# Flow correlation

Large-scale adversaries such as the government and deep state actors are known for attempts at deanonymization and identification of high value targets. But traffic associations cannot be easily made due to the prevalence of so-called "Guard nodes", which were implemented as a means to prevent predecessor attacks and short-flow correlation (FC) attacks. Enter **DeepCorr**:

> such association can not be detected by inspecting the packet contents ... due to onion encryption. The goal of an adversary in this setting is to identify (some or all of) the associated flow pairs (the entry and exit segments) by comparing traffic characteristics, e.g., packet timings and sizes, across all of the ingress and egress flows. Linking associated flow pairs using traffic characteristics is called flow correlation

> DeepCorr learns a correlation function that can be used to link flows on arbitrary circuits, and to arbitrary destinations. In other words, DeepCorr can correlate the two ends of a Tor connection (a pair of *associated flows*) even if the connection destination has not been part of the learning set. Also, DeepCorr can correlate flows even if they are sent over Tor circuits different than the circuits used during the training process.

FC attacks are used to link network flows in the presence of encryption and other content obfuscation mechanisms. Using bridges (e.g., something like meek-azure) only obfuscates a bridge's IP and does *not* deploy traffic obfuscation. It is important to note this factor in traffic FC, but it's even more important when determining attacks, in general, against the Tor network. Because Tor employs Guard nodes and various other protections against timing attacks, it is increasingly difficult to place the nature of extended outages. But it is also easier to place them when several key identifiers are met. Exponentially increasing the volume of these traffic flow pairs doesn't seem to have a large impact on the effectiveness of DeepCorr FC attacks. It is still ten times as fast.

Comparing DeepCorr to something like RAPTOR, which uses the Spearman Correlation to achieve linkability between two arbitrary flows, "the performance gap between [the two] is significantly wider for shorter flow observations." To be precise, "each DeepCorr FC takes 2ms compared to RAPTOR's more than 20ms, when both target a 95% on identical dataset."

> Similar to existing FC techniques overviewed earlier, our FC system uses the timings and sizes of network flows to cross-correlate them. A main advantage of deep learning algorithms over conventional learning techniques is that a deep learning model can be provided with raw data features as opposed to engineered traffic features