Attributing State-backed Information Operations on Twitter through Machine Learning

Connor, James (School: Northport High School)

With the inherent interconnectivity of online communications comes the presence of disinformation. While the utilization of disinformation by individuals is harmful, the tactic is worse when adopted by foreign nations who seek to influence politics, finances, or ideological beliefs around the world. Despite the prevalence of state-backed disinformation campaigns - known as information operations - there remains little means for efficiently attributing an operation to the responsible nation-state. This research confronts that problem by constructing an algorithm for associating a state-backed information operation on Twitter with its nation-state actor. Similar research into nation-state hacking groups known as APTs has proven successful, offering support for a behavior-based approach. Sets of labeled data made available by Twitter were utilized to identify three characteristics of these nation-state actors, aka Advanced Persistent Manipulators (APMs). These characteristics are (1) methods used, (2) motivations, and (3) topics discussed, and each was identified through various methods ranging from predictive models to text analysis. Once each characteristic was identified, the three were quantified and used to train a KNN classification model. The result was a model capable of attributing individual tweets from an information operation with an accuracy of approximately 92%. However, caution is advisable in the model's implementation due to the ethical ramifications of accusing a nation-state of an information operation based solely on a predictive model. That said, the algorithm demonstrates potential for the use of behavioral analysis in information operation attribution, and further proves that attribution of these operations is both possible and automatable.

Awards Won:
Office of Naval Research on behalf of the United States Navy and Marine Corps: The Chief of Naval Research Scholarship Award of $15,000
Fourth Award of $500