Detecting DGA-based Botnet with Outlier Detection

Teppei Fukuda, Tomohiro Ishihara, Akira Kato
The University of Tokyo, Keio University

**Problem**

- **Goal:** Detect DGA-based botnet.
- Each bot dynamically generate numerous random domain names and use a small subset as C&C.
- **Difficulties to detect DGA-based botnets**
  - C&C domain names continue to be updated
  - We don’t know details of the algorithm
  - Algorithm continue also changes

**Proposed Approach**

**Insight**

- DNS queries for DGA-generated domain names increase suddenly only a short period of time (temporal locality)

**Approach**

1. Extract suspicious domains applying outlier detection to total number of requests per day
2. Classify by the statistical features (n-gram, entropy, etc.)

**Methodology**

Input: Non-existent domains which were queried in the academic network.

**Extraction**

1. For each domain, calculate time-series data about total number of requests per day
2. Apply median-absolute-deviation (MAD) based outlier detection for above data
3. Extract domains which have outlier as suspicious domains

**Classification**

Prepare: Build SVM models for DGAs by well-known DGA domains
4. Classify DGA domains by SVM using the statistical features

**Temporal Locality**

**Dataset:** DNS traffic at the Upper DNS Hierarchy
**Span:** 2014/04~2014/05

- Unique number of sender IP address which query non-existent domains generated by DGA or antivirus software.

**Results**

**Dataset:** DNS traffic at cache DNS server
**Span:** 2014/04/21~2014/05/05

1. ROC for SVM from cross validation
   - AUC: 0.982925

2. C&C detection result a day on average
   - Truth: blacklist(DNS-BH)
   - | Predict | C&C | Legitimate |
   - |------:|-----|-----------|
   - |     5 |     1 |           |
   - |    9  | 991679 |           |