**Appendix**

1. The comparison for the aggregated traffic

In this part, we show the comparison between the true traffic values and the ConvLSTM-TM predictions for the aggregated traffic. For a better illustration, only 100 time points are plotted in Figures 1-3 for Abilene, CERNET and GEANT. These figures demonstrate ConvLSTM-TM predicts well for the whole traffic situation.

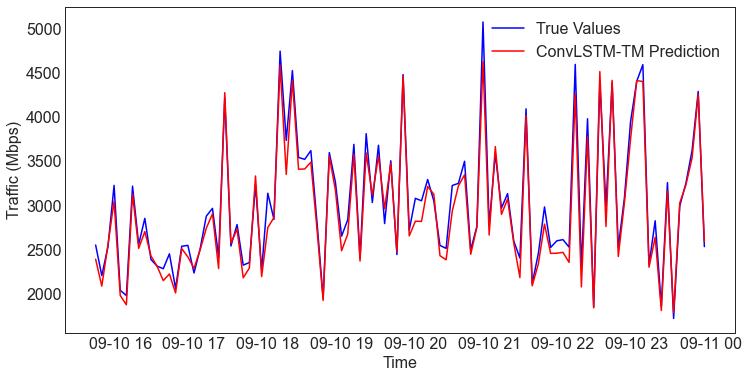


Figure 1. The comparison for the aggregated traffic in the Abilene network.

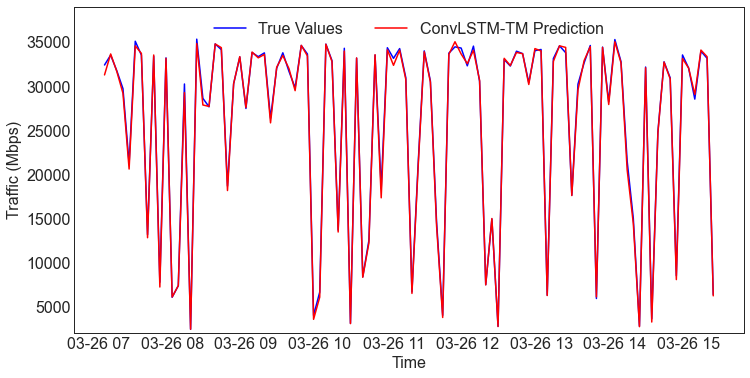


Figure 2. The comparison for the aggregated traffic in the CERNET network.

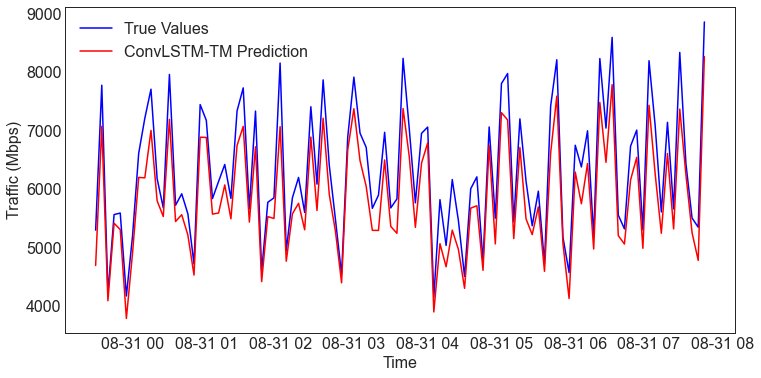


Figure 3. The comparison for the aggregated traffic in the GEANT network.

2. The comparison for the selected node pair

In this part, we show the comparison between the true traffic values and the predictions for a single node pair. For a better illustration, only 20 time points are plotted in Figures 4-6 for Abilene, CERNET and GEANT. These figures demonstrate ConvLSTM-TM predicts no worse than other models for the single node pair situation.

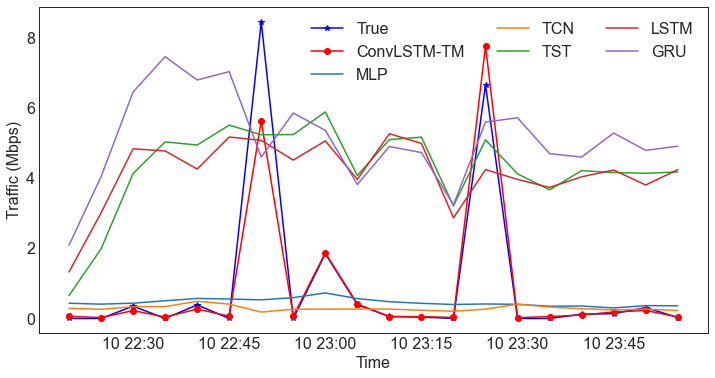


Figure 4. The comparison for a single node pair in the Abilene network.

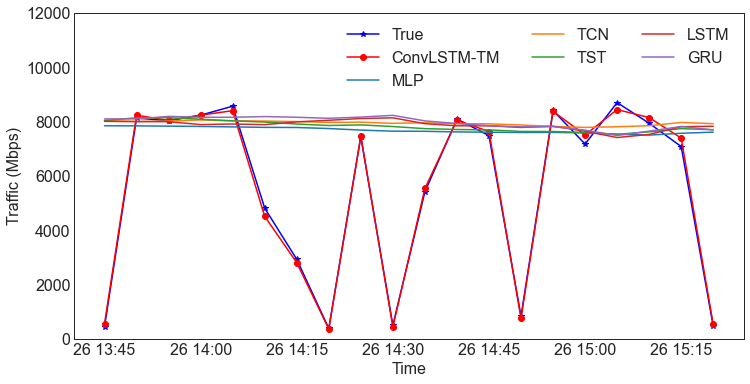


Figure 5. The comparison for a single node pair in the CERNET network.

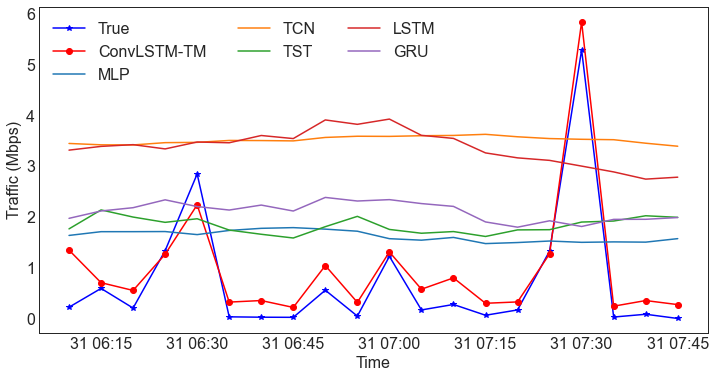


Figure 6. The comparison for a single node pair in the GEANT network.