# Theory

The purpose of the classifier to evaluate an IPv4 address and assign it to a classification based upon the behavior of the content which it is hosting.

## Server Classifications (tentative)

* Sinkhole
* Mail
* Parking
* Shared
* Corporate
* Private
* DynDNS

## Dependencies

* Deteque
* IMS
* Alexa (planned)

# Process

* Only evaluate IPv4 which have at >50 1-day OR >500 3-day Deteque records. This lower bound is used to minimize the requests sent to Actor Infrastructures which typically are found to be hosting minimal content, if any at all.
* Results *typically* fall in one of two main categories:
  + Sinkhole/(Shared/Private/Corporate): Low number of Deteque records
  + Sinkhole/Parking: High number of Deteque records

### Profile

* Sinkhole Servers
  + **Domains**: Unresolvable, 4xx status codes, timeout errors
  + **Wildcards**: Unresolvable, 4xx status codes, timeout errors
* Mail Servers
  + **Domains**: Unresolvable, 4xx status codes, timeout errors
  + **Wildcards**: Resolvable, 2xx status codes, redirects
* Parking Servers
  + **Domains**: Resolvable, 2xx status codes, redirects
  + **Wildcards**: Resolvable, 2xx status codes, redirects
* Shared/Private/Corporate Servers
  + **Domains**: Resolvable, 2xx status codes
  + **Wildcards**: Unresolvable, 4xx status codes, timeouts

# Proposals

## General

* Check for existence of domains in Alexa 1 Million
* Evaluate nameservers and establish a DB of known parking/sinkhole usage.
* Run as daily task to track the daily fluctuation of domains on an IPv4

## Sinkhole:

* Domains resolve to an ip for a long time, DNS A record is persistent
* If a Sinkholed domains moved to another ip, its usally to same org or same ASN
* Numerous domains point to an a sinkhole IP irrespective of the malware family (sharing resources)
* Domains resolving to the same sinkhole ip share the same NS record
* When sent an http request
  + No response
  + 4xx error
  + All domains on the same ip have the same content
* If ip is found to be a neighboring node of at least two sinkholes, we define ip to be a potential sinkhole. Further, ip is added to spot when it is a parent or child node of at least one sinkhole in the same AS and the domains resolving to both IP addresses share the same NS

# Issues

1. Wildcarding: While the use of wildcarding has resulted in identifying many Parking Servers, it is not without complication. We are currently using the assumption that if domain resolution and wildcarding resolution have a high number of 200 status codes with low disparity that we can assume the IPv4 is hosting parked domains. The image pasted below however shows this to not be the case as some sites by default return content with a 200 response code. This behavior is actually disruptive to the expected behavior of the web as they are artificially generating a 404 page.

