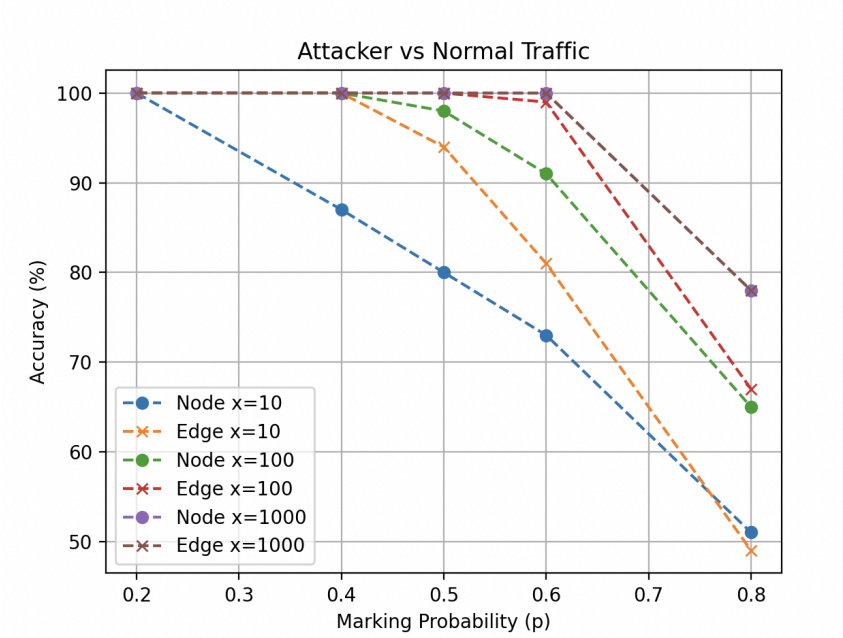
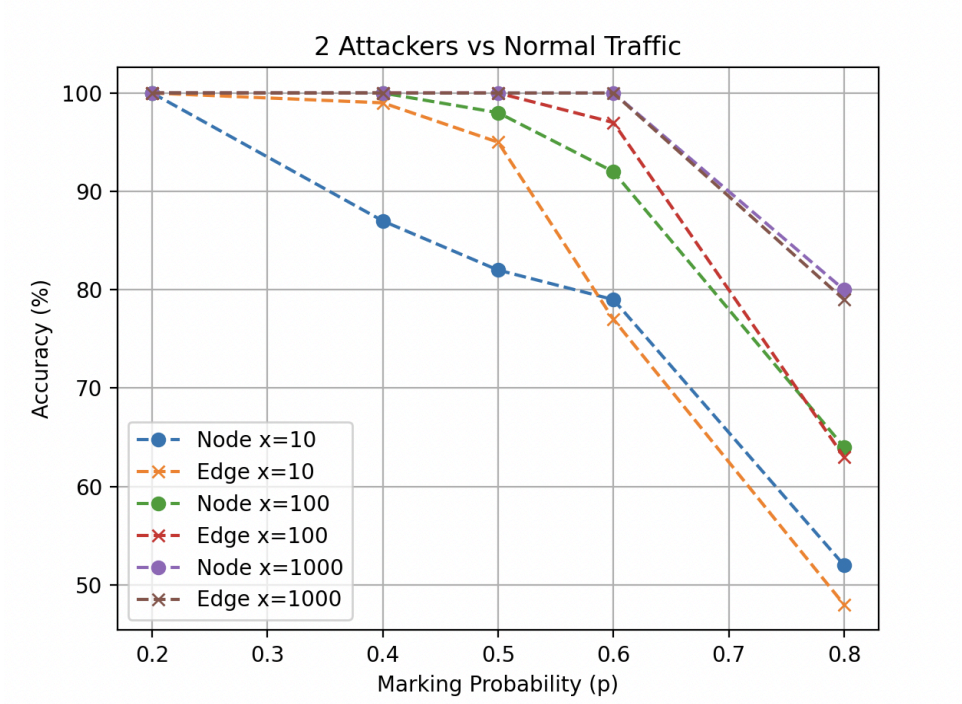
HW1 Report

So in this assignment, we are tasked to simulate the behavior of probabilistic packet marking where we compare the performance of node sampling algorithm and edge sampling algorithm. These algorithms are variations of Probabilistic Packeting Marking (PPM). PPM is a traceback mechanism to identify the source of network attacks such as DOS attacks. It significantly reduces space consumption where every router does not have to append its address to a packet as it travels through the network. Optimally, PPM allows the attack path to be sampled one node at a time. The algorithms share main components including a marking procedure executed by network routers and a path reconstruction procedure implemented by the victim. The marking procedure generates a random number between 0 and 1 as routers process packets. If this random number is less than a predefined marking probability (p), then the router "marks" the packet by writing its own information (e.g. IP address or edge information) into the packet before routing it. The reconstruction procedure mathematically reconstructs the path back to the attacker with algorithms like the aforementioned node sample or edge sampling in this case. The victim receives probabilistically marked packets over time, especially as DoS attackers send numerous packets. Accumulating these packets allows the reconstruction procedure to take place.





Let's discuss the plots where we have simulated a single attacker and two attackers creating DoS attacks against a normal user. In the single attacker vs normal user, we see that as the attack rate increases to 10, 100, or 1000 times higher than the normal user, the traceback accuracy for both algorithms converge toward 100%. The higher the attack rate, the easier it is for the victim to reconstruct the path since the victim has a lot of marked packets, especially since DoS attacks rely on high-volume traffic such as SYN floods or ping floods. We see that certain marking probabilities such as 0.5 (x=10) and 0.6 (x=100, x=1000) provide the optimal accuracy. Lower marking probability means that there are not enough packets for reconstruction procedure. Conversely, higher marking probability shows the accuracy decreasing. This may be due to routers that are physically closer to the victim having frequent overwrites on marks by the router. We can see in the two attackers vs the normal user, we can see that node sampling fails since there are multiple attackers, leading to the node sampling algorithm to drop significantly in the ranges of 60-70% as seen in the plot. This can be due to the algorithm only sampling the path one node at a time, relying only on a tuple of the node and count. Therefore, the victim reconstructing the path cannot differentiate which routers belong to which branch, so the reconstruction is inaccurate. With multiple attackers conspiring together, edge sampling succeeds. In the graph, it maintains high accuracy where structured edge information is added during packet marking because the victim reconstructs the path by building a tree structure and extracting the path by enumerating acyclic paths from tracking the distance from the victim. This allows the attacker branches to be isolated and accurately traced without interference unlike the node sampling algorithm.