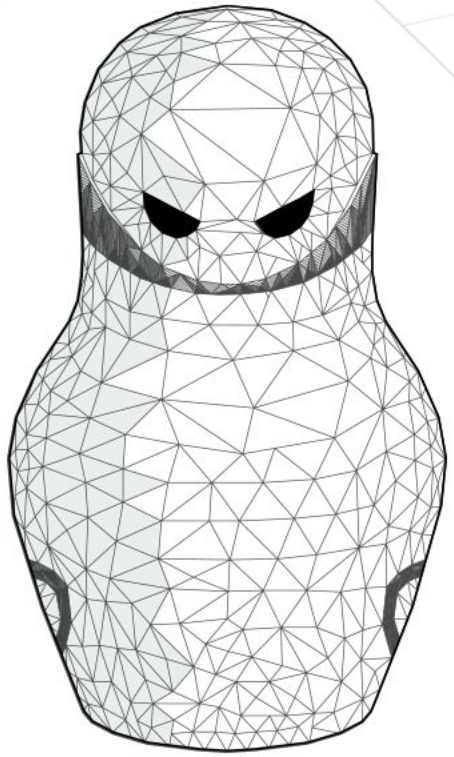


2015 ZERO NIGHTS

Data Mining in the service of nmap

Sergey Ignatov, Omar “beched” Ganiev
sergey@ctf.su, beched@incsecurity.ru

ZERO NIGHTS 2015

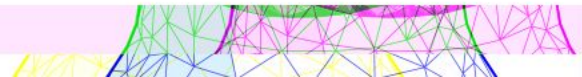
Data Mining in the service of nmap
$who

SECURITY

Проблема

Мир не стоит не месте, все развивается, сетевые сканеры в том числе. Начиная от дедушки nmap 1.0 (September 1, 1997), сканеры развиваются в раличных направлениях.

- Сканирование всего интернета (zmap)
- Сканирование уязвимостей в сервисах (nmap scripting engine и другие коммерческие сканеры безопаности)
- Интегрированный TCP стек (polarbear from IOActive)

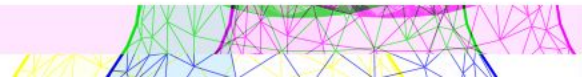


Но сканирования обнаруживаются IDS/IPS :(

Обход

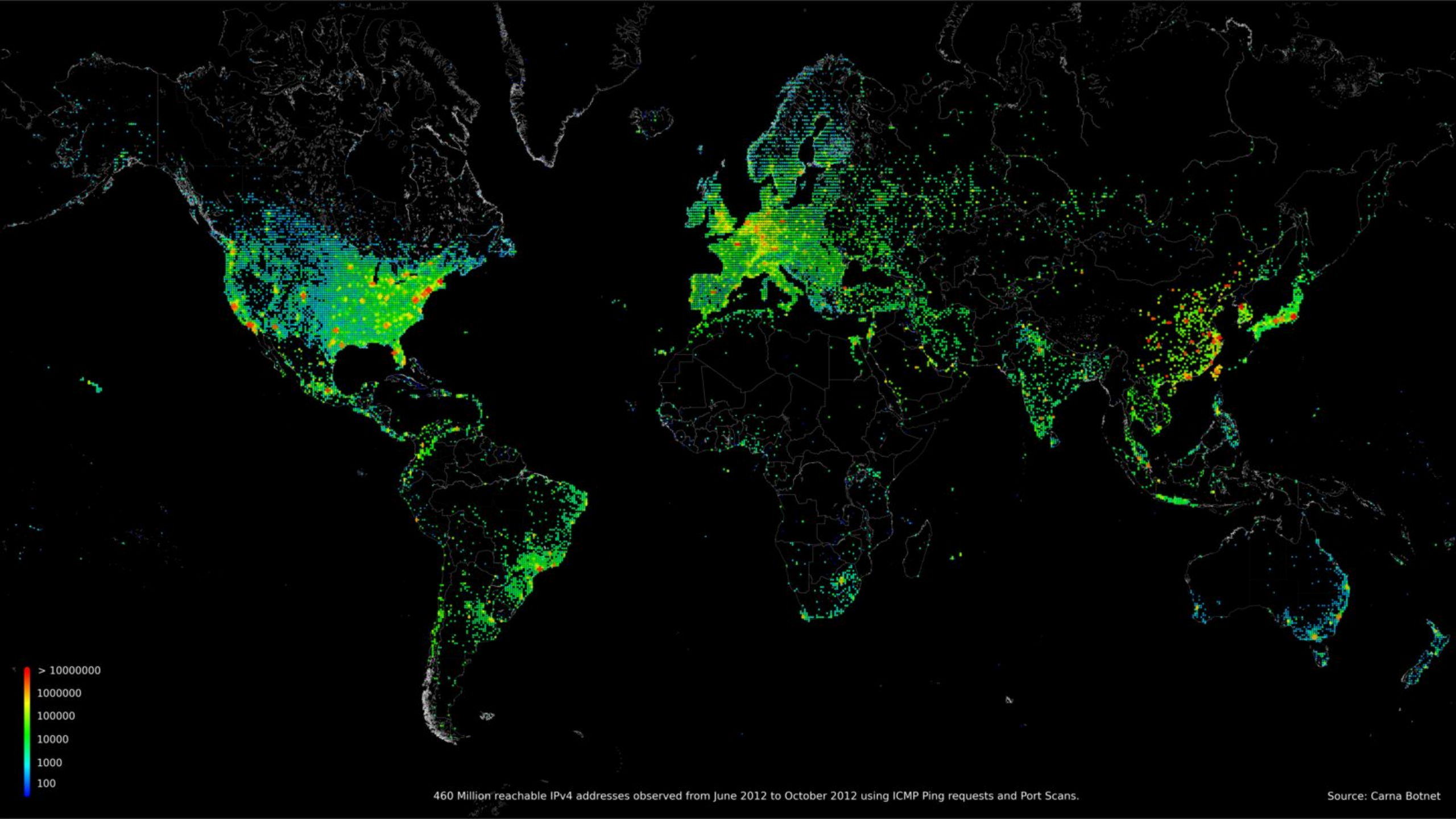
- Сканирование через сторонние сервисы (ftp bounce, proxy)
- Уязвимости в самом IDS/IPS
- Всякие сетевые хитрости (https://nmap.org/book/man-bypass-firewalls-ids.html)
- *Сканировать медленнее чем IDS хранит историю*
- Рассмотрим именно такой подход сканирования с задержками, мы хотим его сделать эффективнее

Как найти максимальное количество портов за минимальное количество запросов?



Data Mining in the service of nmap
Internet Census 2012

- http://internetcensus2012.bitbucket.org
- 420K взломанных домашних маршрутизаторов (пароли по умолчанию)
- Просканировали весь интернет!
- Dataset: 9TB of raw logfiles (500Gb in zpaq)

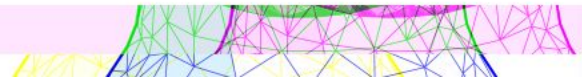


460 Million reachable IPv4 addresses observed from June 2012 to October 2012 using ICMP Ping requests and Port Scans.

Source: Carna Botnet

Методика

- На основе данных о всех хостах в интернете строим оракул, который будет предсказывать наиболее вероятные порты
- Идея очень проста: порты не встречаются сами по себе, они встречаются группами, так давайте хранить не список самых вероятных портов, а список самых вероятных групп



Методика

- По сути группы соответствуют различным классам сетевых устройств/сервисов
- Если нашли открытые порты X,Y и закрытый Z, то вероятнее всего будут открыты порты из групп где есть порты X или Y, и нет Z

Методика

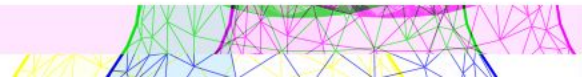
- Сначала мы пытались применить всякие умные алгоритмы и machine learning, делая вид, что умеем это
- Но дерево решений для сканирования будет экспоненциального размера, значит, его построение тоже экспоненциально
- Алгоритмы классификации не в тему

Методика

- Что ещё можно? Можно пробовать кластеризовать результаты скана и получить классы сервисов (“сайт”, “циска”, “винда”, …)
- Проблема: в датасете интернет-скана много NAT-узлов, отравляющих обучение
- Проблема: в датасете много зафаерволенных хостов, в которых виден только порт 80

Методика

- Для кластеризации нужно метрическое пространство
- Можно взять какой-нибудь simhash, или можно взять длину LCS, которая динамически вычисляется за произведение длин последовательностей
- Но у нас около 10^7 уникальных последовательностей, а подходящая кластеризация работает не лучше чем за квадрат
- К тому же из NAT-последовательностей надо вычленять разумные подпоследовательности портов

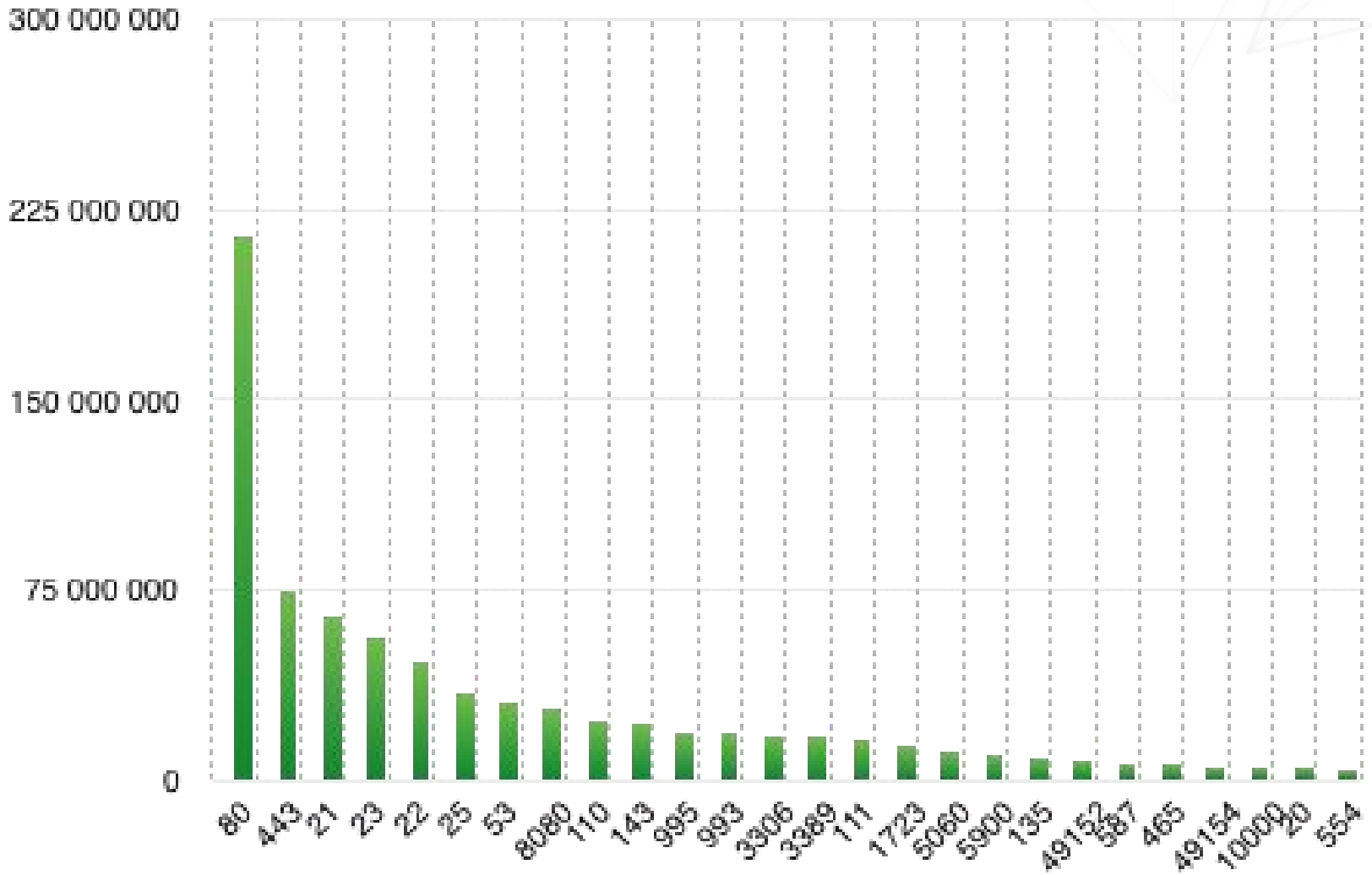


Методика

- Keep It Simple, Stupid!
- Давайте просто сортировать статистику =)
- В качестве базовой статистики, по которой начинаем сканировать без предварительных знаний, для начала возьмём топ портов

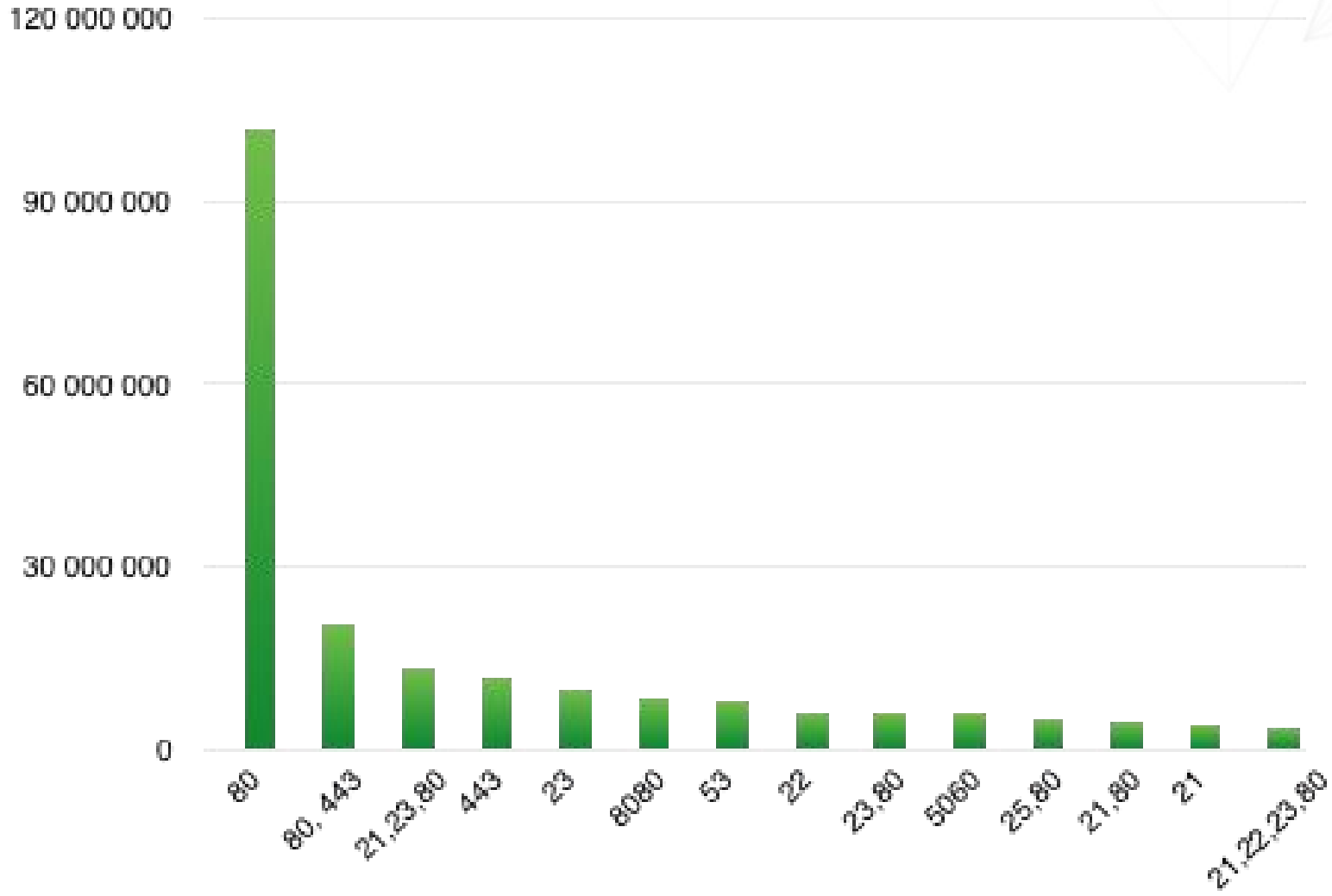
Data Mining in the service of nmap

Сколько открытых портов в интернете?



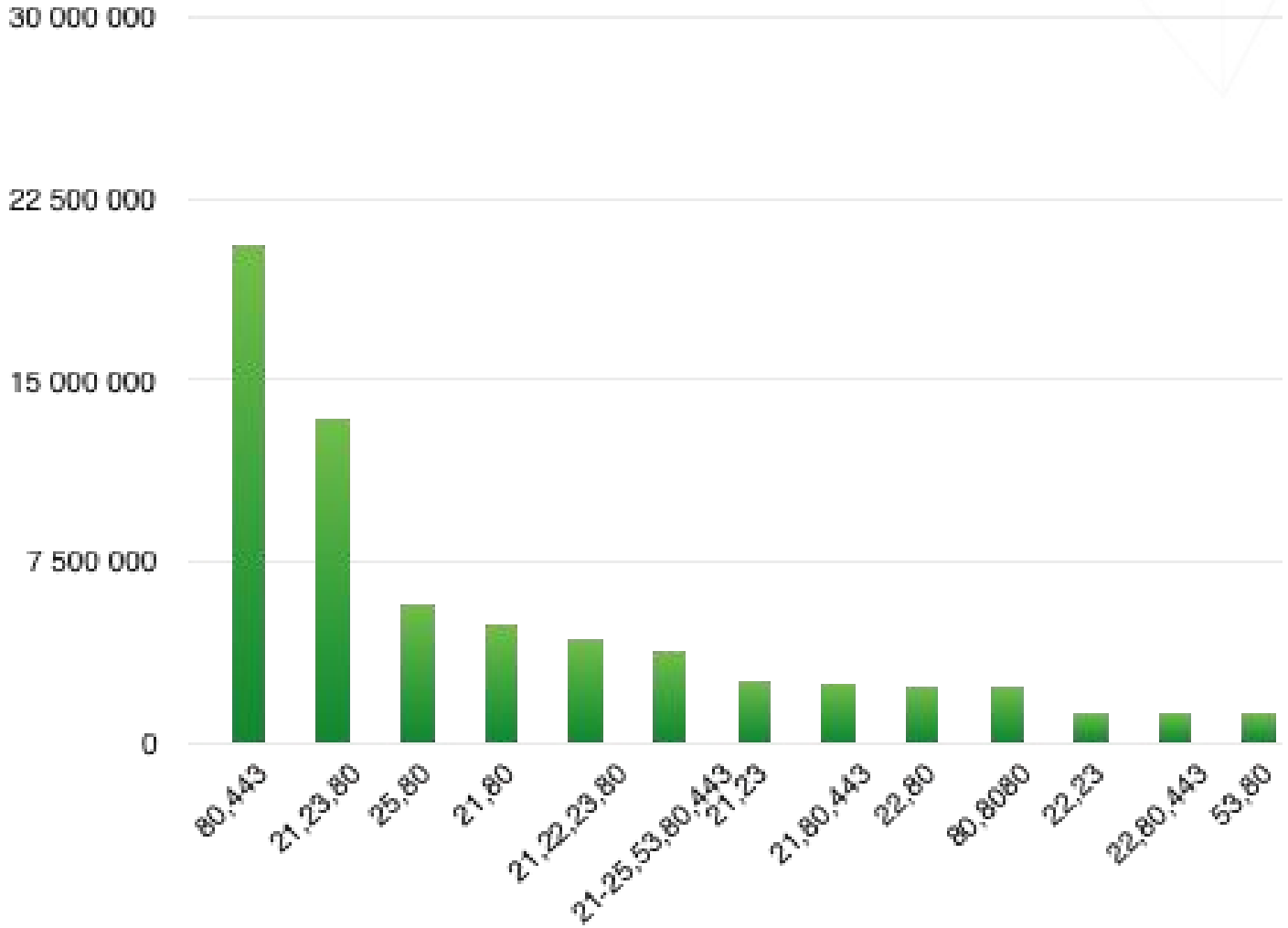
Data Mining in the service of nmap

TOP13 самых встречаемых групп



Data Mining in the service of nmap

TOP13 самых встречаемых групп



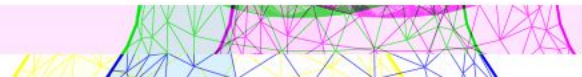
Чем плох NMap?

- Чем плох NMap? Много чем, там много наследия прошлого
- Главное - он медленный
- По умолчанию NMap сканит по топу портов (4243 порта), при полном наборе сканит случайным образом

Data Mining in the service of nmap

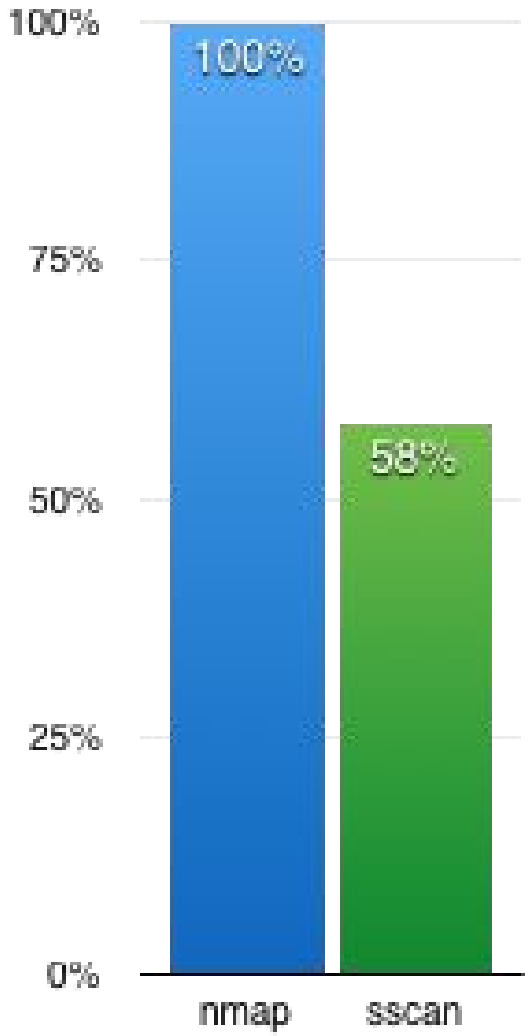
Результаты

- Пока что с лобовым алгоритмом и предварительной статой уже получилось ускорение на 42% относительно NMap
- Статистику измеряли путём имитации сканирования
- При этом сравнивалось, за какое количество запросов наш алгоритм находит столько же портов, сколько всего нашёл NMap (по умолчанию он не ищет дальше своего топа)



Data Mining in the service of nmap

Результаты



Планы на будущее

- Патч для nmap, изменяемая стратегия сканирования
- Данные по интранету (внутренние пентесты)
- Компактный словарь
  (сейчас 150Мб и самый полный 666Мб)
- Более точная кластеризация на основе данных протокола
  (Тип сервиса, версия, ОС и т.д.)
- ...

Data Mining in the service of nmap
Вопросы

Q?

sergey@ctf.su

beched@incsecurity.ru