Time Series Analysis of Anonymized Communication Channels

Summary
Collaborating groups of highly sophisticated transnational criminals use anonymous communication networks to coordinate their activities, and investigators need tools to uncover such activity. This project leverages recent advances in time-series data mining to analyze anonymized and encrypted network traffic and unmask coordinated criminal actions, ultimately providing investigators with a set of methodologies, algorithms, and software tools that can be used to classify, correlate, and thereby discover and understand the operation and structure of such groups.

Problem addressed
This project will provide the tools to uncover well-organized transnational criminal and terrorist enterprises that use anonymizing communications. The toolkit is meant to operate using publicly available – open source – data logs and repositories that record communication patterns and requests for network information, such as those sent to the global Domain Name System (DNS). The purpose of the project is to identify collaborating groups of highly sophisticated transnational criminals that use anonymous communication networks to coordinate their activities. We assume that the structure of the transnational criminal groups we target will be specialized, have subgroups that develop and maintain sophisticated, global communication strategies, and possess tactical and technical abilities similar to global terrorist groups. We further assume that the subgroups we examine have both ongoing and opportunistic, “one-off” collaborative relationships.

Approach
Our approach is to use recent advances in time-series data mining to analyze anonymized and encrypted network traffic by using both supervised and unsupervised machine learning methods and to use a novel methodology that discovers network-service information by correlating passive network measurements leaked across multiple types of protocols in order to unmask coordinated criminal actions. Our time series analysis will be enhanced by the incorporation of our recent work in protocol-level scoped privacy. Previous work logically intersected specific information that is leaked by protocols, where it is leaked (i.e. scopes), whether to trust/distrust the scopes that it is leaked into, and what actions (privacy exploits or security protections) could be overcome using the specific leaked data in those specific scopes. We will use suspected substandard privacy settings as a filtering mechanism to focus analysis on suspected “weak” usages of anonymized communication infrastructures.

Anticipated Impact for DHS

This work directly supports DHS efforts to identify and combat transnational criminal and terrorist organizations. The tools and techniques developed in this work will enable DHS to discover and understand the operations and structure of criminal organizations that use anonymizing communication channels.