Problem Def

ML Paper (Sept 26, 2016)

- detection is based on:
  - URL
  - flow duration
  - number of bytes trasferred from client to server & other way
  - user agent
  - referer
  - MIME-type
  - HTTP Status

- The n-dimensional feature vector represents each proxy log & used to differentiate b/w legit & malicious traffic

- Paper model only analysis single proxy log, & skips temporal features

- Attacking domains change frequently but behaviour doesn't.

How

- The proxy logs originating at a particular user machine are grouped into bags based on the domain in the URL.

- The bags are labeled according to the domain.
  - ↳ if domain is in any blacklist, the bag has a +ve label
  - ↳ if not, bag has a -ve label

MIL

- flow is described by a vector of features

$x \in X \subseteq \mathbb{R}^d$ & a label $y \in Y = \{+1, -1\}$

(+1: malicious, -1: not)

- Network traffic monitored in a given period is fully described by the completed annotated data

$$D_{cmp} = \{(x_1, y_1) \dots (x_m, y_m)\} \in (X \times Y)^m$$

assumed to be generated from i.i.d. (independant, identically distributed) random vars with unknown dist

$$p(x, y)$$

- Annotating everything is expensive, thus we use bags of flows
- The weakly annotated data

$$D_{bag} = \{\underbrace{x_1, \dots, x_m}_{\text{features}}, \underbrace{(B_1, z_1), \dots (B_n, z_n)}_{\text{assignment to labled bags}}\}$$

$$\{(B_1, z_1), \dots (B_n, z_n)\} \in (P \times Y)^m$$

P = set of all partitions of indeces $\{1, \dots m\}$.

- The ith bag is a set of flow features $\{x_j \mid j \in B_i\}$ labeled by $z_i \in Y$.

- D_{bag} carries partial info about D_{comp}.

Assumptions:

1) Flow features $\{x_1, \ldots x_m\}$ are the same in both.

2) Negative bags contains a single instance, & the label is correct.

$\Rightarrow z_i = -1$ implies $|B_i| = 1$ & $y_i = -1$

3) +ve bags have a variable size & at least 1 instance is positive

$\Rightarrow z_i = +1$ implies $\exists j \in B_i$ s.t. $y_j = +1$