CS 294-28 / Network Security / Fall 2009

Scams

Scribed by Michael Zhang

- Bot Defenses
    - Defenses for Bots are dissatisfying
    - Comes in several categories
        - Usual malware detection problem
            - Nothing particularly special about how infections occur
            - Can try to prevent using various means i.e. intrusion detection, tripwires, etc...
        - Vulnerabilities in the associated command and control
        - Vulnerabilities in the associated market
        - Correlate attack activity with C&C activity
            - This is the angle unique to botnets
            - Has proven hard in practice (not much additional leverage beyond just detecting either attack or C&C activity by itself)

- Modern Cyber-Attack Ecosystem
    - Underground economy
    - Measuring sending of spam
        - Inside *Storm* – Botnet infiltration
    - Measuring the fruits of spam
        - Scam infrastructure
    - Profitability
        - Phishing
        - Spam conversion
- How big is the problem?
    - Are we dealing with a $100 billion drain on the worldwide economy? Is it something quite minor? Do our foes have awesome resources or are they losers that are just squeaking by?
    - How many players?
    - What sort of fish? Big or little?
    - Uber-hackers? Whack-a-mole?
- How to soundly measure
    - Get scammers to reveal
    - Run scams
-
- In marketplaces there's a phenomena called a *Ripper*
    - Someone who rips you off
    - Potentially makes the marketplace inefficient / difficult to measure/assess
- The *Storm* Botnet
    - Uses the Overnet peer-to-peer system for part of its C&C

- OID: Overnet ID
    - 32-bit identifier

- Estimating population sizes: **Mark and Recapture**
    - Scheme that came out of wildlife management
        - i.e. You want to know how many bald eagles are in the park
    - Capture a sample group, size C0
        - Mark (i.e. tag them) and release them
        - Capture a sample group again, size C1
        - R are marked
        - Population estimate: (C0*C1)/R
    - What does this require?
        - *Independence*
            - If the probabilities are correlated, then this is no good
        - *Stationarity* (closed system)
            - You can't have your eagles flying off elsewhere or new ones fly in
    - Variance: $[C0*C1*(C0-R)(C1-R)] / R^3$
    - Example - In one campaign...
        - C0 = 1.8 million addresses seen by crawler
        - C1 = 3015 of those spams seen in spamtrap (arguably independent)
        - R = 8
        - Population estimate = 677 million addresses
        - Variance = 239 million