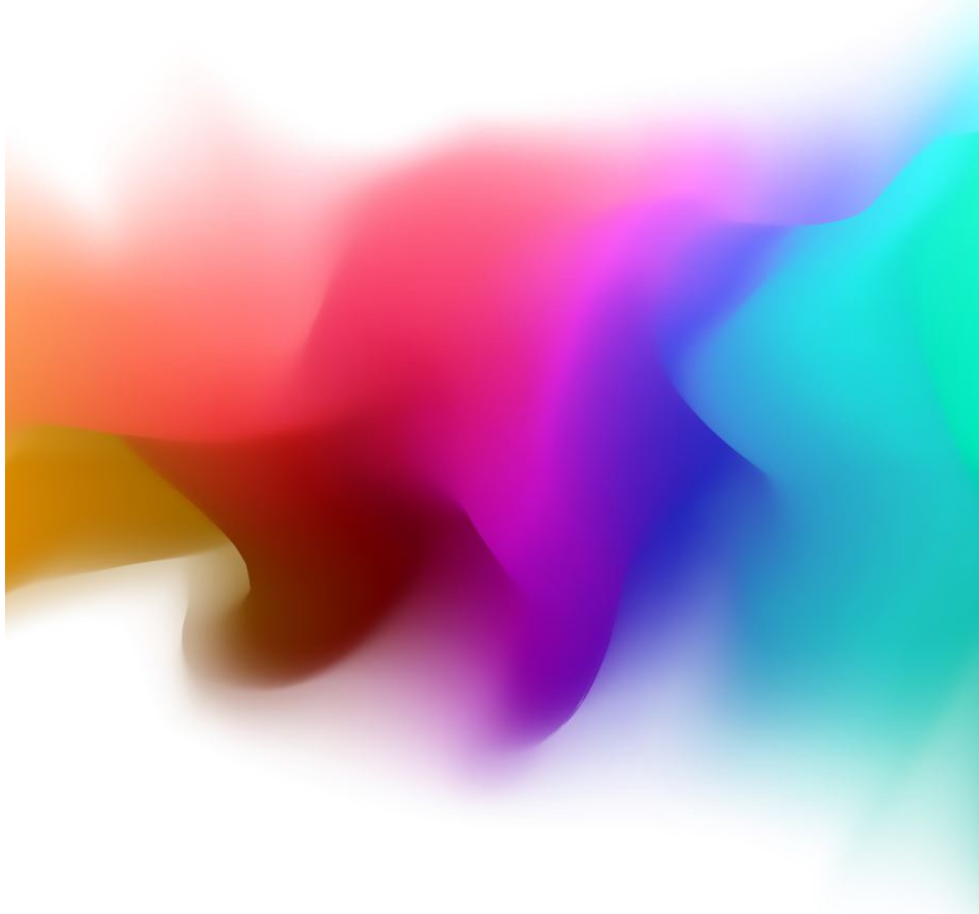
Find Me If You Can: Uncovering and Protecting Anonymized Communication Channels

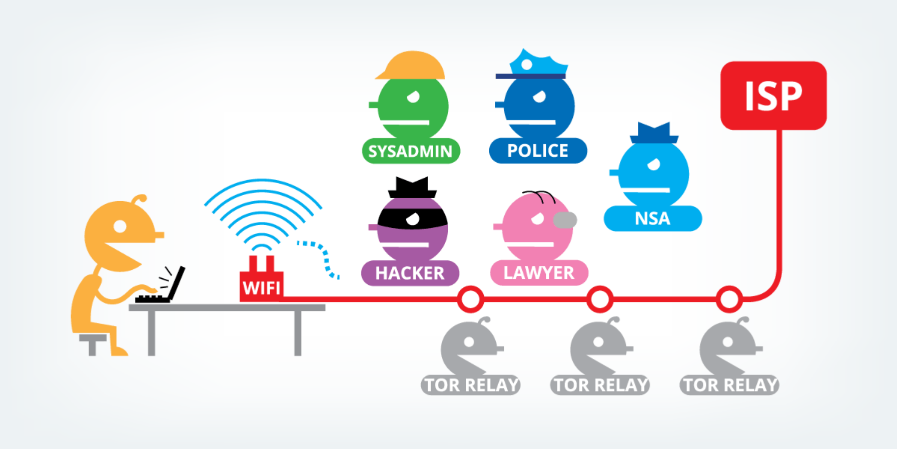
PROF. BOB SIMON
COMPUTER SCIENCE DEPARTMENT
C4I & CYBER CENTER



Why anonymous communication?

The good

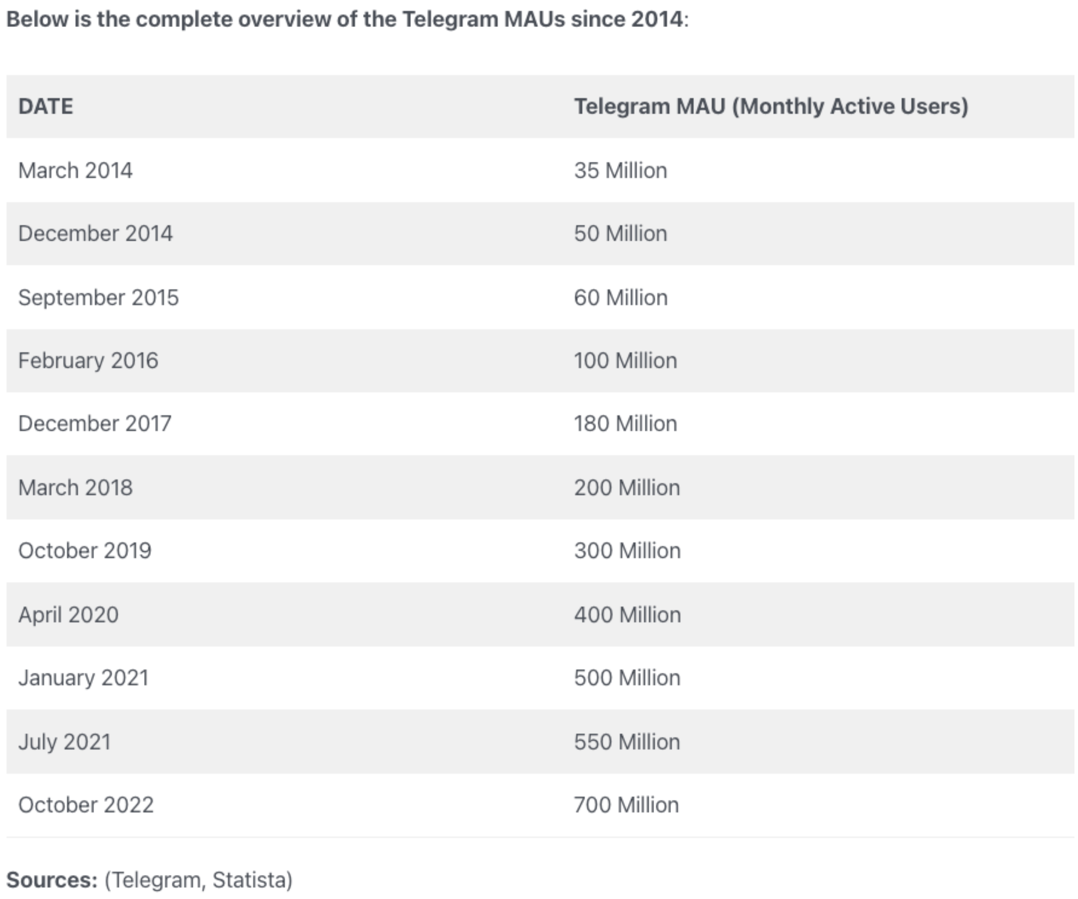
To advance human rights and freedoms by creating and deploying free and open source anonymity and privacy technologies, supporting their unrestricted availability and use, and furthering their scientific and popular understanding. – **Tor website**



The bad

- Dark Web services
- Internet trolling and misinformation
- Botnet command and control
- Cyber attacks

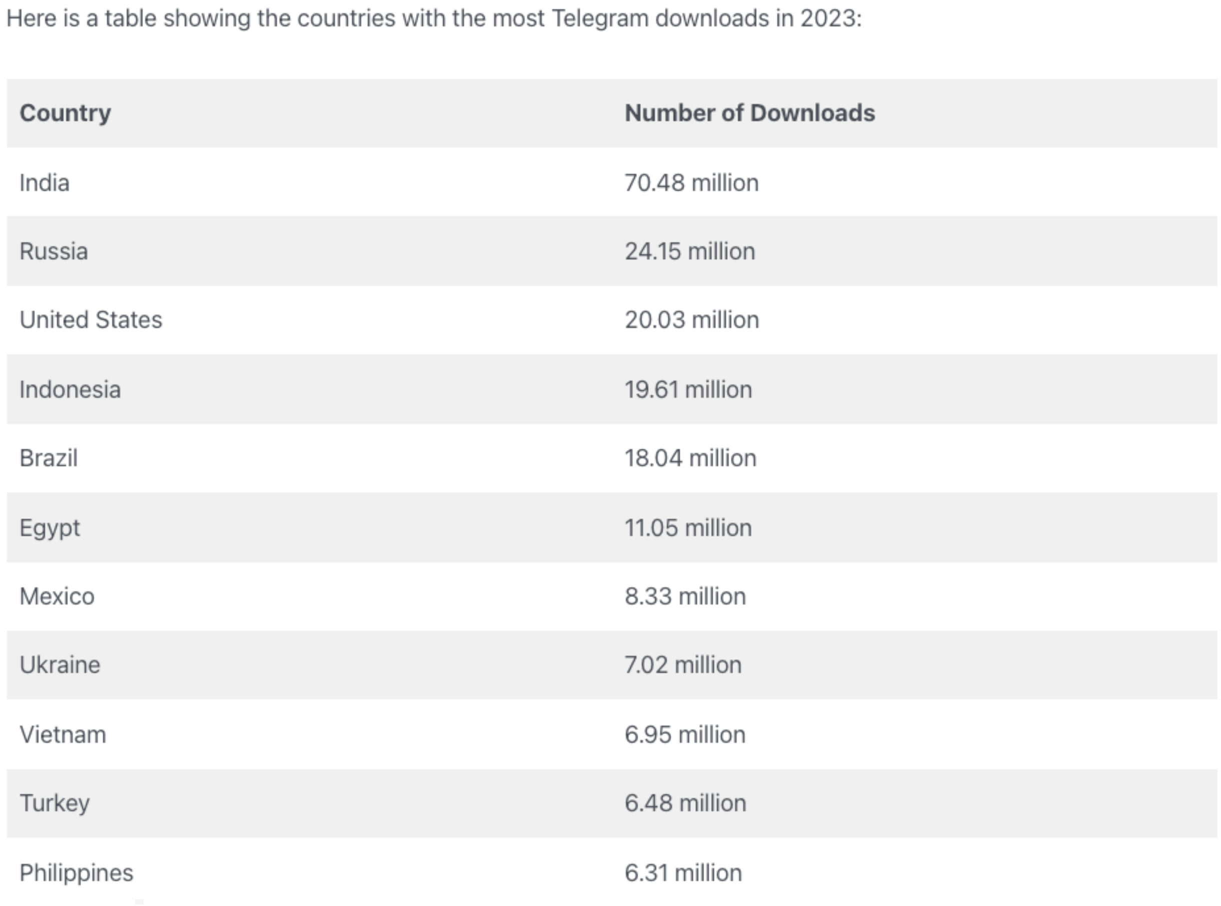
Example: (self-reported) Growth of Telegram



Below is the complete overview of the Telegram MAUs since 2014:

DATE	Telegram MAU (Monthly Active Users)
March 2014	35 Million
December 2014	50 Million
September 2015	60 Million
February 2016	100 Million
December 2017	180 Million
March 2018	200 Million
October 2019	300 Million
April 2020	400 Million
January 2021	500 Million
July 2021	550 Million
October 2022	700 Million

Sources: (Telegram, Statista)



Here is a table showing the countries with the most Telegram downloads in 2023:

Country	Number of Downloads
India	70.48 million
Russia	24.15 million
United States	20.03 million
Indonesia	19.61 million
Brazil	18.04 million
Egypt	11.05 million
Mexico	8.33 million
Ukraine	7.02 million
Vietnam	6.95 million
Turkey	6.48 million
Philippines	6.31 million

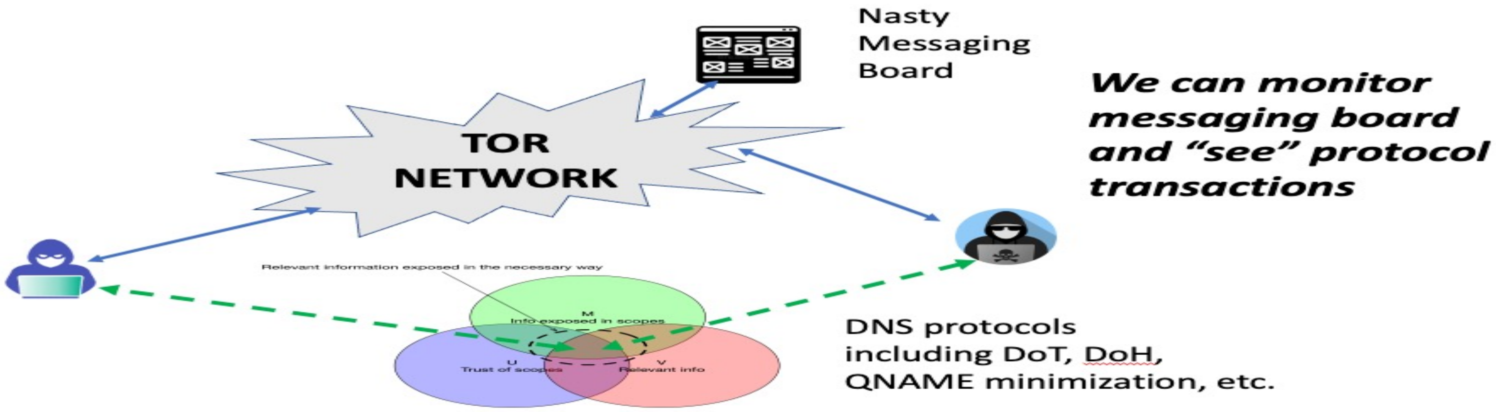
Project: Time Series Analysis of Anonymized Communication Channels

Co-PI: Prof. Eric Osterweil

Sophisticated transnational criminal organizations (TCOs) use global anonymizing networks to manage their activities

It is possible to de-anonymize some communication to identify participants and understand behavior

Approach is to uncover TCO pattern-of-life in anonymous networks



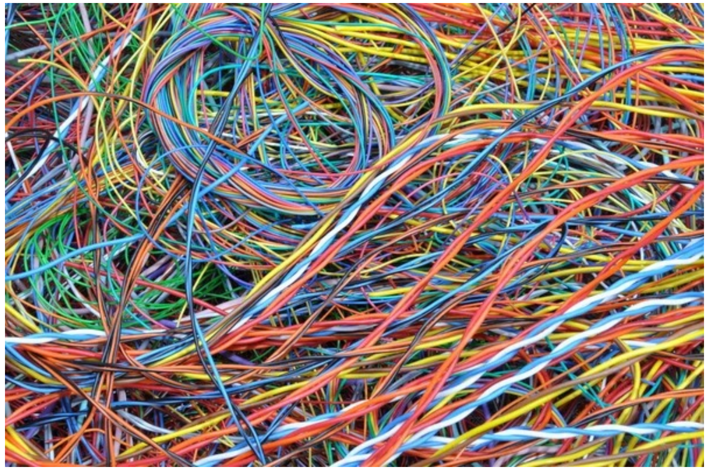
Nasty Messaging Board

We can monitor messaging board and "see" protocol transactions

DNS protocols including DoT, DoH, QNAME minimization, etc.

Criminal Investigations and Network Analysis
A DHS CENTER OF EXCELLENCE

Challenge and Research Question



This is a big noisy system

Need to rely upon “Pattern-of-Life” assumptions

Basic Research Questions

What are the minimal sets of network sensor locations and/or access to network management logs necessary in order to “find the hidden collaborator(s)?”

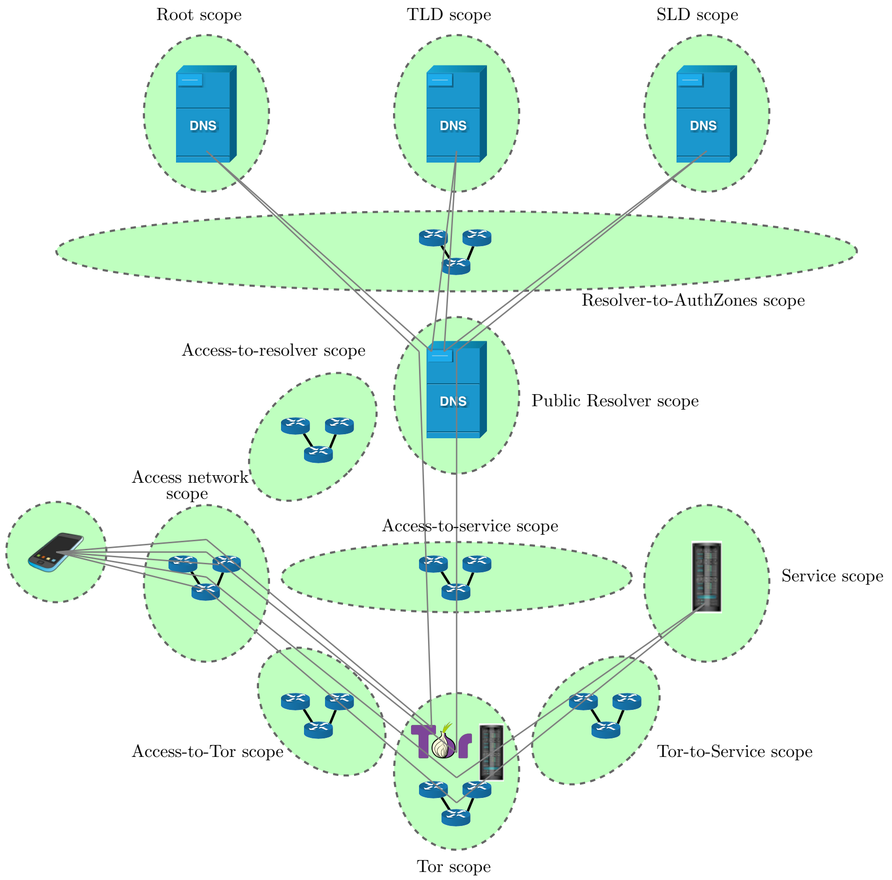
What features do we need to extract for our ML models?

Technical Approach

Build out “realistic” testbed consisting of DNS, Internet simulator (GNS3) and Tor Software

Algorithm

1. PCAPs turn into dataframes on a per IP basis if IP is seen in packet or if cache entry was valid when packet was sent
2. 1 second sliding-sum over features extracted for each scope
3. Feature selection based on scopes available
4. Signals are transformed using Topological Data Analysis-Persistent Landscape to capture multivariate PoL
5. Above steps are repeated for features of the service of interest
6. PoLs seen in scopes is compared in PoL of the users on the service of interest using cross correlation for the time the persona was using the service
7. Signals with the highest cross correlation are assumed to be the same persona



Experimental results

Used a year long message log of a well-known large scale social network application. The dataset consists of 948,169 topic-driven interaction sites. The **database is fully anonymized.**

The dataset uses timestamps to log users and conversational threads.

Objective is to cluster users into groups of “interest” or “not-of-interest.”

Leading candidates for feature selection are relative interarrival time and message length. We reached this conclusion after a Grid search of 30 features

Topological Data Analysis using Persistent Landscapes is an effective preprocessing step

81% accuracy with vanilla DNS pre tor

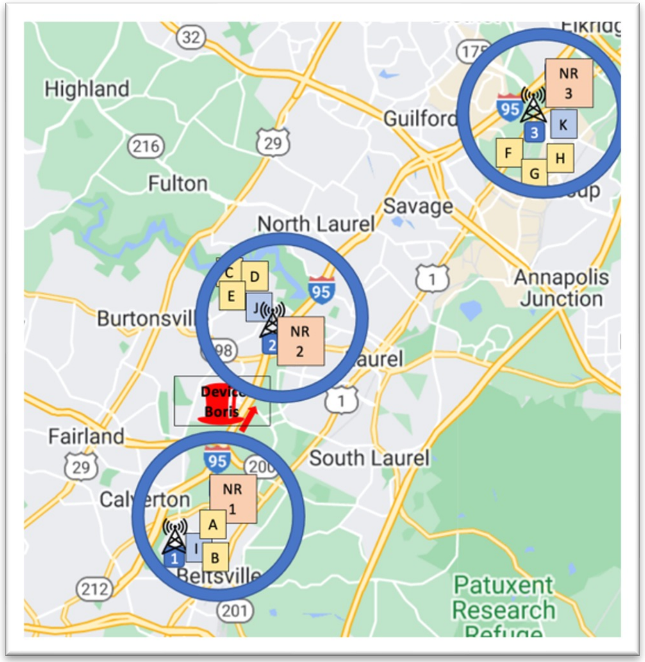
56% accuracy with DoT pre tor

Project: 5G Network Level Awareness

Global growth of 5G networking

Increasingly common for military operations, first responders or routine public safety patrols to use communication infrastructure managed or accessible by neutral or even hostile entities

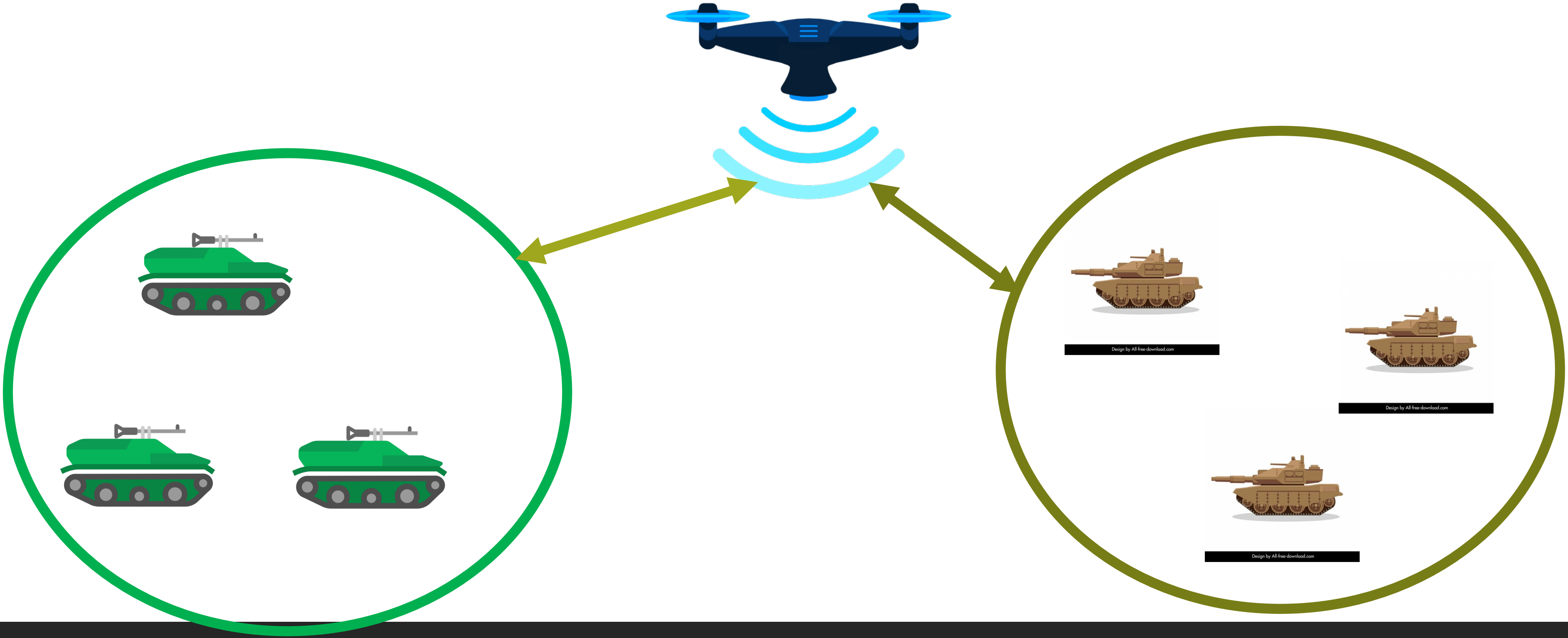
Question: how to protect against a range of cyber attacks?



Axiom, Inc

Overlay networking and group communication

Rapidly form hierarchy of communicating entities


Design by All-free-download.com
Design by All-free-download.com
Design by All-free-download.com

Moving Target Defense

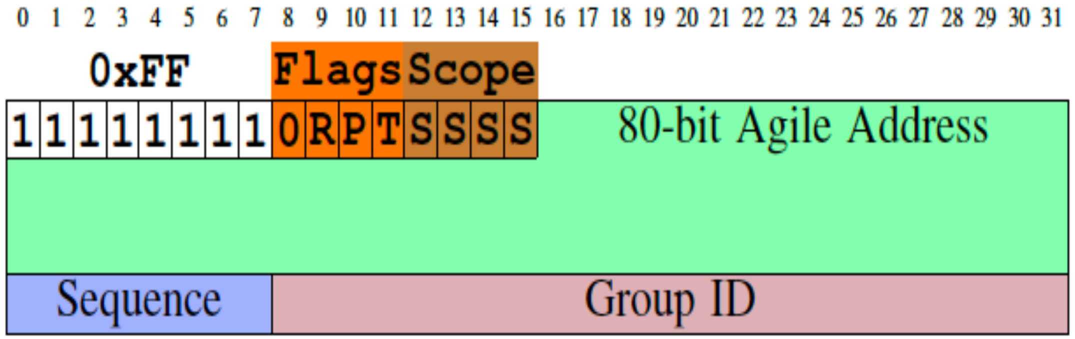
Cryptographically rotate addresses

Two levels

- Network Level (IP Addresses)
- Embed keyed nonce inside message

Developed and implemented attack taxonomy

Run analytics to detect attack and rekey group



SARCAST Addressing Scheme in IPv6

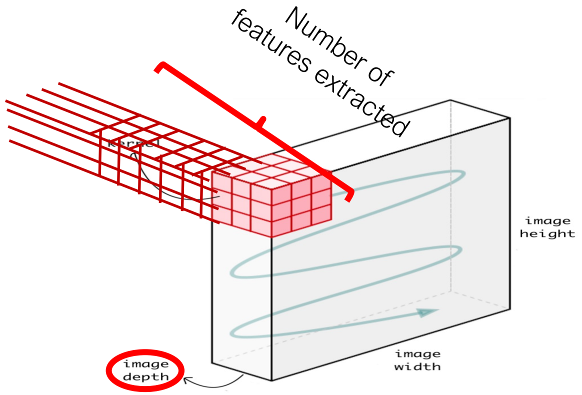
Project: Detection of Misbehaving Internet of Things Devices

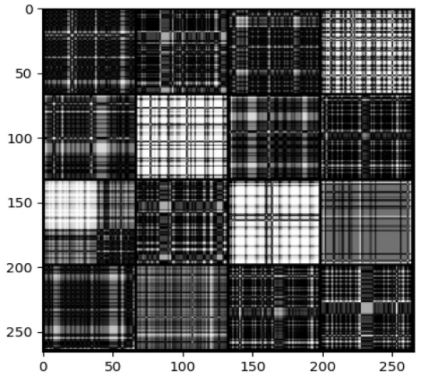
IoT devices are network reachable, hard to patch and resource constrained

We use conversational-based packet captures used to identify misbehaving devices

Transform into images using stacked Gramian Angular Fields

Transfer Learning for multiple protocols (Bluetooth, Z-Wave, Zigbee, LoRA, to name several)





Resulting Image for single “conversation”

Classification and anomaly detection by convolutional neural network

Observations

Really fun to work in C4I and Cyber Space

Project up next: Hierarchical UAS (Joint with Prof. P. Pathak)

Questions?