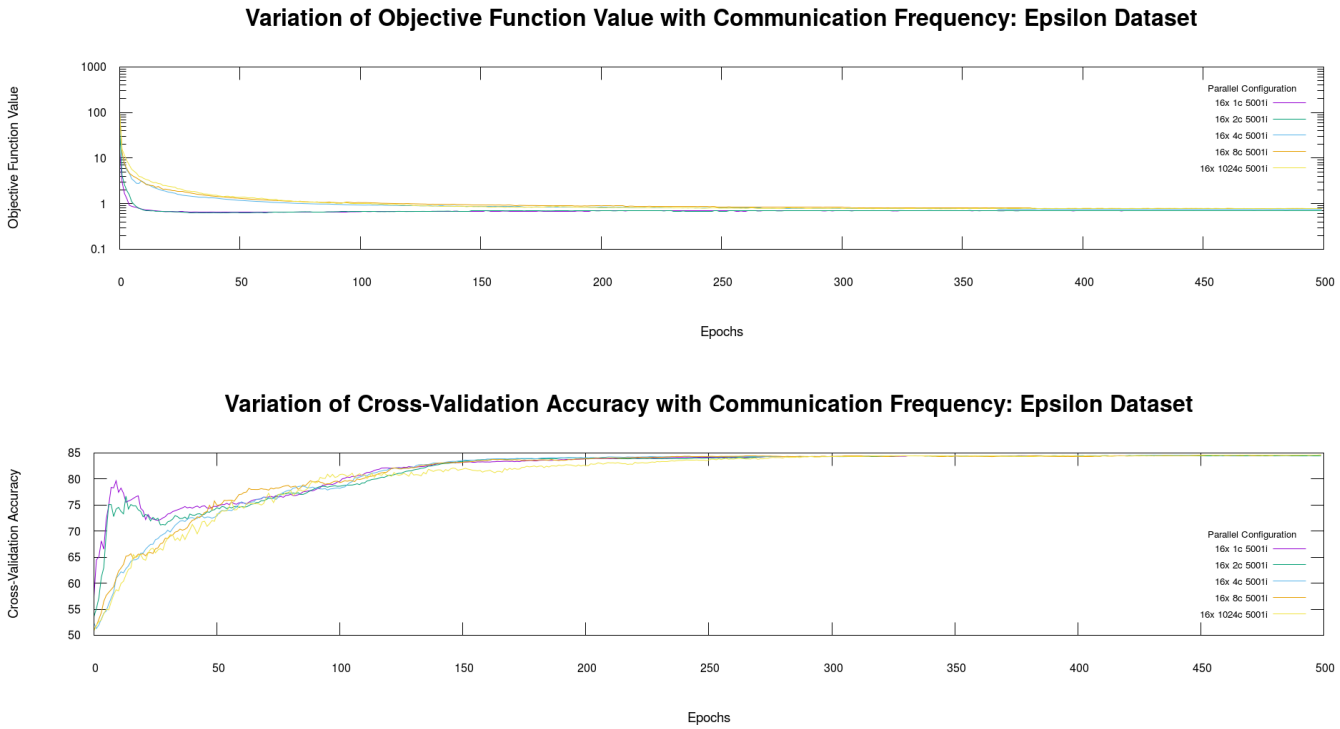

Fig. 101. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 16

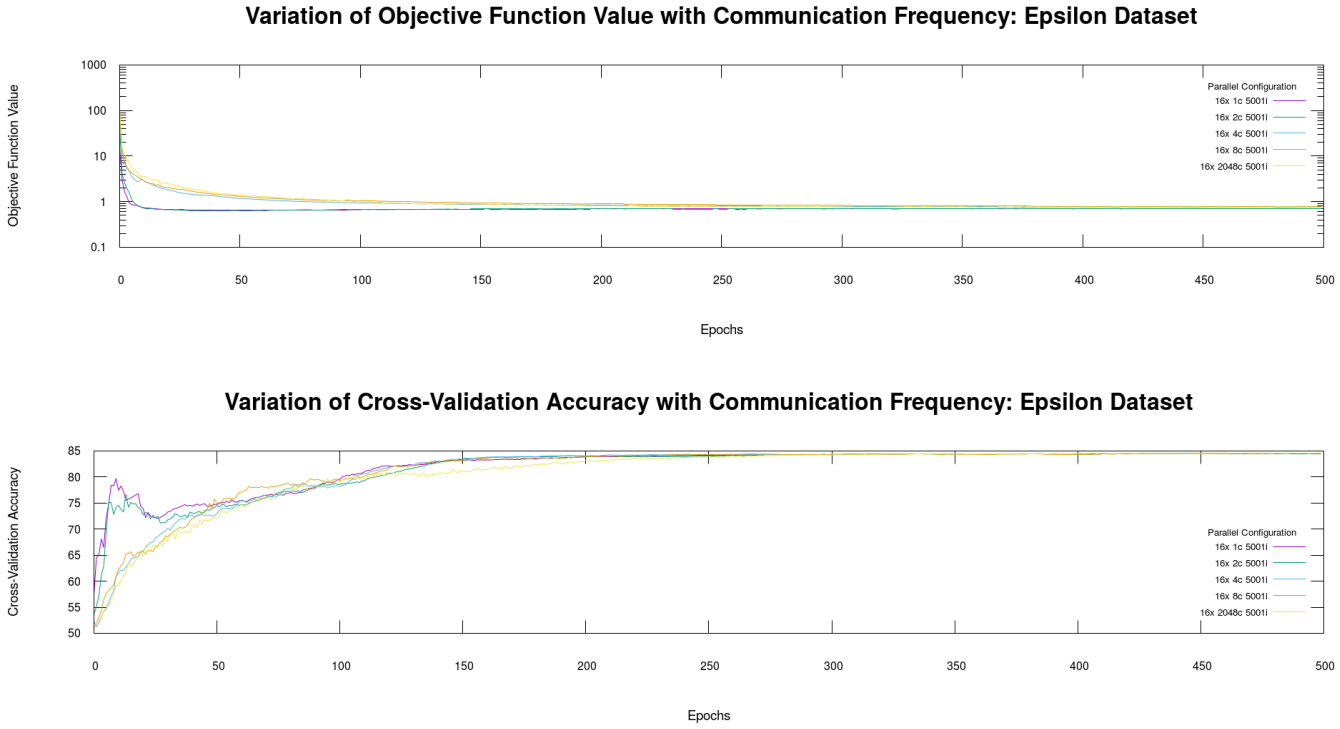

Fig. 102. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,2048,], Parallelism = 16

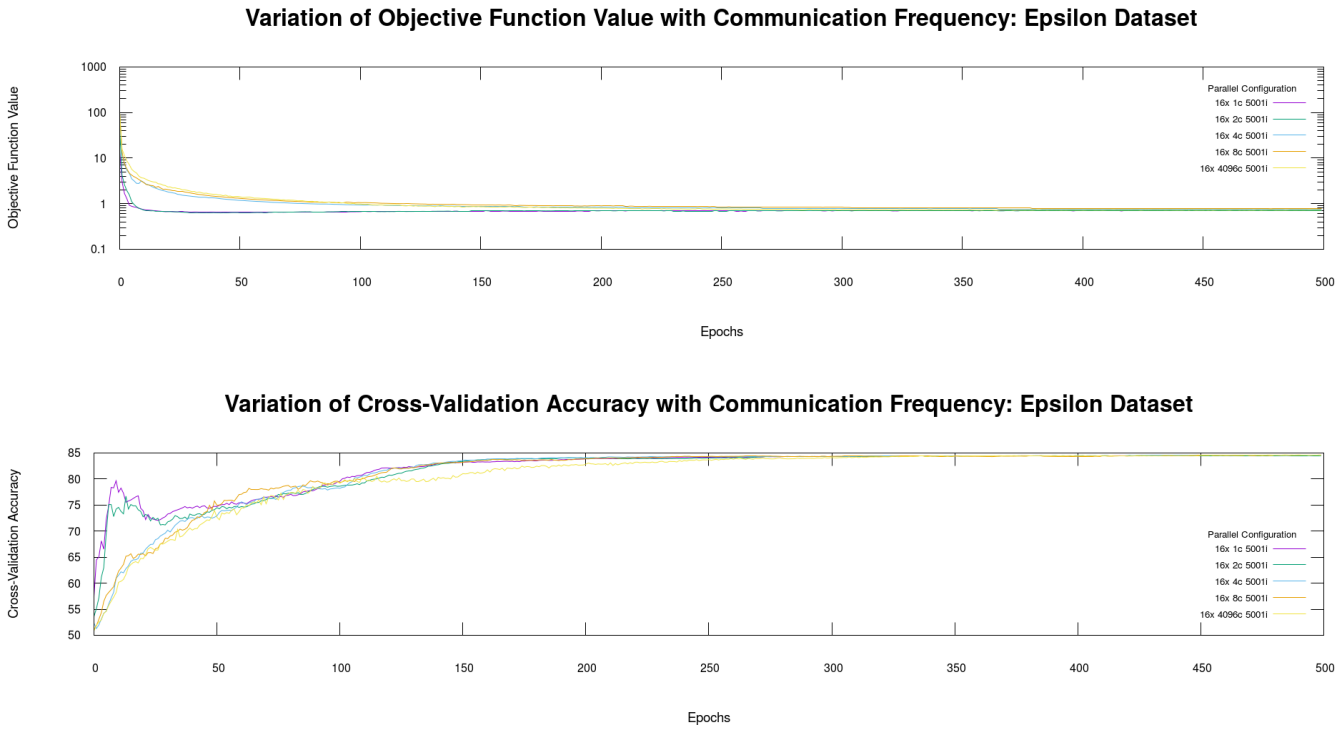

Fig. 103. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,4096,], Parallelism = 16

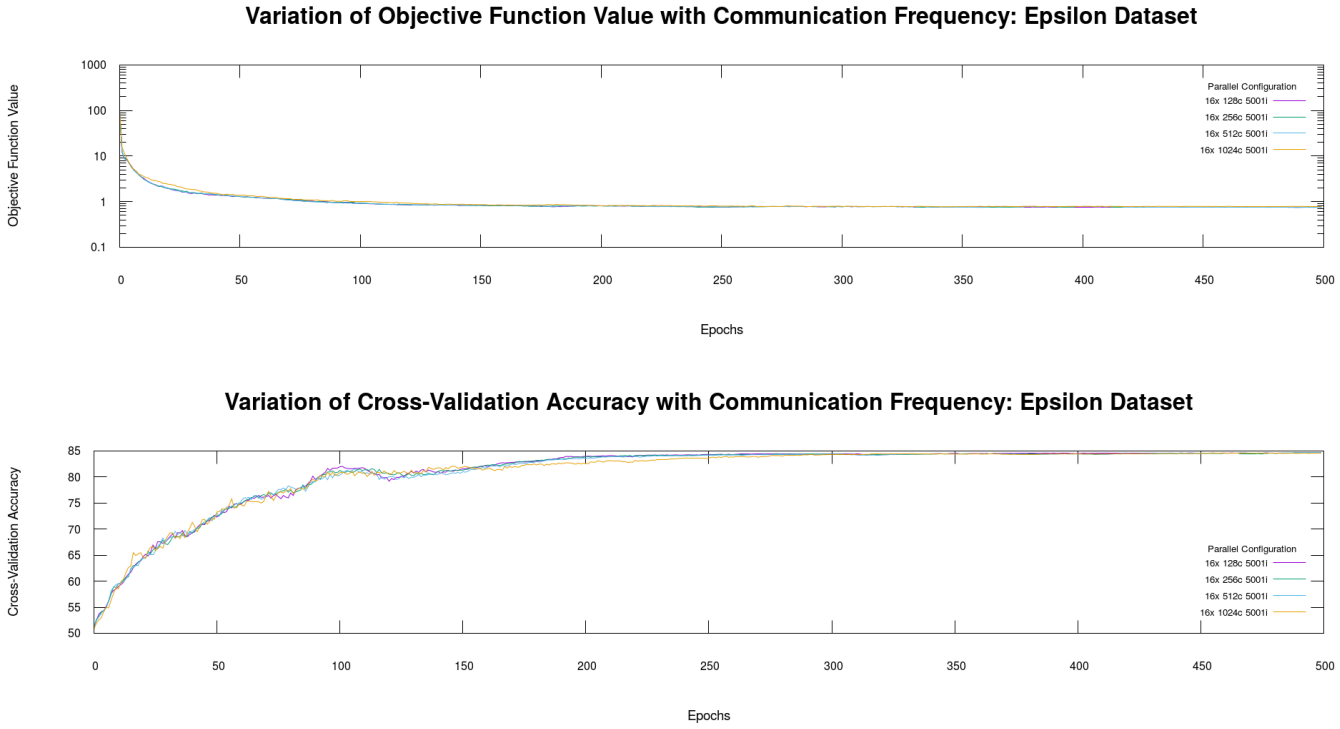

Fig. 104. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [128,256,512,1024,], Parallelism = 16

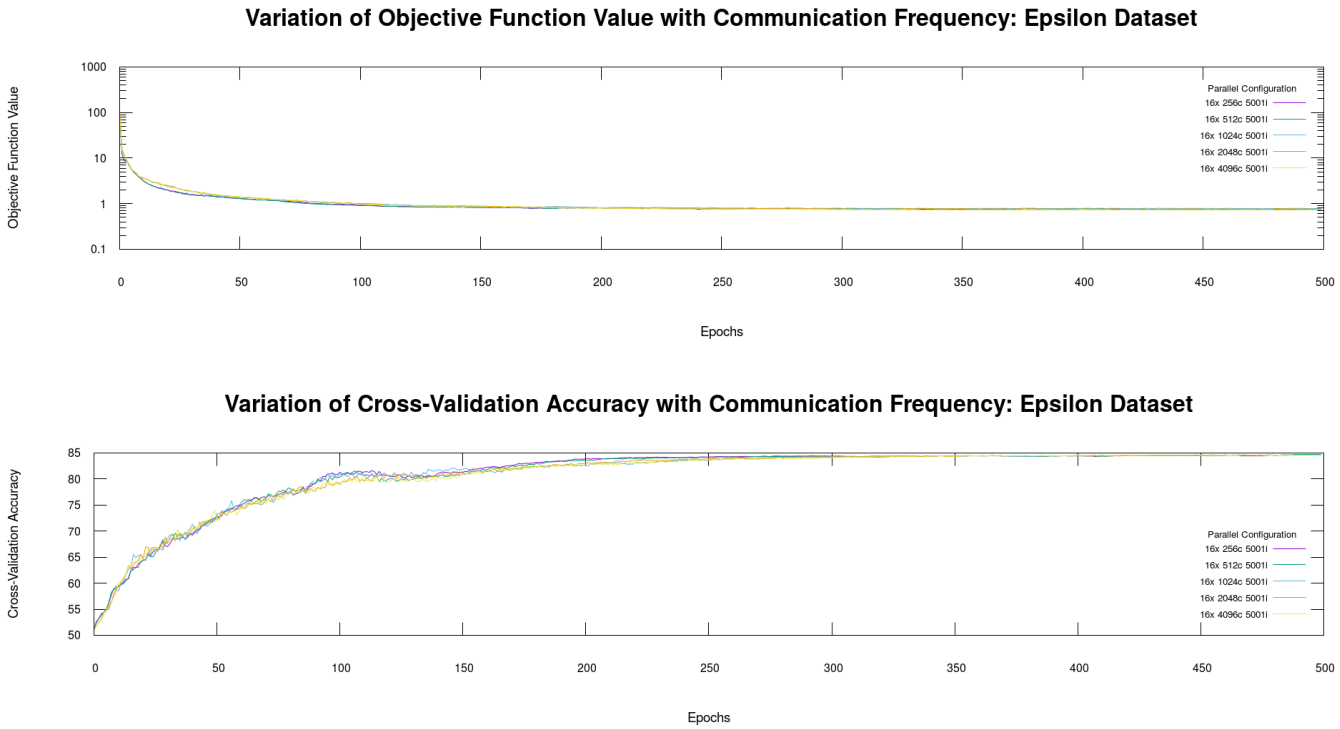

Fig. 105. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [256,512,1024,2048,4096,], Parallelism = 16

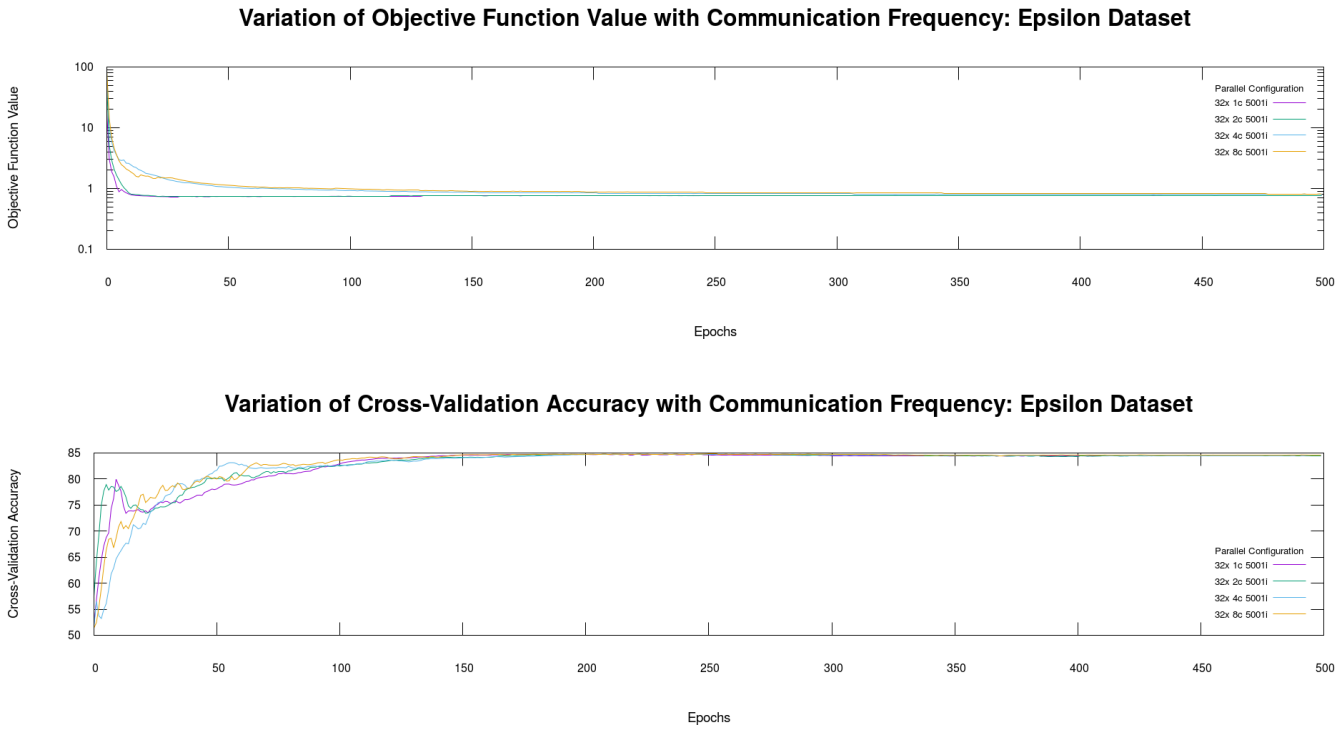

Fig. 106. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,], Parallelism = 32

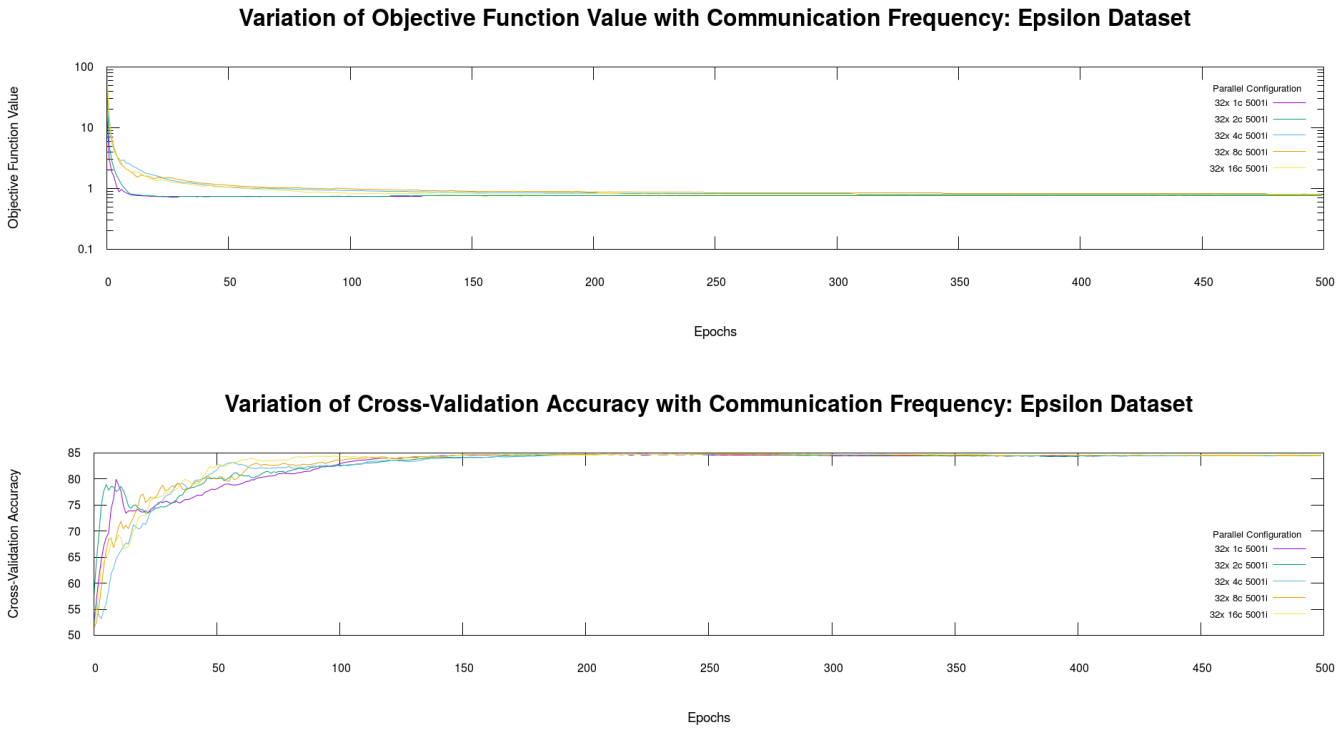

Fig. 107. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,16,], Parallelism = 32

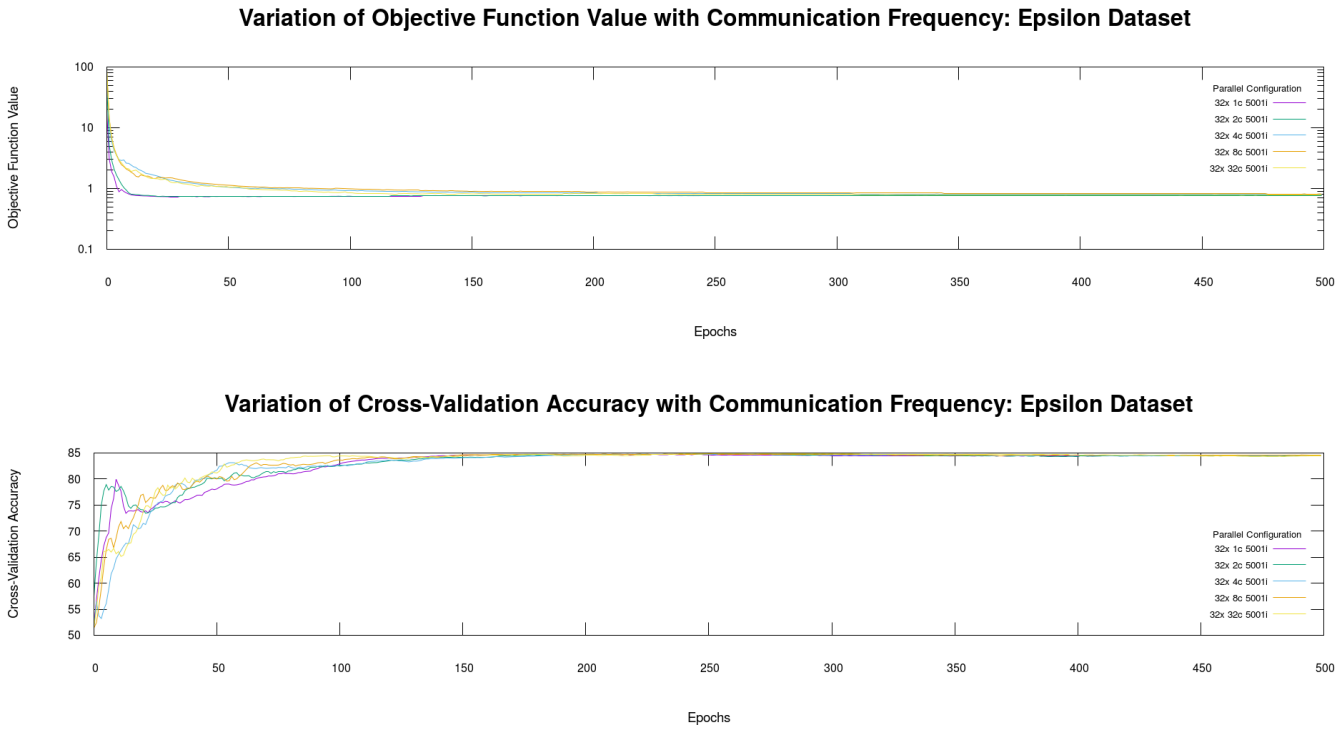

Fig. 108. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,32,], Parallelism = 32

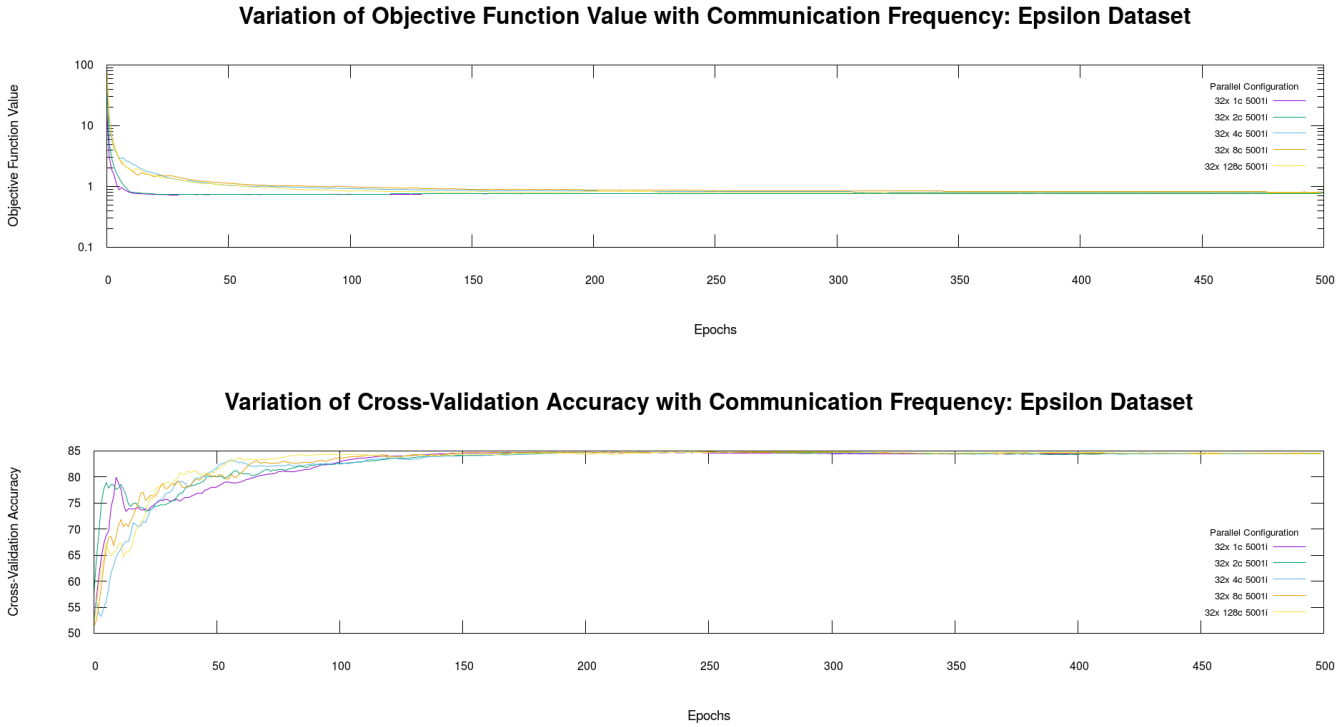

Fig. 109. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,128,], Parallelism = 32

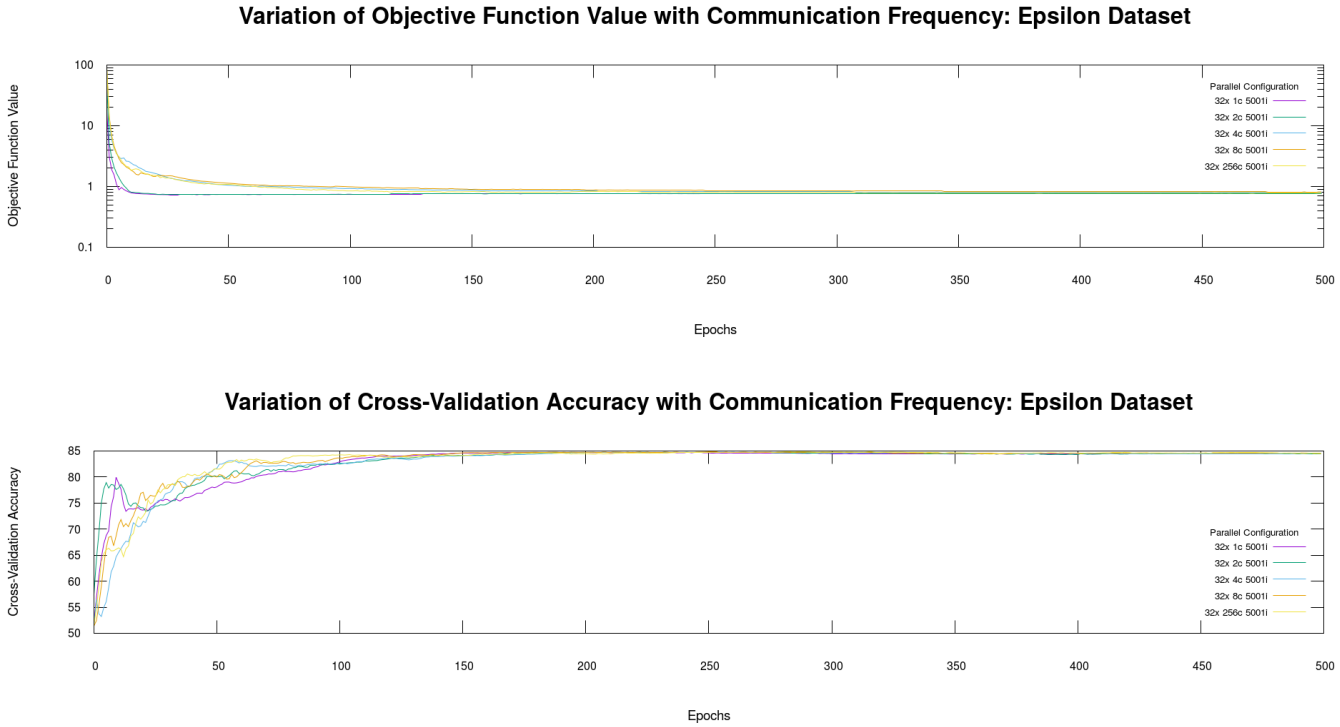

Fig. 110. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,256,], Parallelism = 32

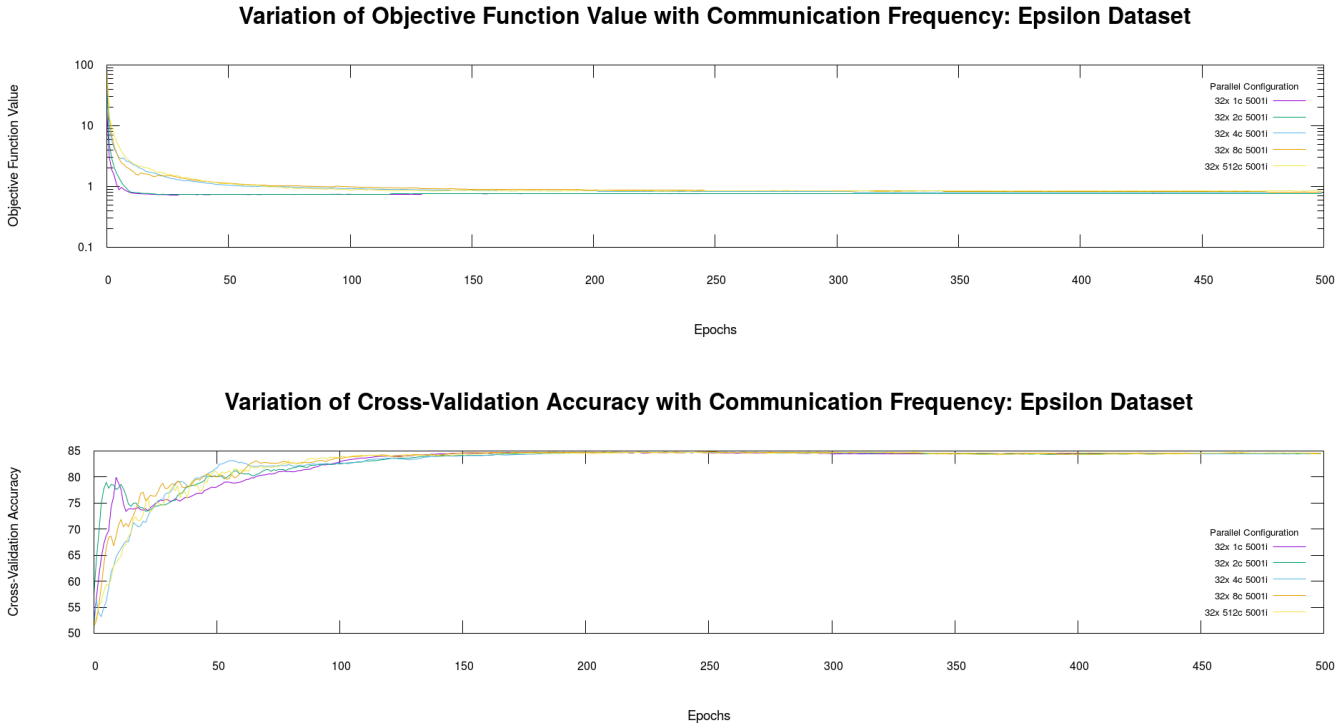

Fig. 111. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,512,], Parallelism = 32

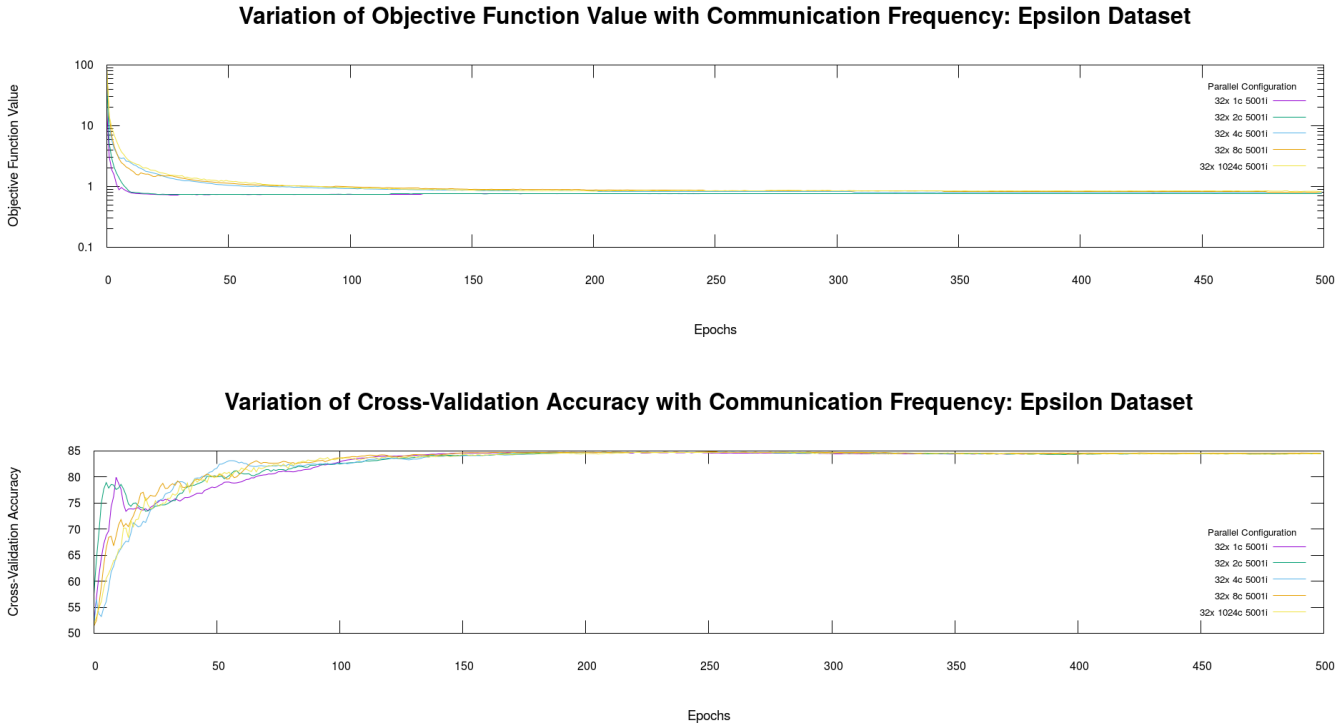

Fig. 112. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 32

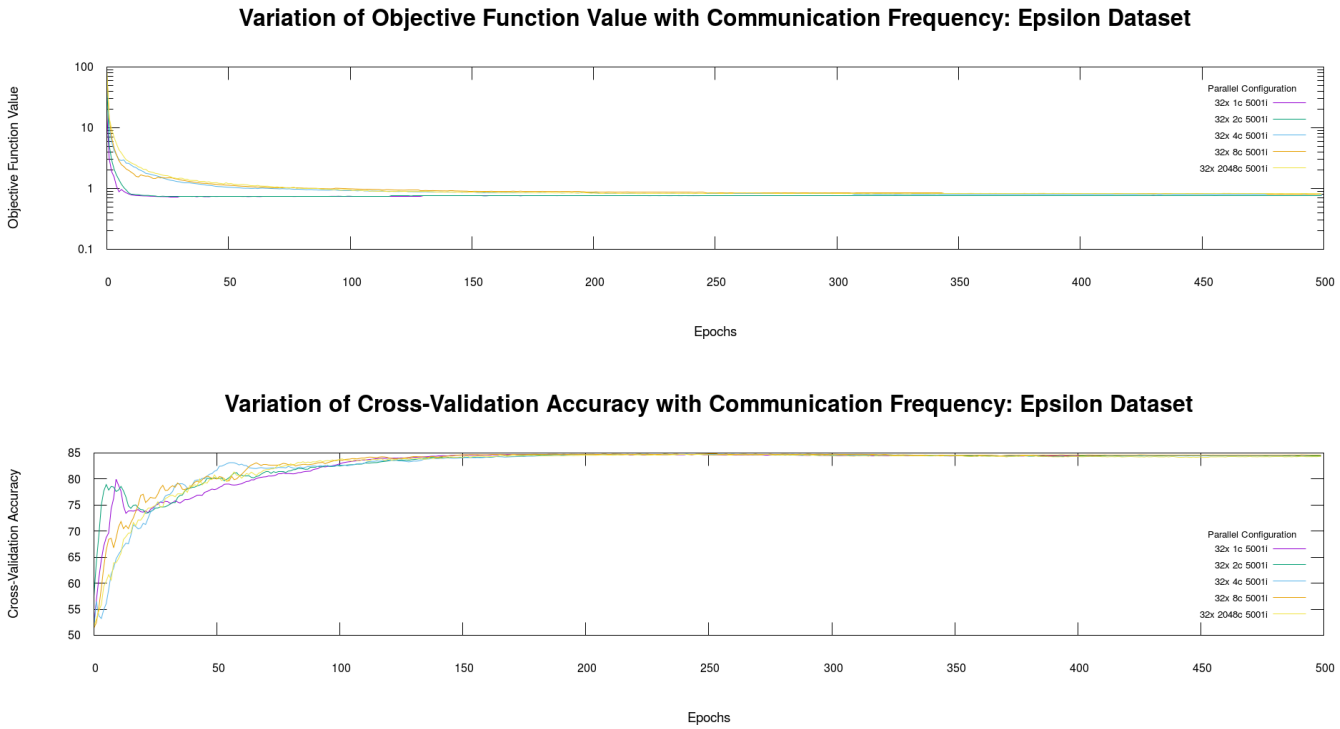

Fig. 113. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,2048,], Parallelism = 32

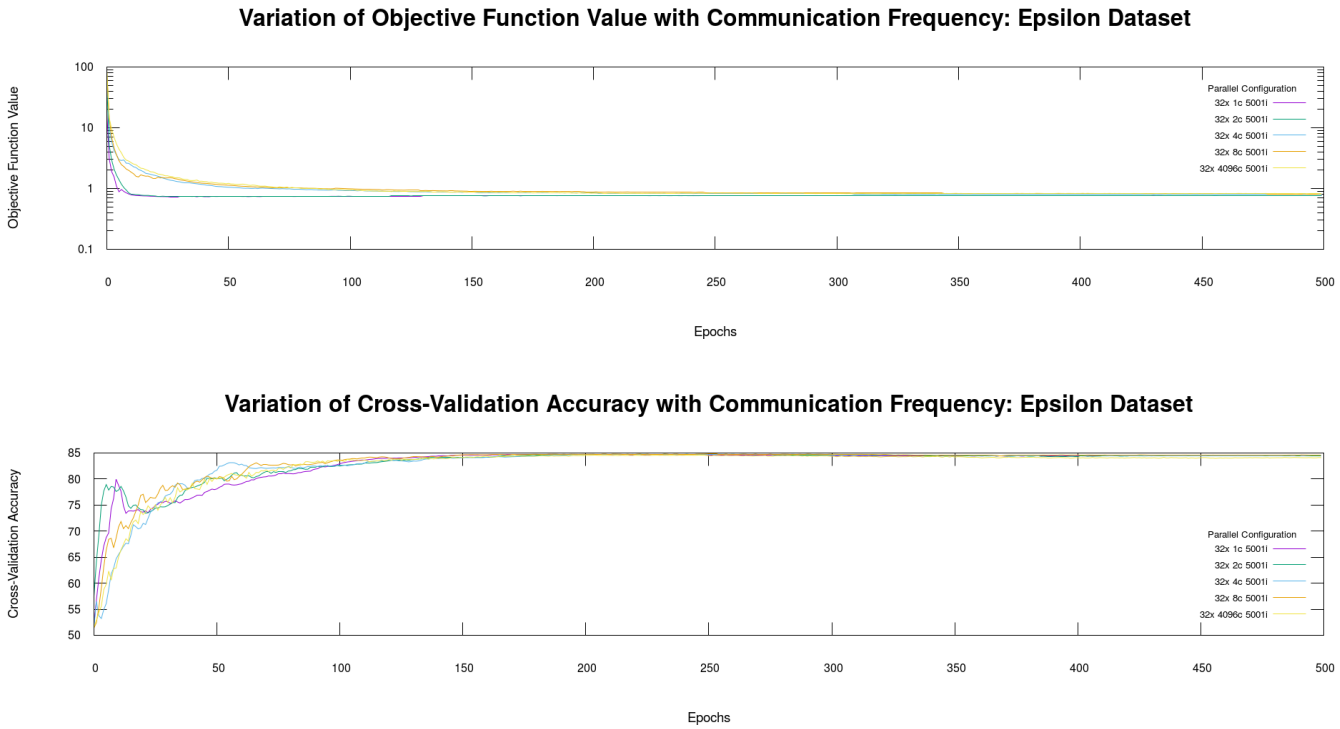

Fig. 114. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [1,2,4,8,4096,], Parallelism = 32

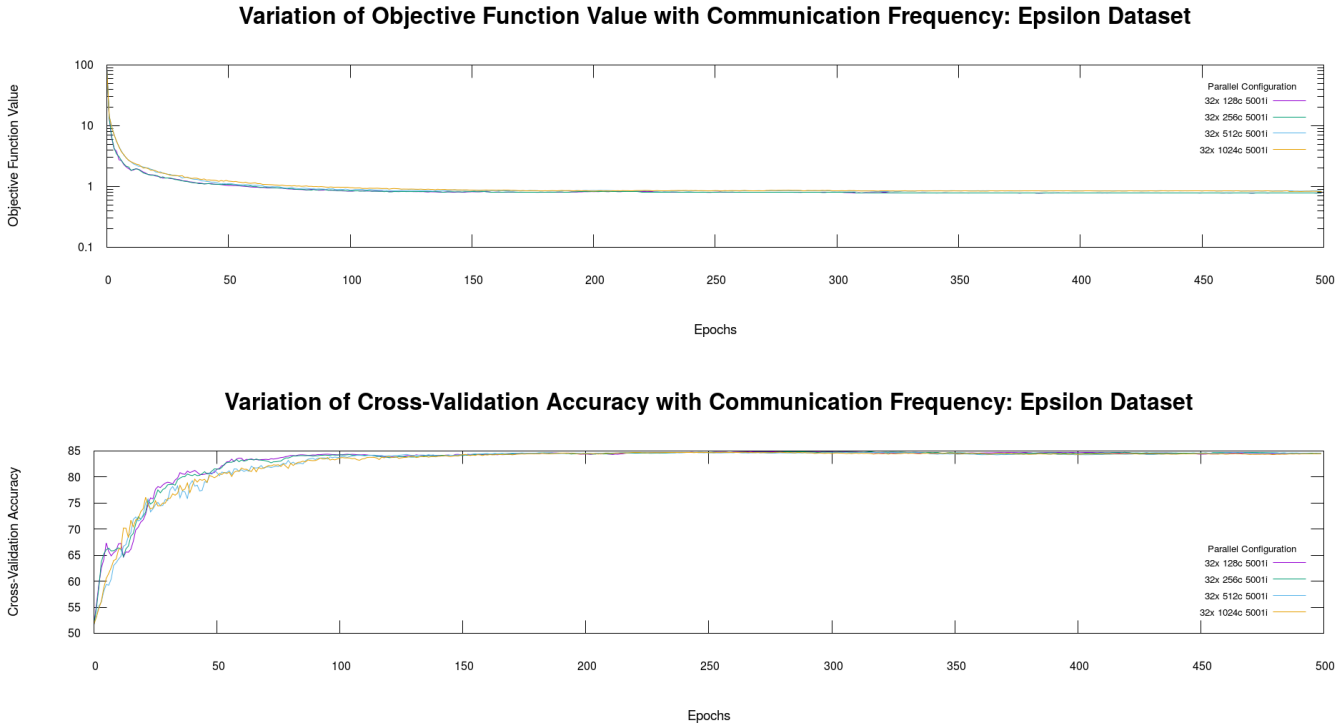

Fig. 115. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [128,256,512,1024,], Parallelism = 32

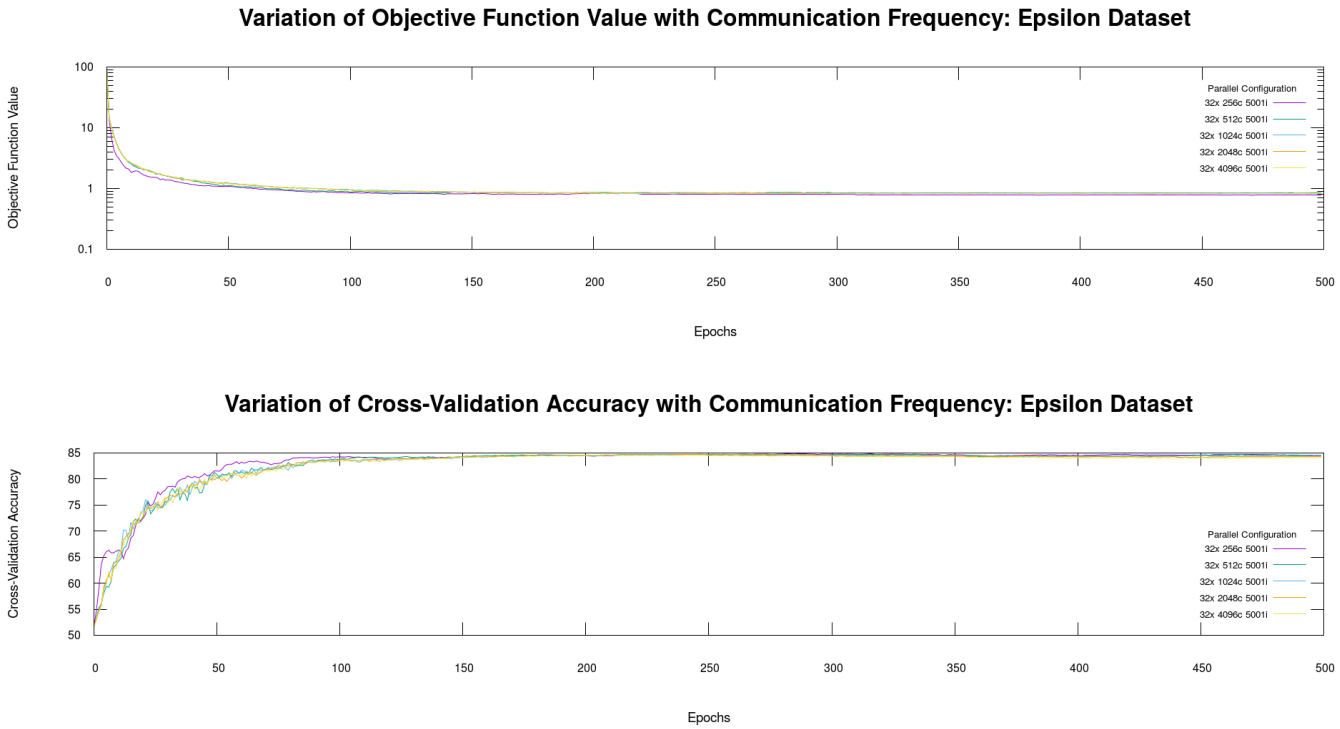

Fig. 116. Distributed Training Time : Dataset Epsilon , Configuration : MSF = [256,512,1024,2048,4096,], Parallelism = 32

C. Webspam All Cross Validation Accuracy and Objective Function Value Variation agains MSFs

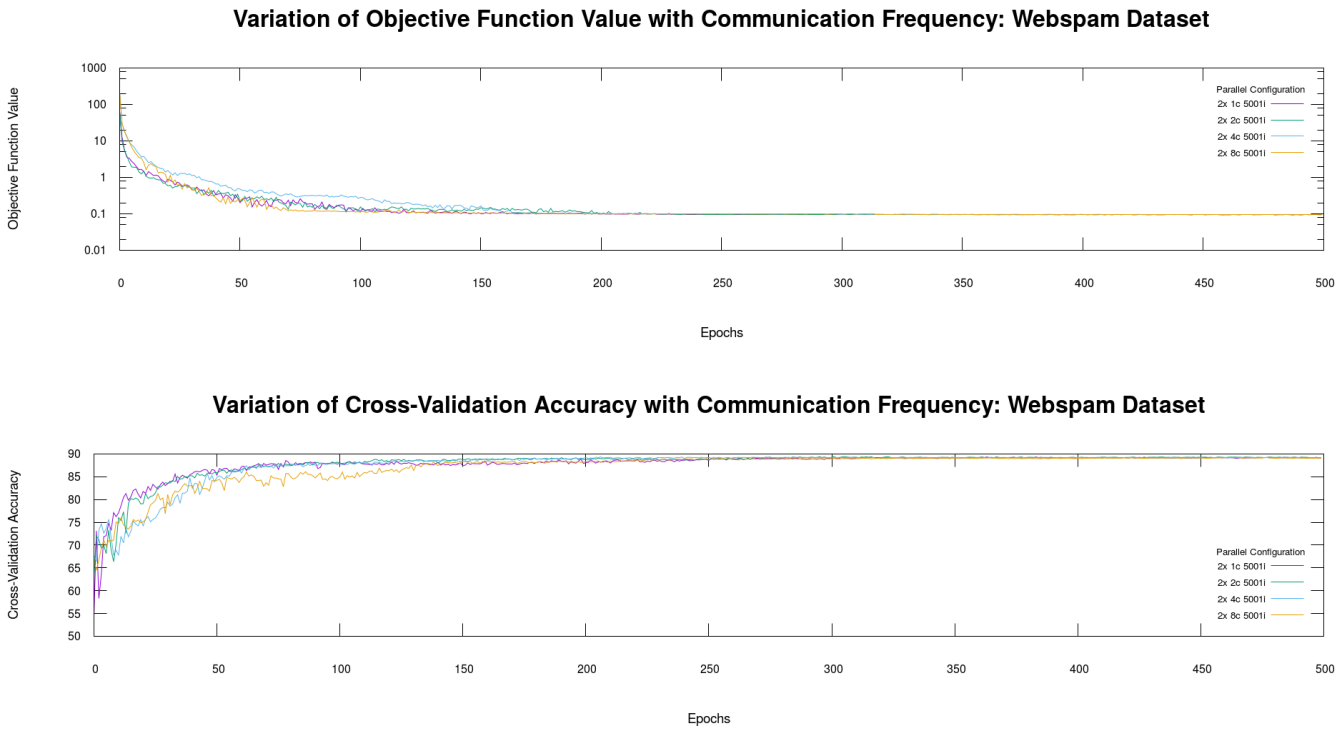

Fig. 117. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,], Parallelism = 2

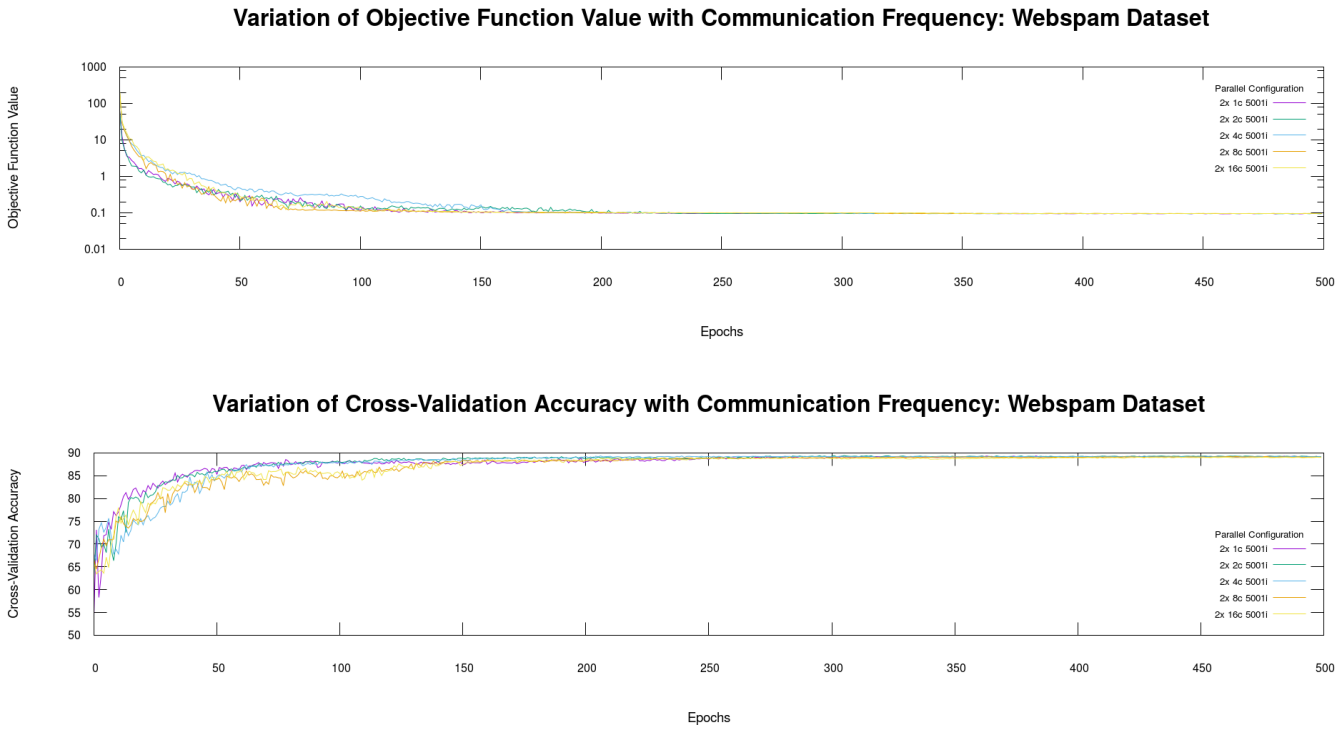

Fig. 118. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,16,], Parallelism = 2

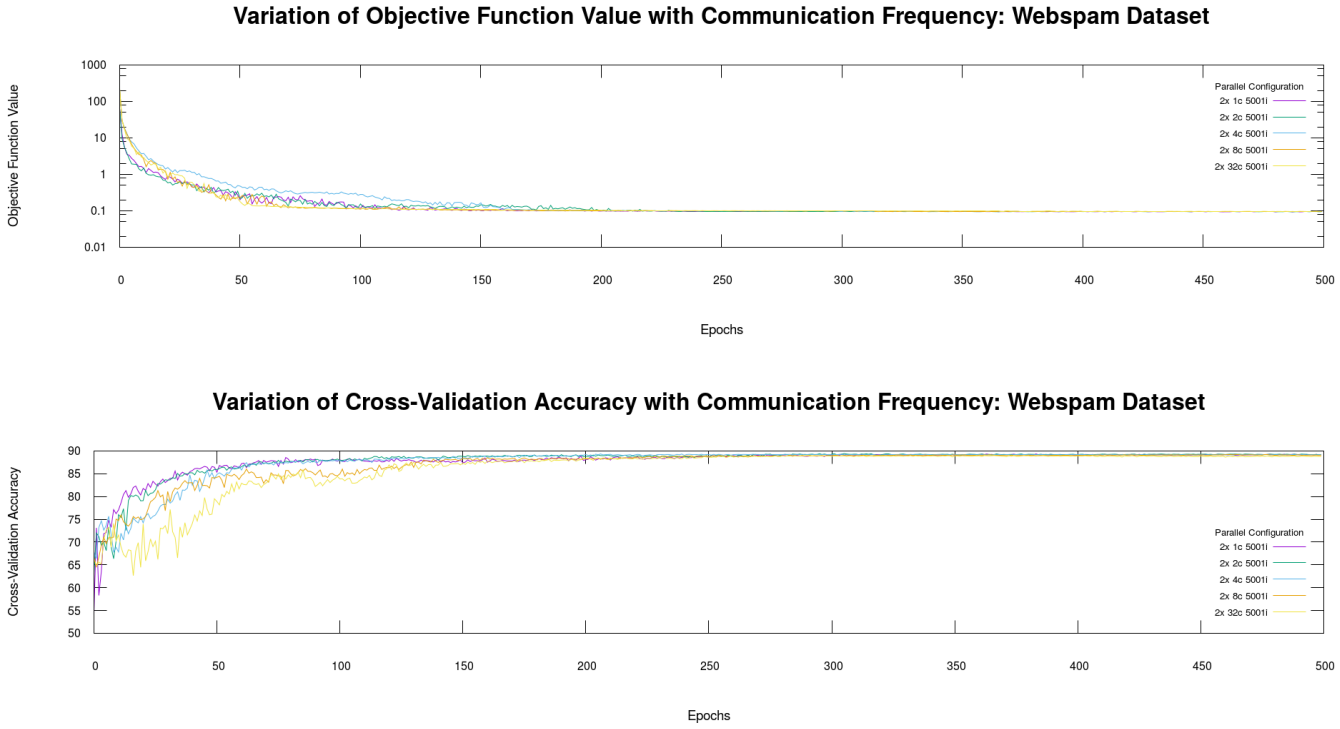

Fig. 119. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,32,], Parallelism = 2

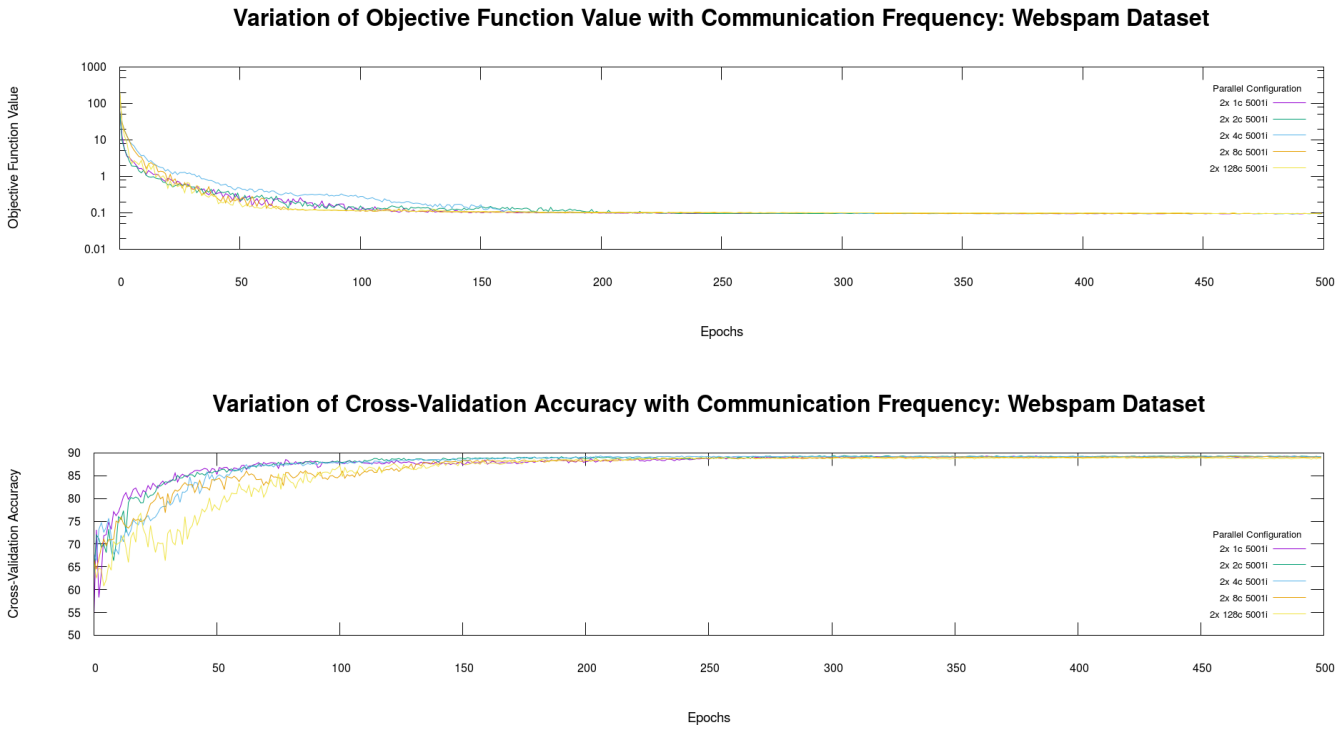

Fig. 120. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,128,], Parallelism = 2

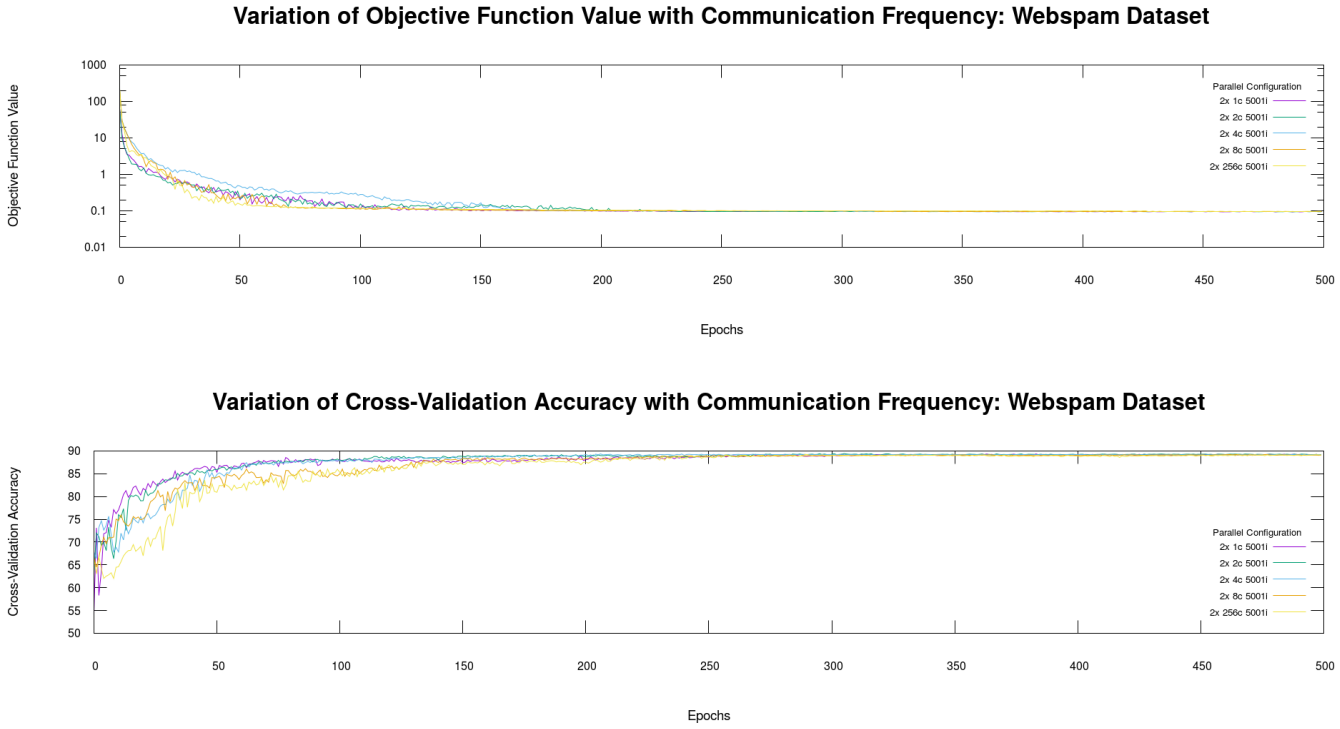

Fig. 121. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,256,], Parallelism = 2

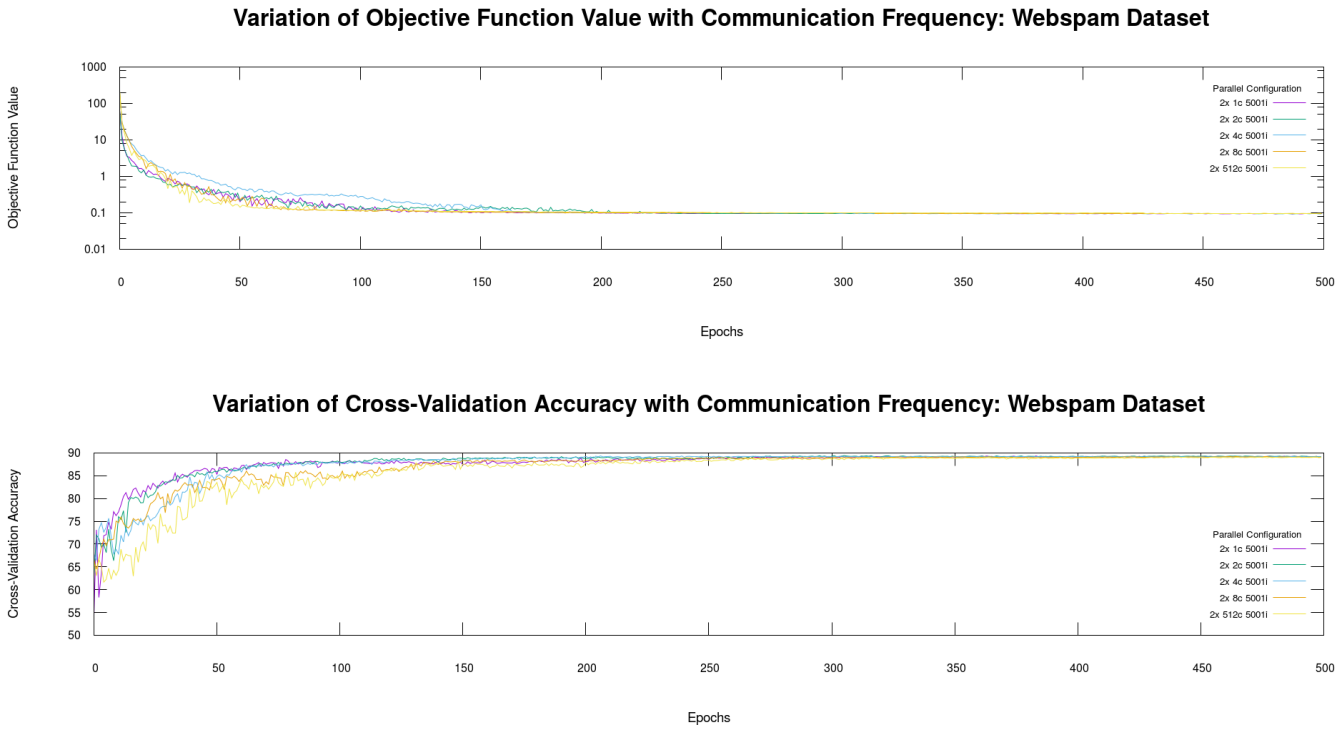

Fig. 122. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,512,], Parallelism = 2

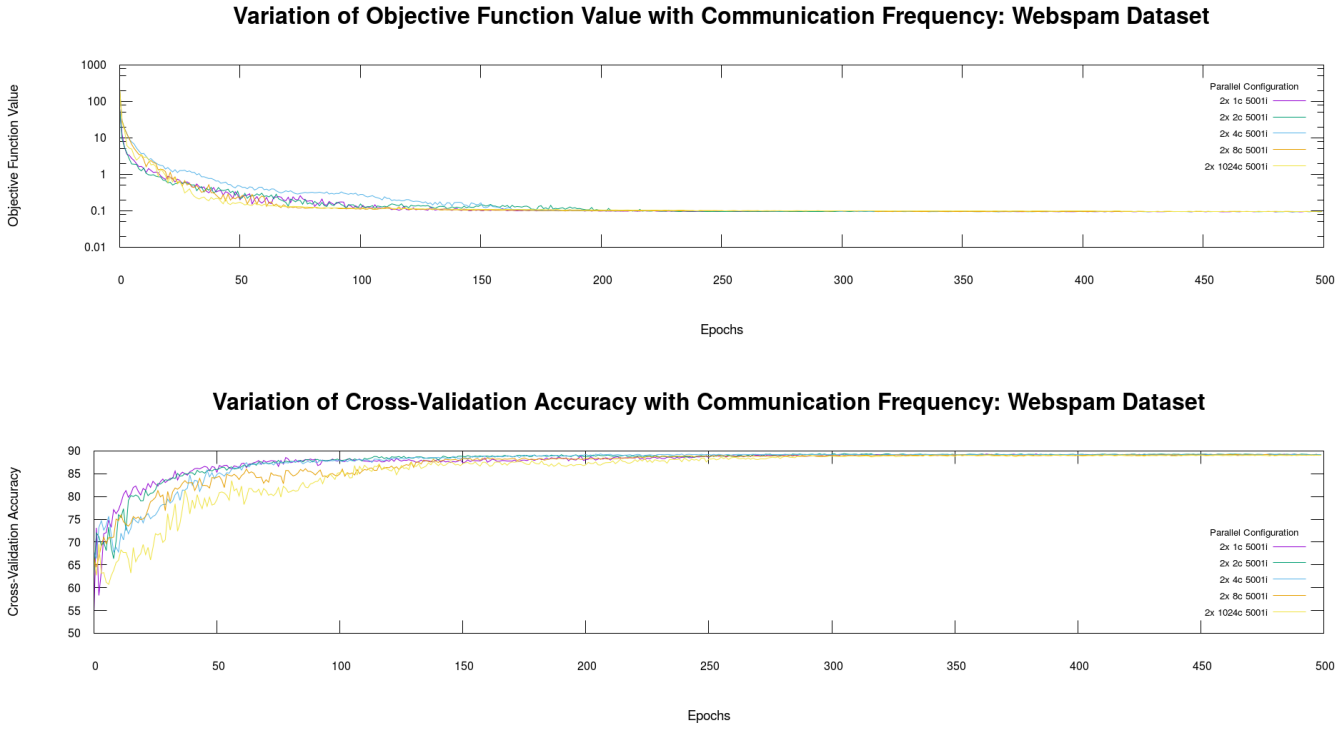

Fig. 123. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 2

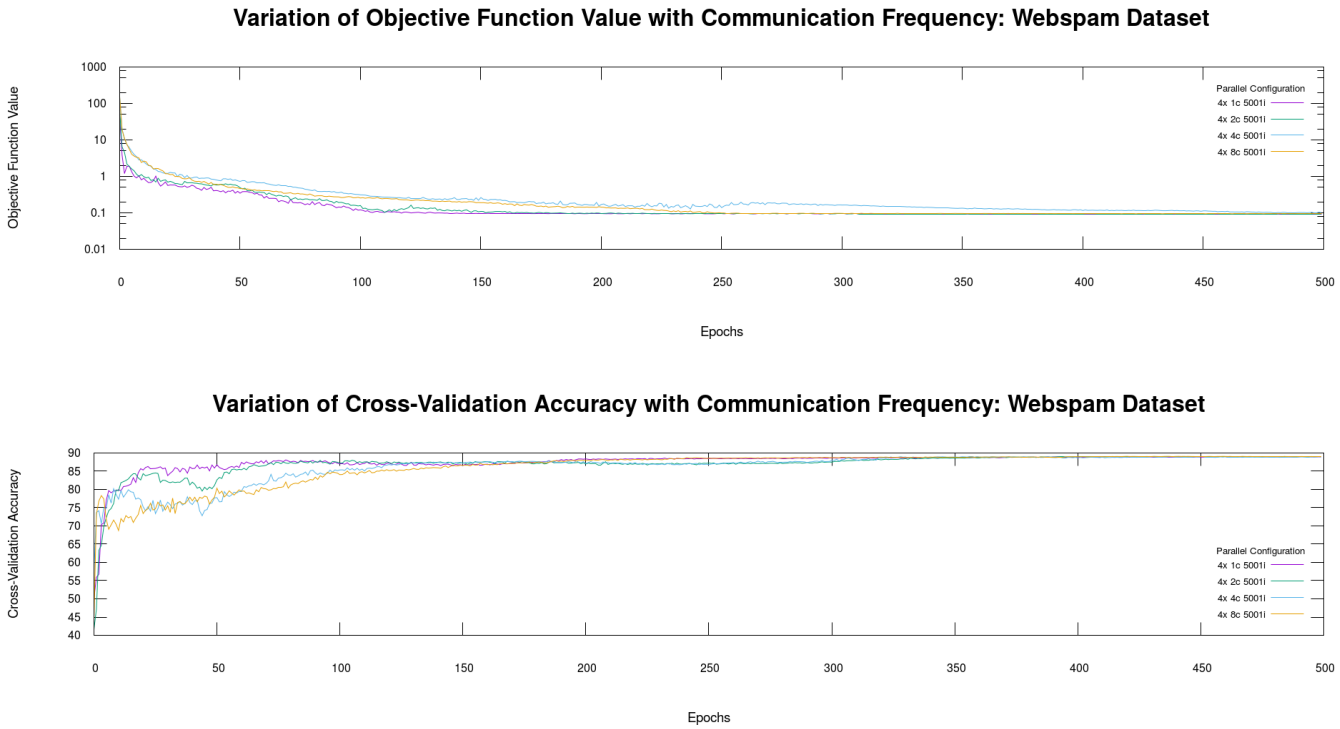

Fig. 124. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,], Parallelism = 4

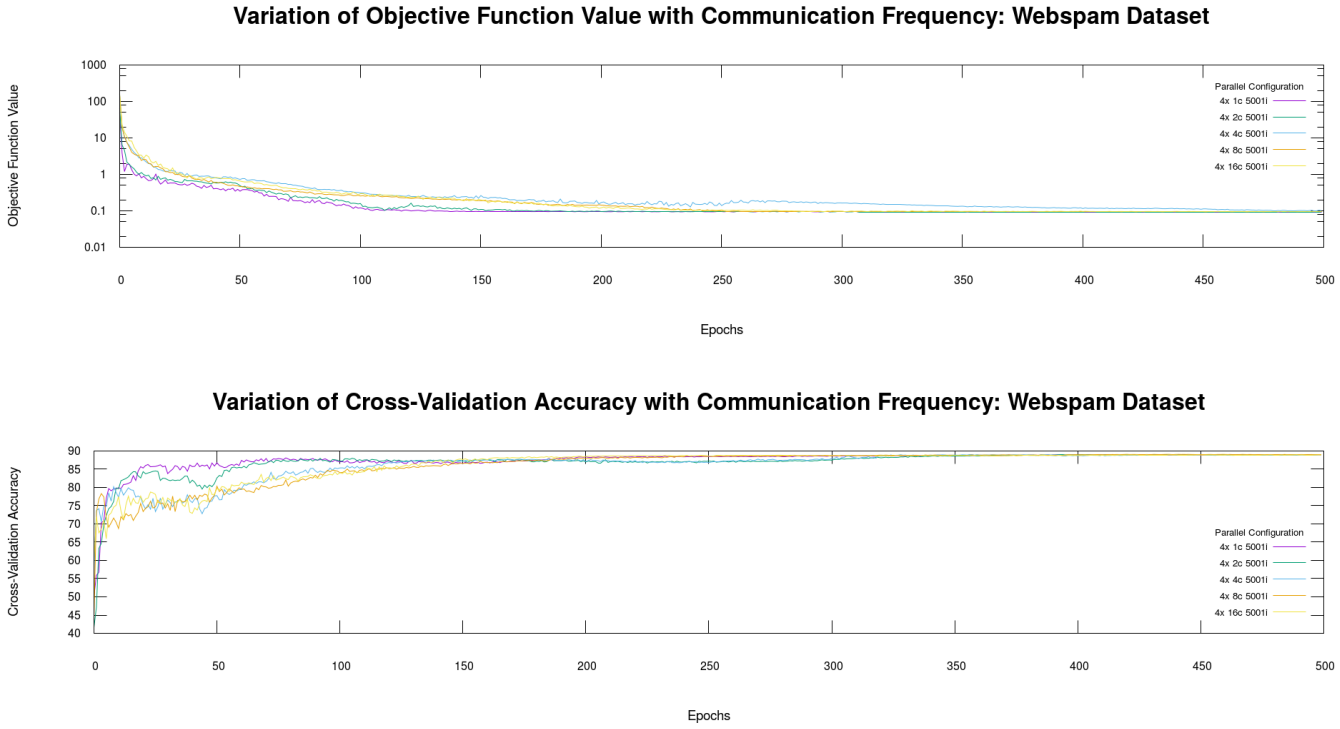

Fig. 125. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,16,], Parallelism = 4

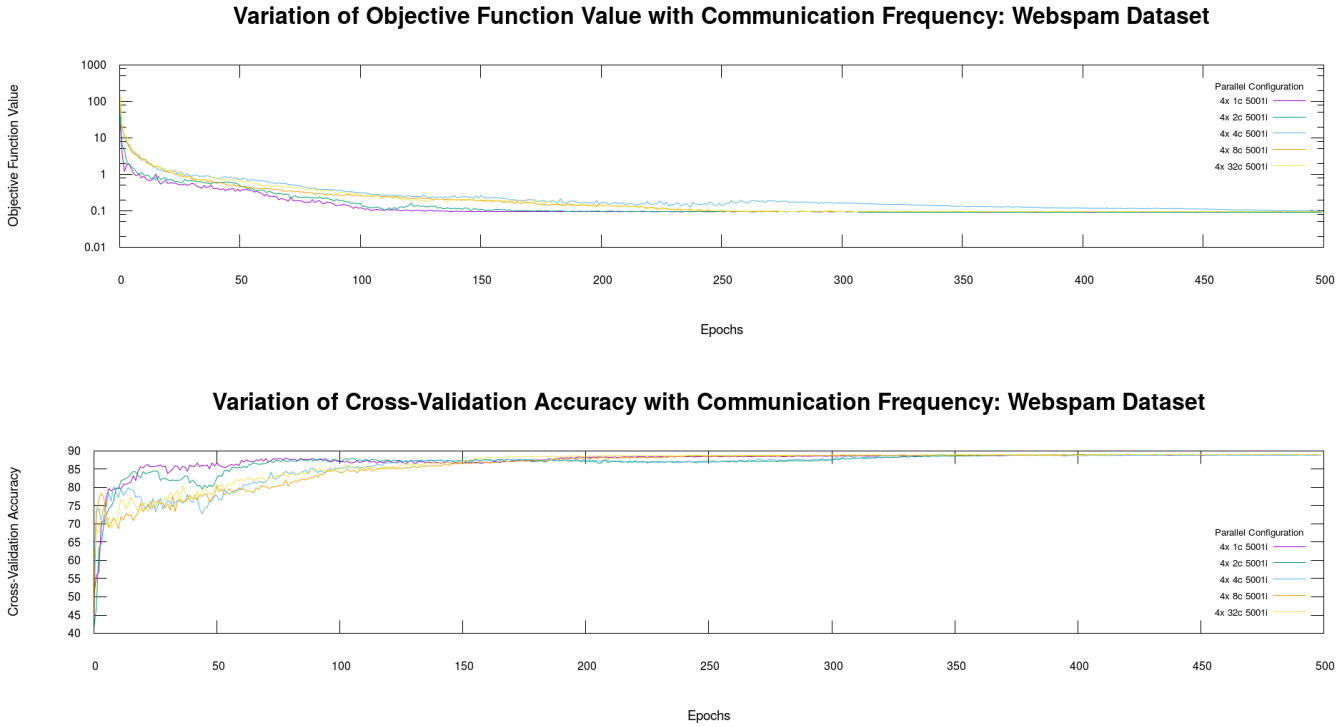

Fig. 126. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,32,], Parallelism = 4

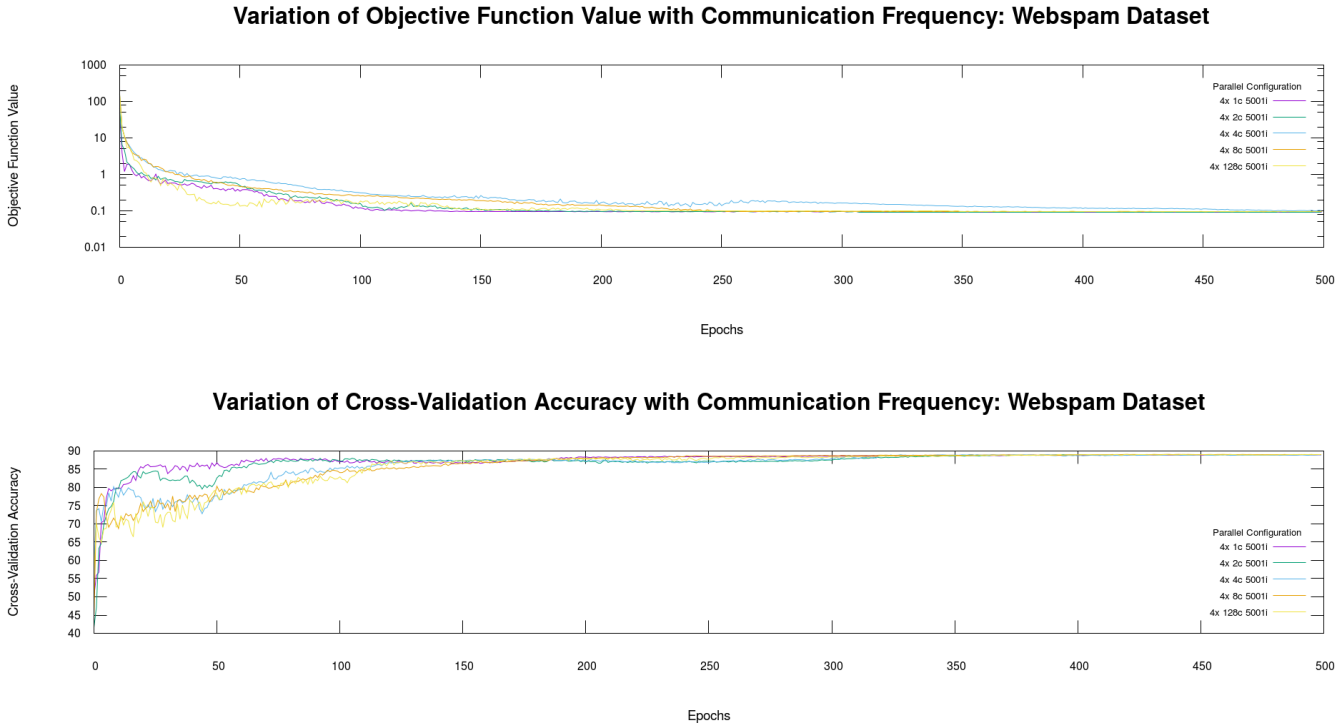

Fig. 127. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,128,], Parallelism = 4

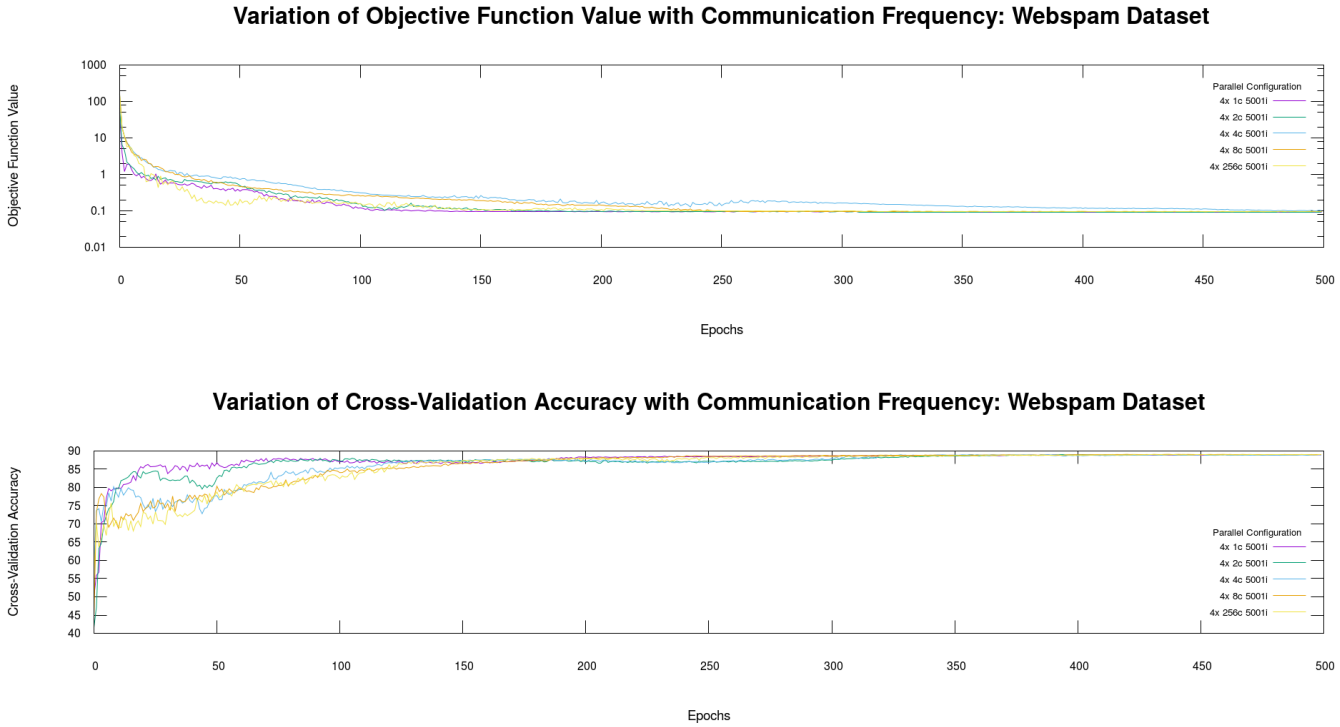

Fig. 128. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,256,], Parallelism = 4

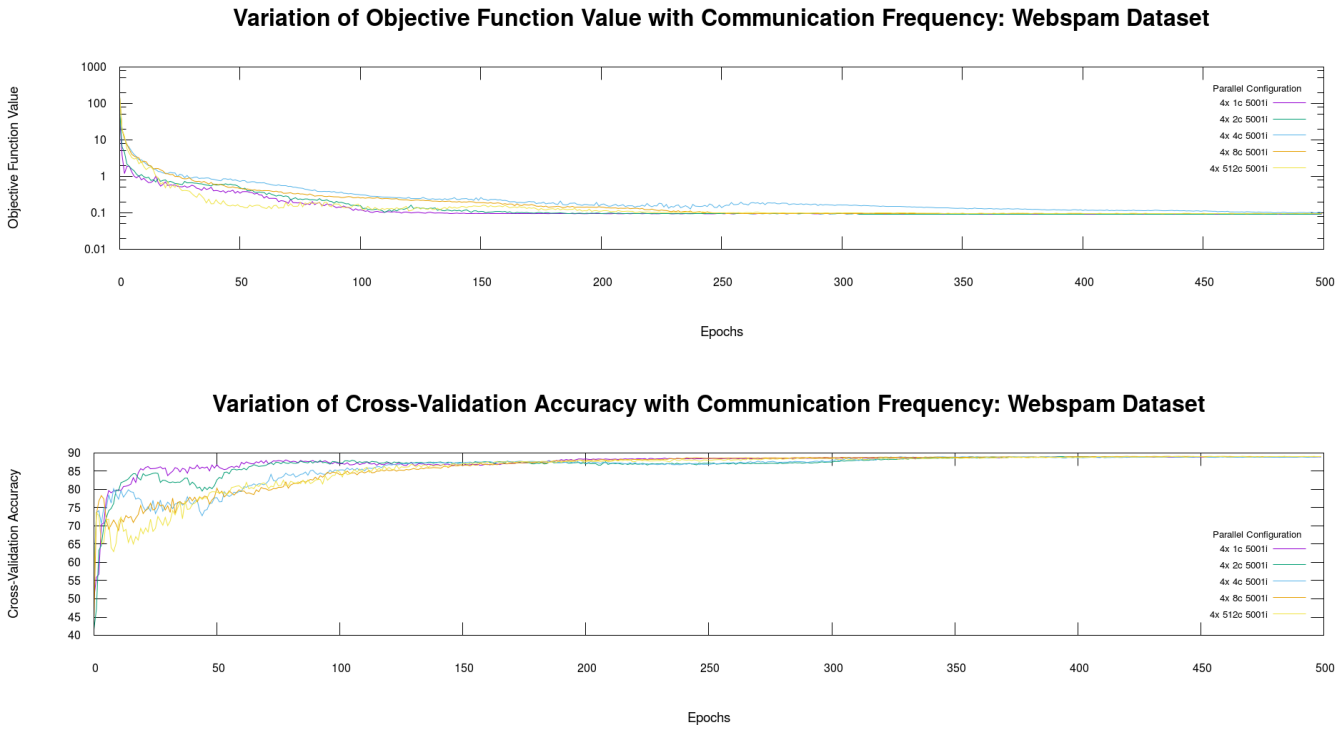

Fig. 129. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,512,], Parallelism = 4

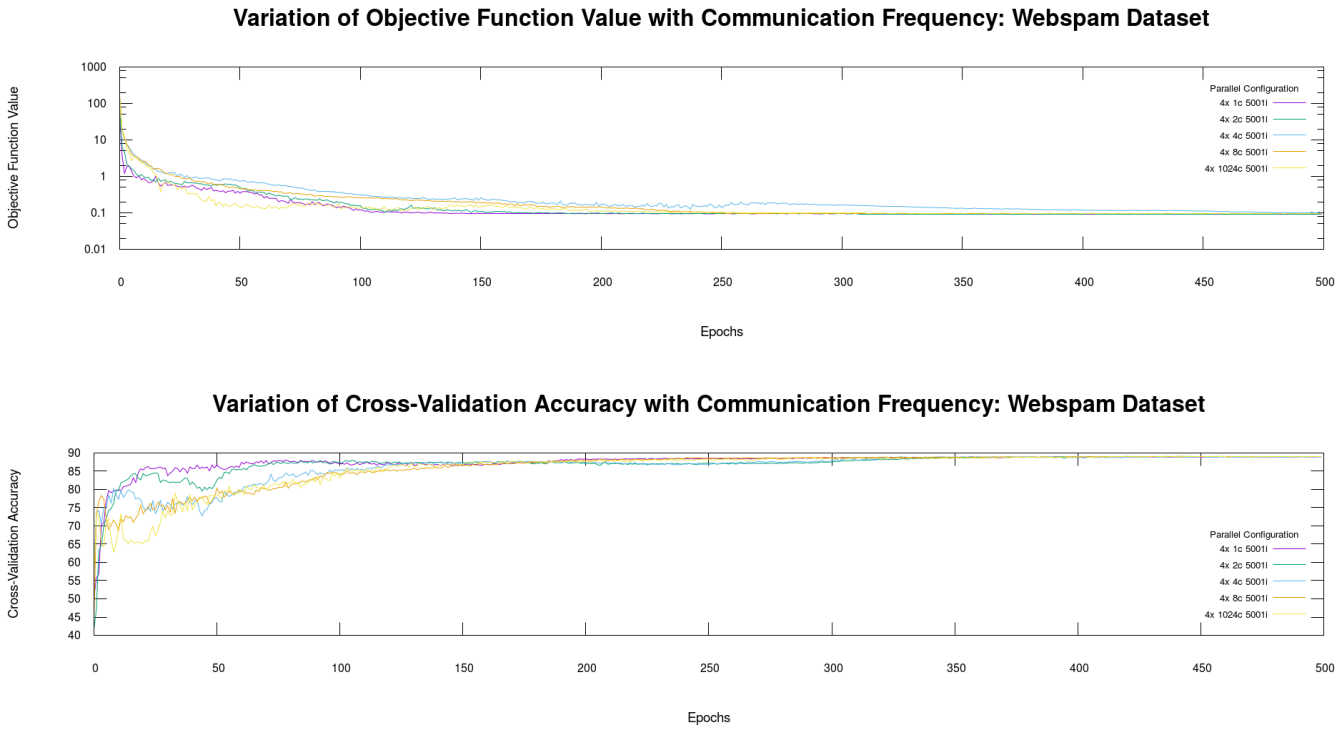

Fig. 130. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 4

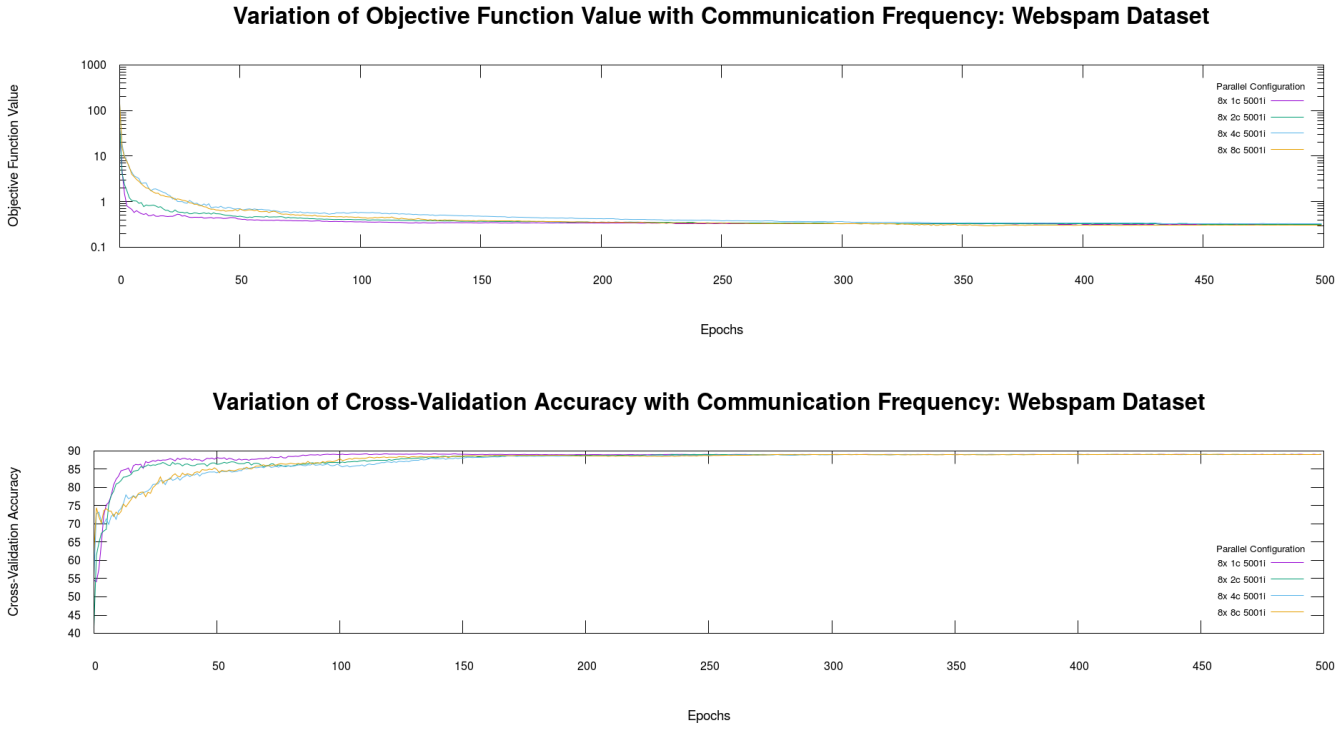

Fig. 131. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,], Parallelism = 8

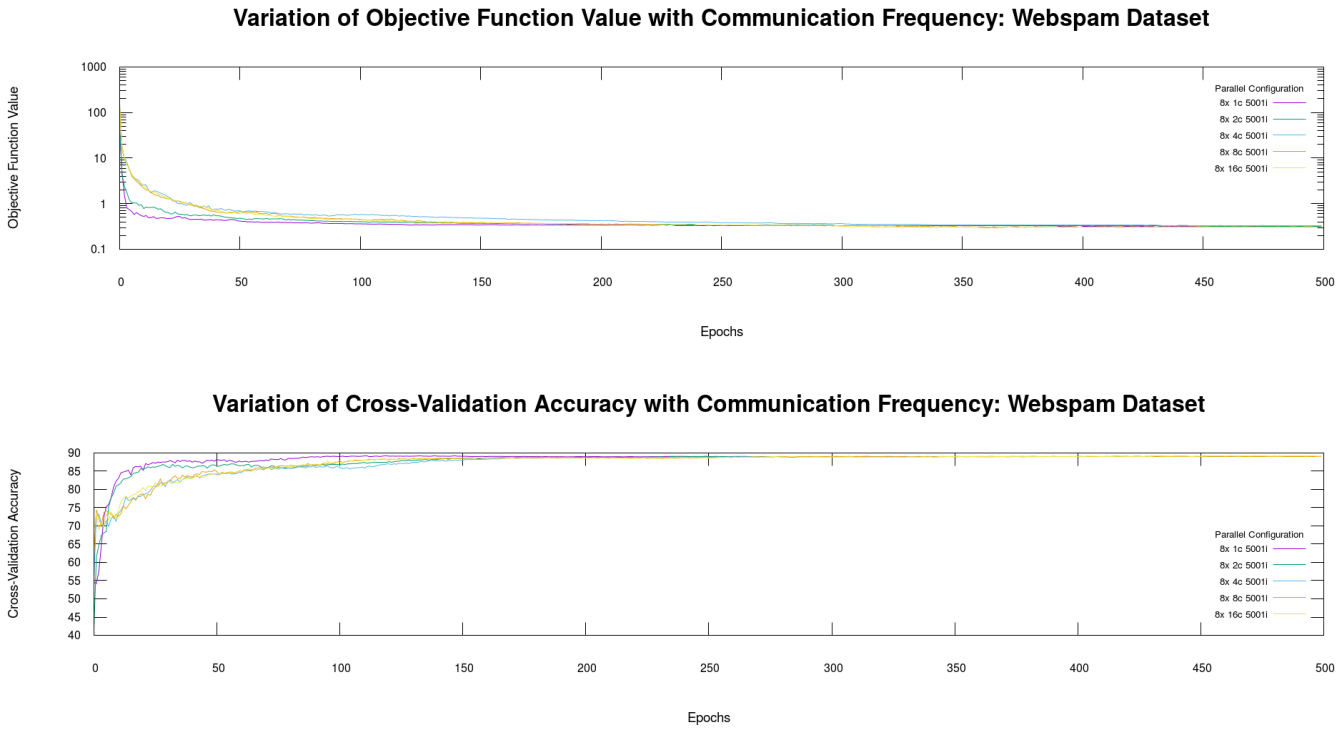

Fig. 132. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,16,], Parallelism = 8

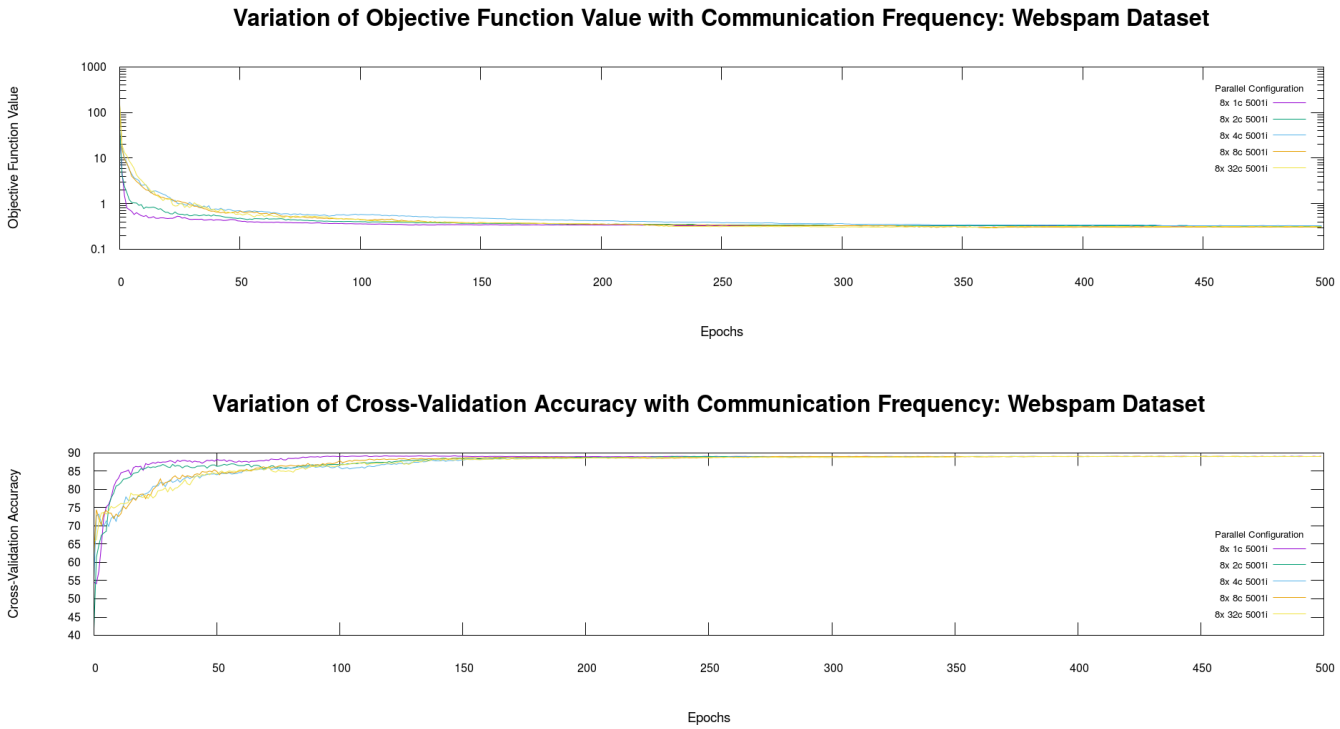

Fig. 133. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,32,], Parallelism = 8

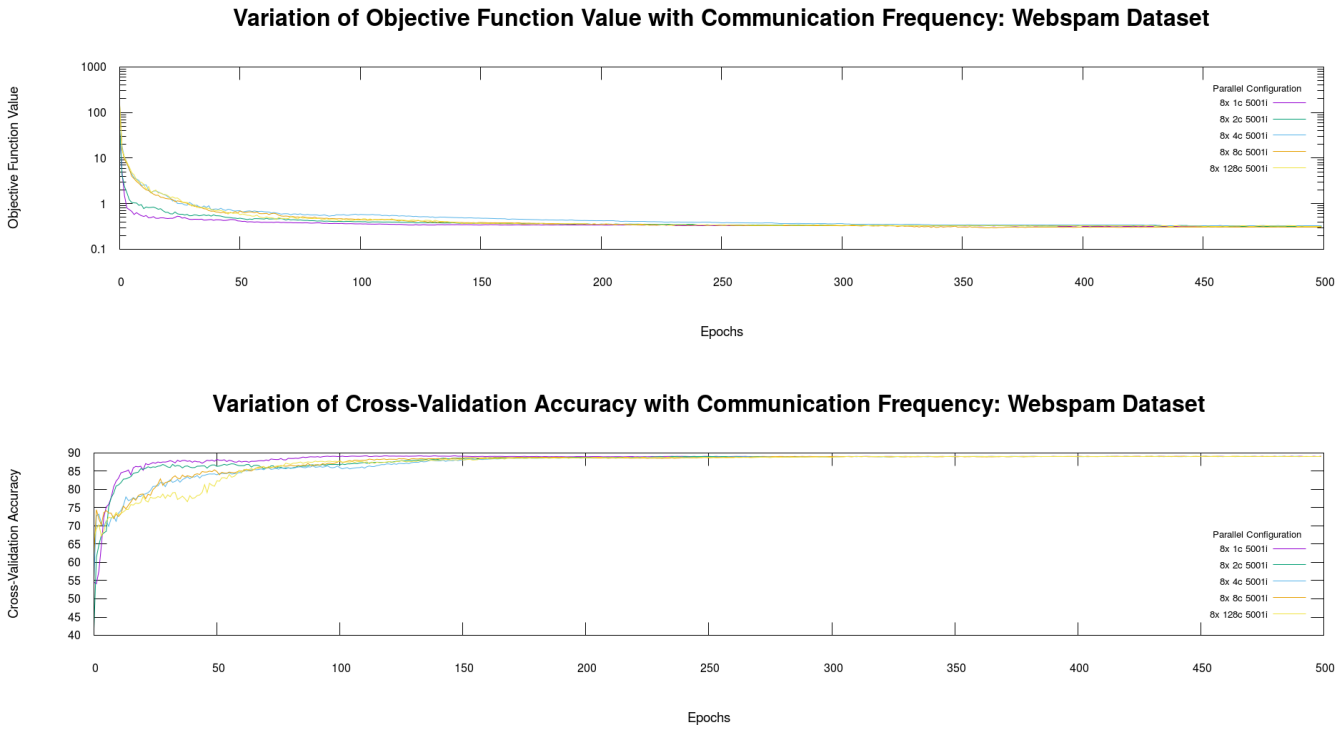

Fig. 134. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,128,], Parallelism = 8

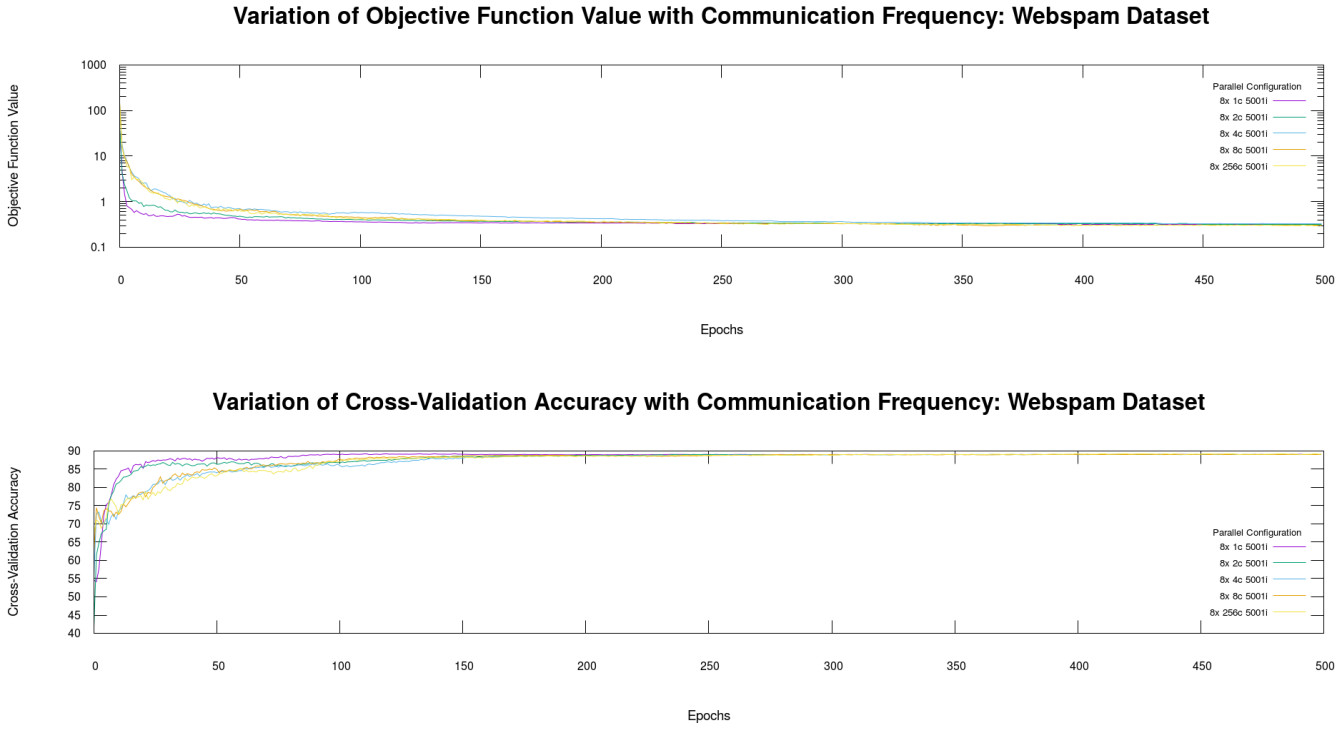

Fig. 135. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,256,], Parallelism = 8

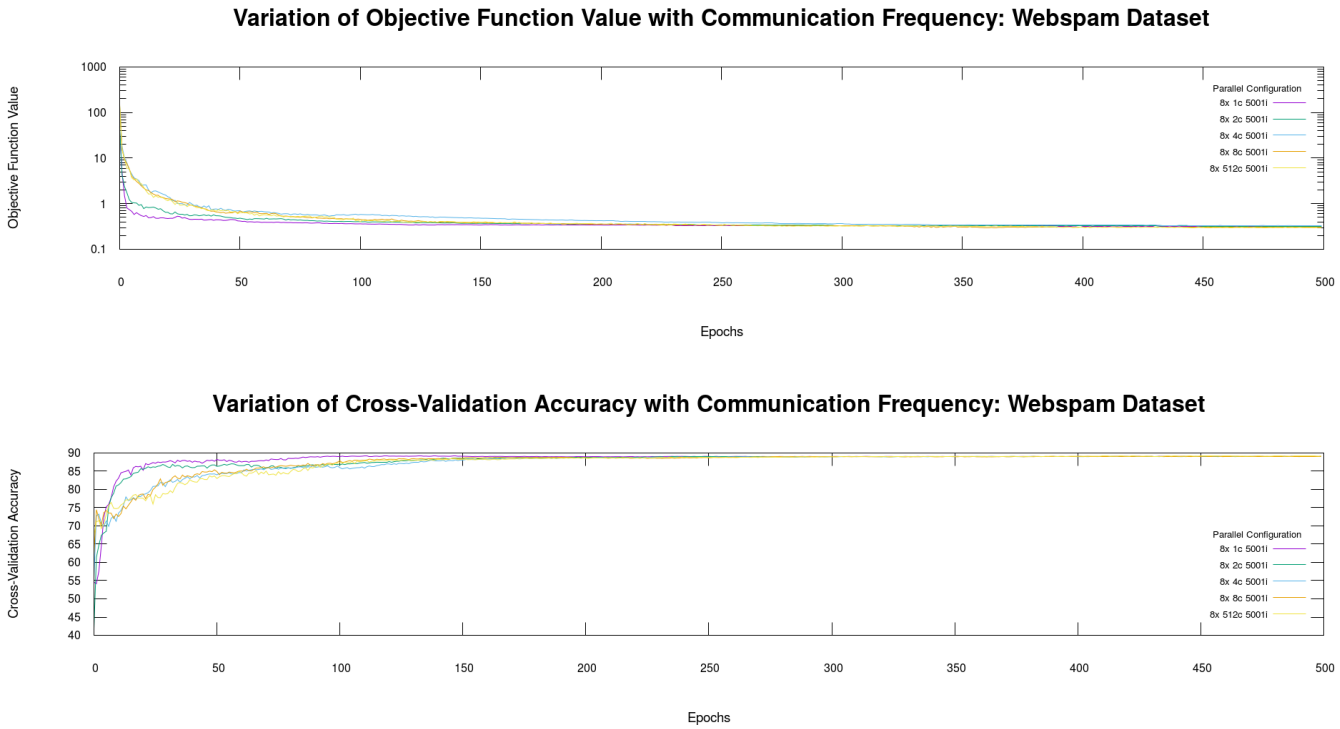

Fig. 136. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,512,], Parallelism = 8

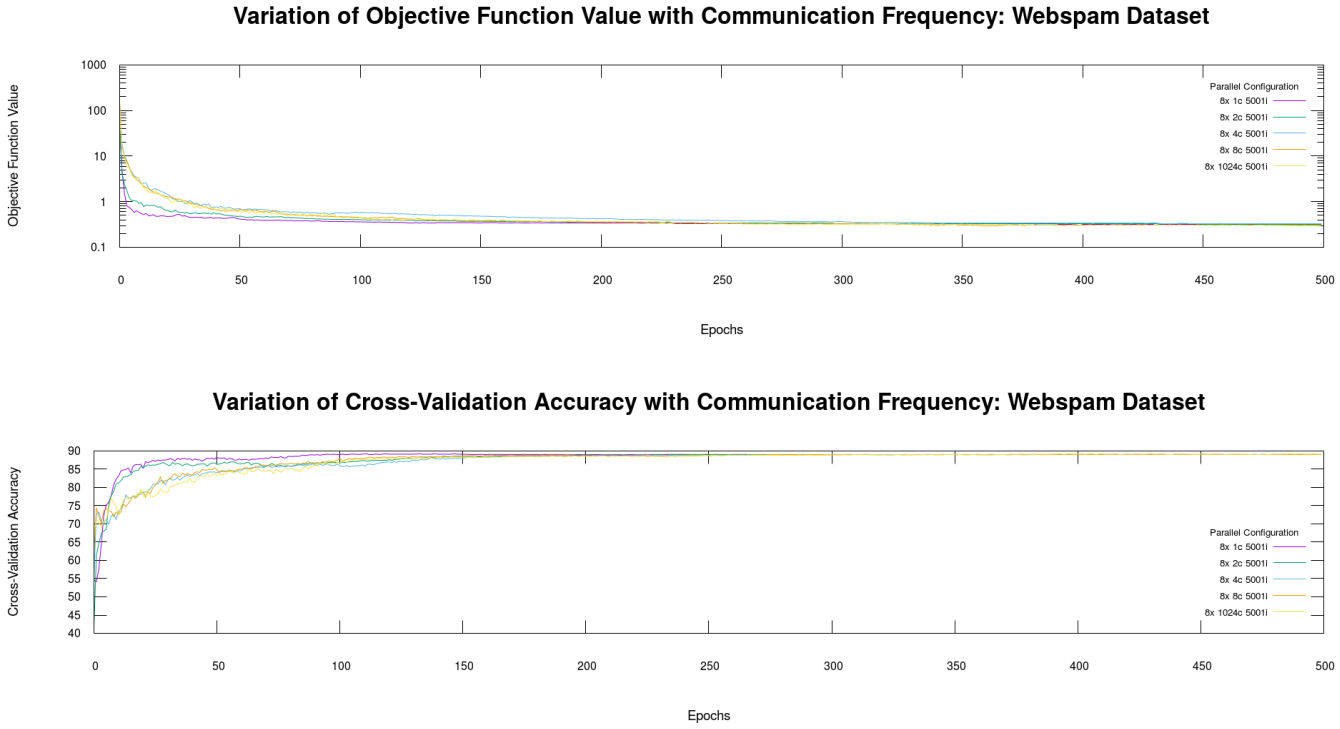

Fig. 137. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 8

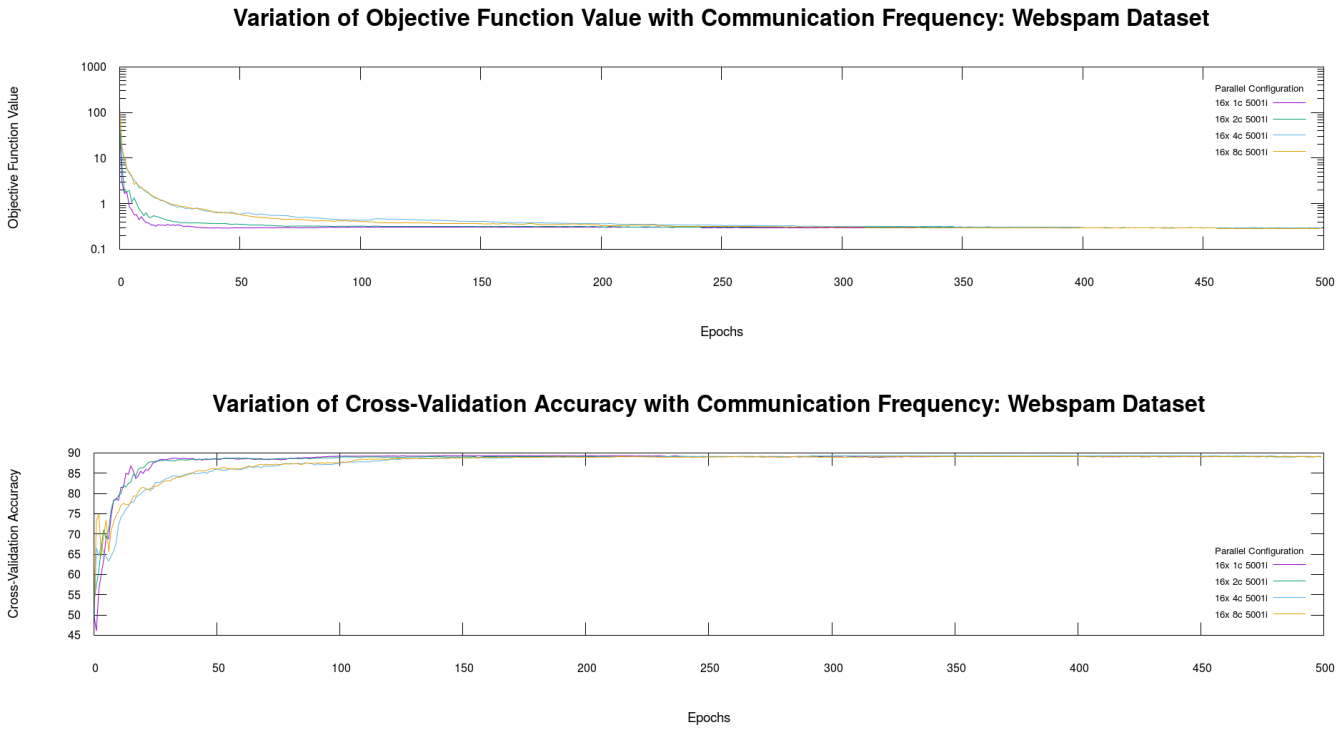

Fig. 138. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,], Parallelism = 16

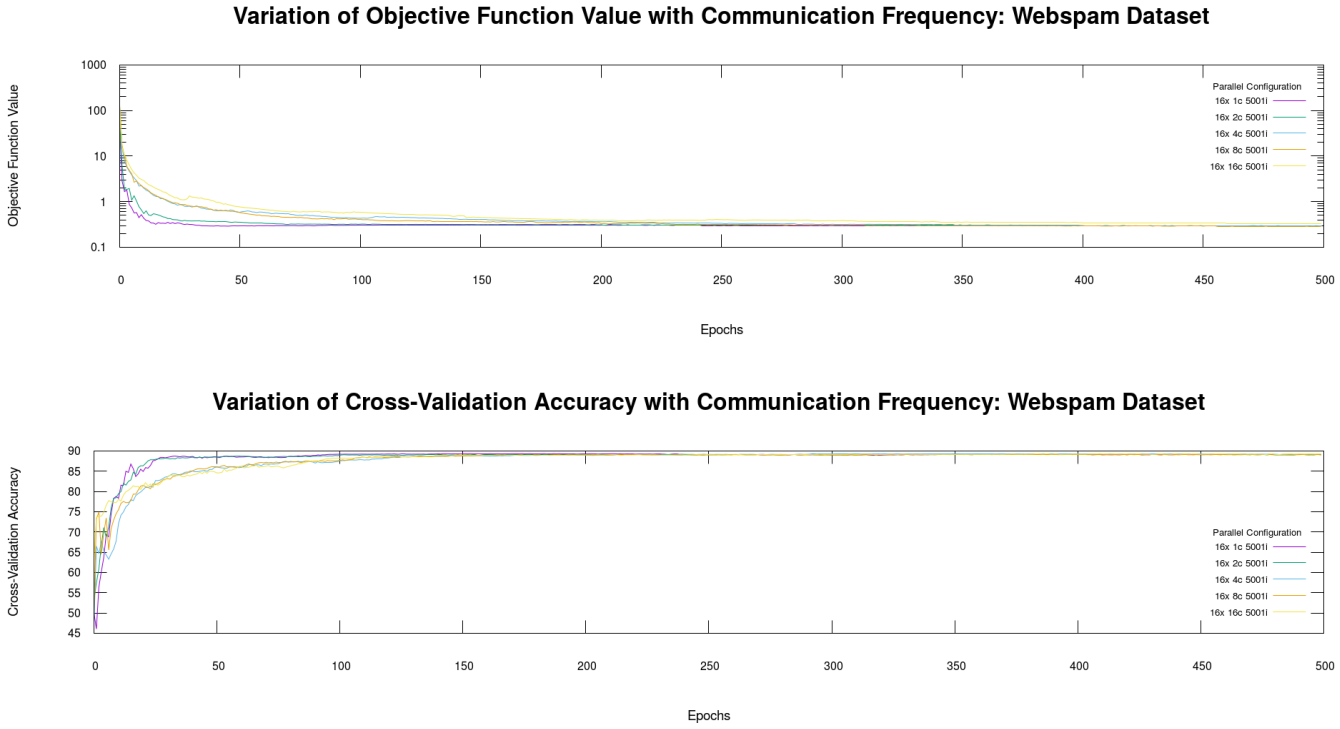

Fig. 139. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,16,], Parallelism = 16

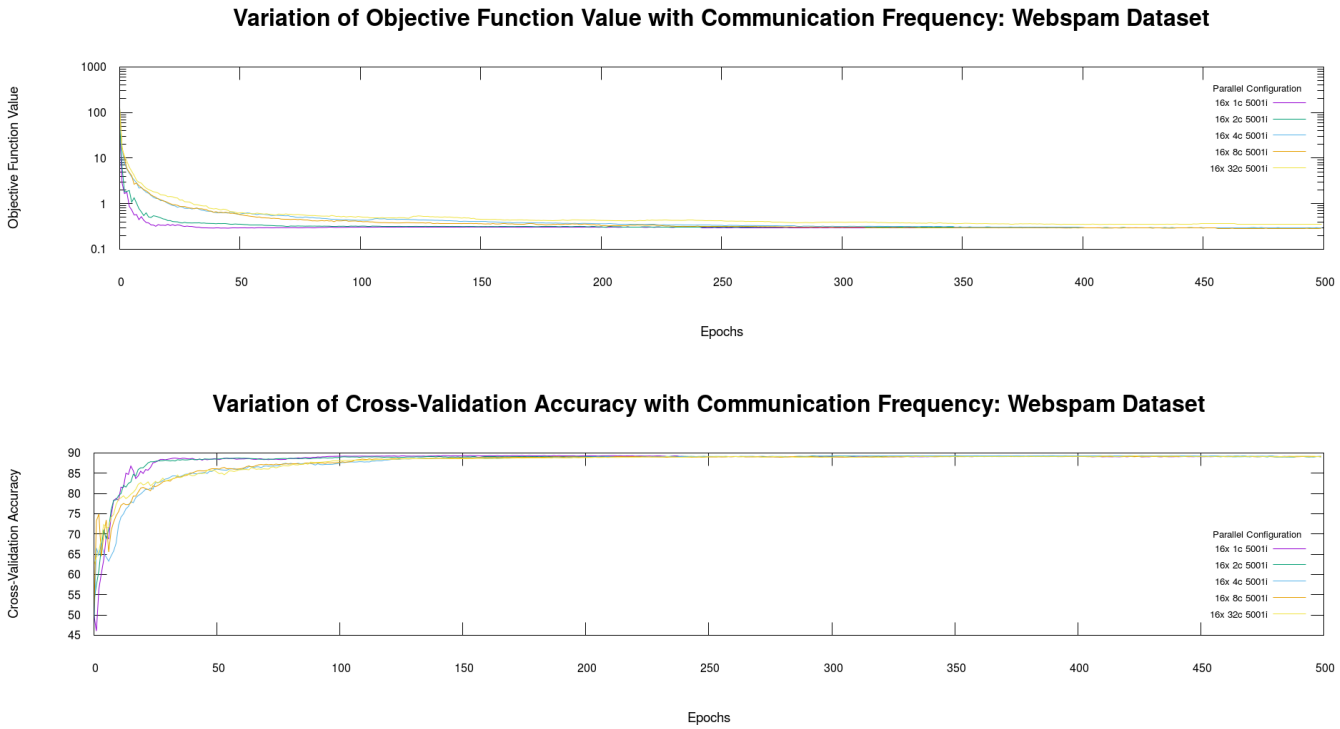

Fig. 140. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,32,], Parallelism = 16

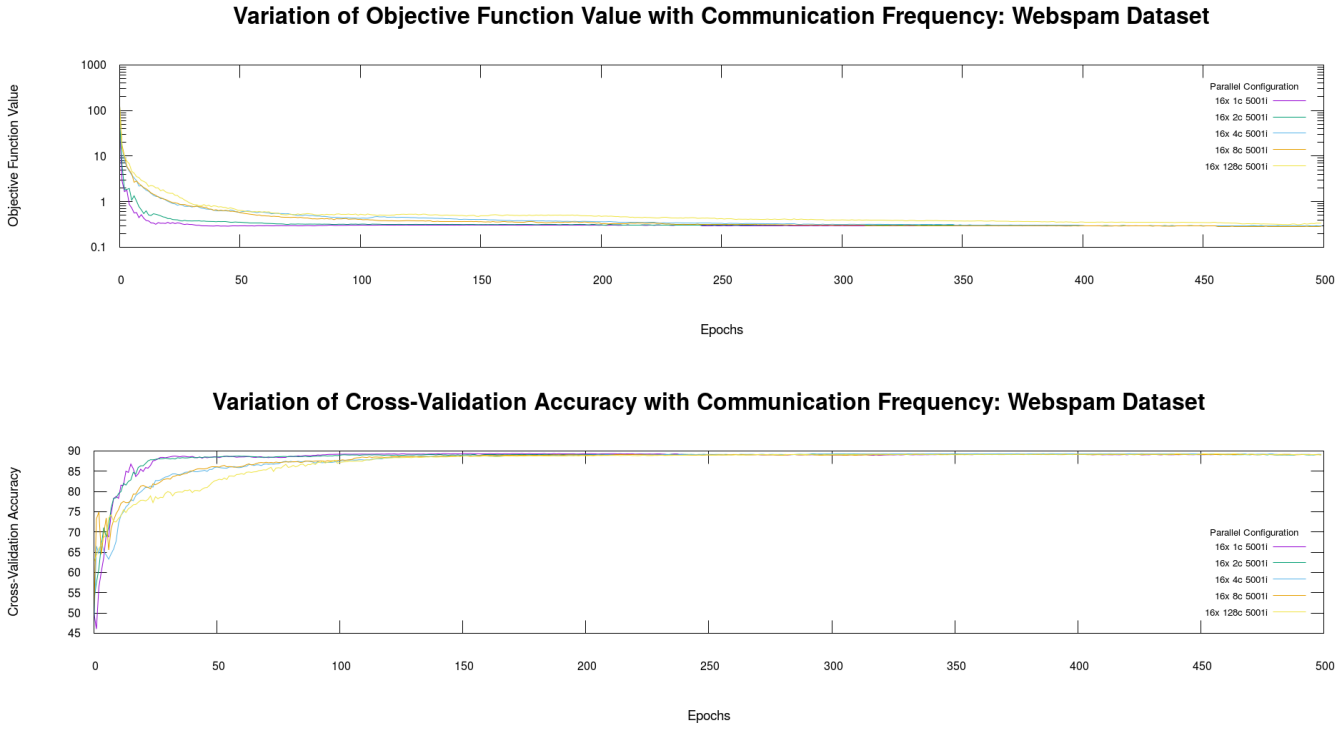

Fig. 141. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,128,], Parallelism = 16

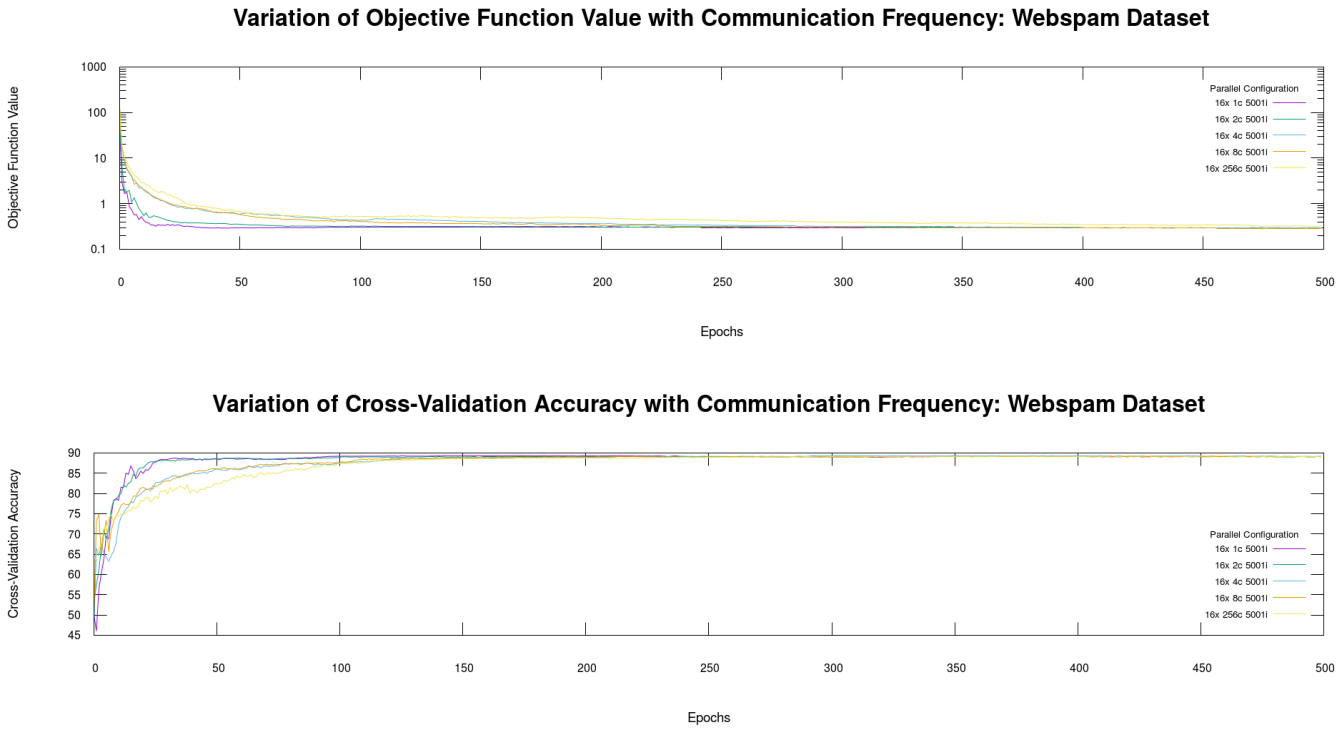

Fig. 142. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,256,], Parallelism = 16

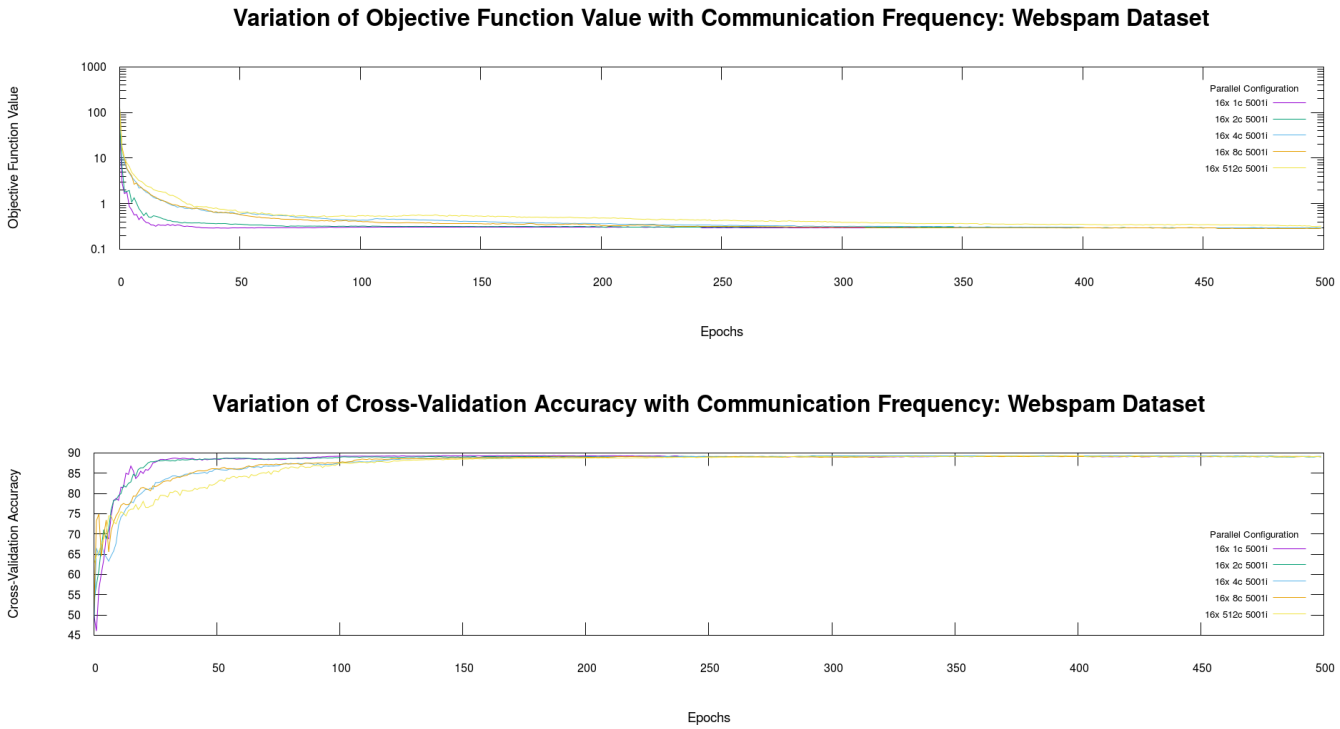

Fig. 143. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,512,], Parallelism = 16

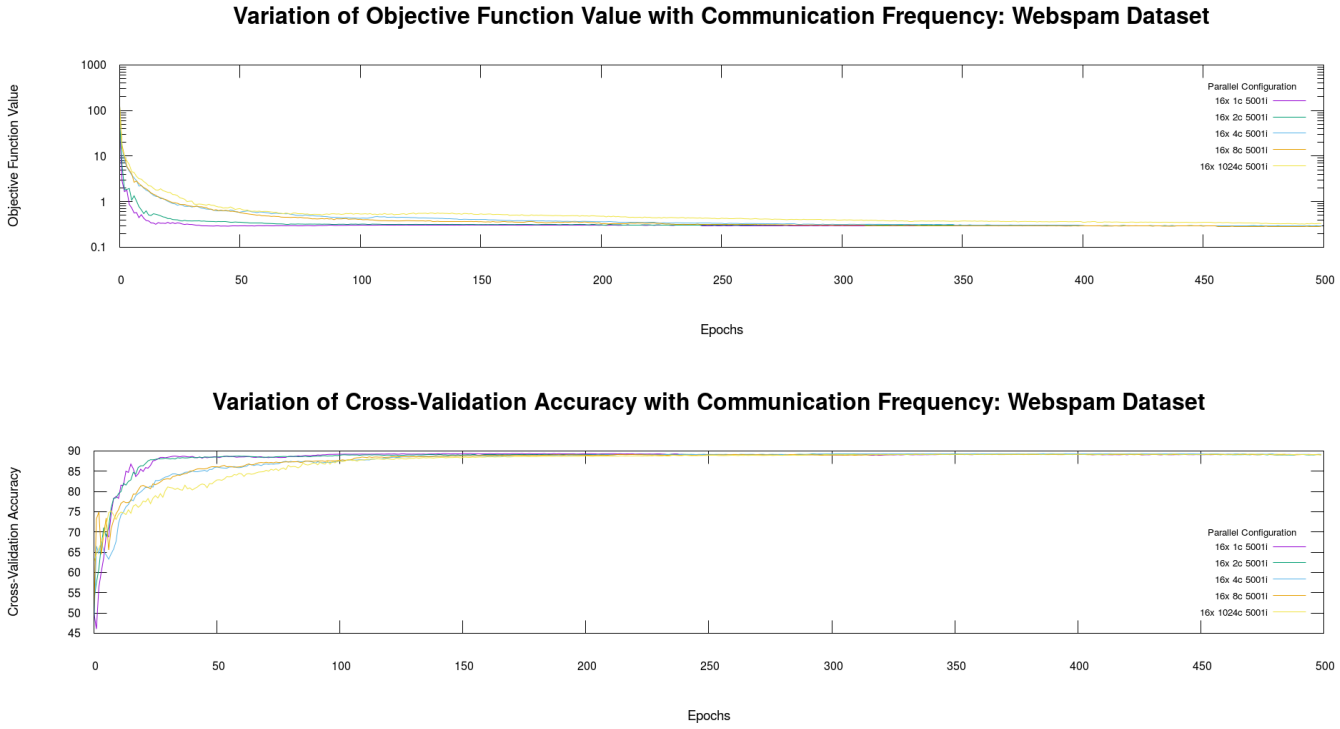

Fig. 144. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 16

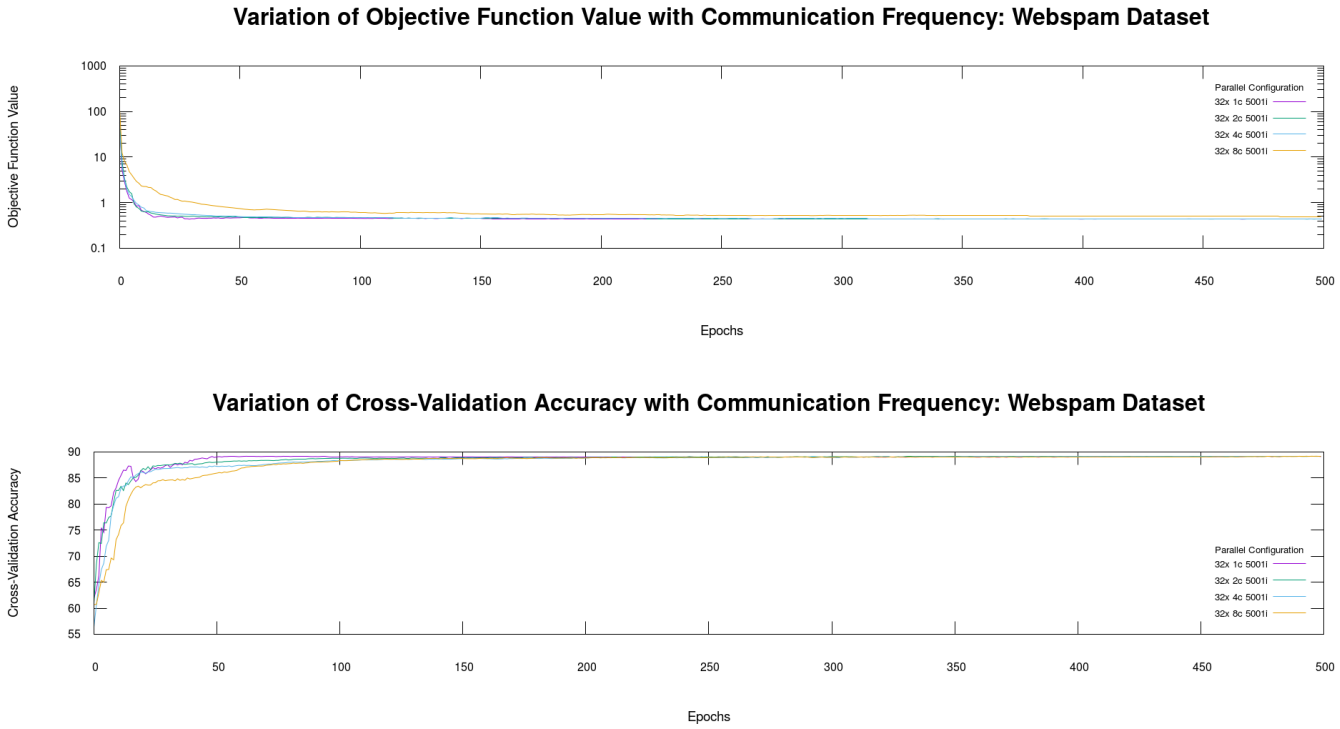

Fig. 145. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,], Parallelism = 32

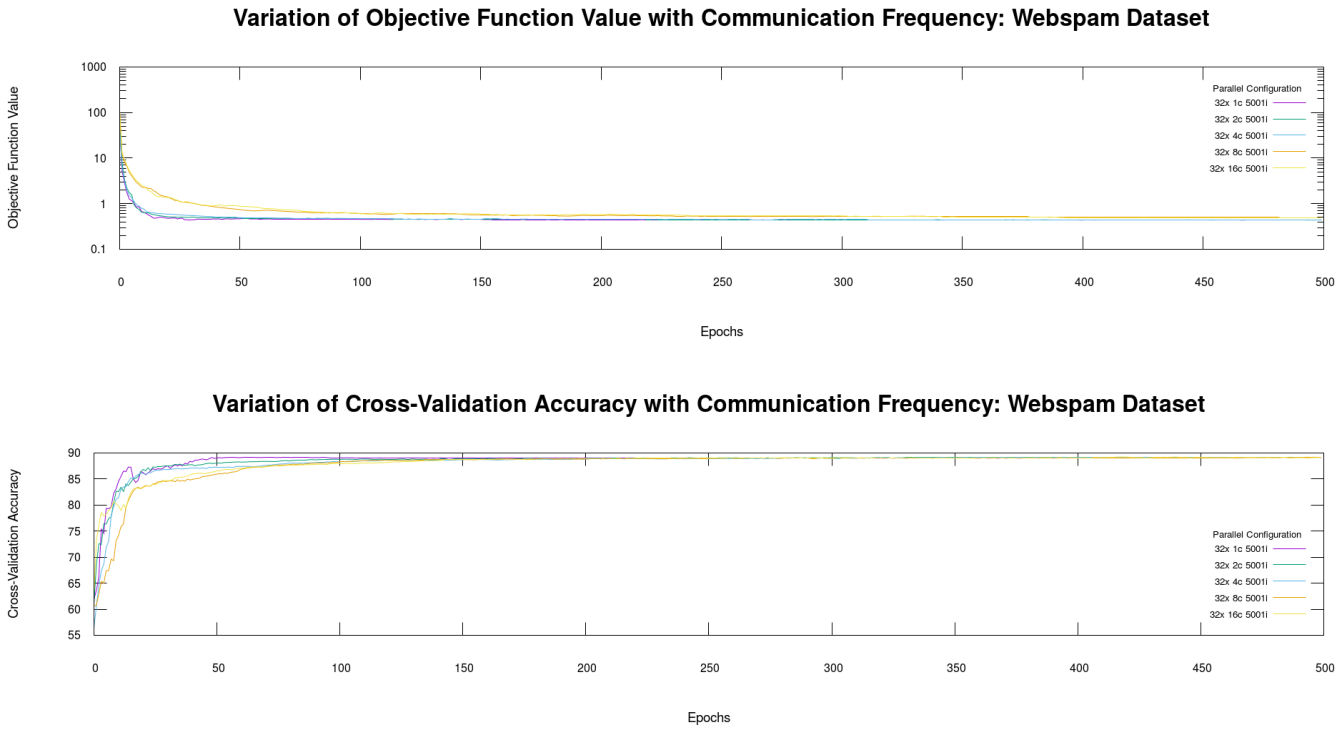

Fig. 146. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,16,], Parallelism = 32

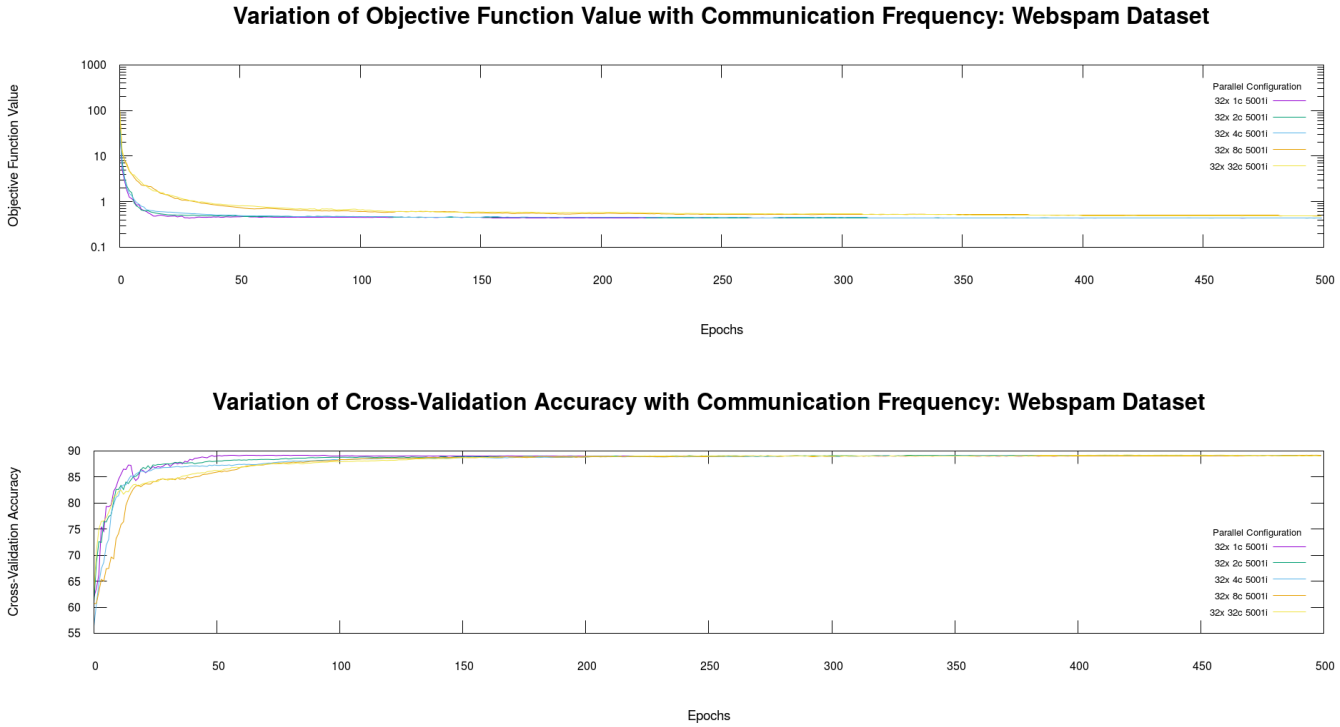

Fig. 147. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,32,], Parallelism = 32

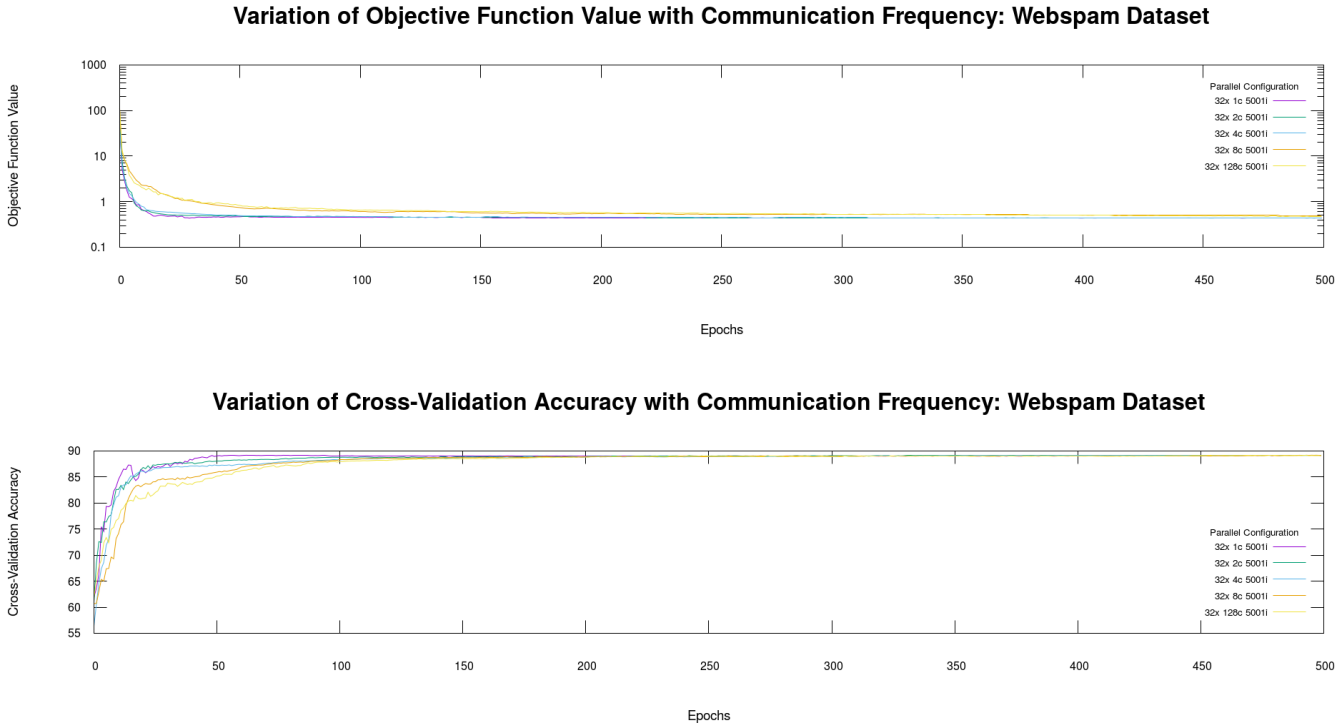

Fig. 148. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,128,], Parallelism = 32

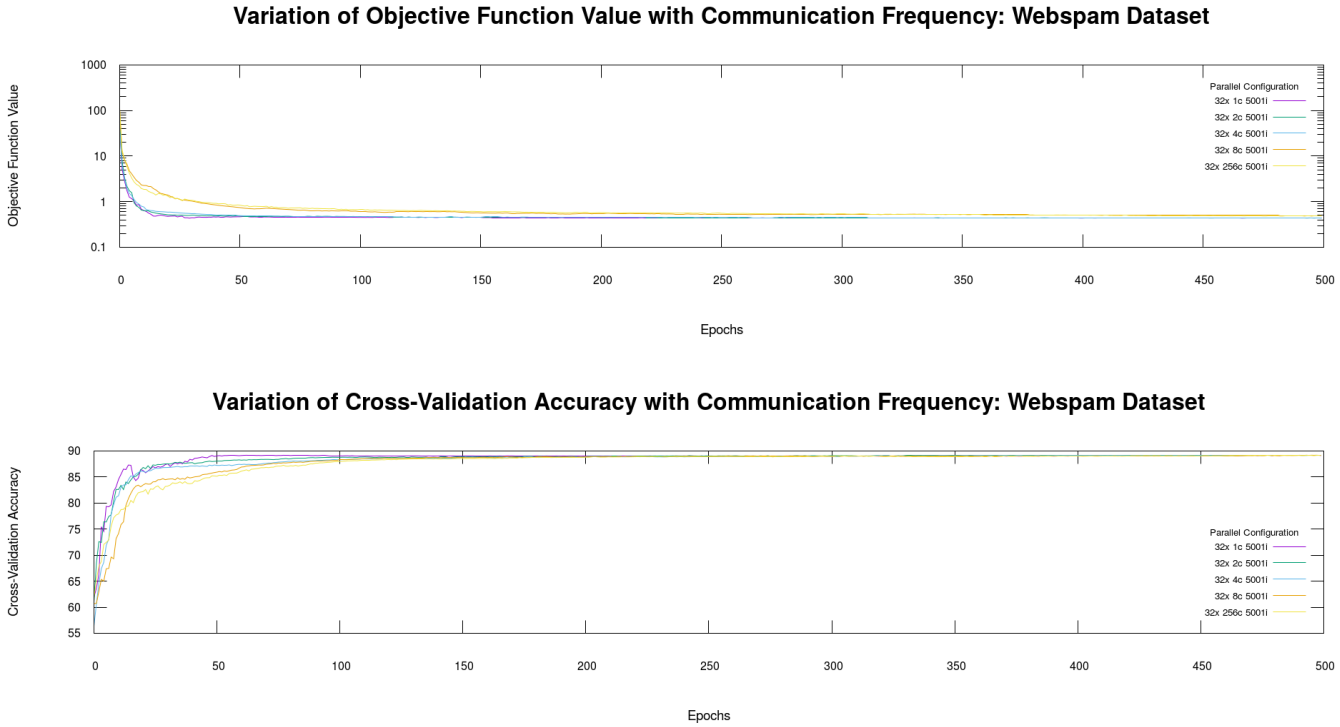

Fig. 149. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,256,], Parallelism = 32

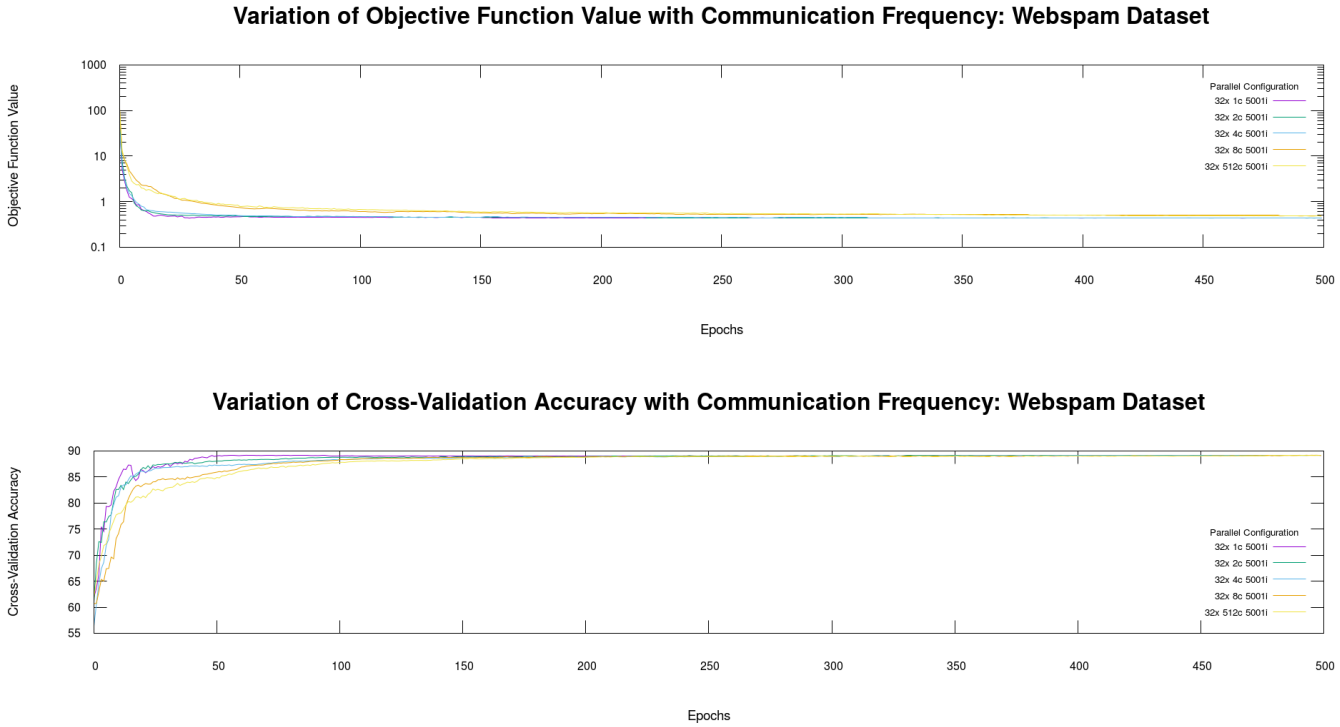

Fig. 150. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,512,], Parallelism = 32

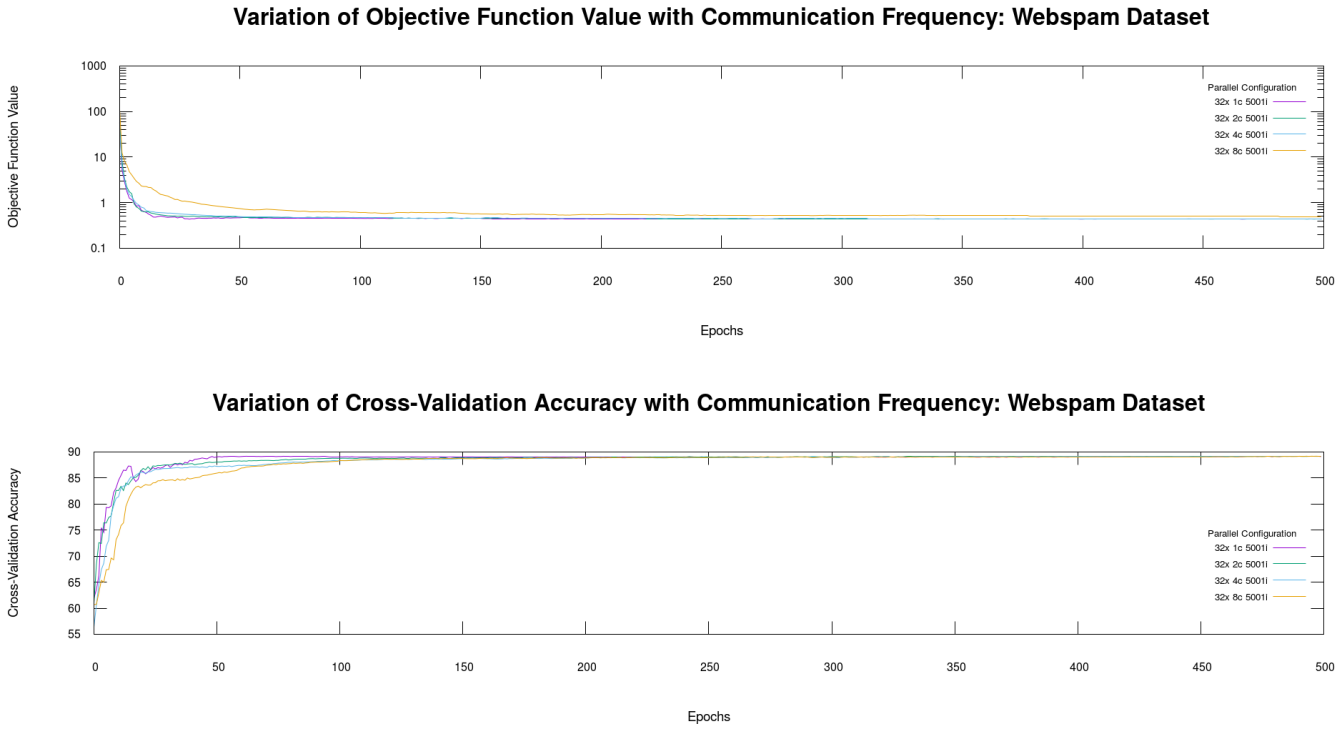

Fig. 151. Distributed Training Time : Dataset Webspam , Configuration : MSF = [1,2,4,8,1024,], Parallelism = 32

D. Summary Ijcnn1 MSF = 1,2,4,8 , Parallelism = [2,4,8,16,32]

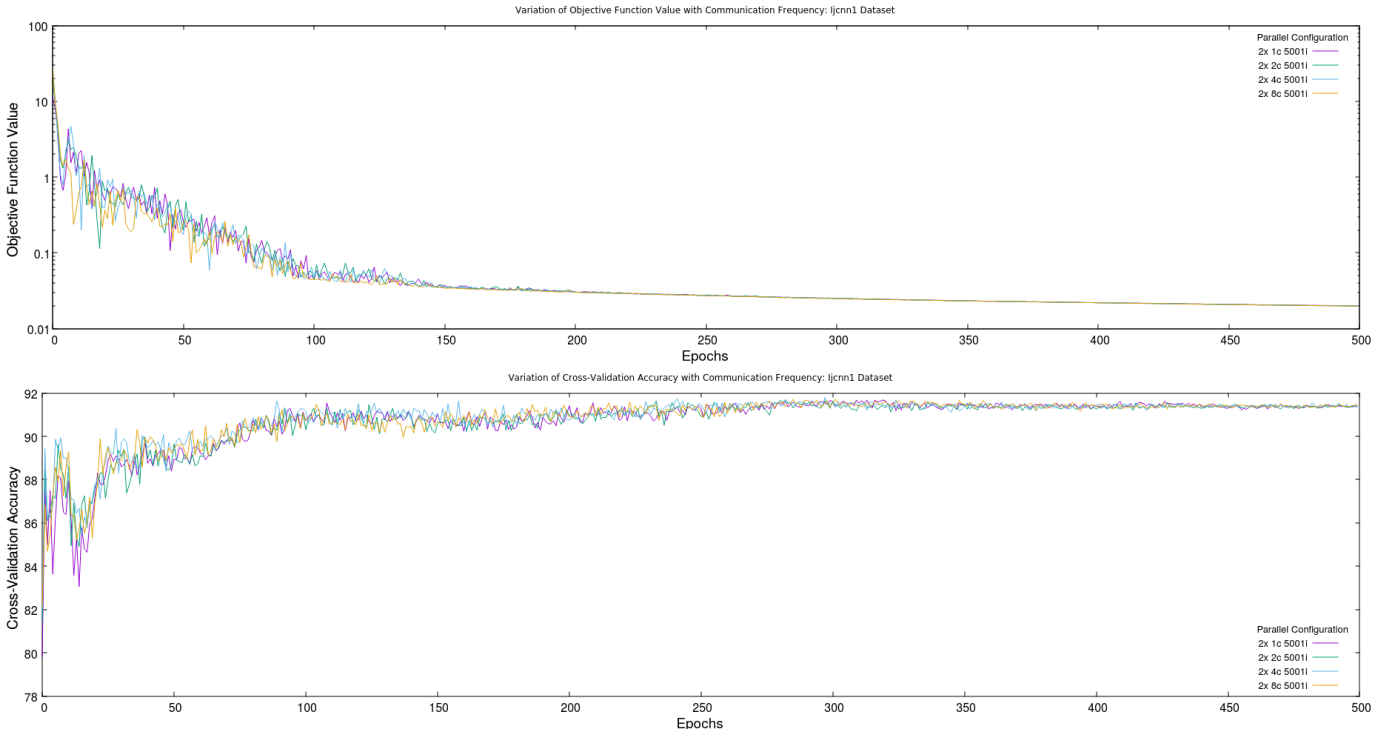

Fig. 152. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 2 and MSF = [1,2,4,8]

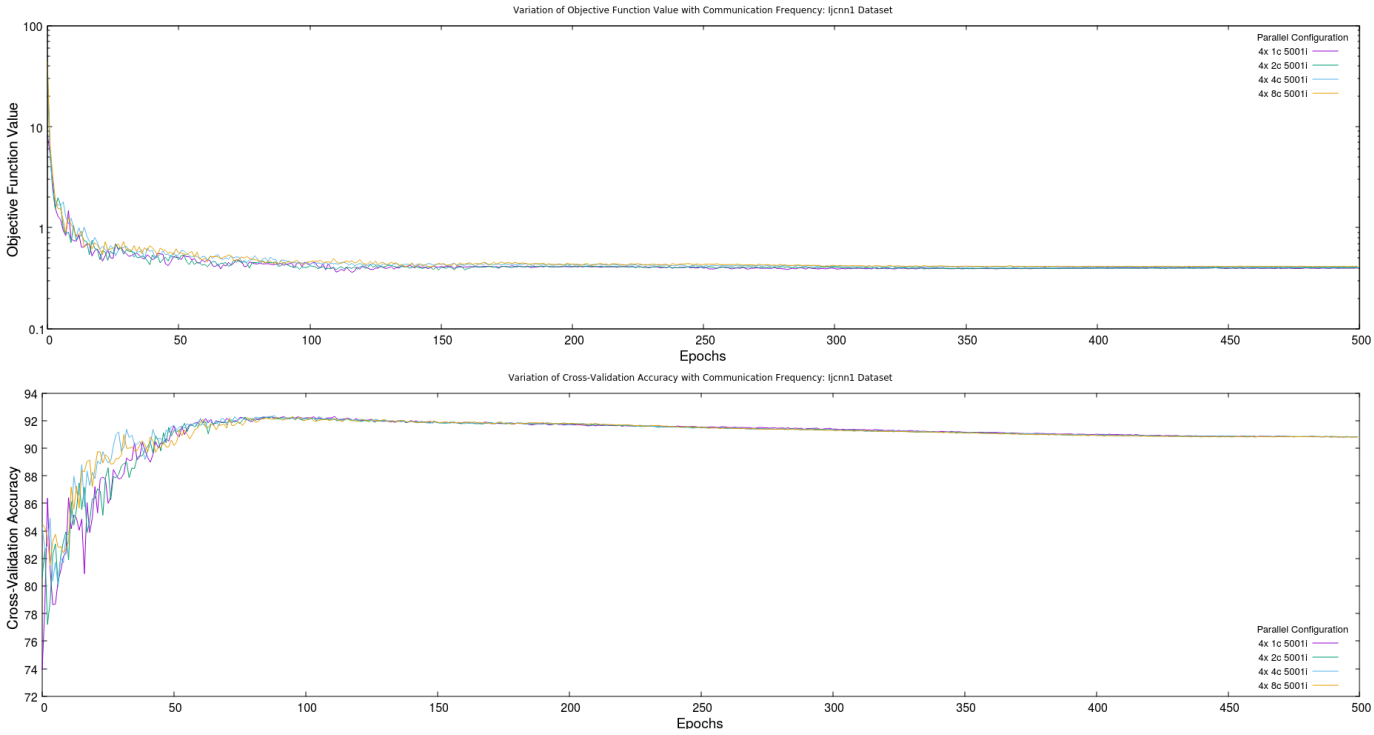

Fig. 153. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 4 and MSF = [1,2,4,8]

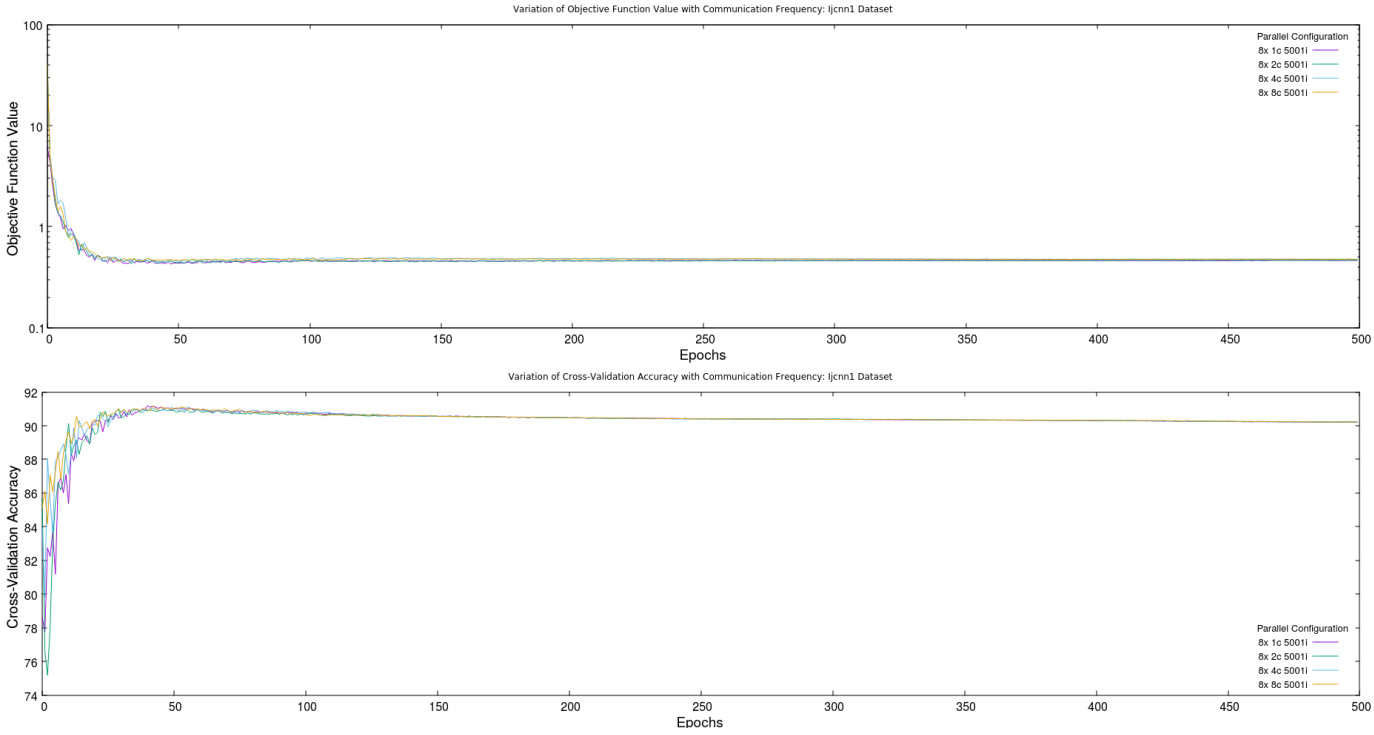

Fig. 154. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 8 and MSF = [1,2,4,8]

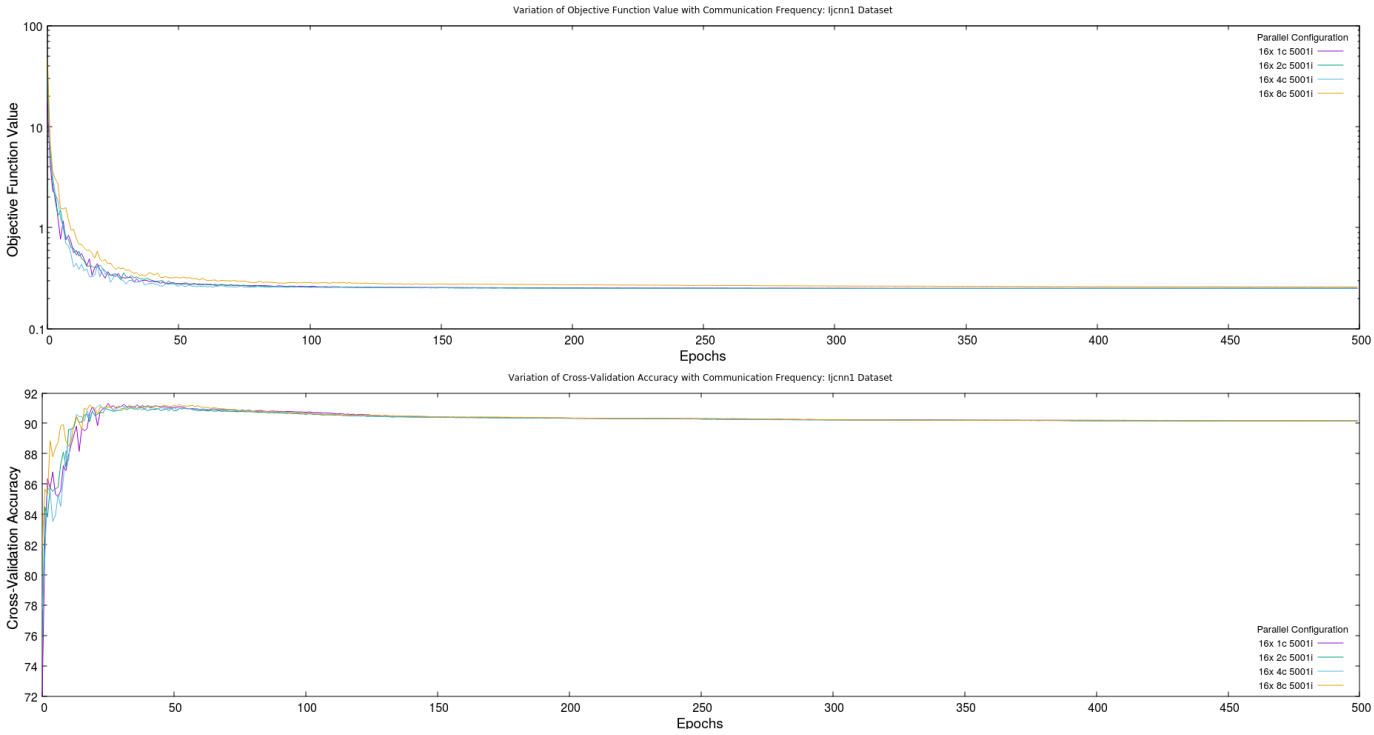

Fig. 155. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 16 and MSF = [1,2,4,8]

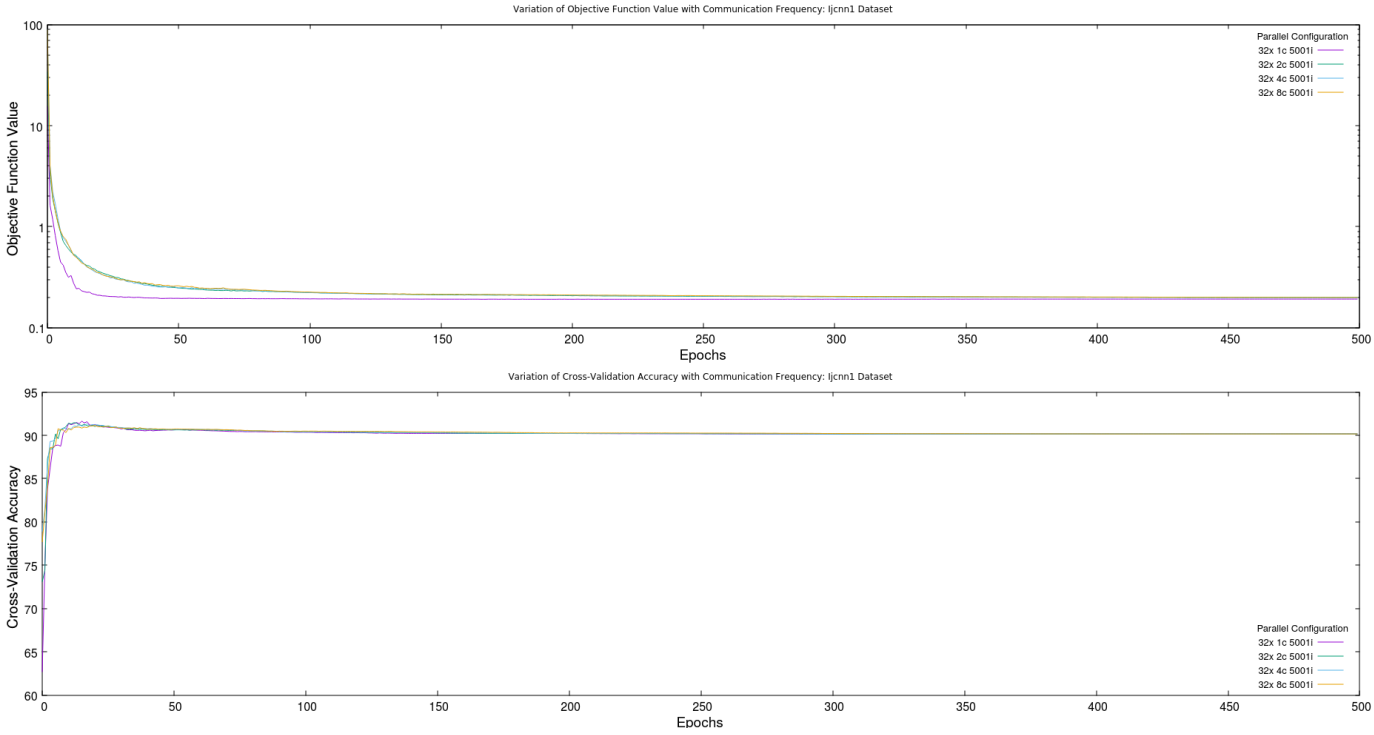

Fig. 156. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 32 and MSF = [1,2,4,8]

E. Ijcnn1 MSF = [128-4096] , Parallelism = [2,4,8,16,32]

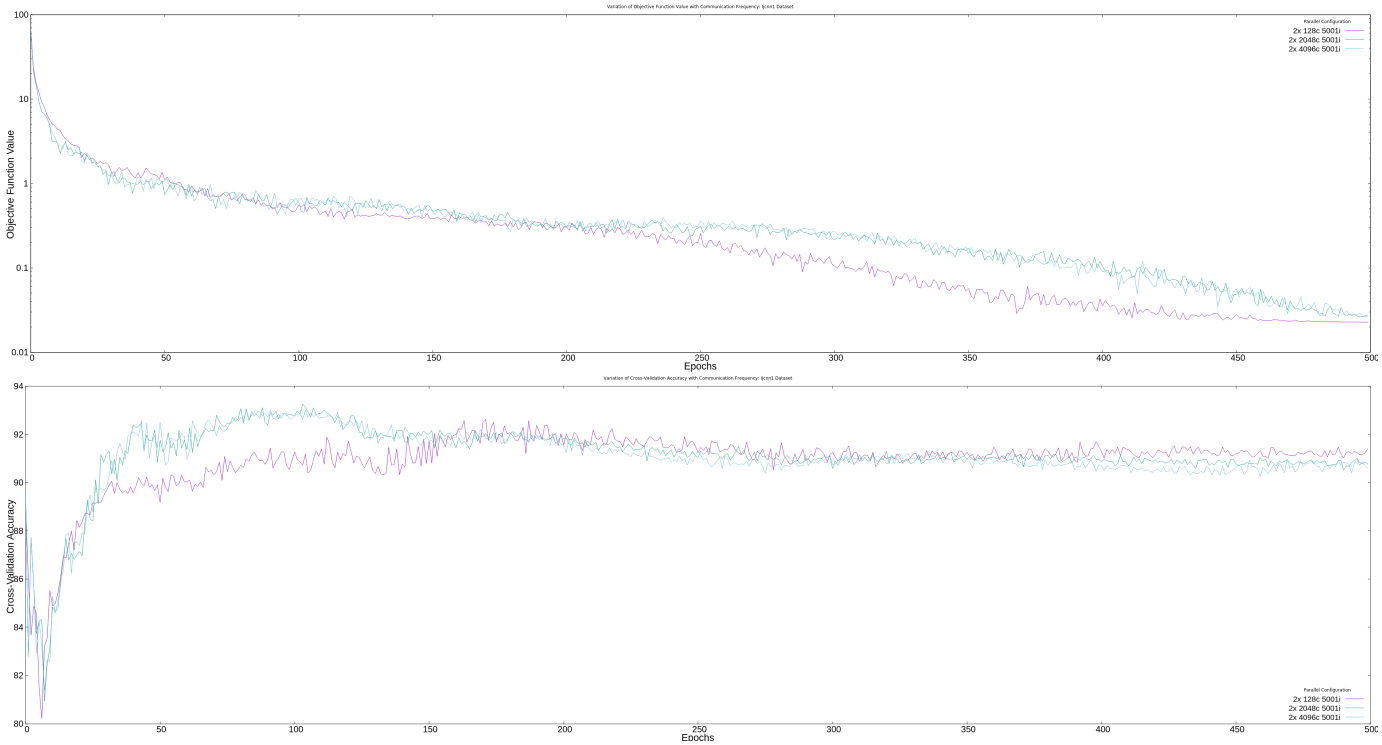

Fig. 157. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 2 and MSF = [128,2048,4096]

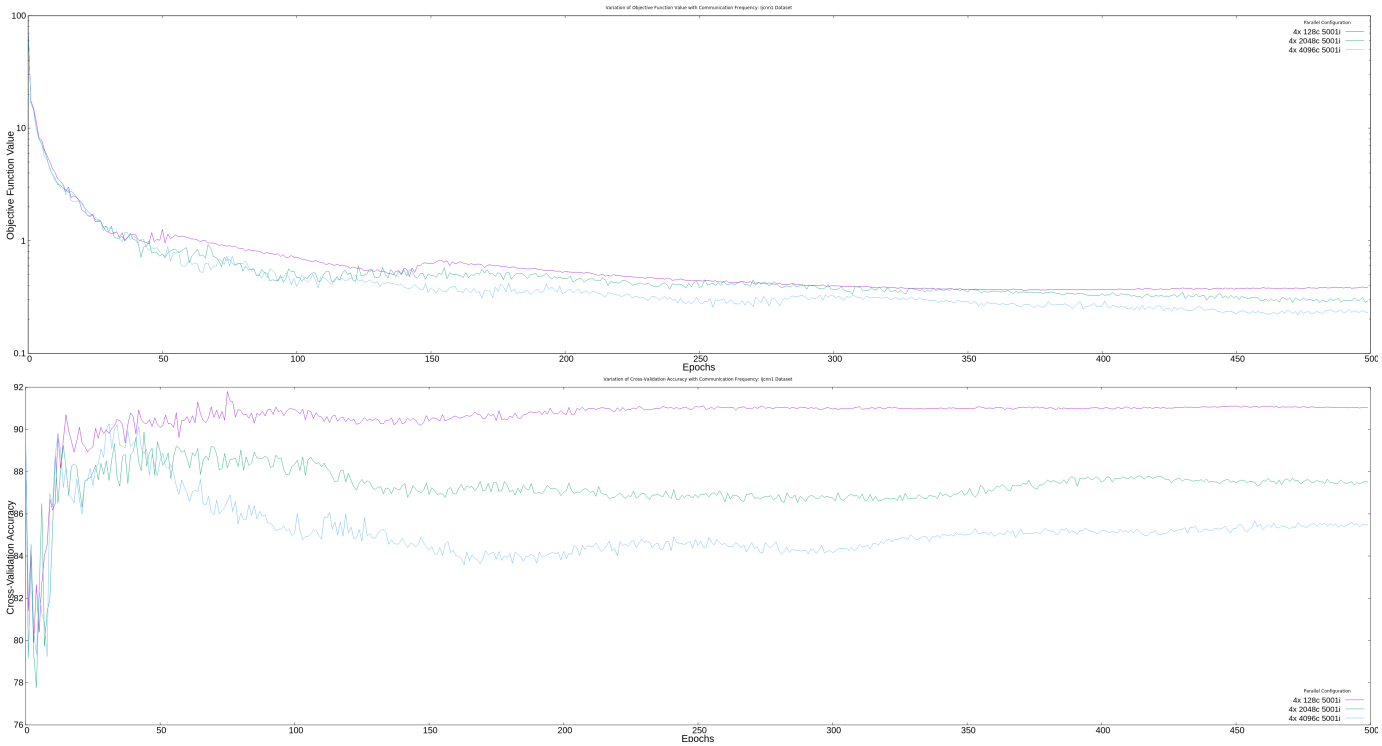

Fig. 158. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 4 and MSF = [128,2048,4096]

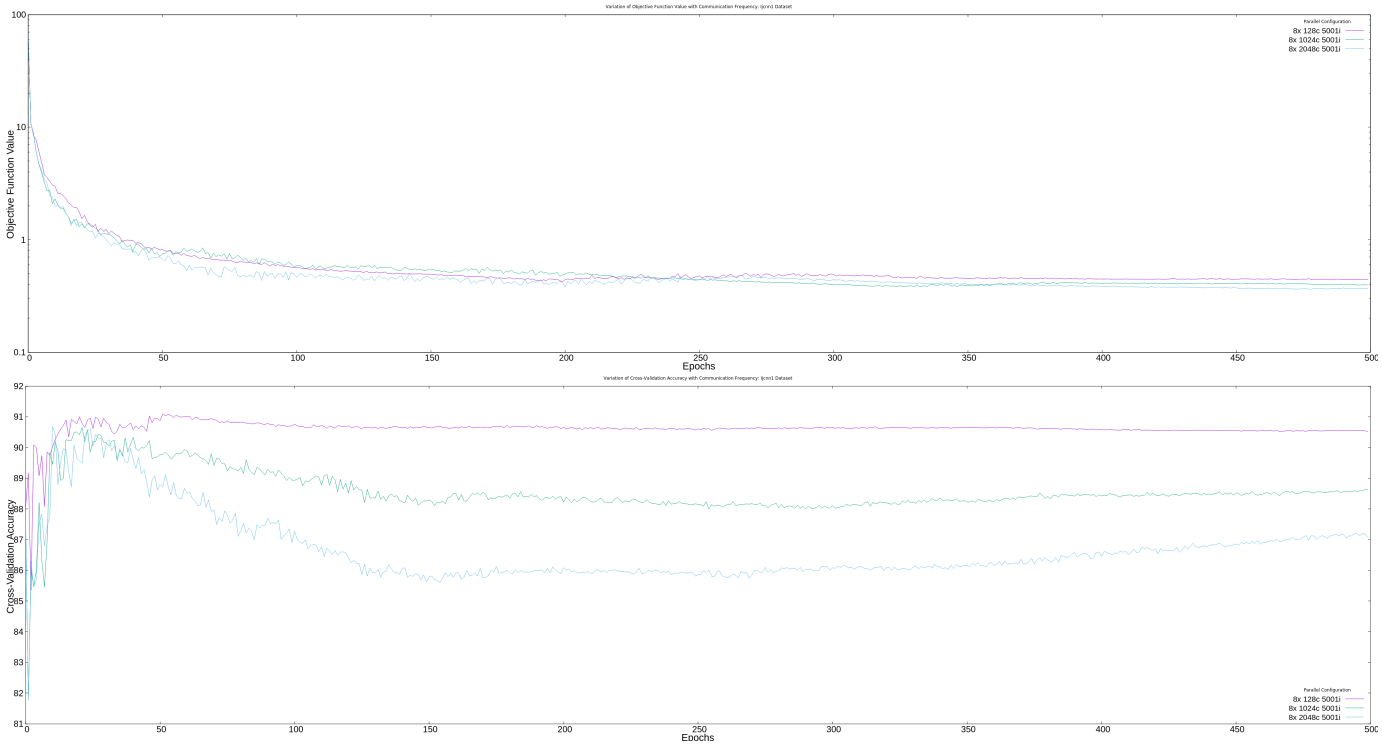

Fig. 159. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 8 and MSF = [128,1024,2048]

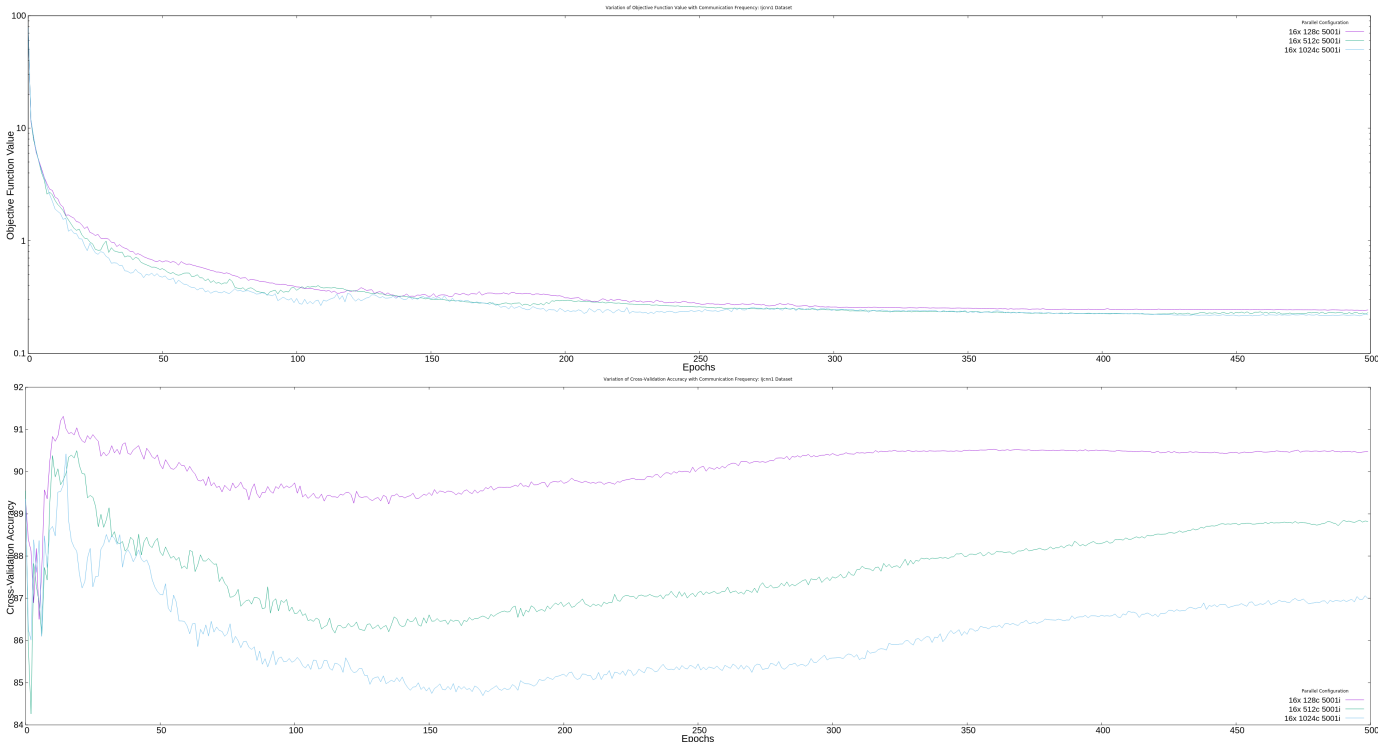

Fig. 160. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 16 and MSF = [128,512,1024]

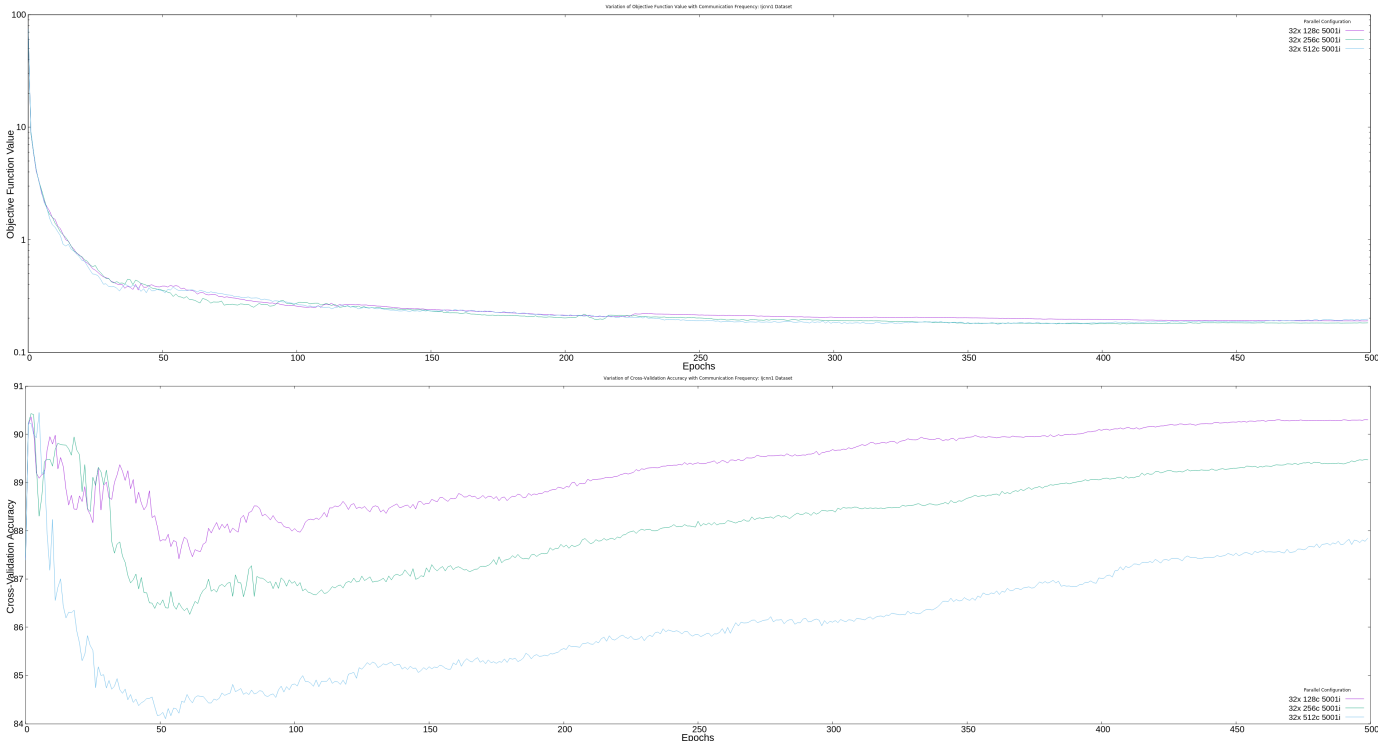

Fig. 161. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms 32 and MSF = [128,256,512]

F. Ijcnn1 MSF = 1,2,4,512 , Parallelism = [2,4,8,16,32]

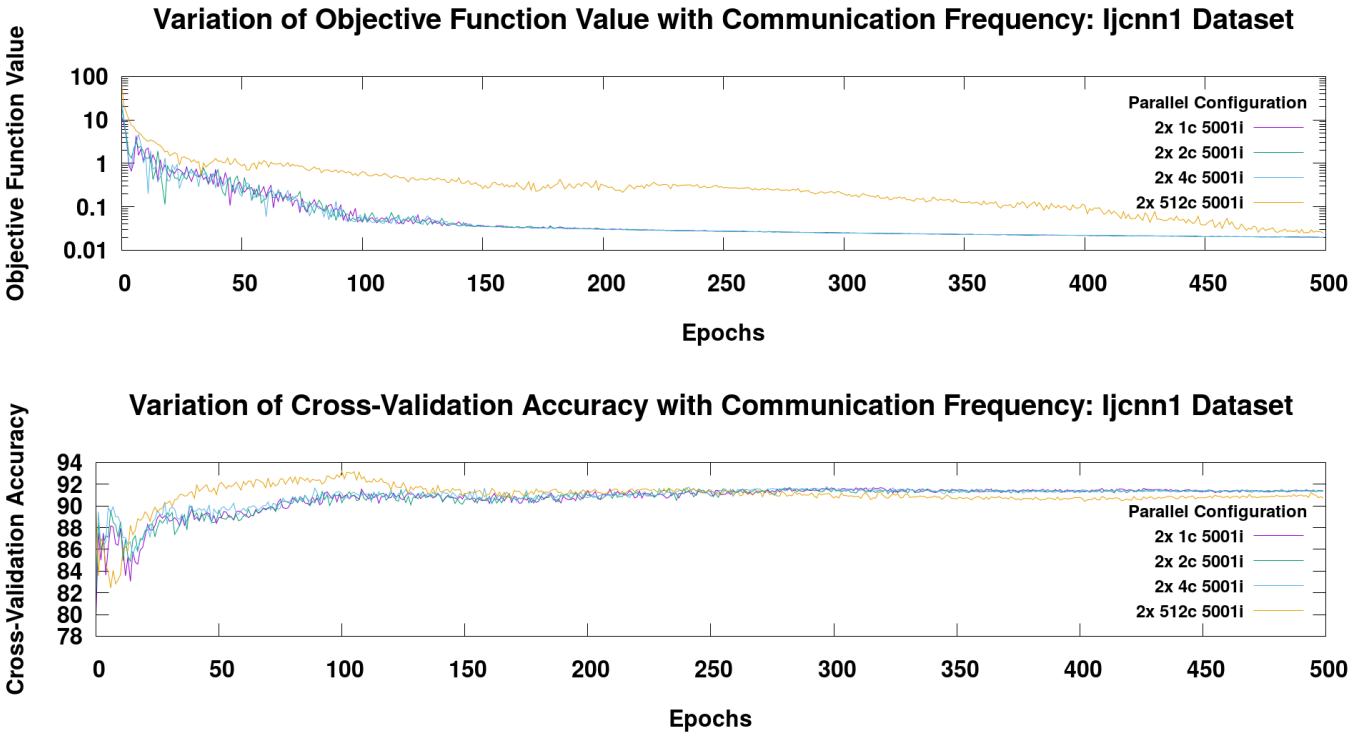

Fig. 162. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms [2] and MSF = [1,2,4,512]

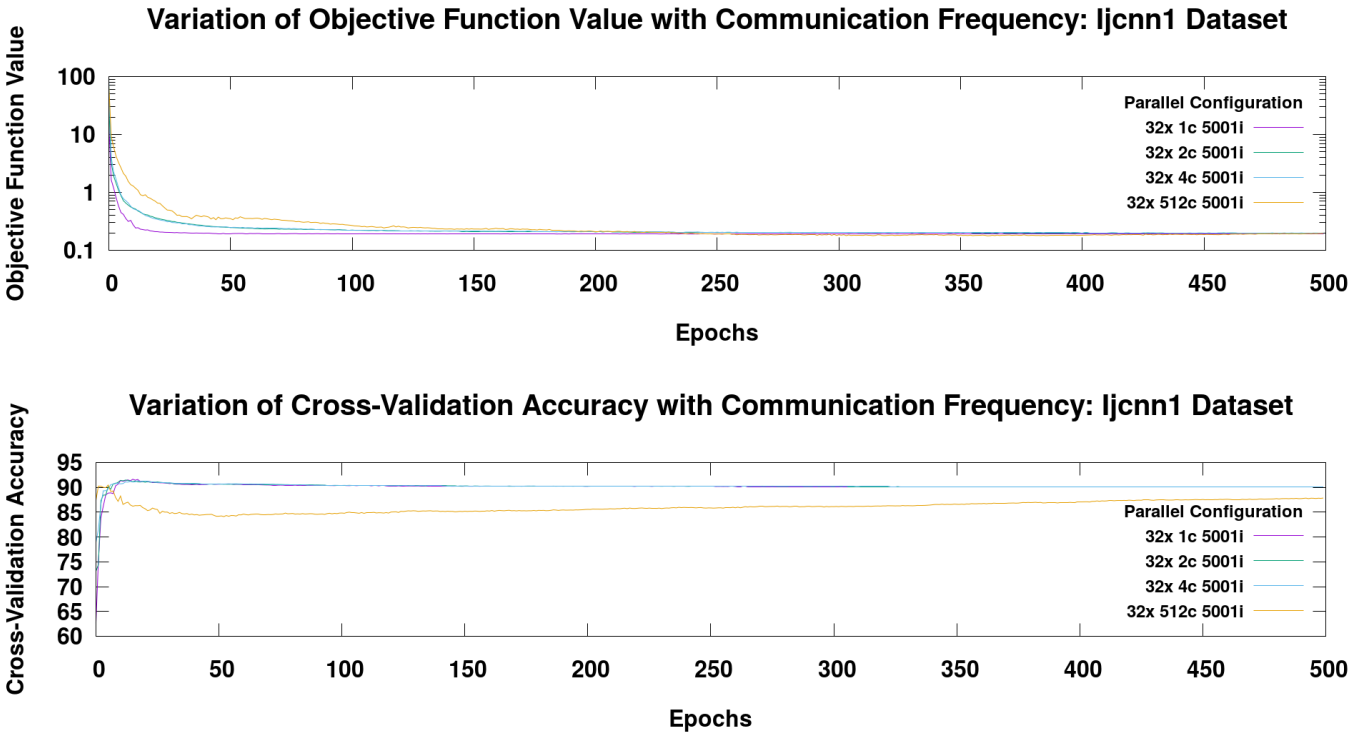

Fig. 163. Distributed Model Synchronizing Algorithm on Ijcnn1 Dataset for Parallelisms [32] and MSF = [1,2,4,512]

G. Webspam MSF = 1,2,4,8 , Parallelism = [2,4,8,16,32]

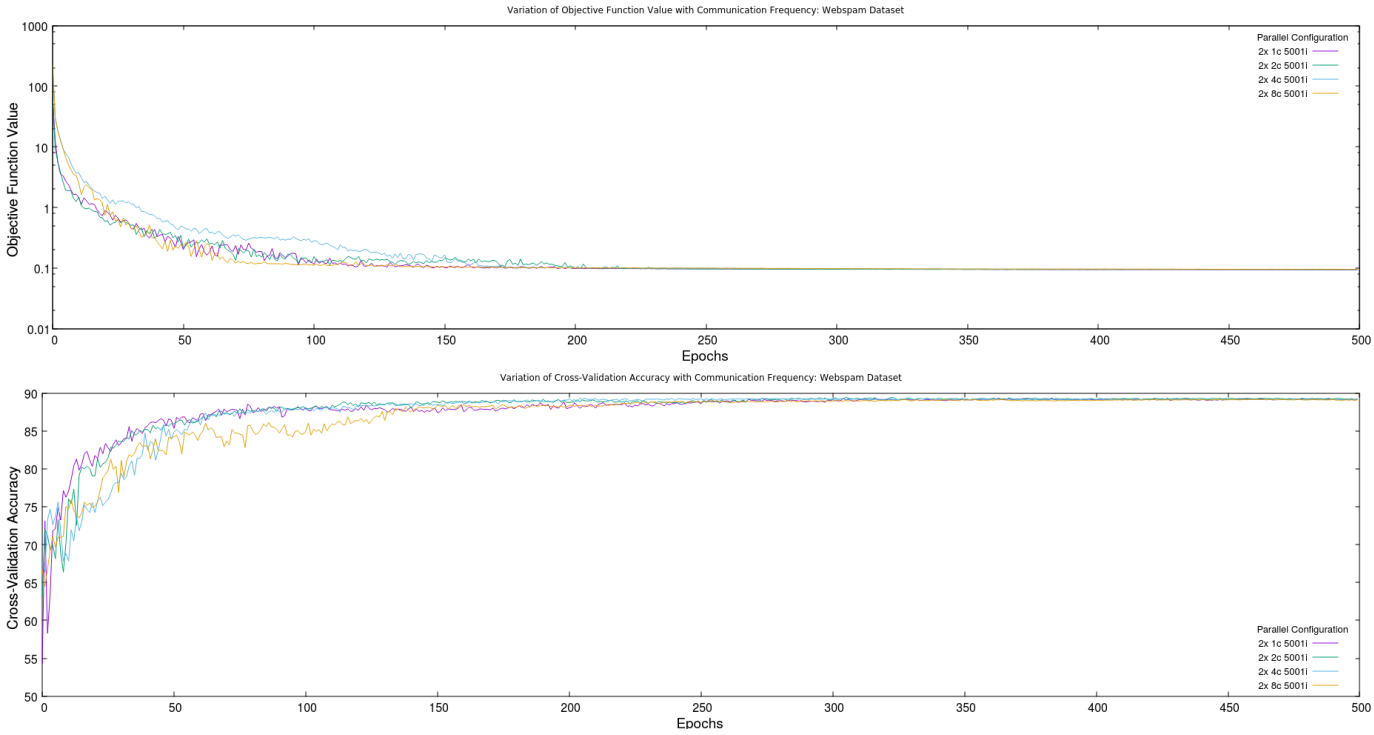

Fig. 164. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 2 and MSF = [1,2,4,8]

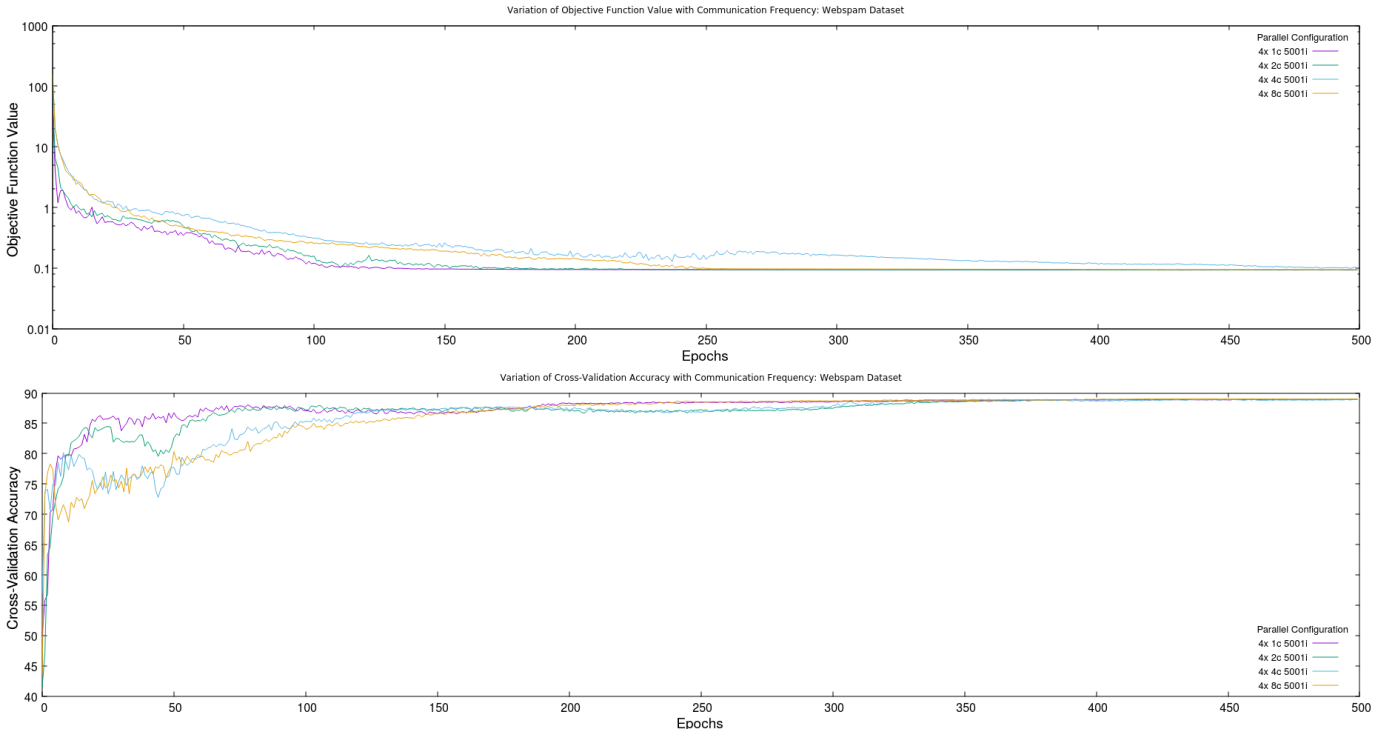

Fig. 165. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 4 and MSF = [1,2,4,8]

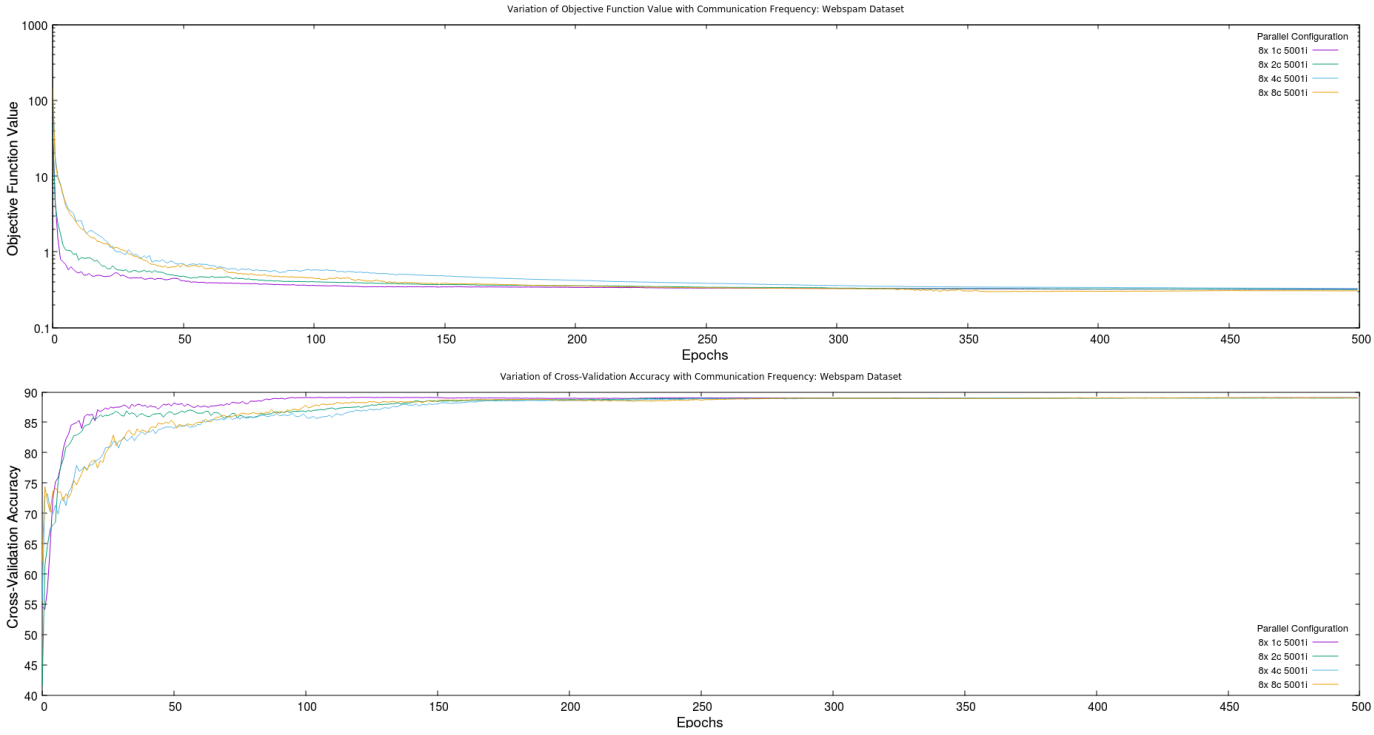

Fig. 166. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 8 and MSF = [1,2,4,8]

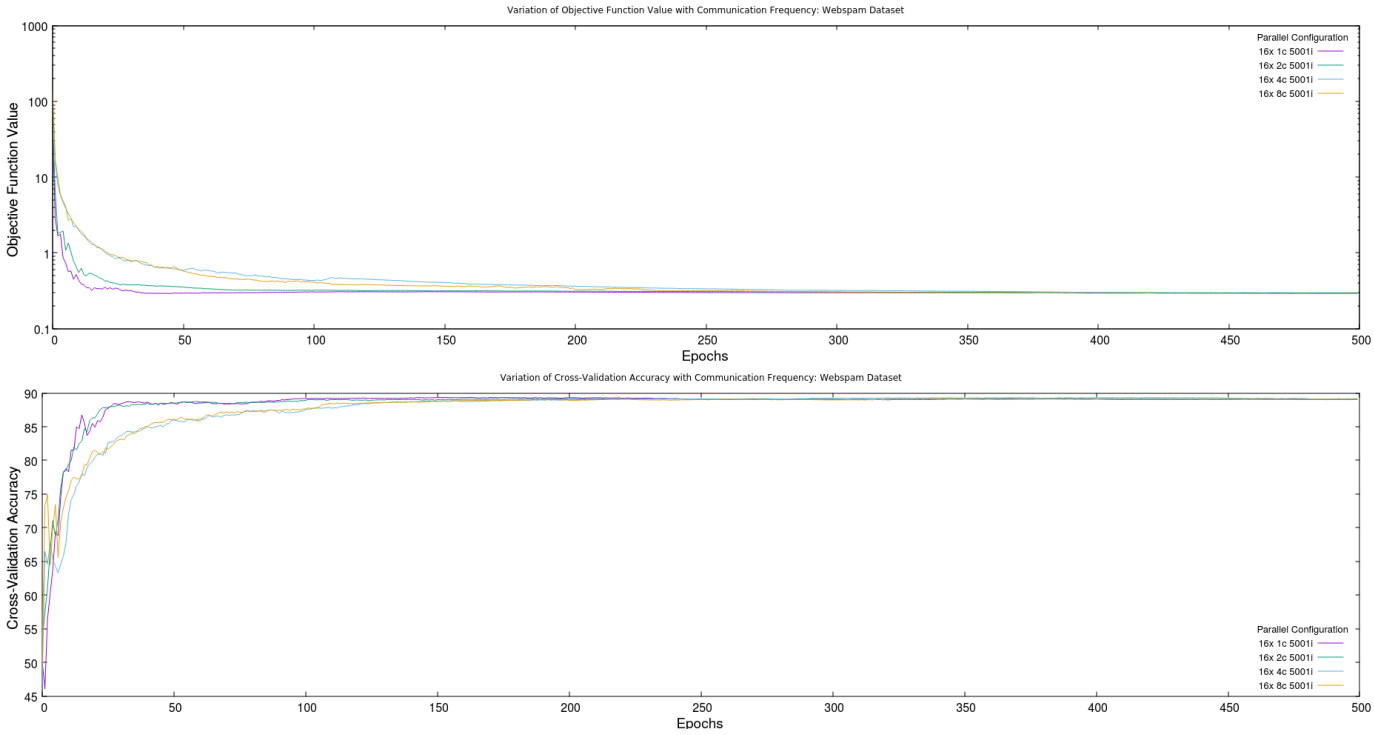

Fig. 167. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 16 and MSF = [1,2,4,8]

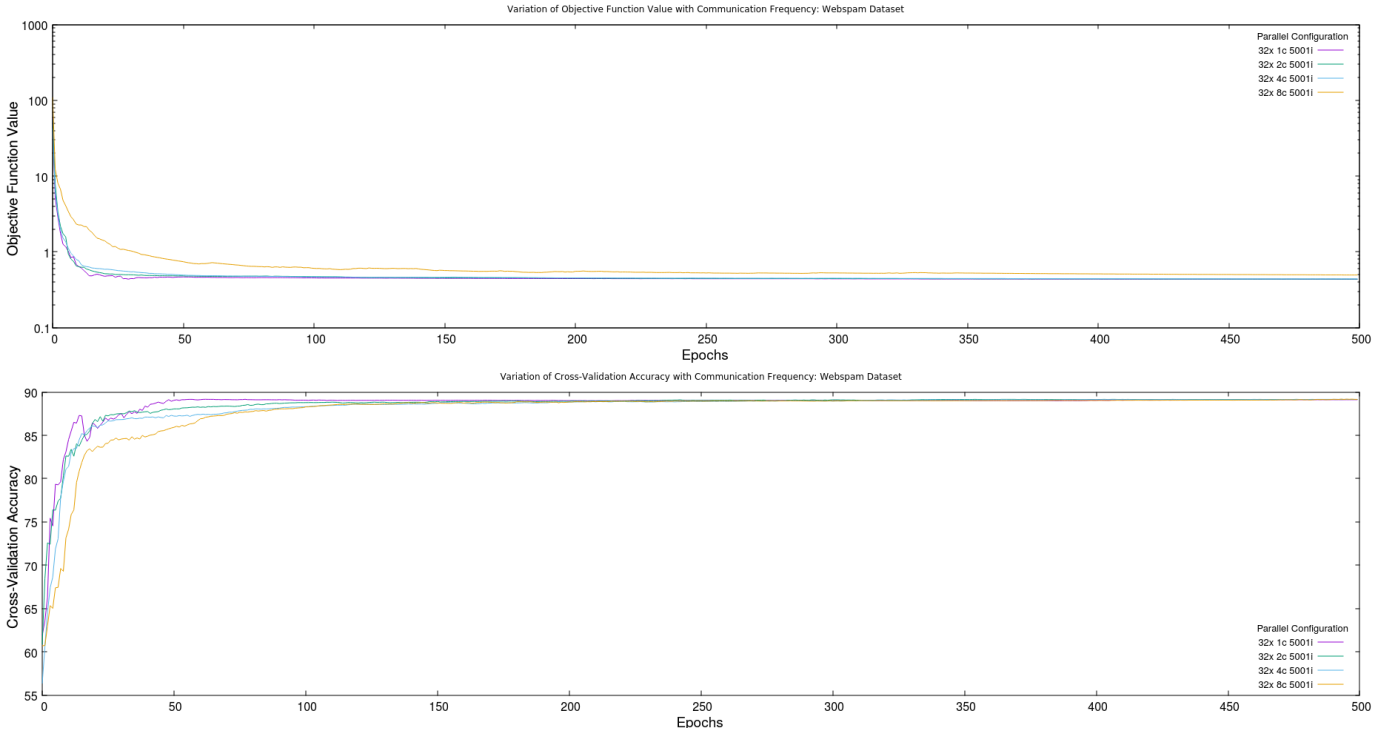

Fig. 168. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 32 and MSF = [1,2,4,8]

H. Webspam MSF = [128-4096] , Parallelism = [2,4,8,16,32]

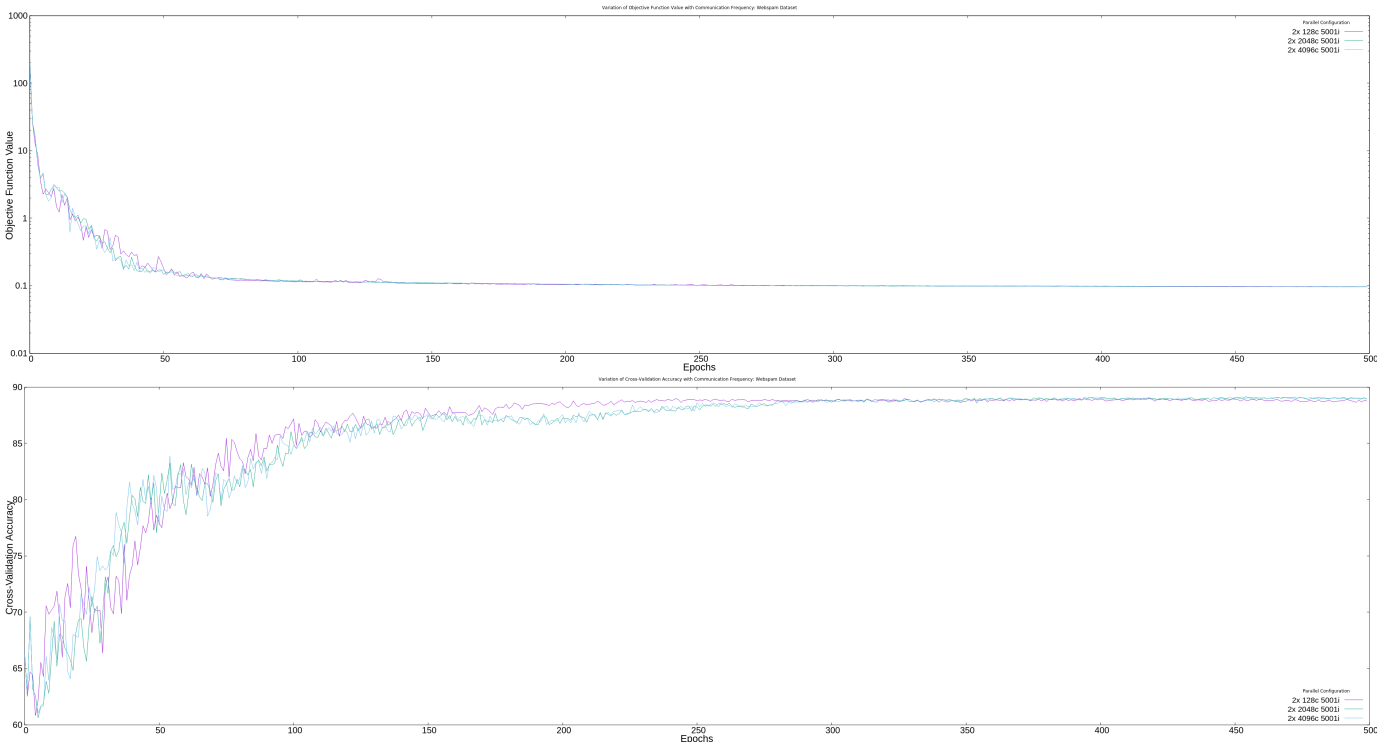

Fig. 169. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 2 and MSF = [128,2048,4096]

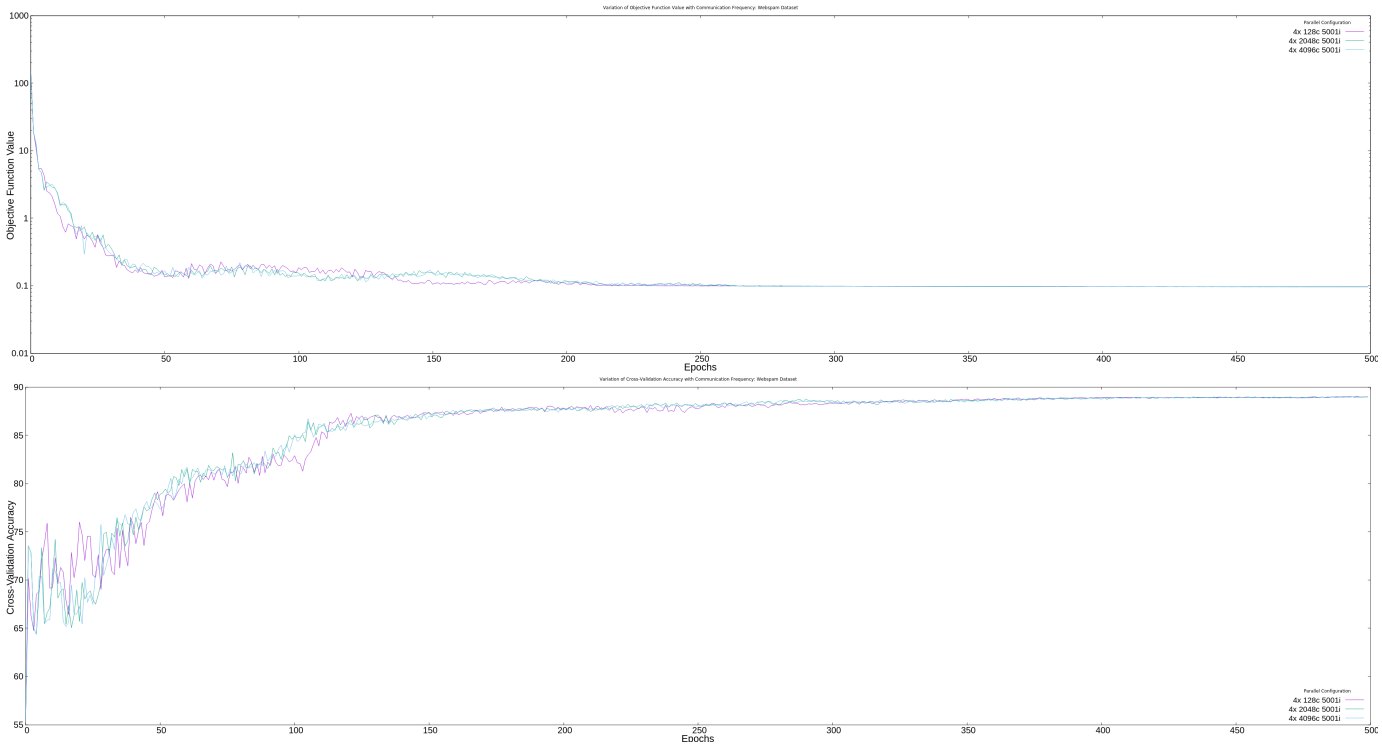

Fig. 170. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 4 and MSF = [128,2048,4096]

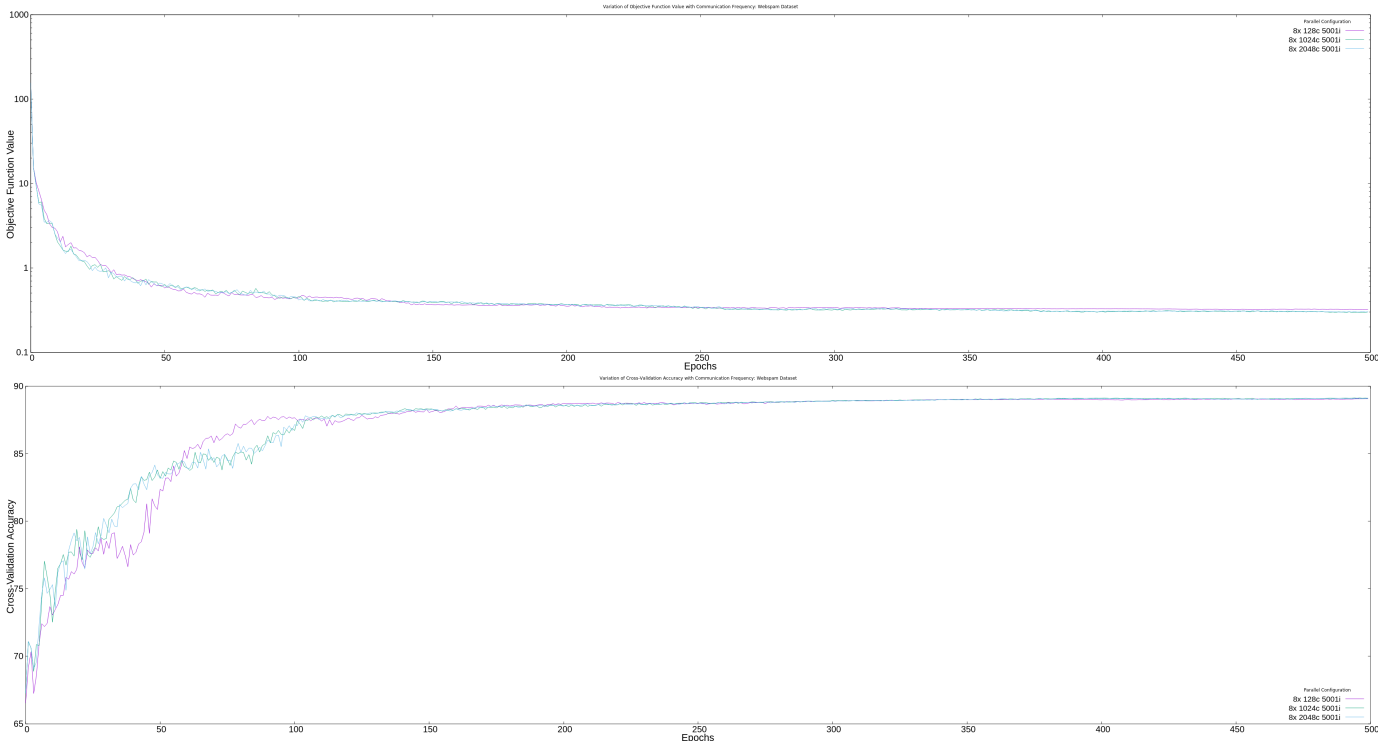

Fig. 171. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 8 and MSF = [128,1024,2048]

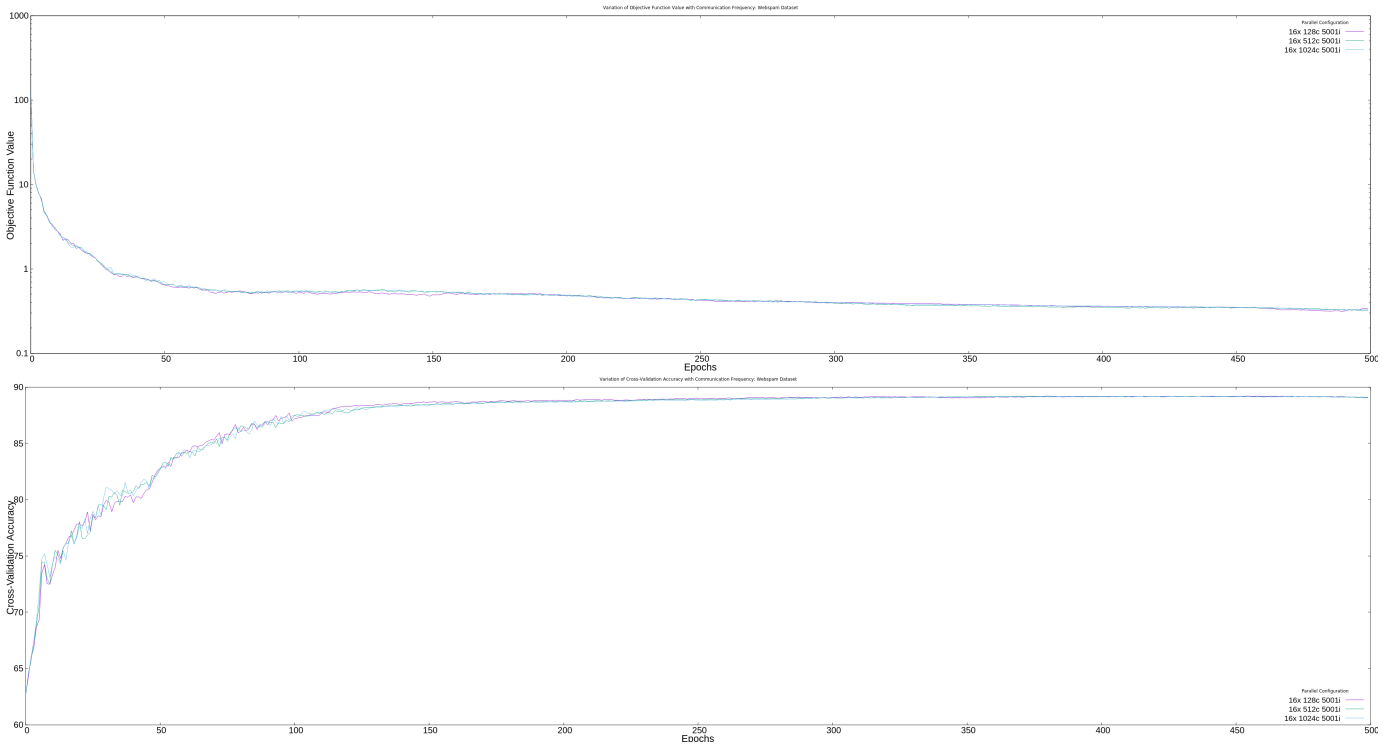

Fig. 172. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 16 and MSF = [128,512,1024]

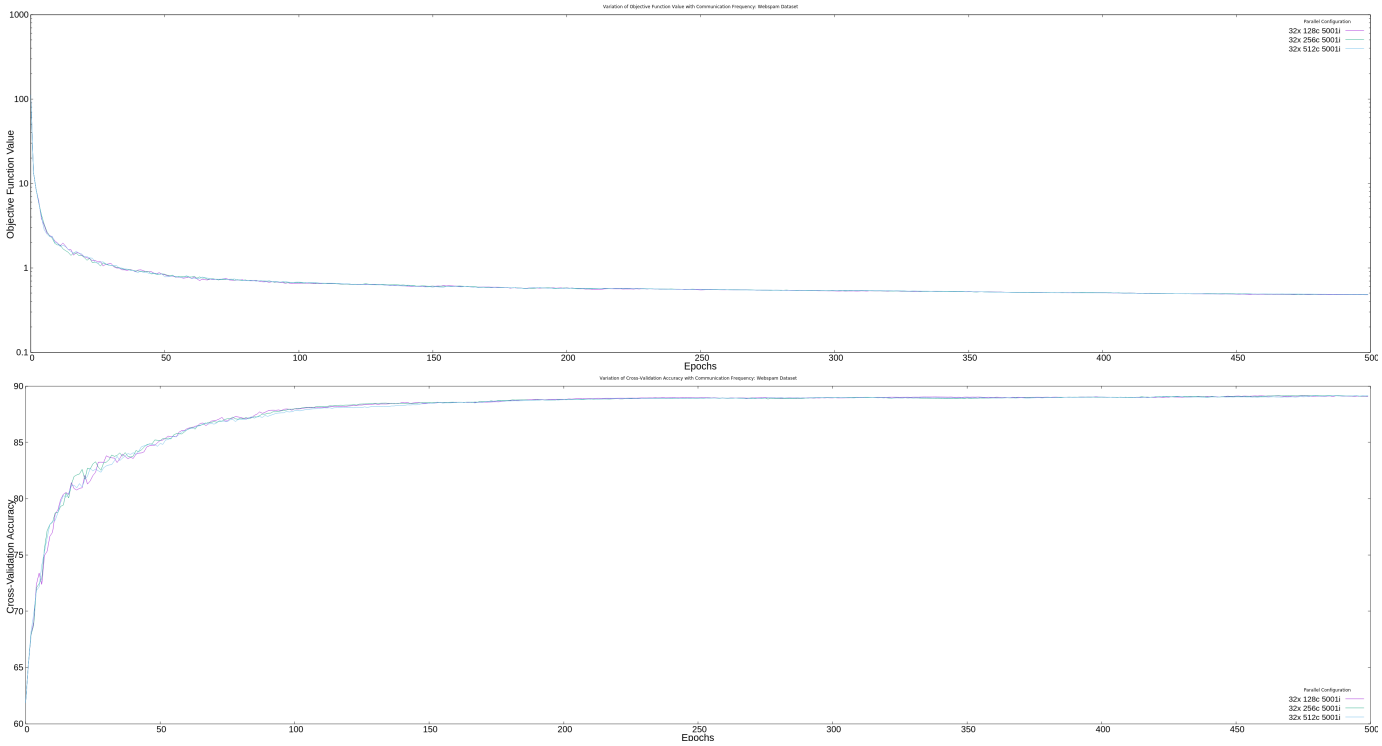

Fig. 173. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms 32 and MSF = [128,256,512]

I. Webspam MSF = 1,2,4,512 , Parallelism = [2,4,8,16,32]

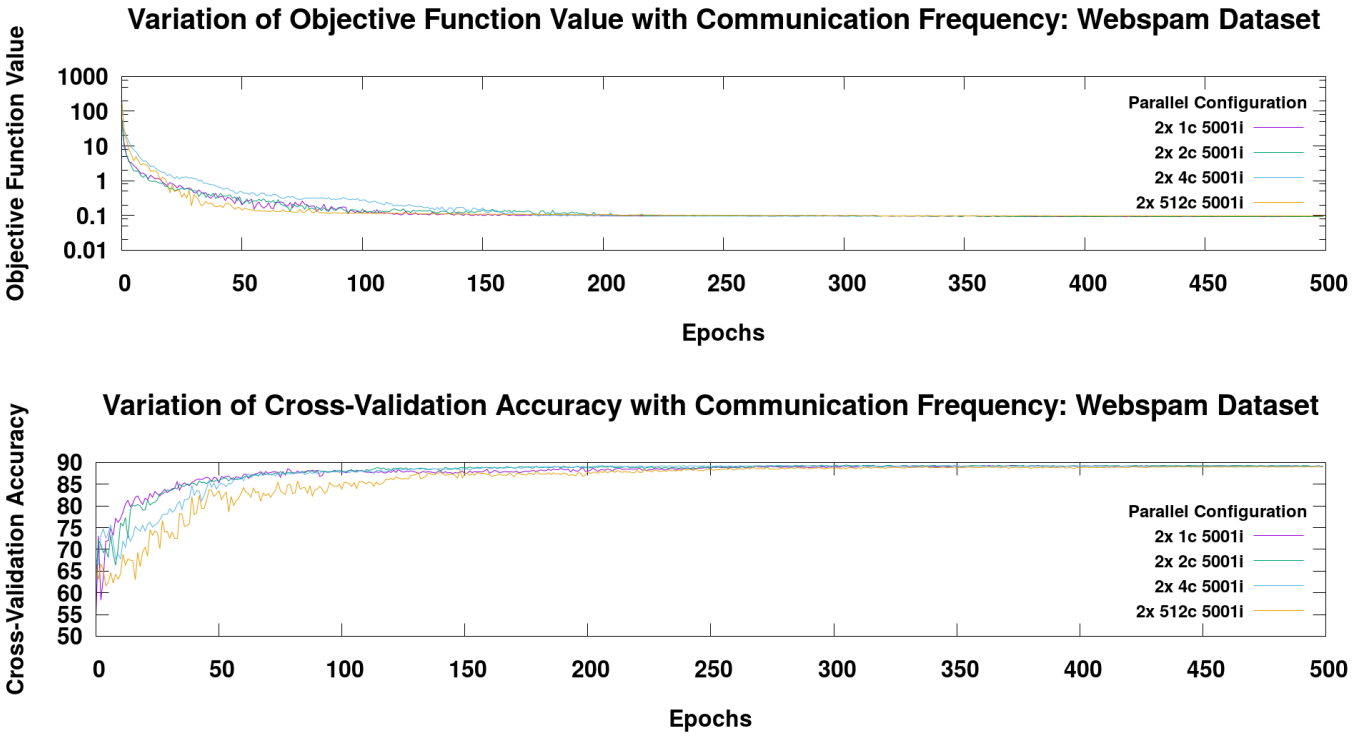

Fig. 174. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms [2] and MSF = [1,2,4,512]

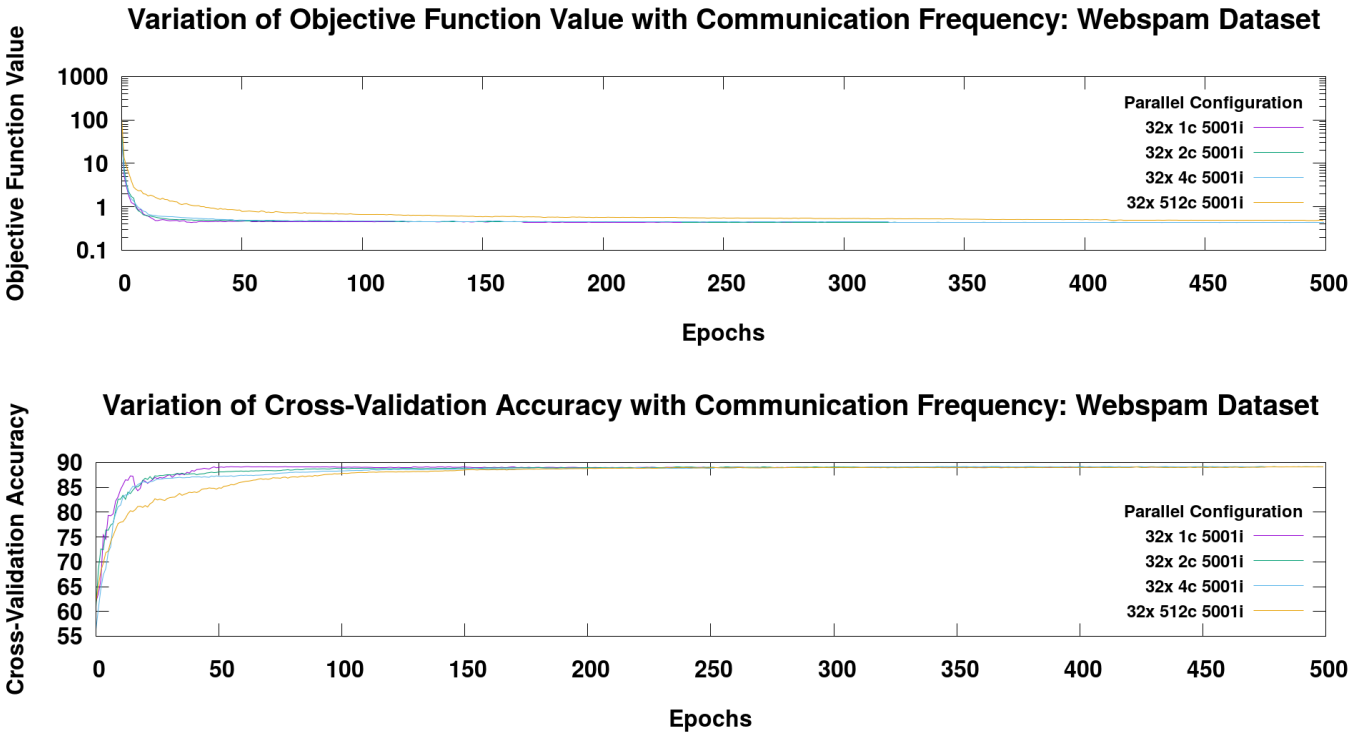

Fig. 175. Distributed Model Synchronizing Algorithm on Webspam Dataset for Parallelisms [32] and MSF = [1,2,4,512]

J. Epsilon MSF = 1,2,4,8 , Parallelism = [2,4,8,16,32]

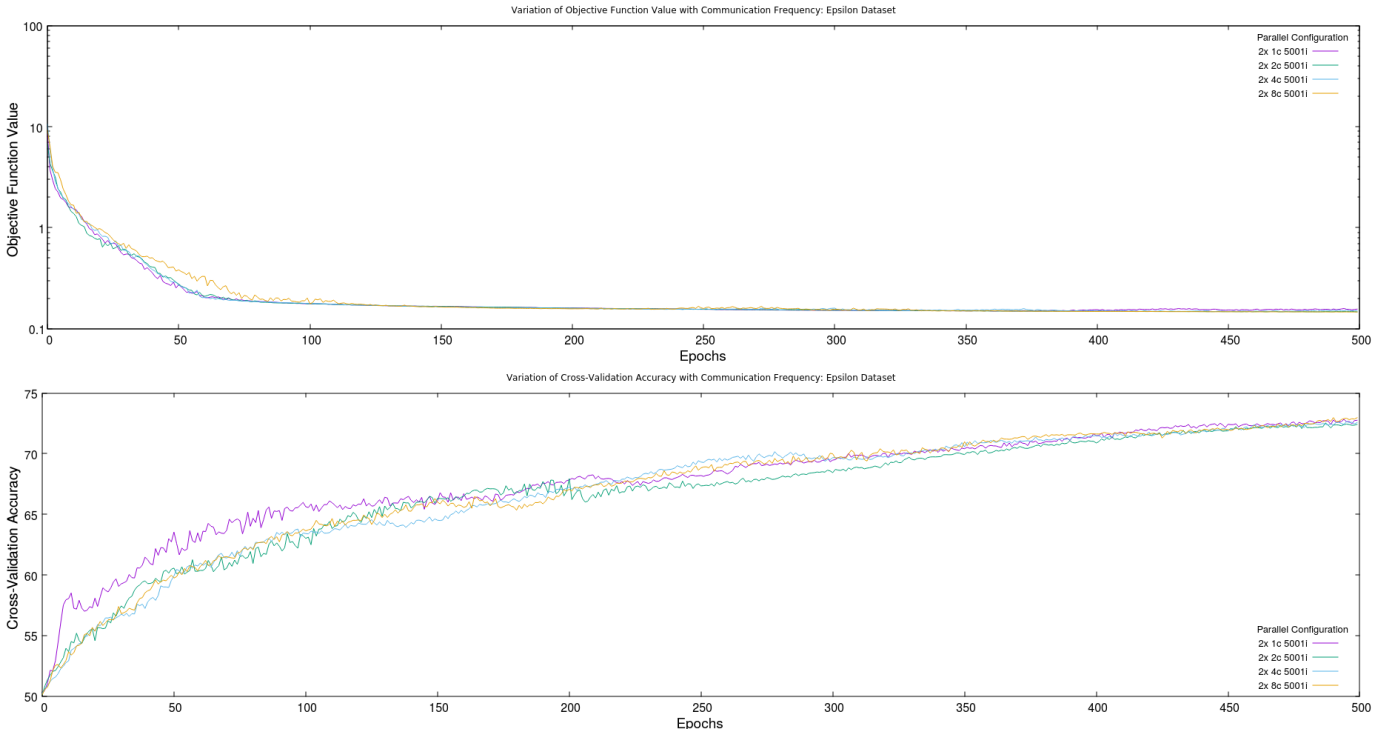

Fig. 176. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [2] and MSF = [1,2,4,8]

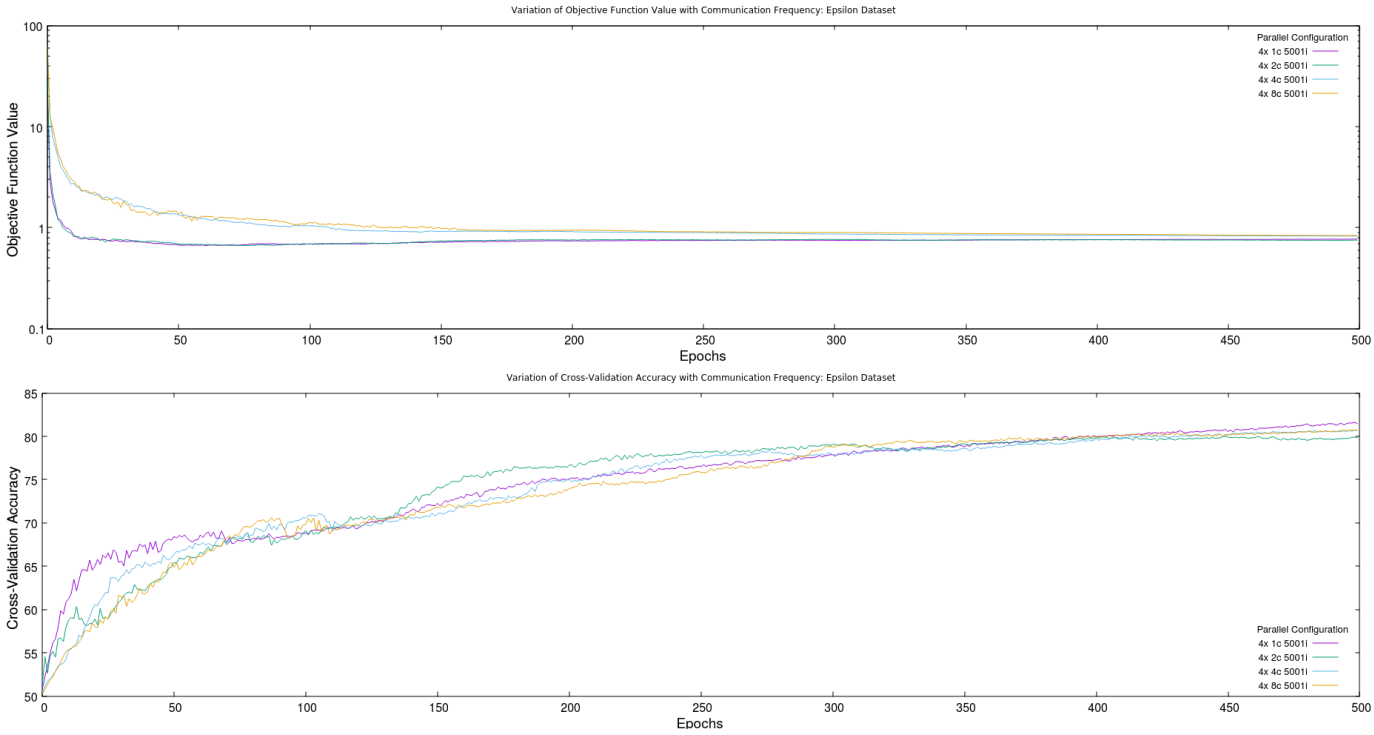

Fig. 177. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [4] and MSF = [1,2,4,8]

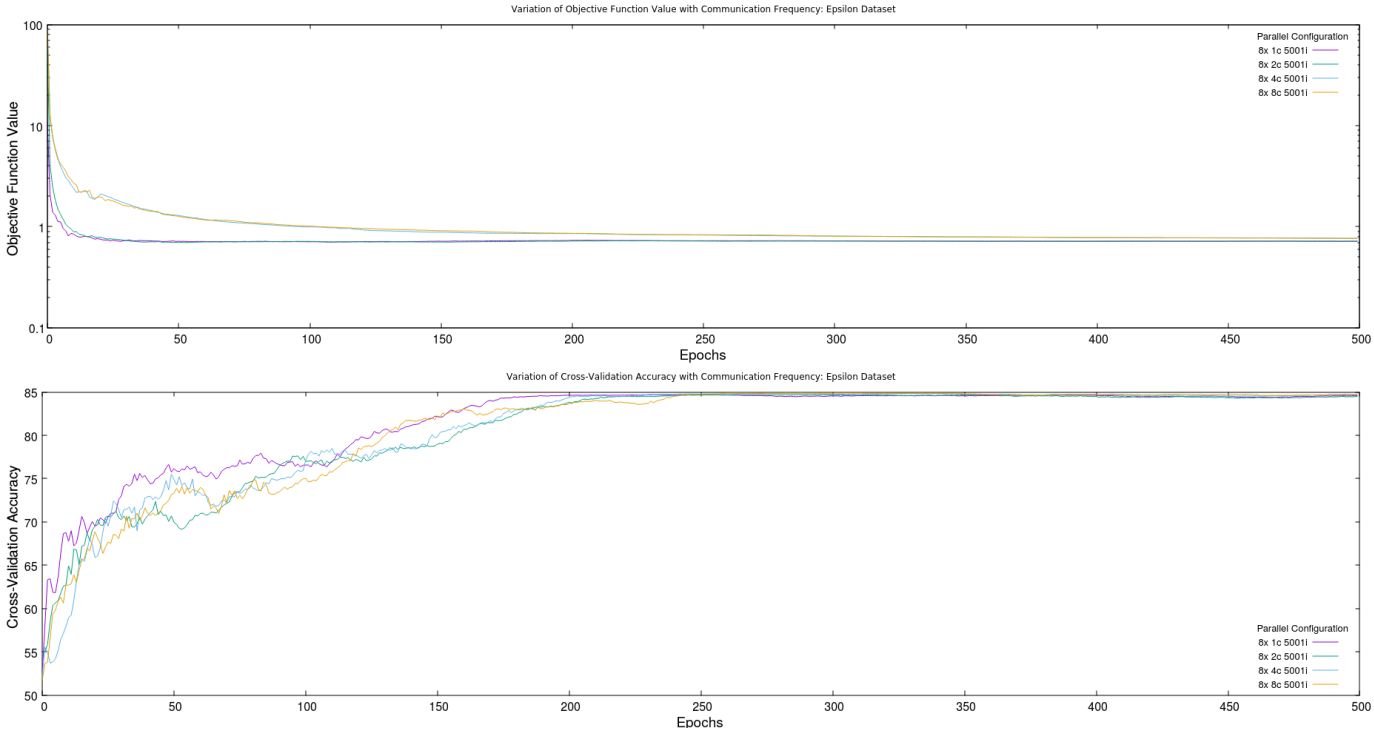

Fig. 178. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [8] and MSF = [1,2,4,8]

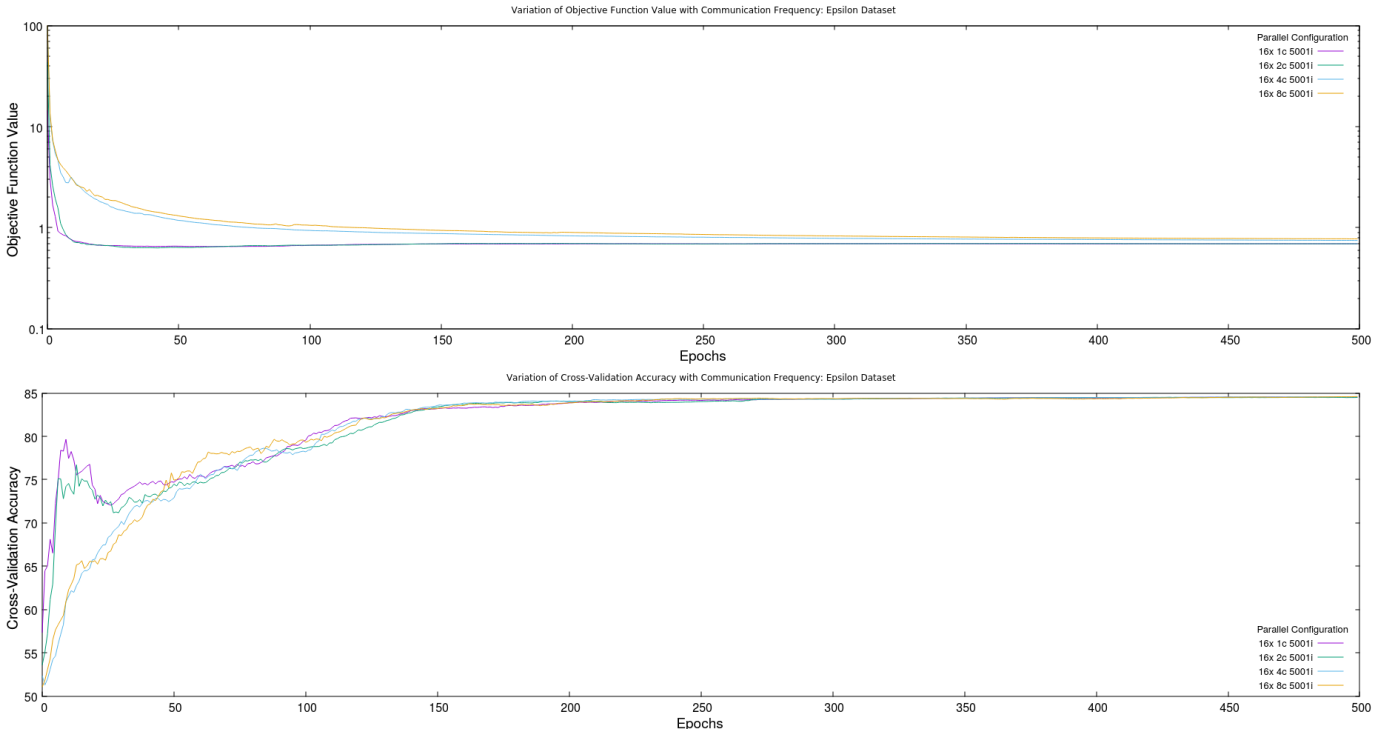

Fig. 179. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [16] and MSF = [1,2,4,8]

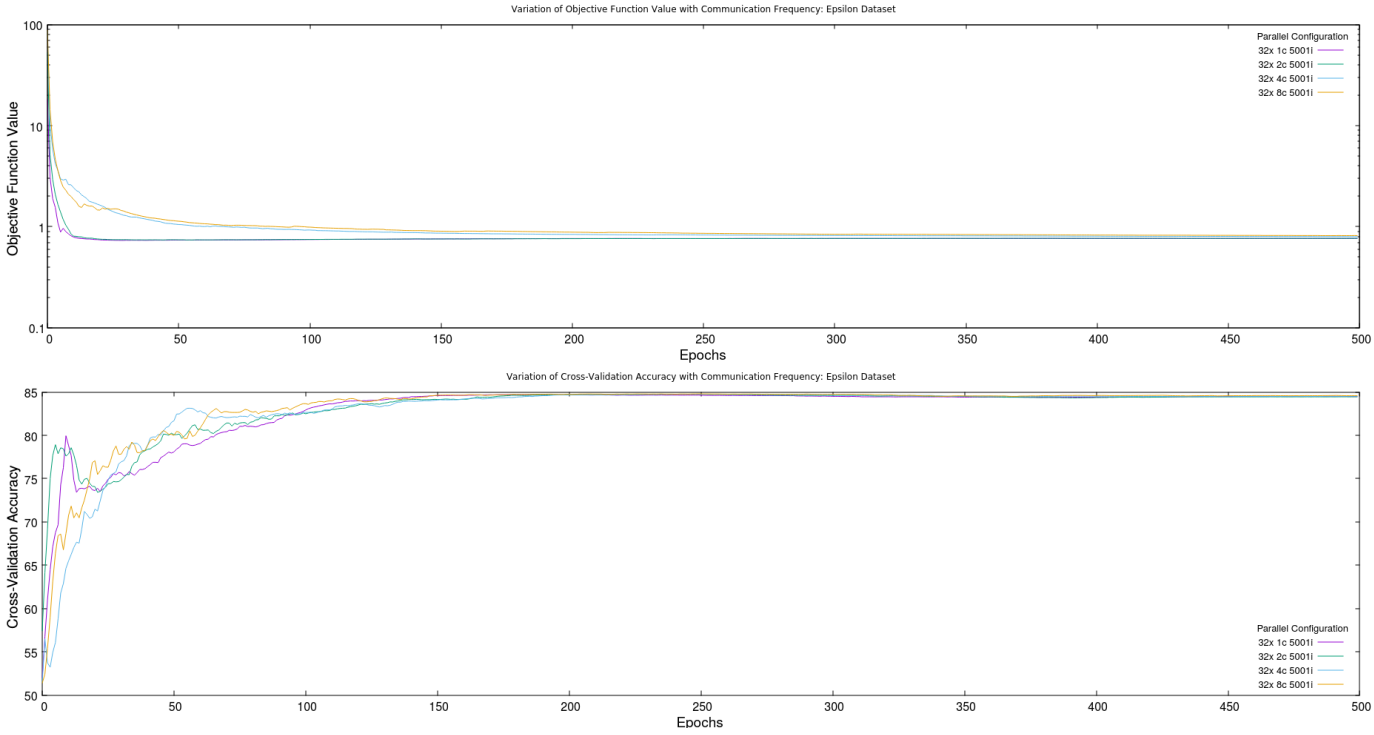

Fig. 180. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [32] and MSF = [1,2,4,8]

K. Epsilon MSF = 1,2,4,512 , Parallelism = [2,4,8,16,32]

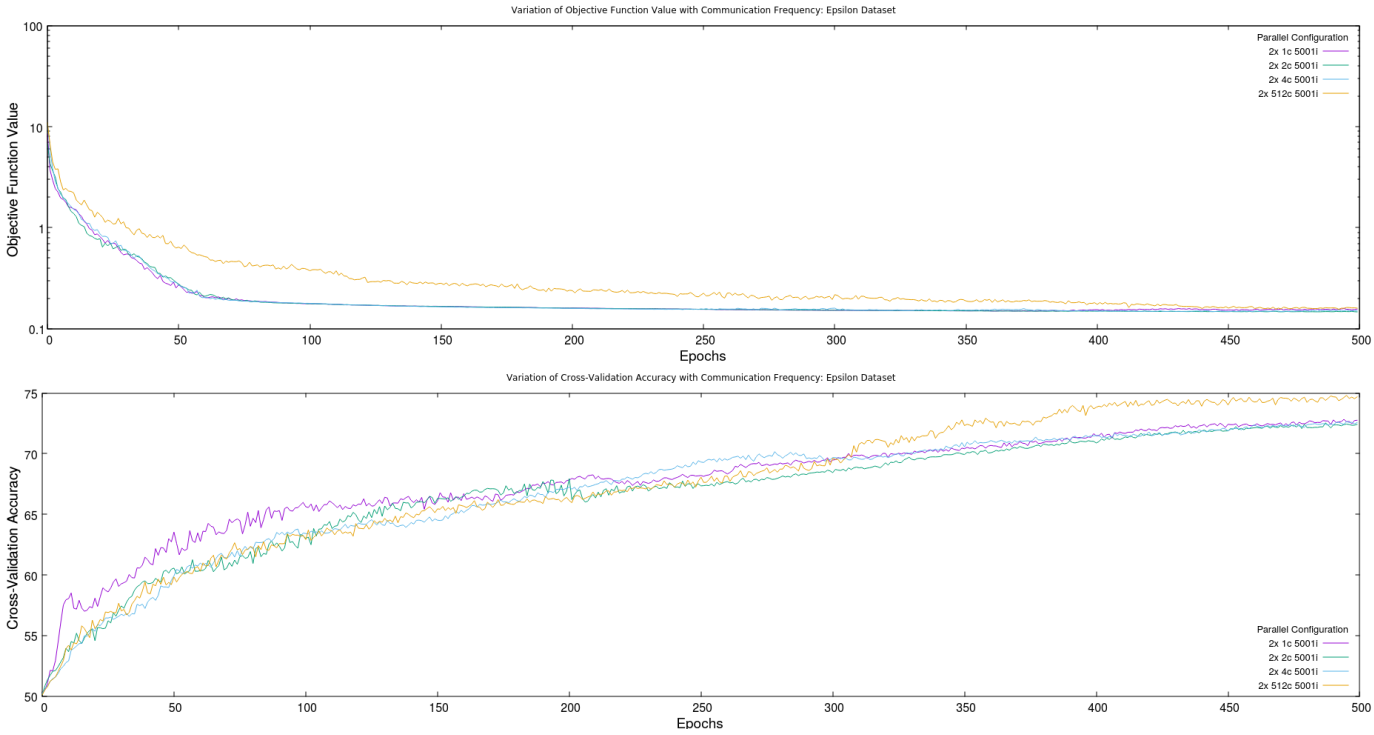

Fig. 181. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [2] and MSF = [1,2,4,512]

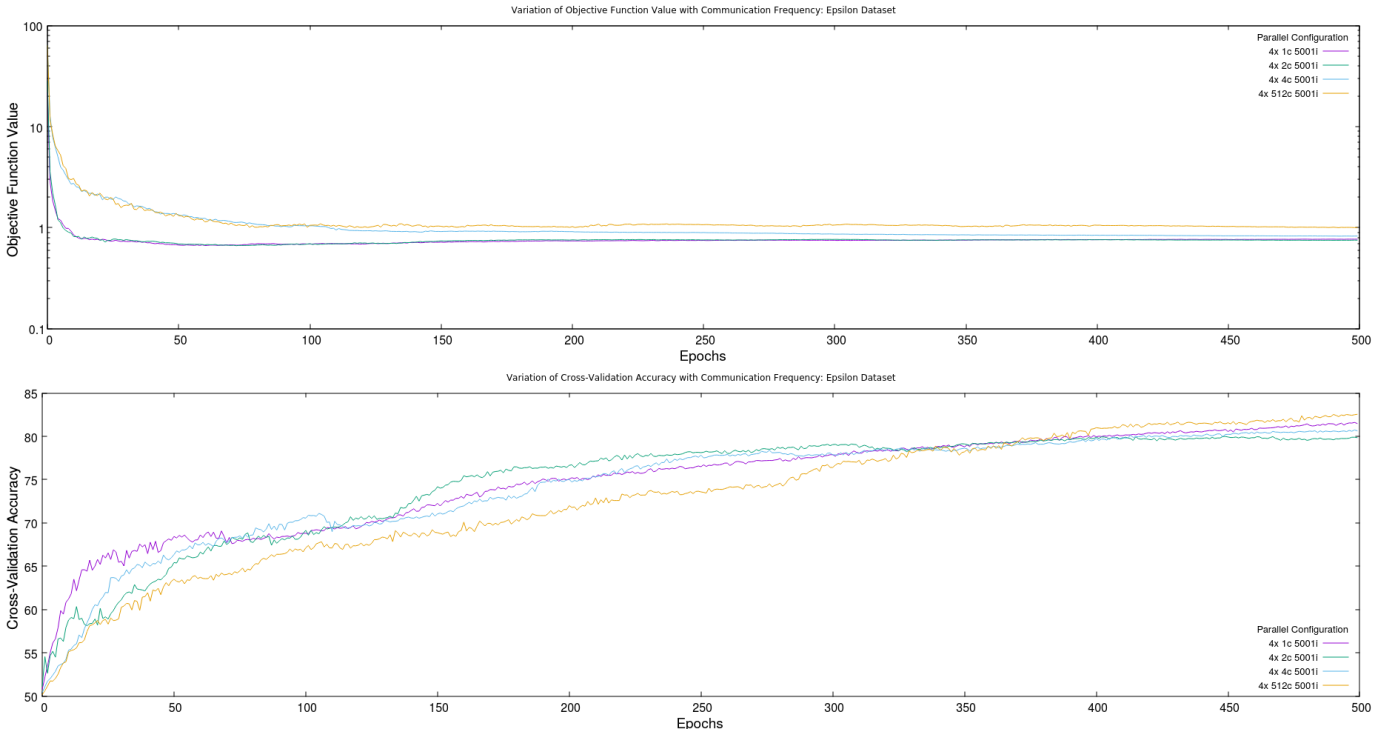

Fig. 182. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [4] and MSF = [1,2,4,512]

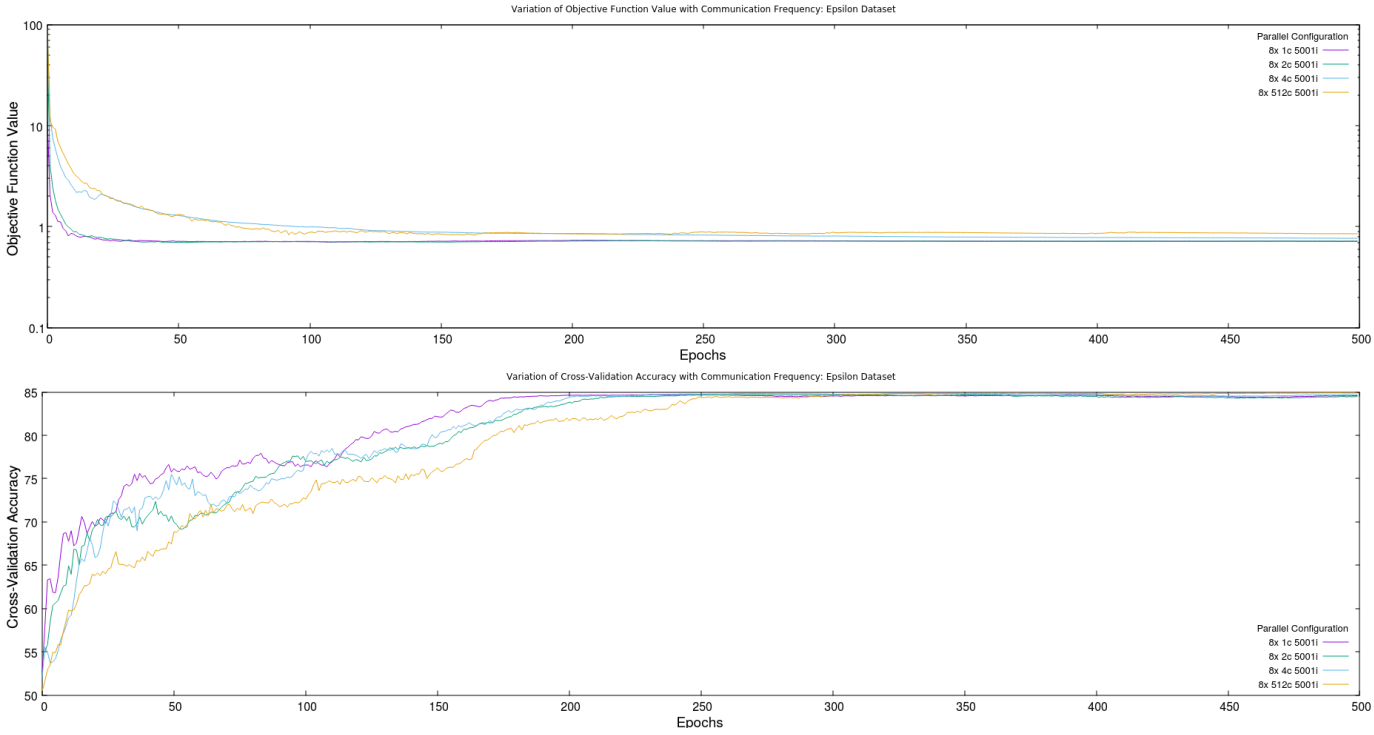

Fig. 183. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [8] and MSF = [1,2,4,512]

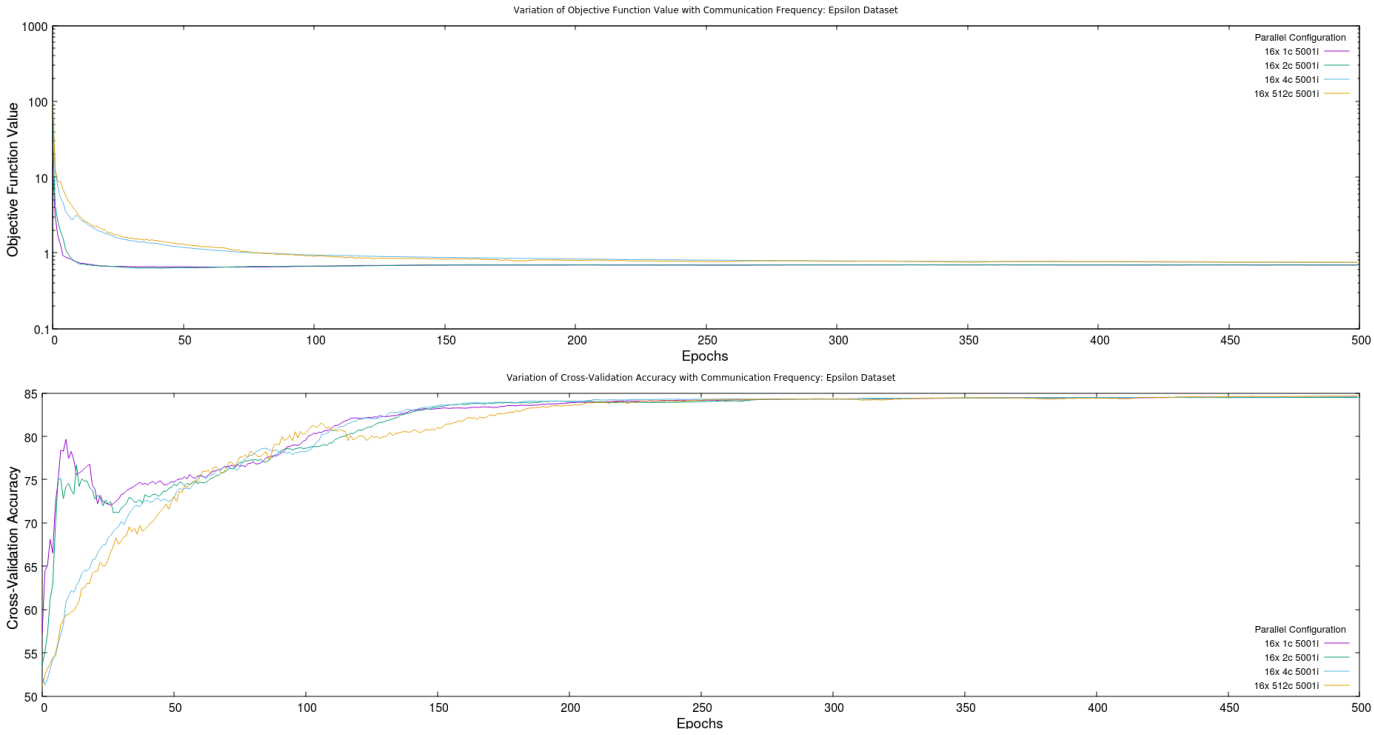

Fig. 184. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [16] and MSF = [1,2,4,512]

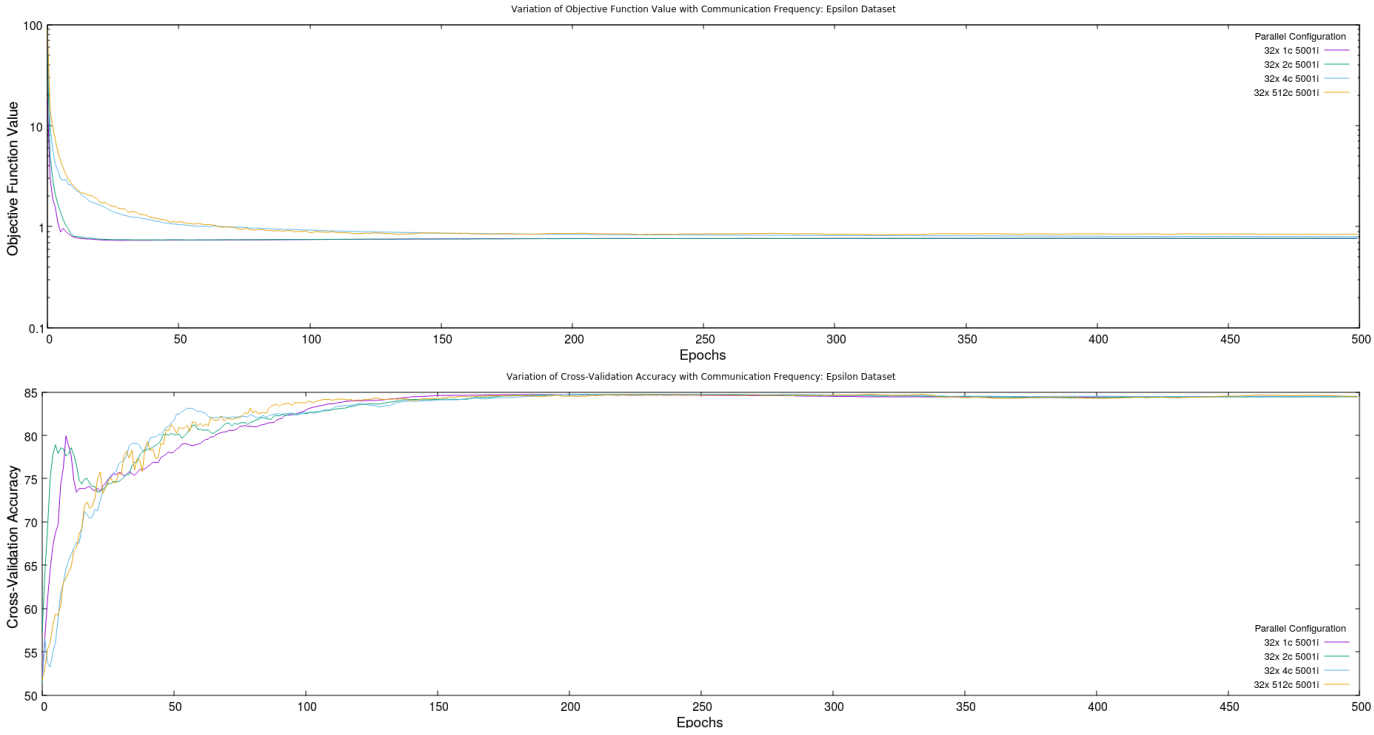

Fig. 185. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [32] and MSF = [1,2,4,512]

L. Epsilon MSF = 1,2,4,4096 , Parallelism = [2,4,8,16,32]

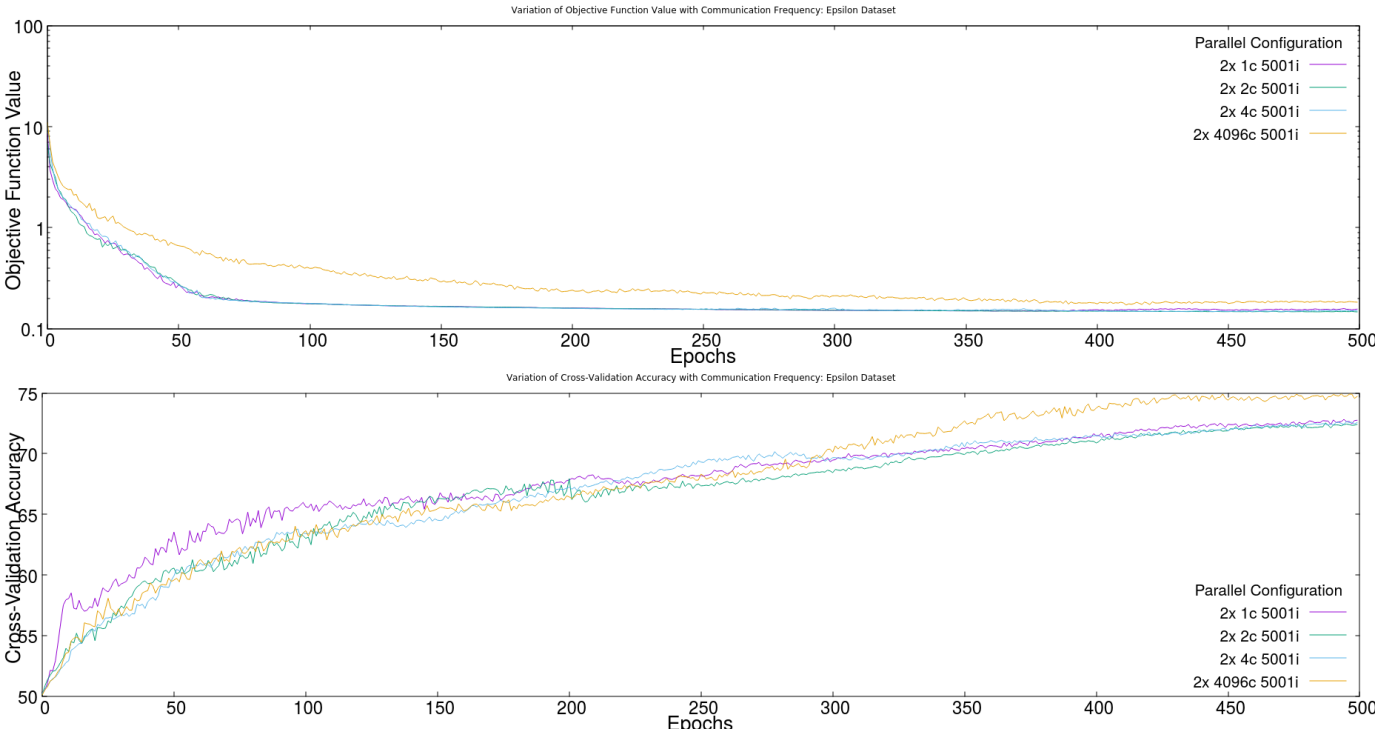

Fig. 186. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [2] and MSF = [1,2,4,4096]

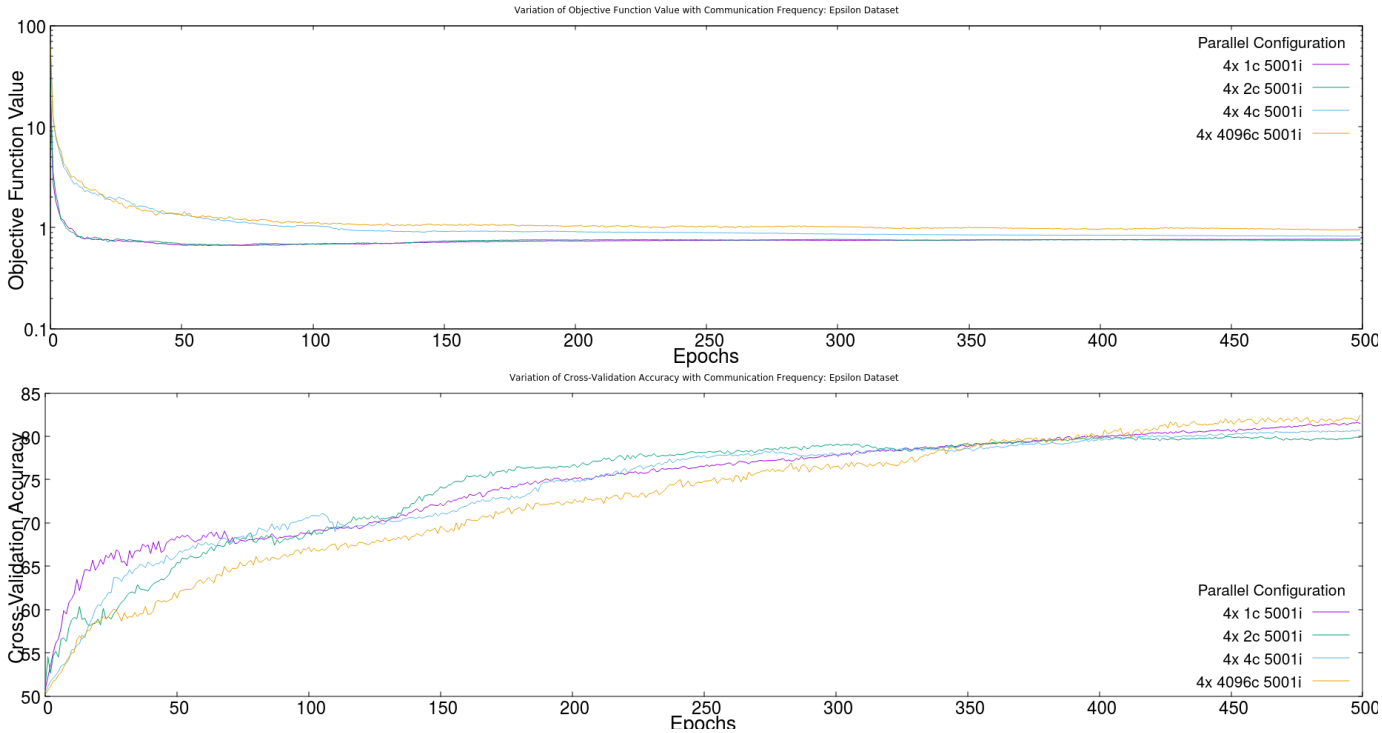

Fig. 187. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [4] and MSF = [1,2,4,4096]

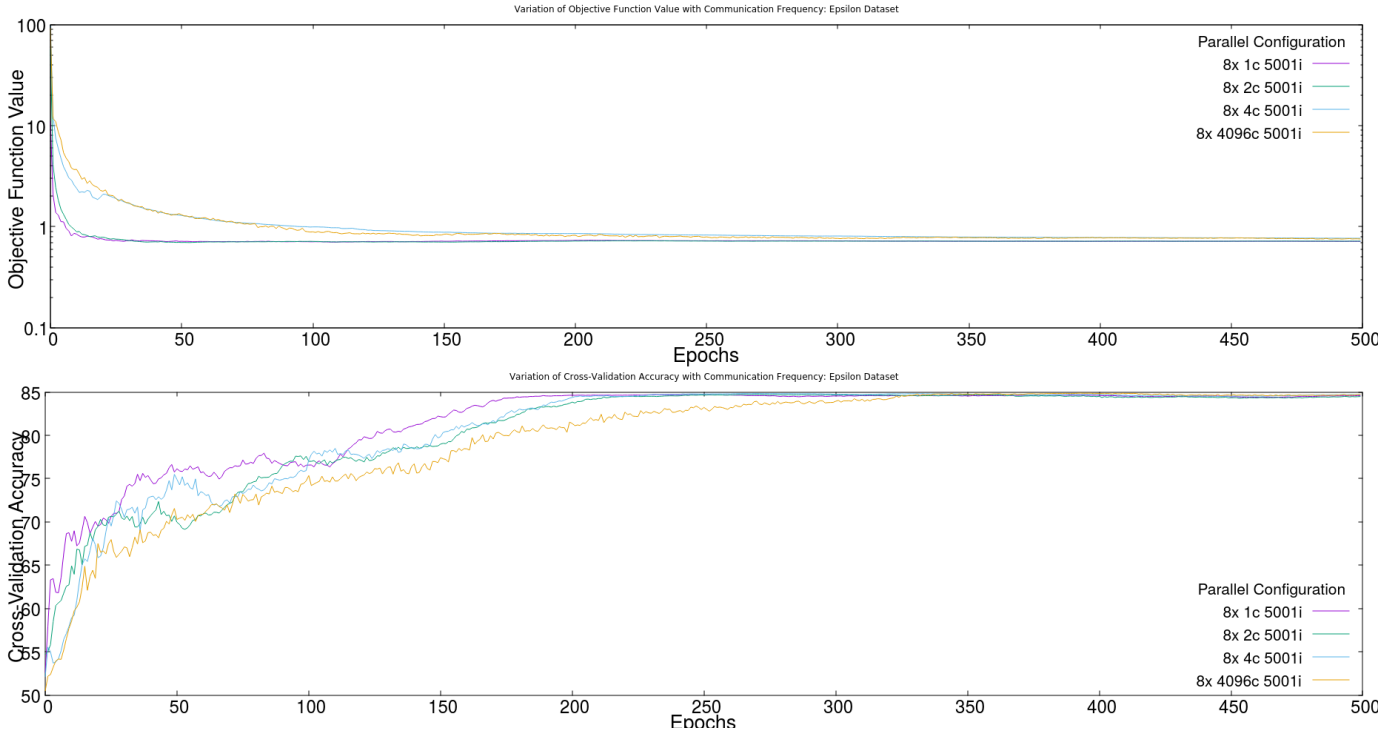

Fig. 188. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [8] and MSF = [1,2,4,4096]

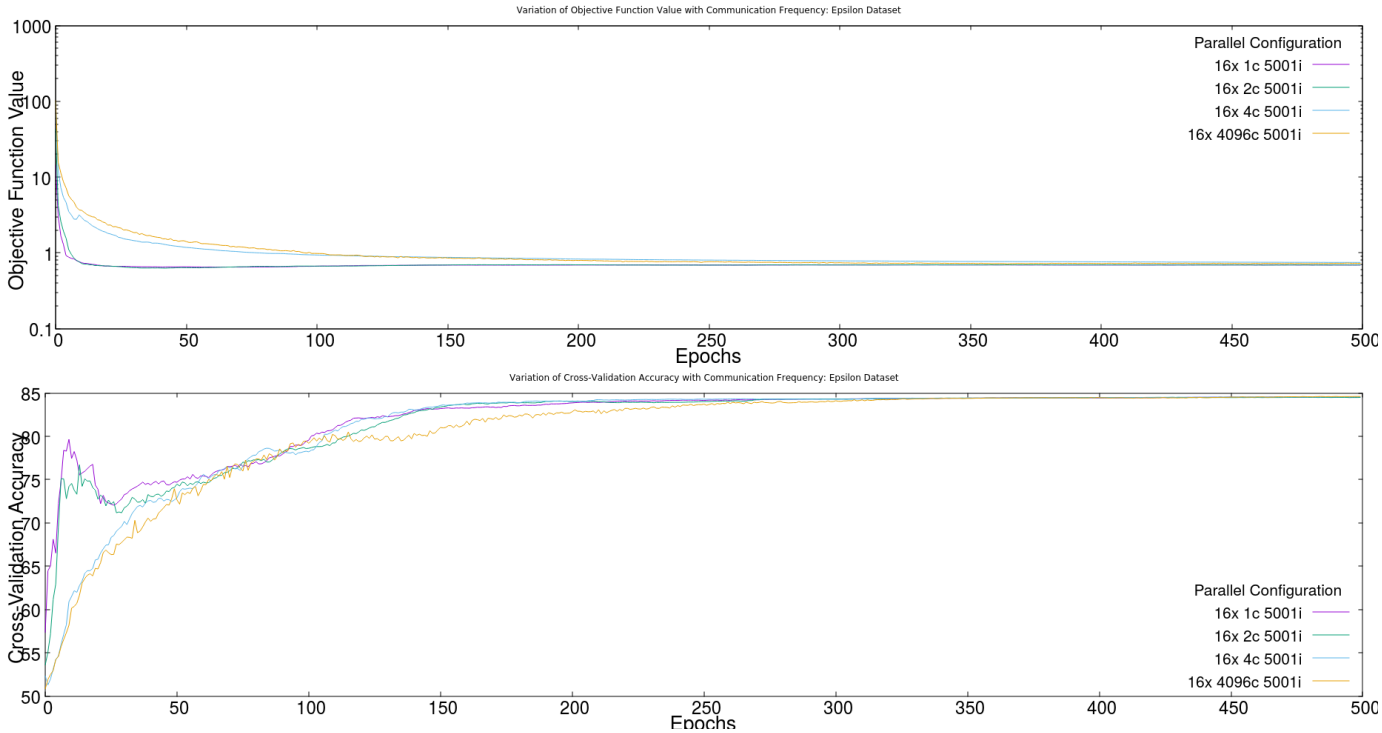

Fig. 189. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [16] and MSF = [1,2,4,4096]

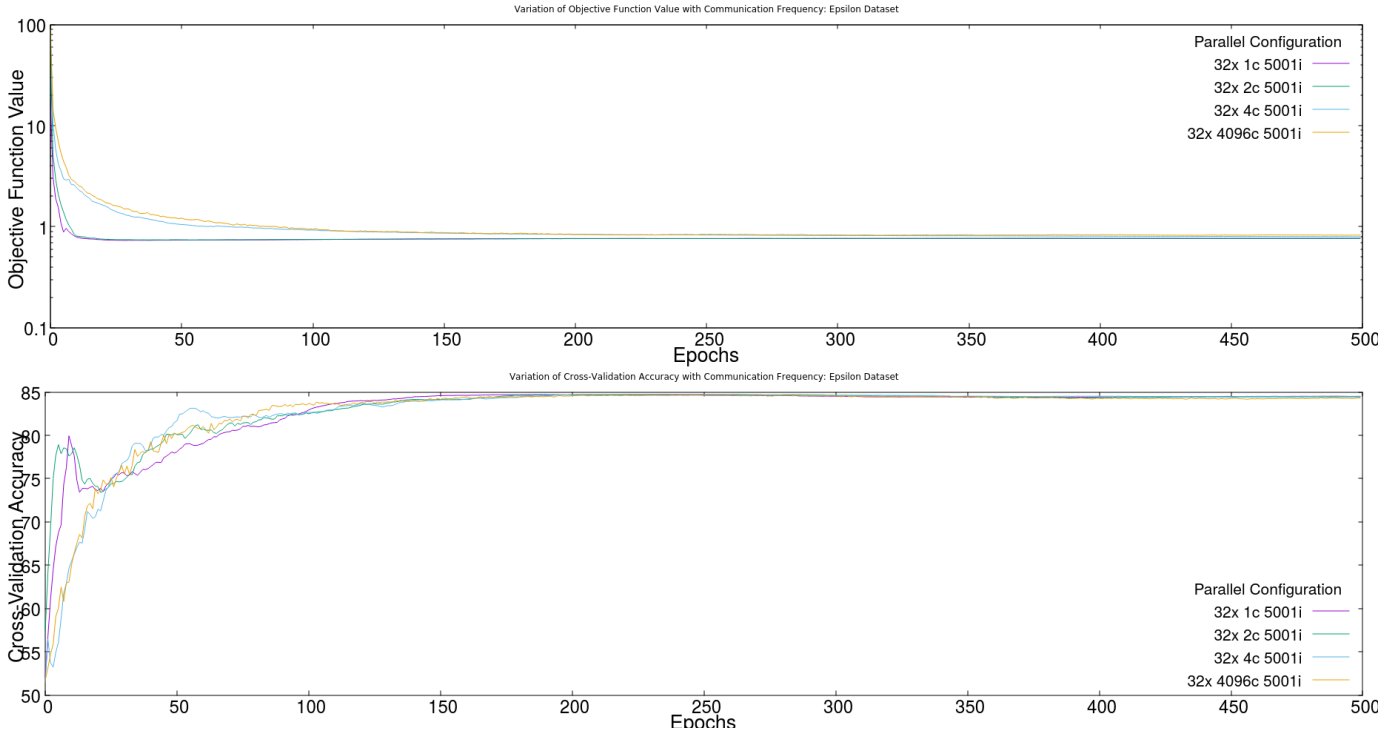

Fig. 190. Distributed Model Synchronizing Algorithm on Epsilon Dataset for Parallelisms [32] and MSF = [1,2,4,4096]

V. All Training Time

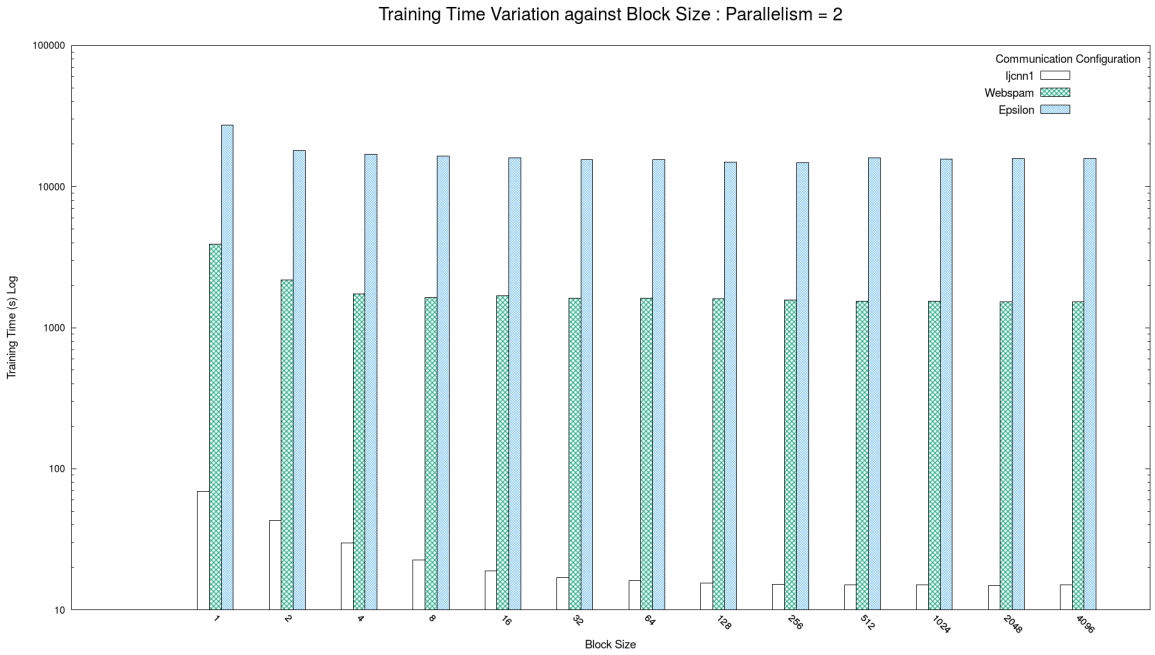

Fig. 191. Distributed Training Time of All Datasets : Parallelism = 2

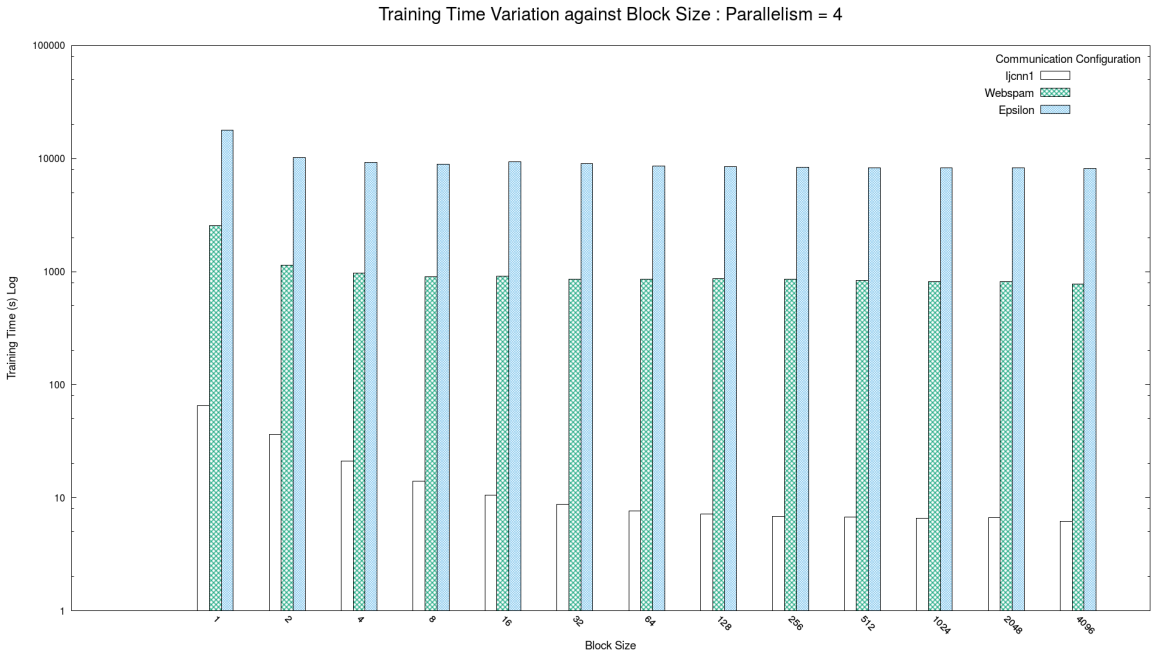

Fig. 192. Distributed Training Time of All Datasets : Parallelism = 4

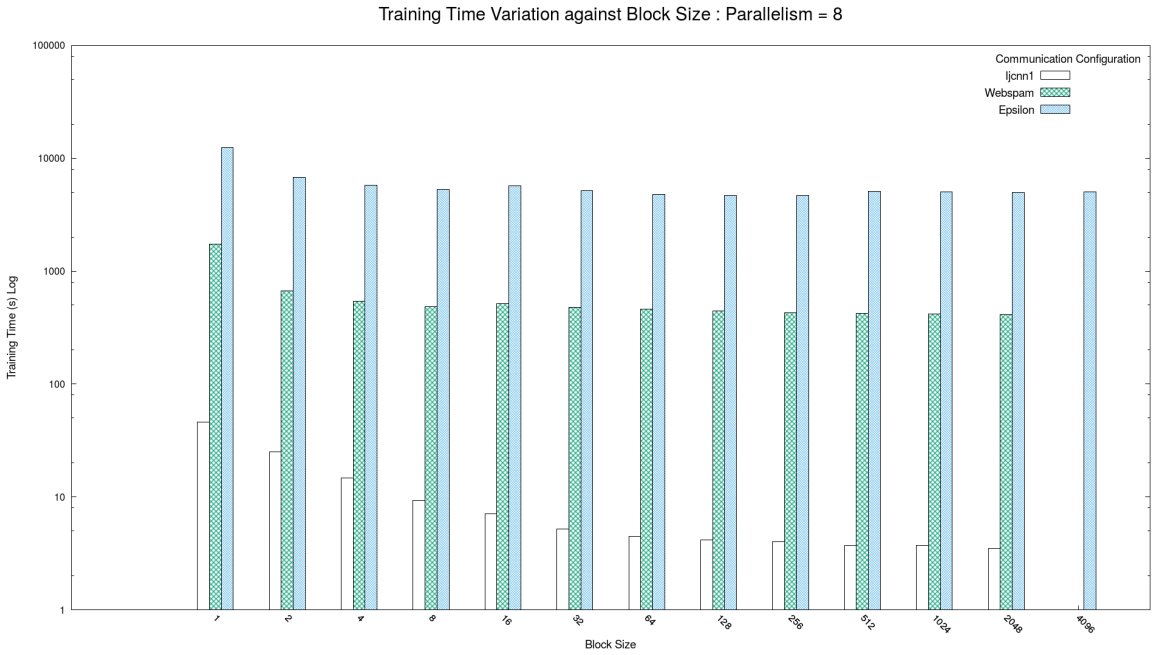

Fig. 193. Distributed Training Time of All Datasets : Parallelism = 8

Fig. 194. Distributed Training Time of All Datasets : Parallelism = 16

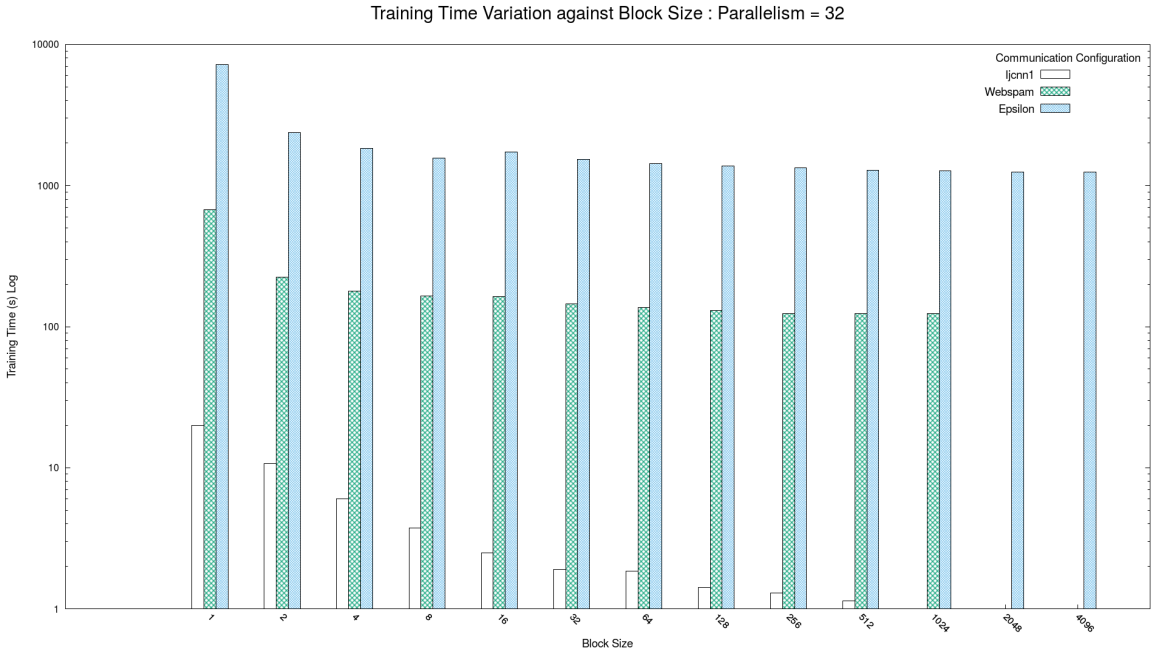

Fig. 195. Distributed Training Time of All Datasets : Parallelism = 32

VI. Training Time Breakdown

A. Ijcnn1

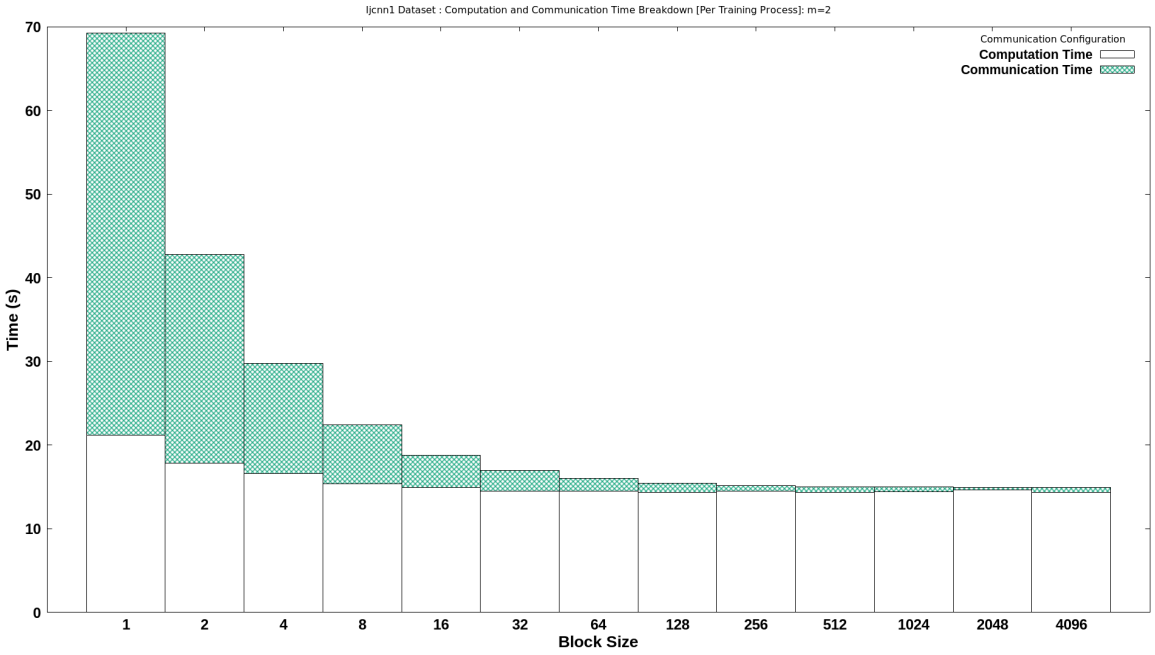

Fig. 196. Distributed Training Time of Dataset Ijcnn1 : Parallelism = 2

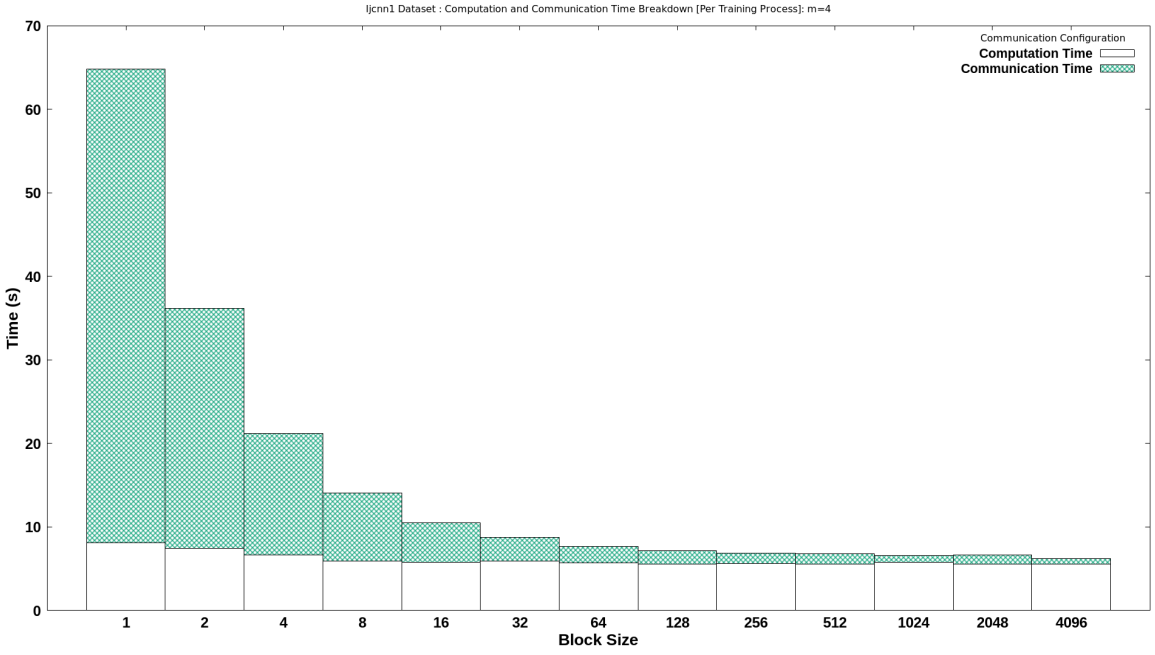

Fig. 197. Distributed Training Time of Dataset Ijcnn1 : Parallelism = 4

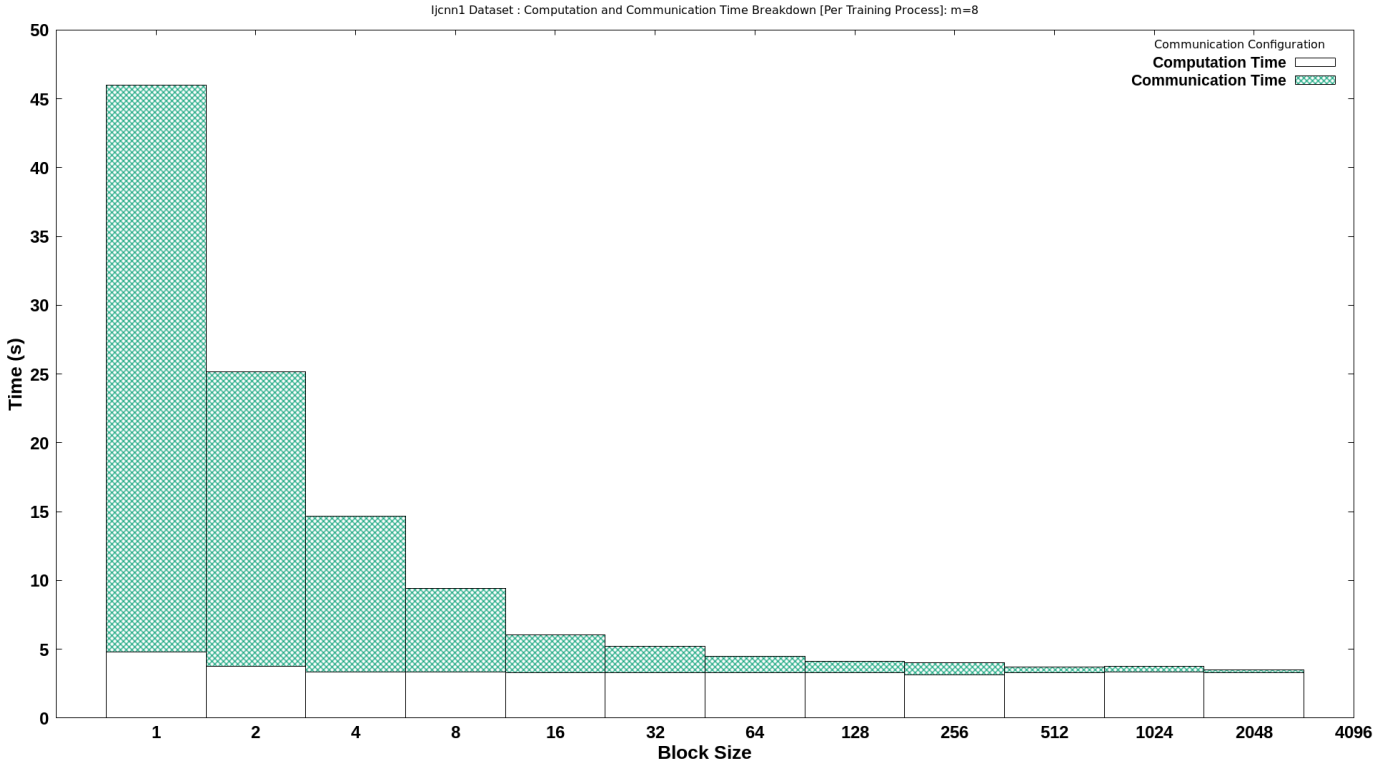

Fig. 198. Distributed Training Time of Dataset Ijcnn1 : Parallelism = 8

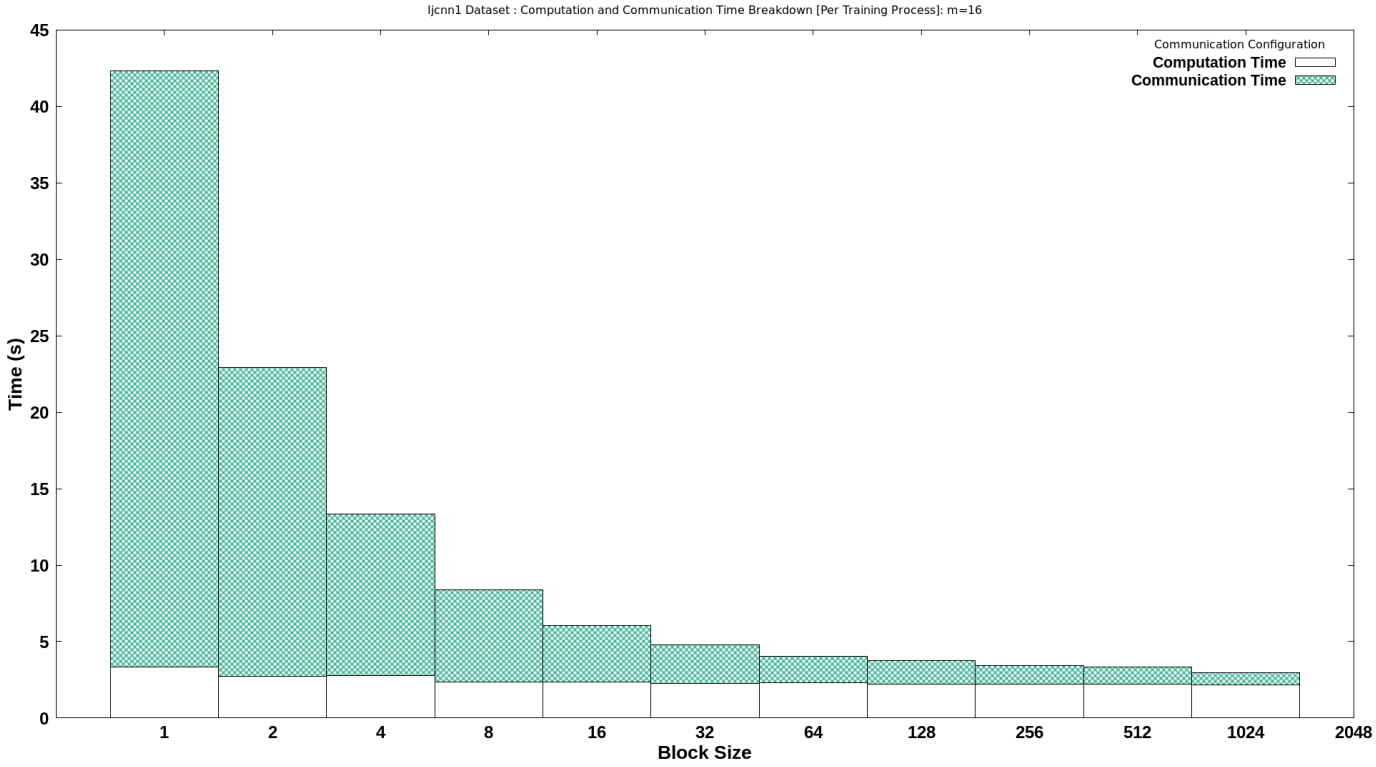

Fig. 199. Distributed Training Time of Dataset Ijcnn1 : Parallelism = 16

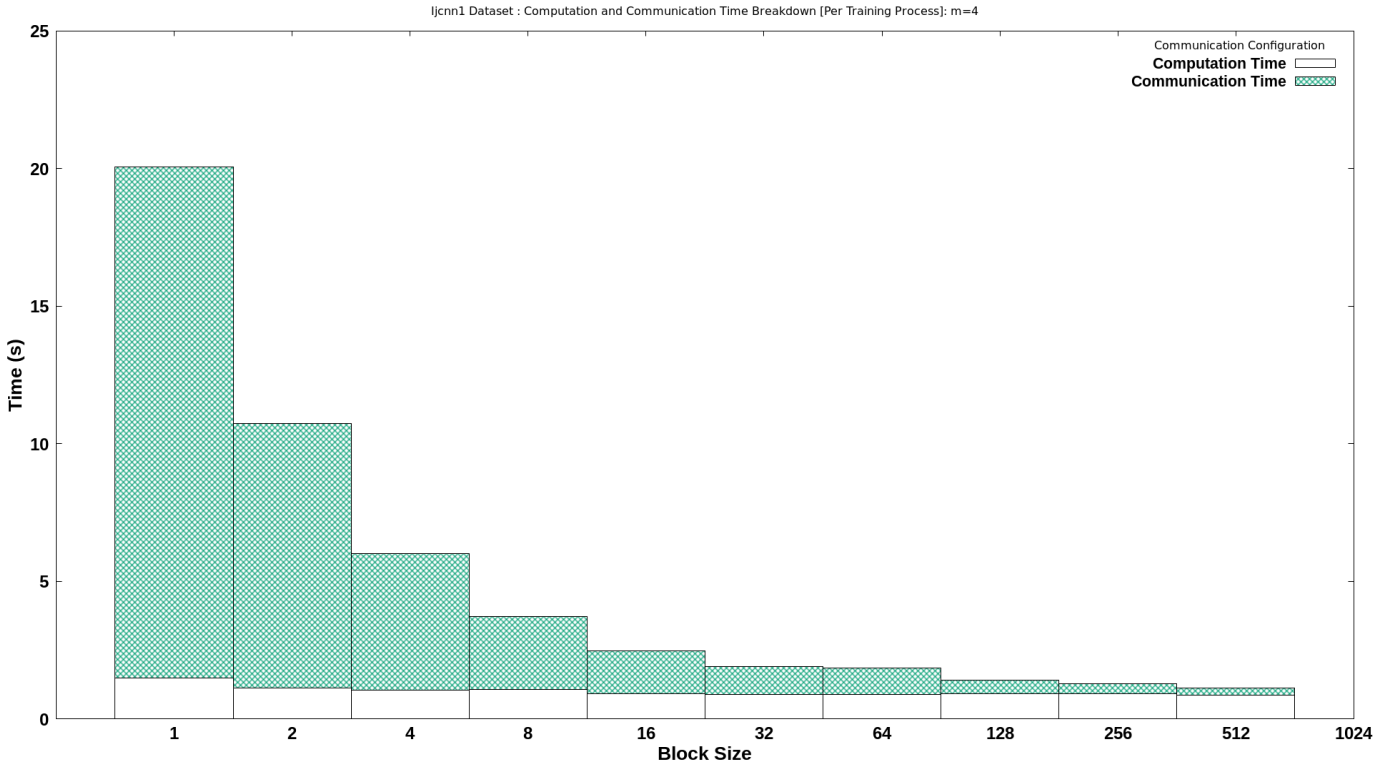

Fig. 200. Distributed Training Time of Dataset Ijcnn1 : Parallelism = 32

B. Webspam

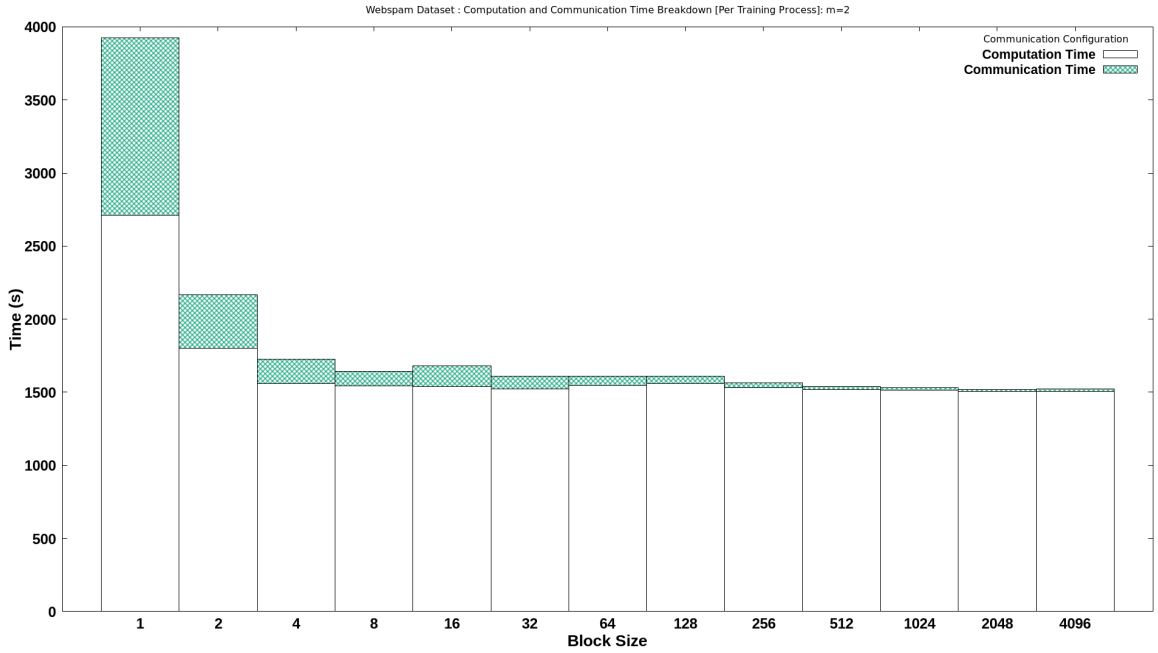

Fig. 201. Distributed Training Time of Dataset Webspam : Parallelism = 2

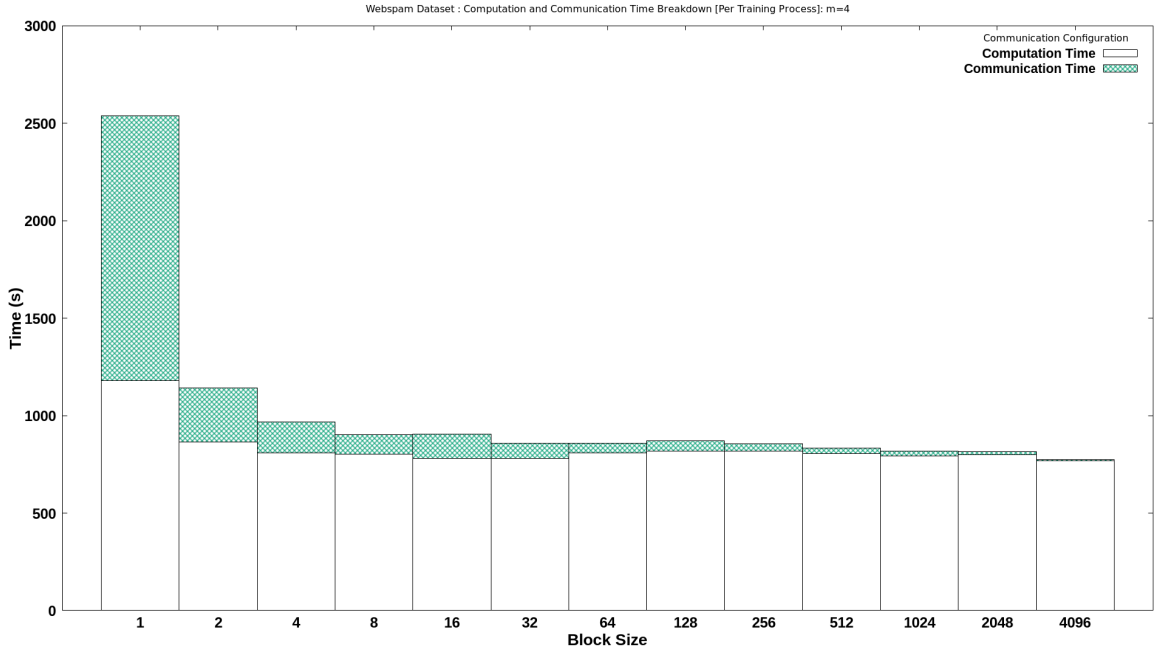

Fig. 202. Distributed Training Time of Dataset Webspam : Parallelism = 4

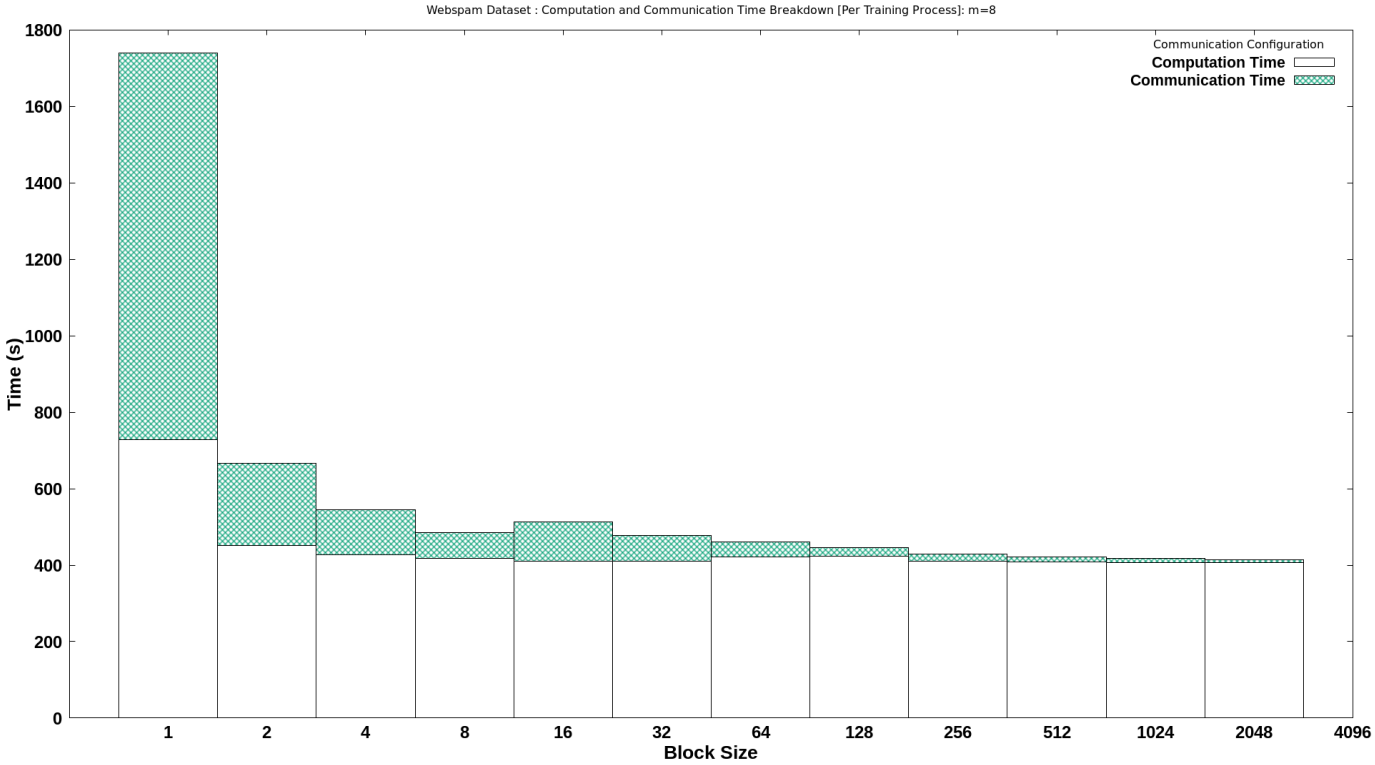

Fig. 203. Distributed Training Time of Dataset Webspam : Parallelism = 8

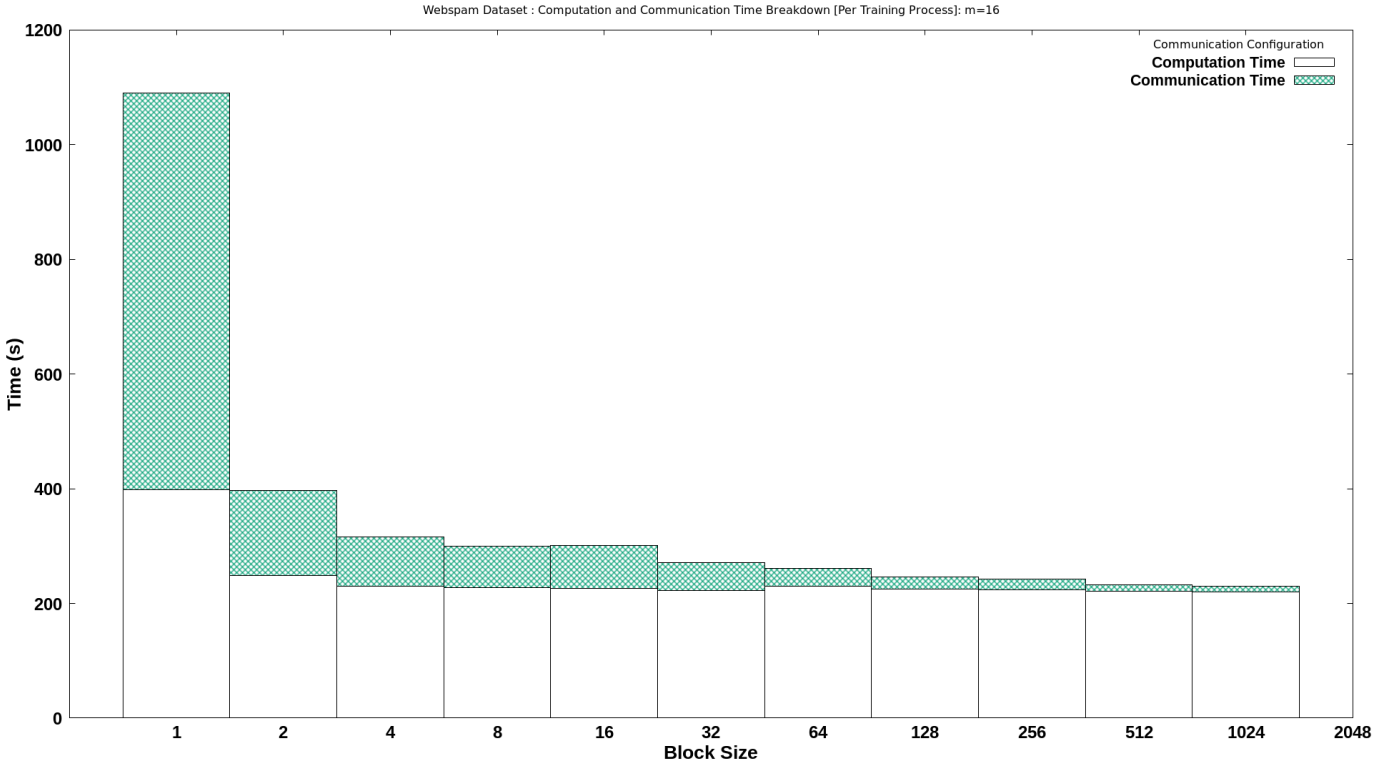

Fig. 204. Distributed Training Time of Dataset Webspam : Parallelism = 16

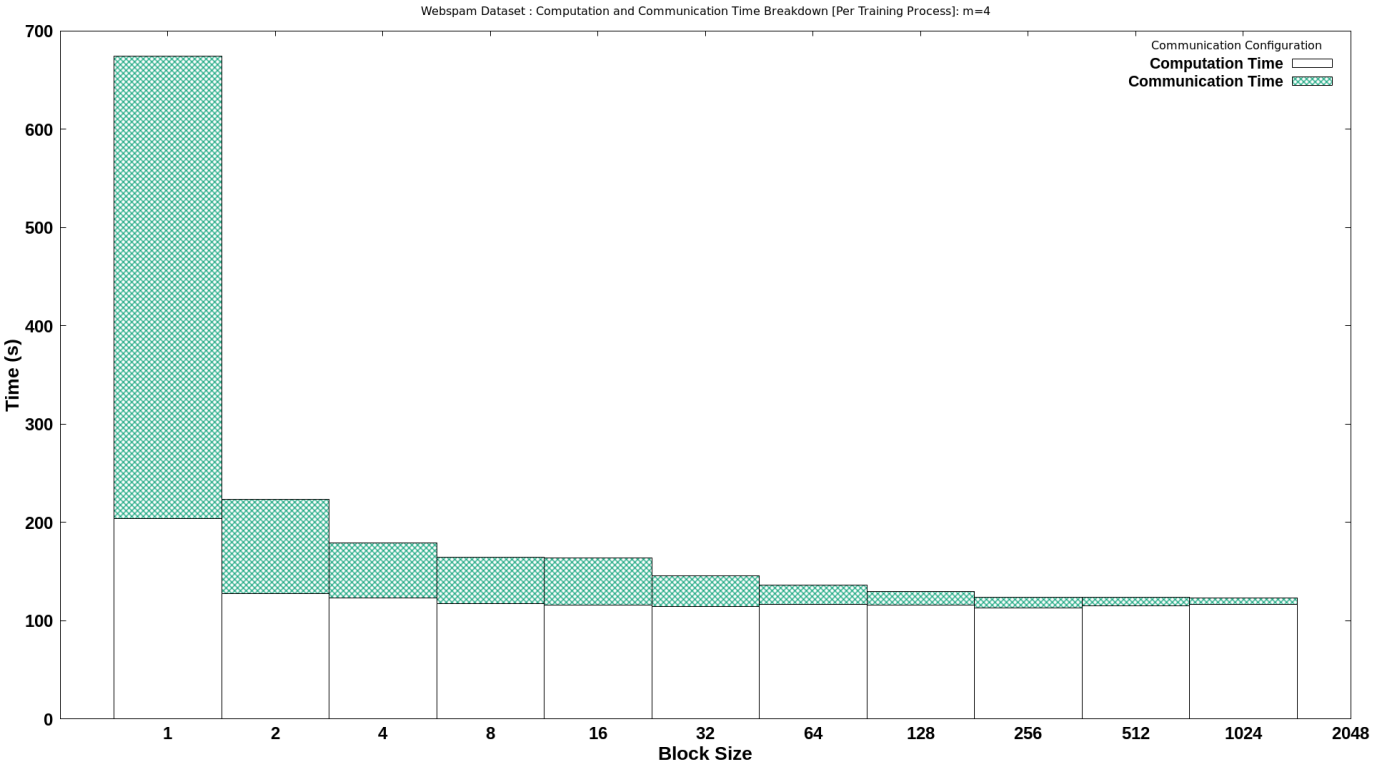

Fig. 205. Distributed Training Time of Dataset Webspam : Parallelism = 32

C. Epsilon

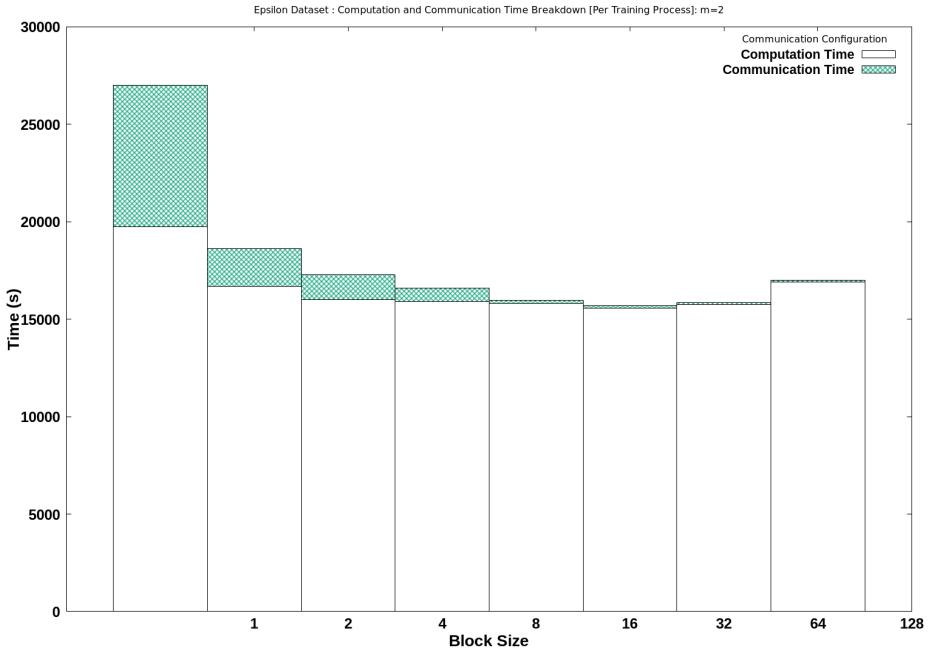

Fig. 206. Distributed Training Time of Dataset Epsilon : Parallelism = 2

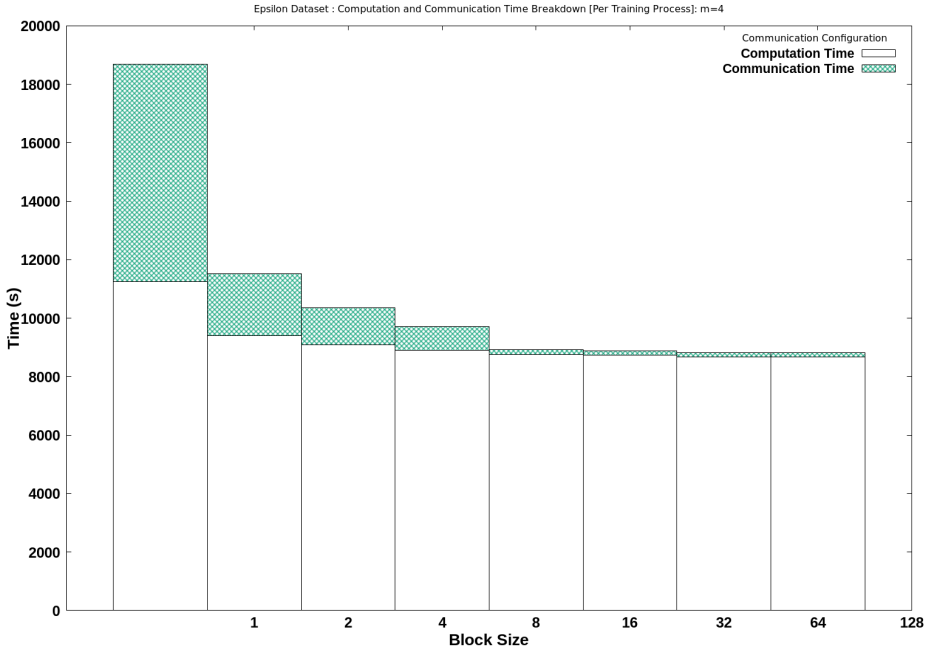

Fig. 207. Distributed Training Time of Dataset Epsilon : Parallelism = 4

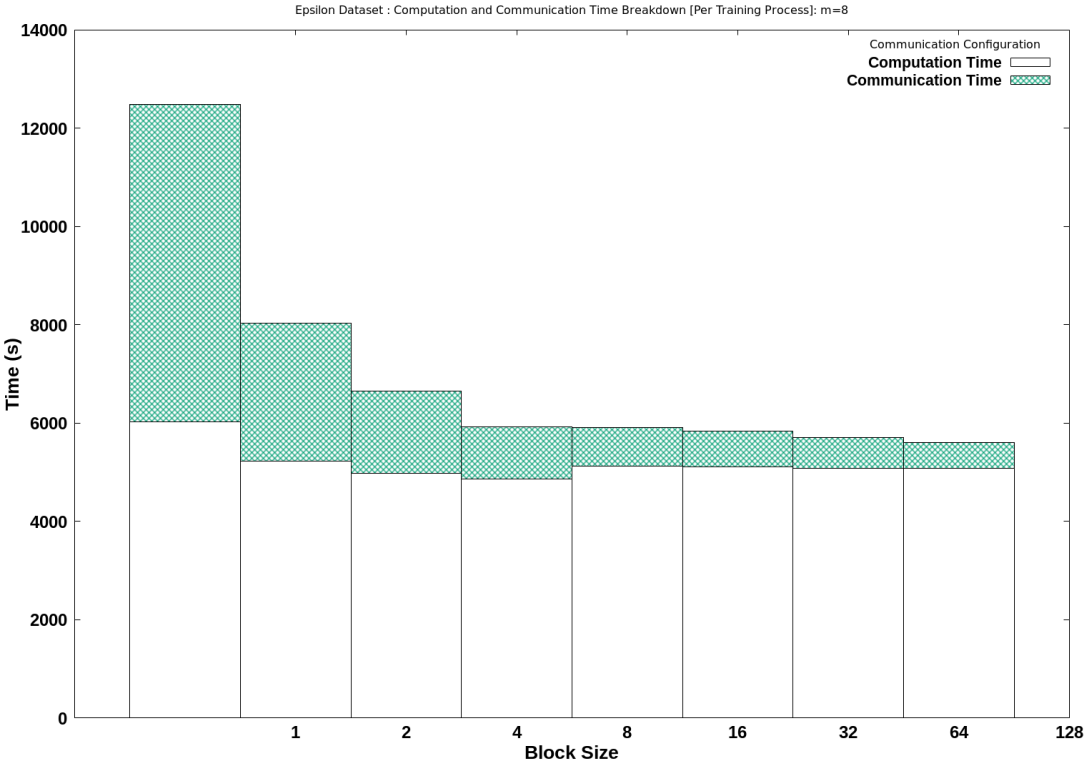

Fig. 208. Distributed Training Time of Dataset Epsilon : Parallelism = 8

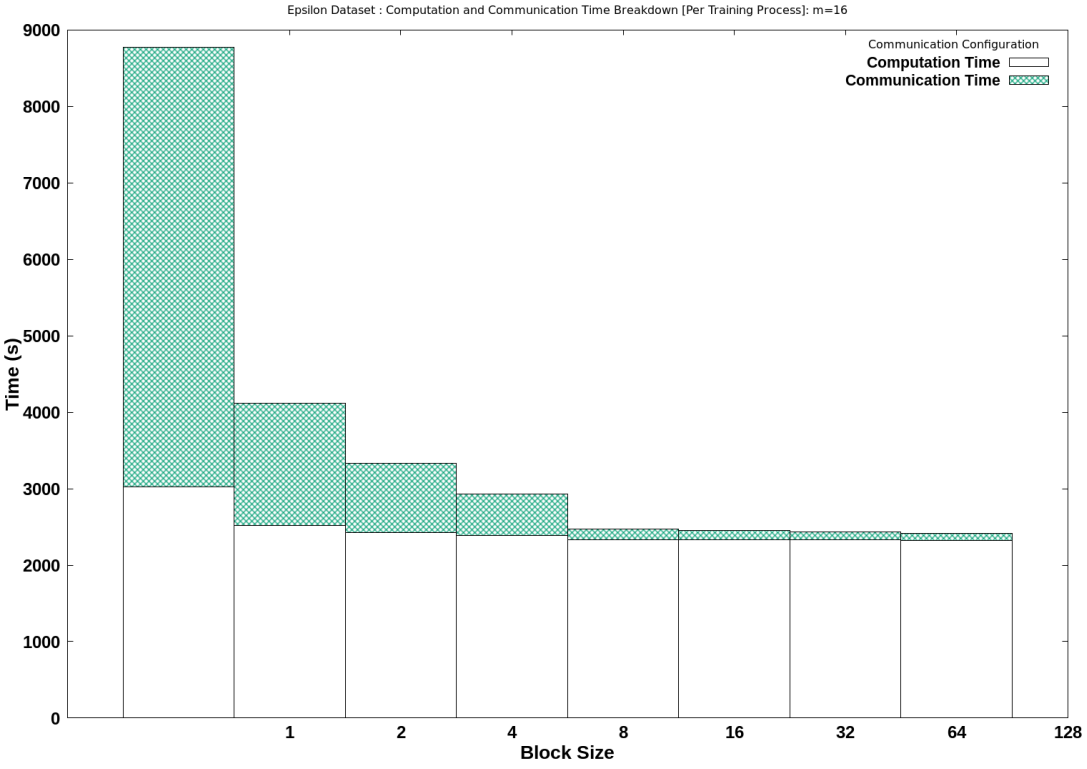

Fig. 209. Distributed Training Time of Dataset Epsilon : Parallelism = 16

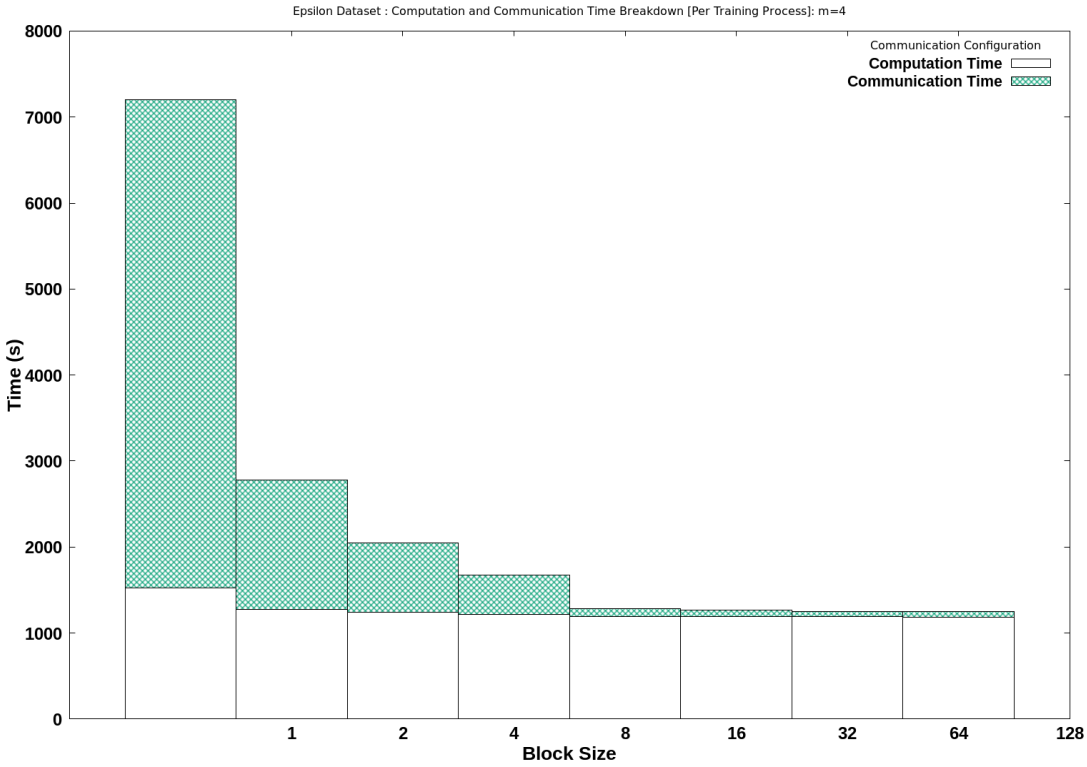

Fig. 210. Distributed Training Time of Dataset Epsilon : Parallelism = 32

VII. Summary of All Training Time and Training Time Breakdown

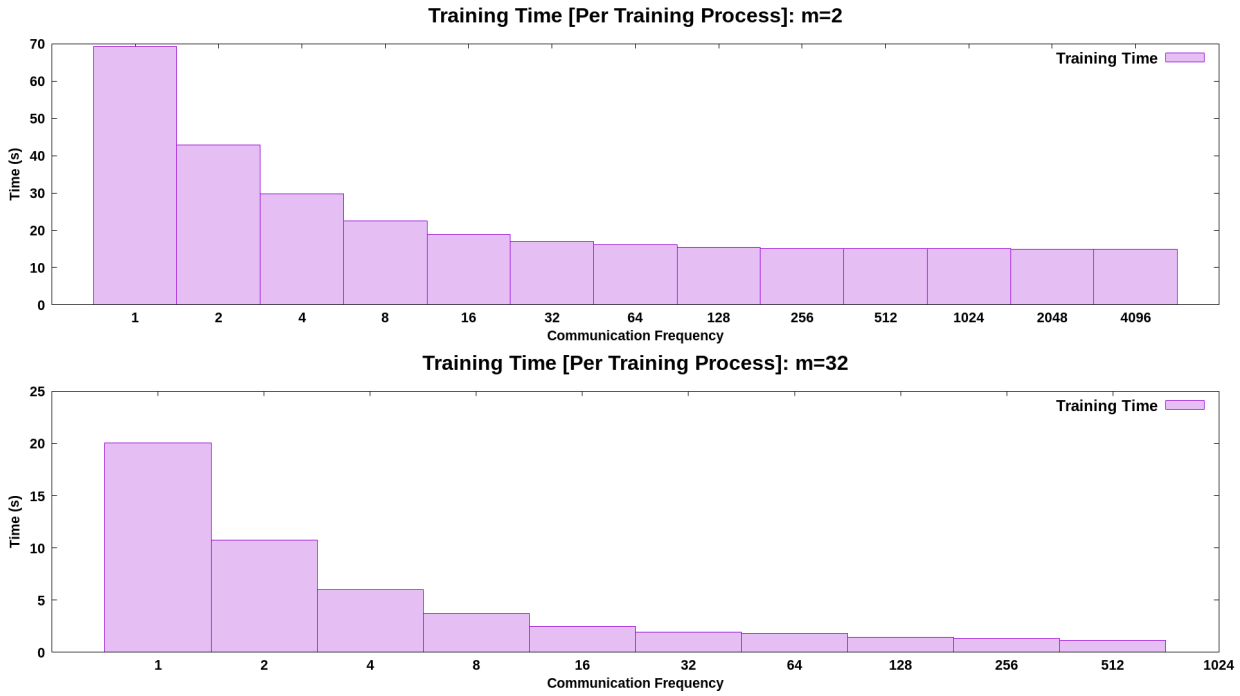

Fig. 211. Distributed Training Time of Dataset Ijcnn1

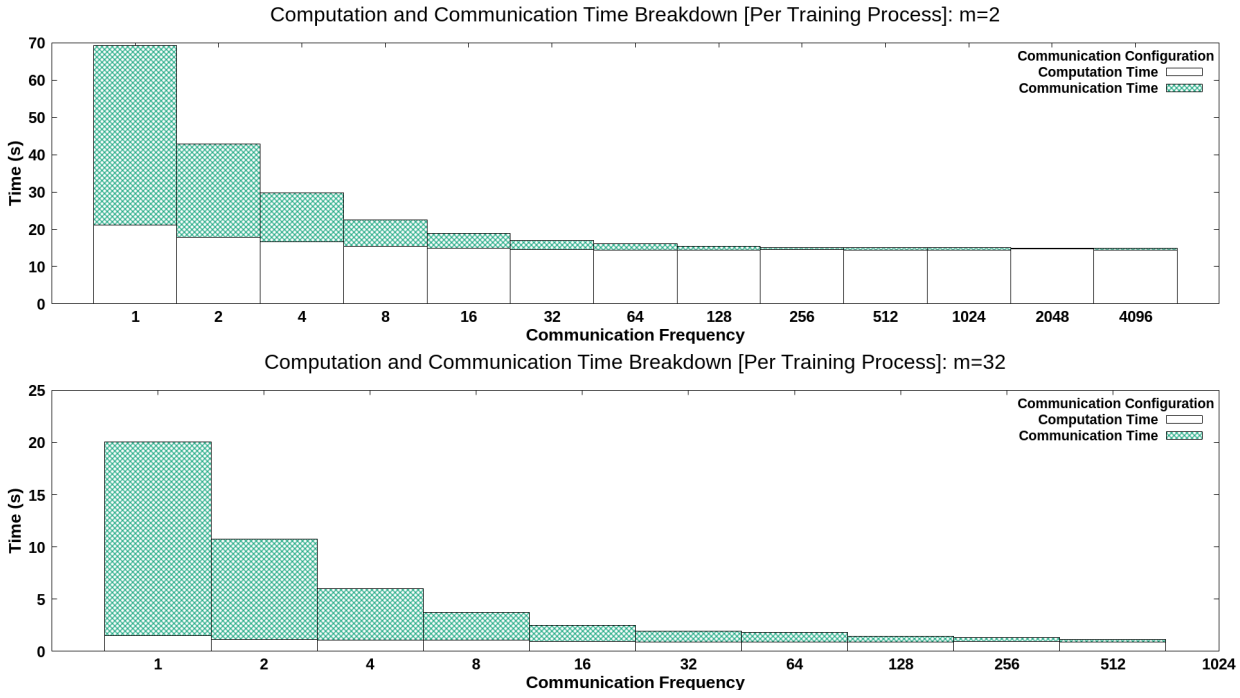

Fig. 212. Distributed Training Time Breakdown of Dataset Ijcnn1

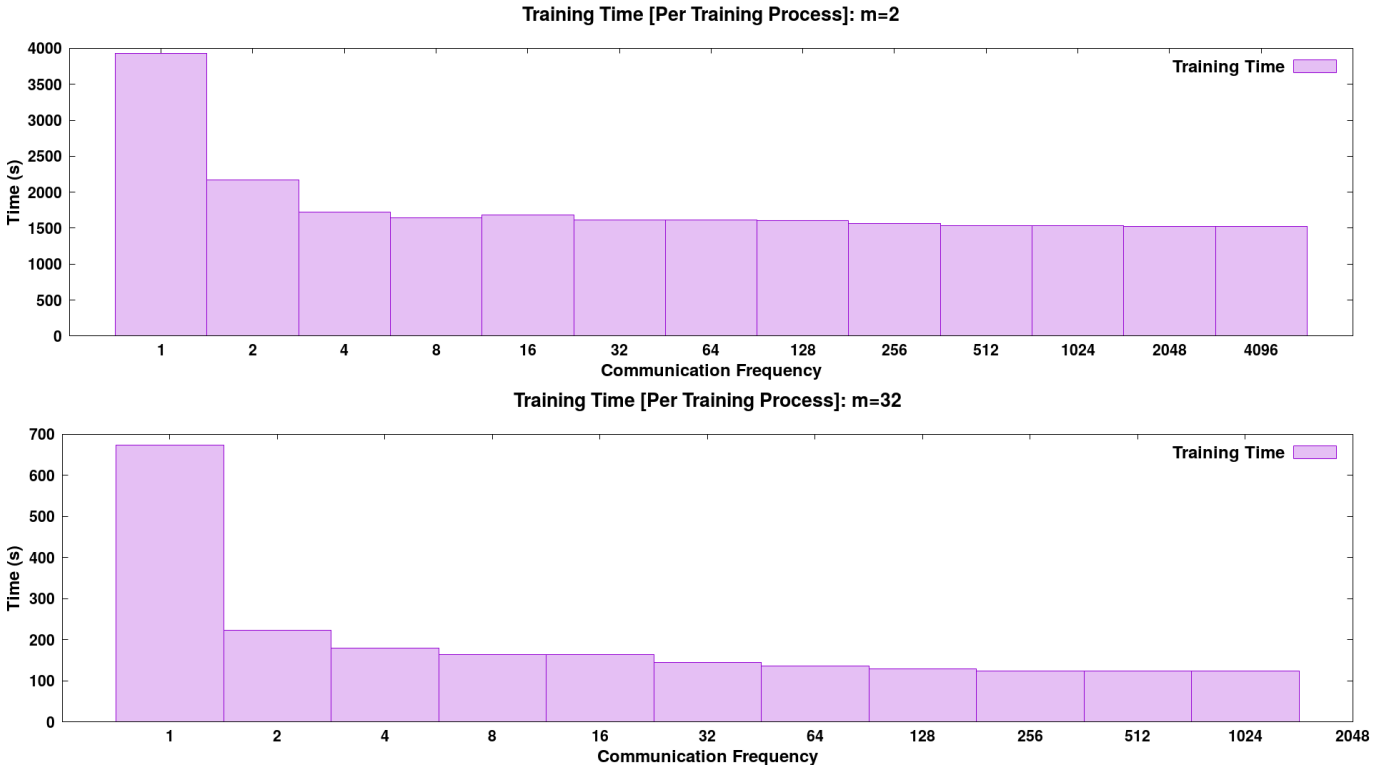

Fig. 213. Distributed Training Time of Dataset Webspam

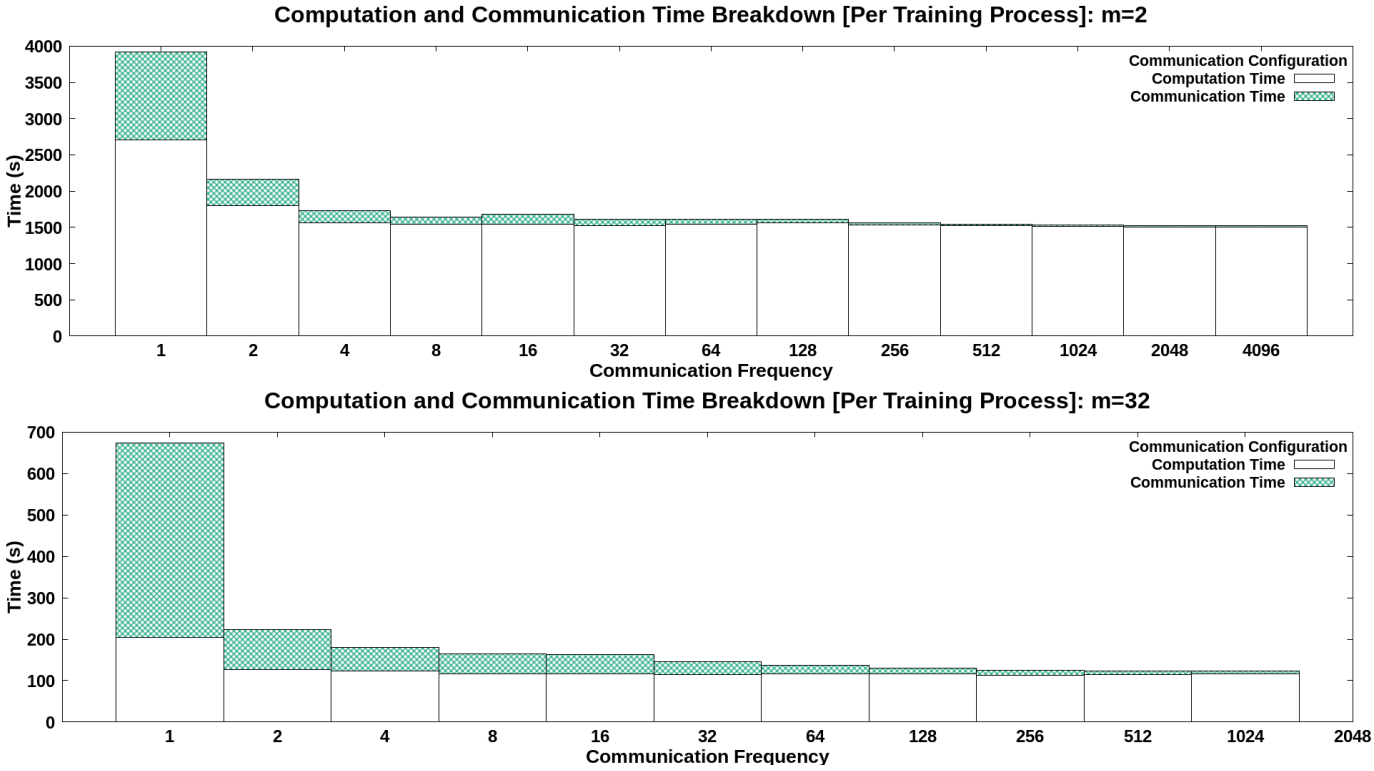

Fig. 214. Distributed Training Time Breakdown of Dataset Webspam

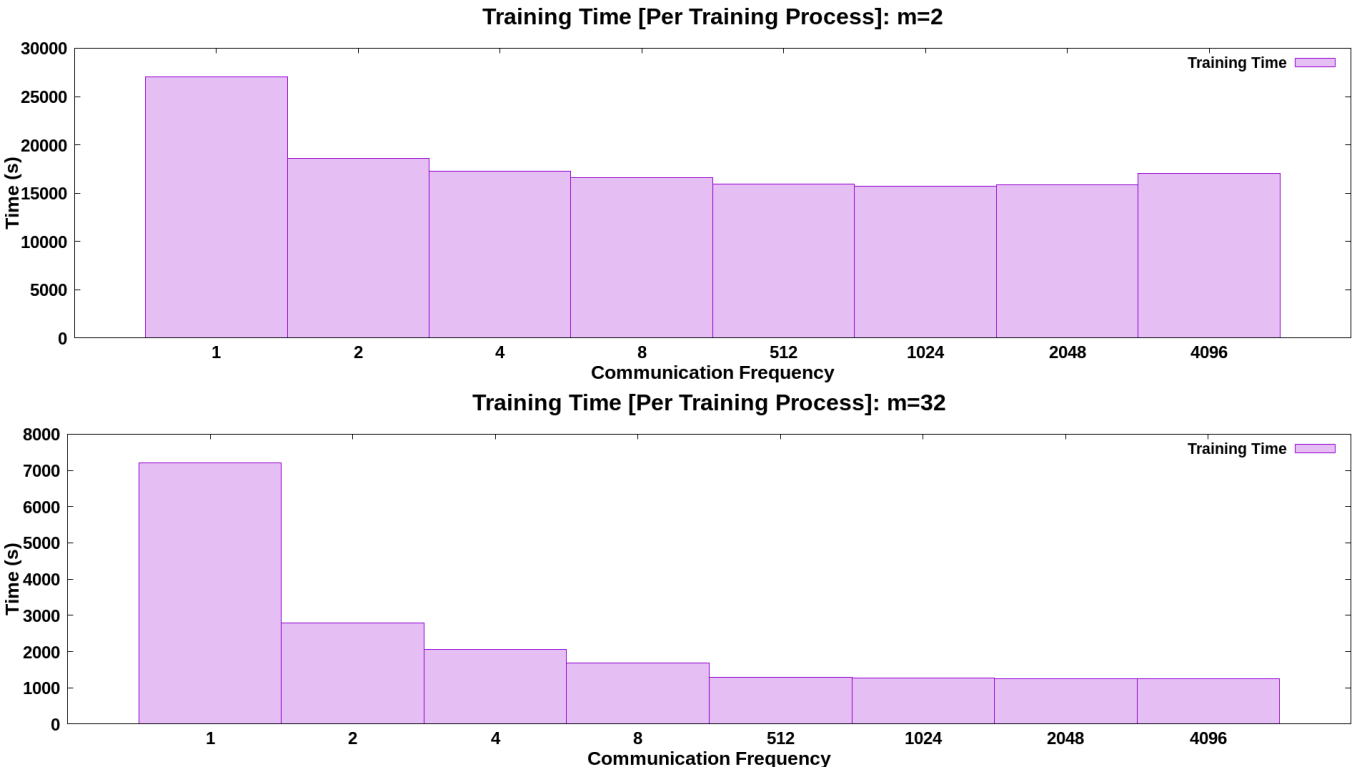

Fig. 215. Distributed Training Time of Dataset Epsilon

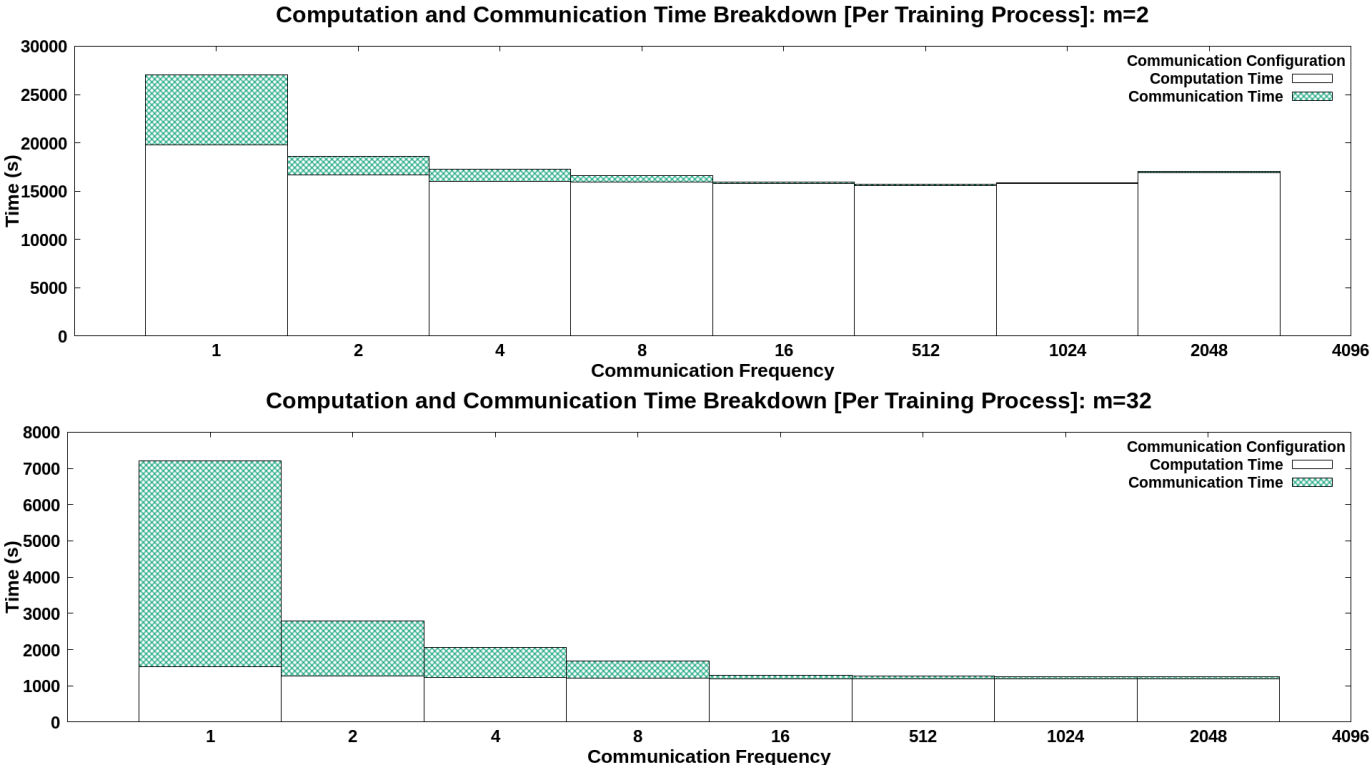

Fig. 216. Distributed Training Time Breakdown of Dataset Epsilon

ACKNOWLEDGMENT

This work was partially supported by NSF CIF21 DIBBS 1443054 and we extend our gratitude to the FutureSystems team and Digital Science Center for their support with the infrastructure.