We consider a parallel service system with customer types $c_1, ..., c_I$, and servers $s_1, ..., s_J$ and a bipartite compatibility graph $G$, operated under the policy of first come first served (FCFS) assign to longest idle server (ALIS). We study this system under fluid scaling, when time and space are scaled by a factor $n$. We derive properties of fluid limits, in an attempt to verify stability and complete resource pooling. We characterize fluid limits for the case when service rates are server dependent, for the case when they are customer dependent, and for the case that the compatibility graph is a tree. We formulate a static planning linear program and obtain maximum throughput compatibility tree, and show that FCFS using this compatibility tree is throughput optimal. We study matching rates, and show by simulation that they are dependent on the shape of the service time distribution.