We are CloudLinux

CloudLinux.com

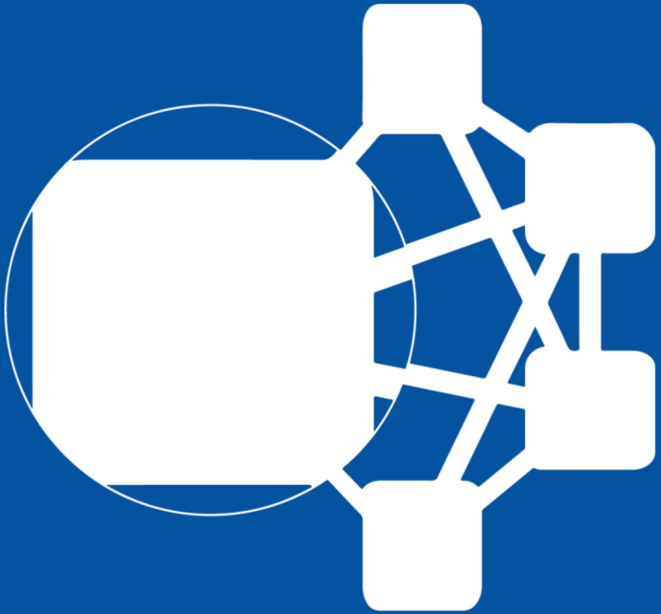
How big data helps secure cloud servers

Igor Seletskiy,
Founder and CEO of CloudLinux

What is a lot of data

- ~4M events a day
  - 4GB of data (with metadata)
  - ~20M events if not for IP blocking
  - Thousands of servers
  - ~1M websites
  - 500K distinct IPs / day

Infrastructure



We are CloudLinux

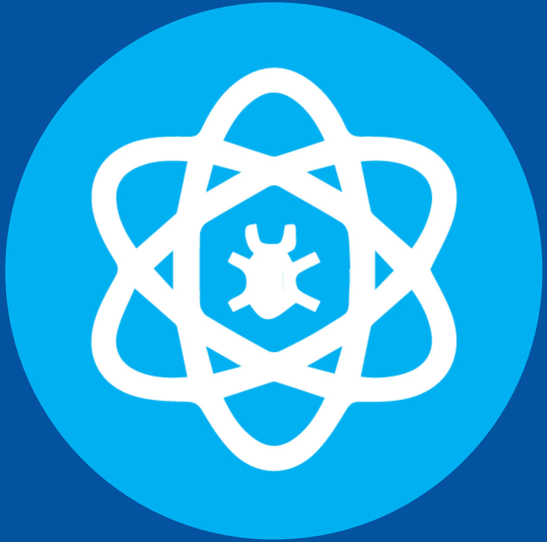
- Distributed Ruleset
- Spark Cluster
  - ~1T RAM, 100s cores
- HDFS
  - 3 replicas (going to 6)

- Brute force is #1
  - WordPress → huge
  - Mail → as big
  - SSH
  - FTP
  - Other CMS → Insignificant

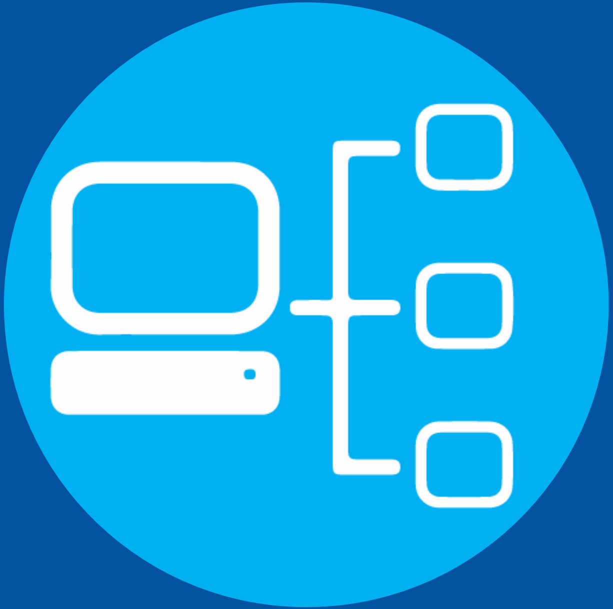
Some things are easy to see

Some things are easy to see

- Exploits → small, very spread out
    - WordPress attacks #1
    - Blind scans are common



# of IPs involved



- Mailbox brute force → subnets
- WordPress → everything
- Exploits → mixed
  - Subnets
  - Specific / 'stable' IPs
  - Datacenters
  - ISPs (DHCP)
  - TOR/VPN/proxy

# of IPs involved

- ~ 15-20K IPs blacklisted daily
  - 50% of IPs change their profile within 5 days
- More servers → more attacking IPs we see. Non linear…
- **Stopping 50x the attacks than a year ago**

We are CloudLinux

Subnets...

- Not a single Captcha passed
  - WordPress or Mail
- Rarely rotated

103.28.132.* WordPressCrackerBotNet
103.28.133.* WordPressCrackerBotNet
114.223.60.* MailCrackerBotNet
114.223.61.* MailCrackerBotNet
114.223.62.* MailCrackerBotNet
114.223.63.* MailCrackerBotNet
114.224.29.* MailCrackerBotNet
114.225.55.* MailCrackerBotNet
114.225.83.* MailCrackerBotNet
117.68.172.* MailCrackerBotNet
117.68.173.* MailCrackerBotNet
117.68.174.* MailCrackerBotNet
117.68.175.* MailCrackerBotNet
117.84.210.* MailCrackerBotNet
117.90.1.* MailCrackerBotNet
117.90.2.* MailCrackerBotNet
117.90.3.* MailCrackerBotNet
117.90.4.* MailCrackerBotNet
117.90.5.* MailCrackerBotNet
117.90.6.* MailCrackerBotNet
…..

Exploits Hunters

- Specific vulnerabilities
  - Often blind
- Generic exploits
  - SQL Injections
  - PHP Injections
    - HTTP Headers injections
- Dumb vs Smart
  - Some mimic real users

WAF

- No rules are really good
  - Too specific (catch just a few, that are also caught by other things)
  - Too generic (a lot of false positives)
- Some rules are really bad
  - Only false positives
    - *3rd party rules*
- Generic rules / catch anything
  - Correlate across many data points

TRACK BAD GUYS/DEVELOP NEW RULES

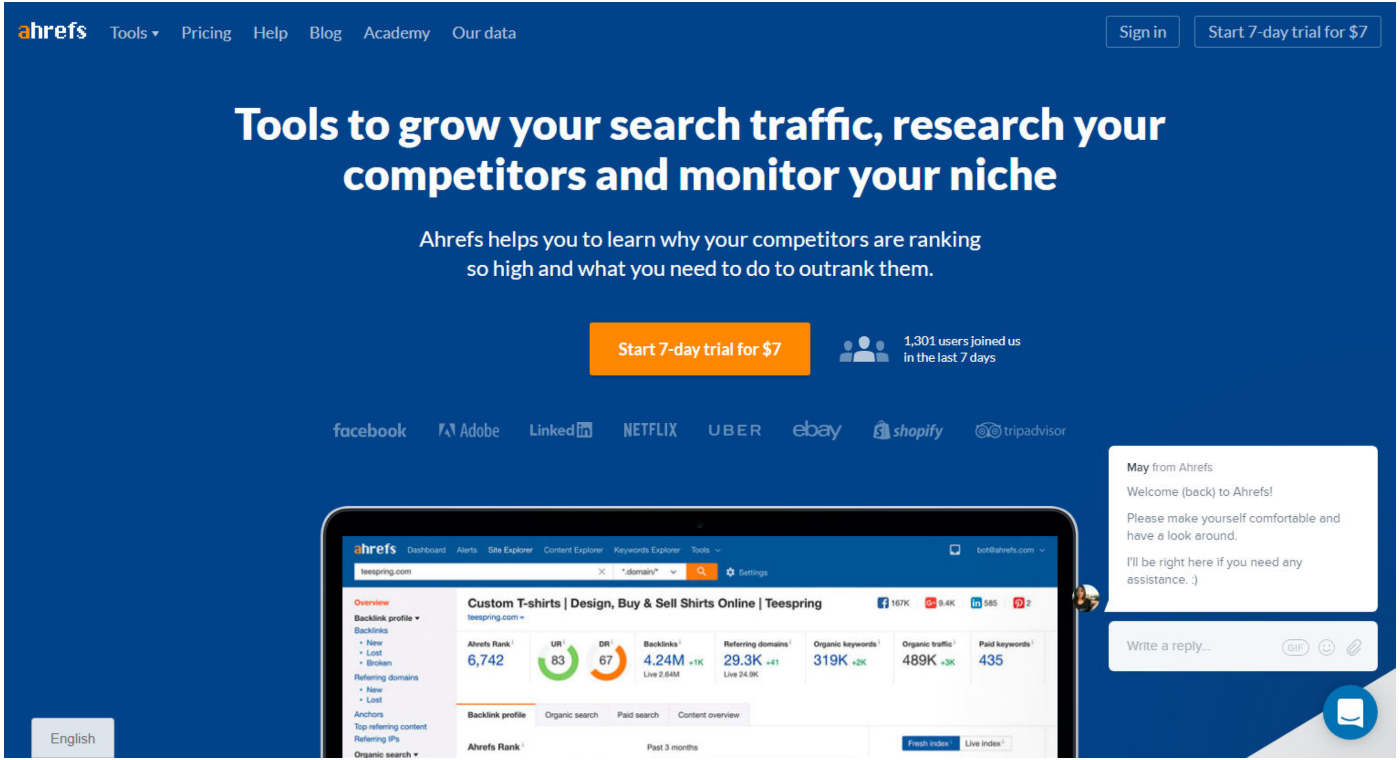
Captcha

- Was useful when we started
  - < 0.1% of blocked IPs would pass captcha
- Usefulness dropping
  - ~5% of blocked IPs pass it
    - anti-captcha.com, 1000 captchas for 50c
- JS/splash screens are no help
  - PhantomJS, selenium, etc...

✓ I'm not a robot

Captcha - Good Bots

Even ahrefs bot now passes captcha



Tools to grow your search traffic, research your competitors and monitor your niche

Ahrefs helps you to learn why your competitors are ranking so high and what you need to do to outrank them.

1,301 users joined us in the last 7 days

teespring.com | *.domain/*

Custom T-shirts | Design, Buy & Sell Shirts Online | Teespring
teespring.com ▾

167K 9.4K 585 2

Ahrefs Rank	UR	DR	Backlinks	Referring domains	Organic keywords	Organic traffic	Paid keywords
6,742	83	67	4.24M +1K Live 2.64M	29.3K +41 Live 24.9K	319K +2K	489K +3K	435

Ahrefs Rank   Past 3 months

May from Ahrefs

Welcome (back) to Ahrefs!

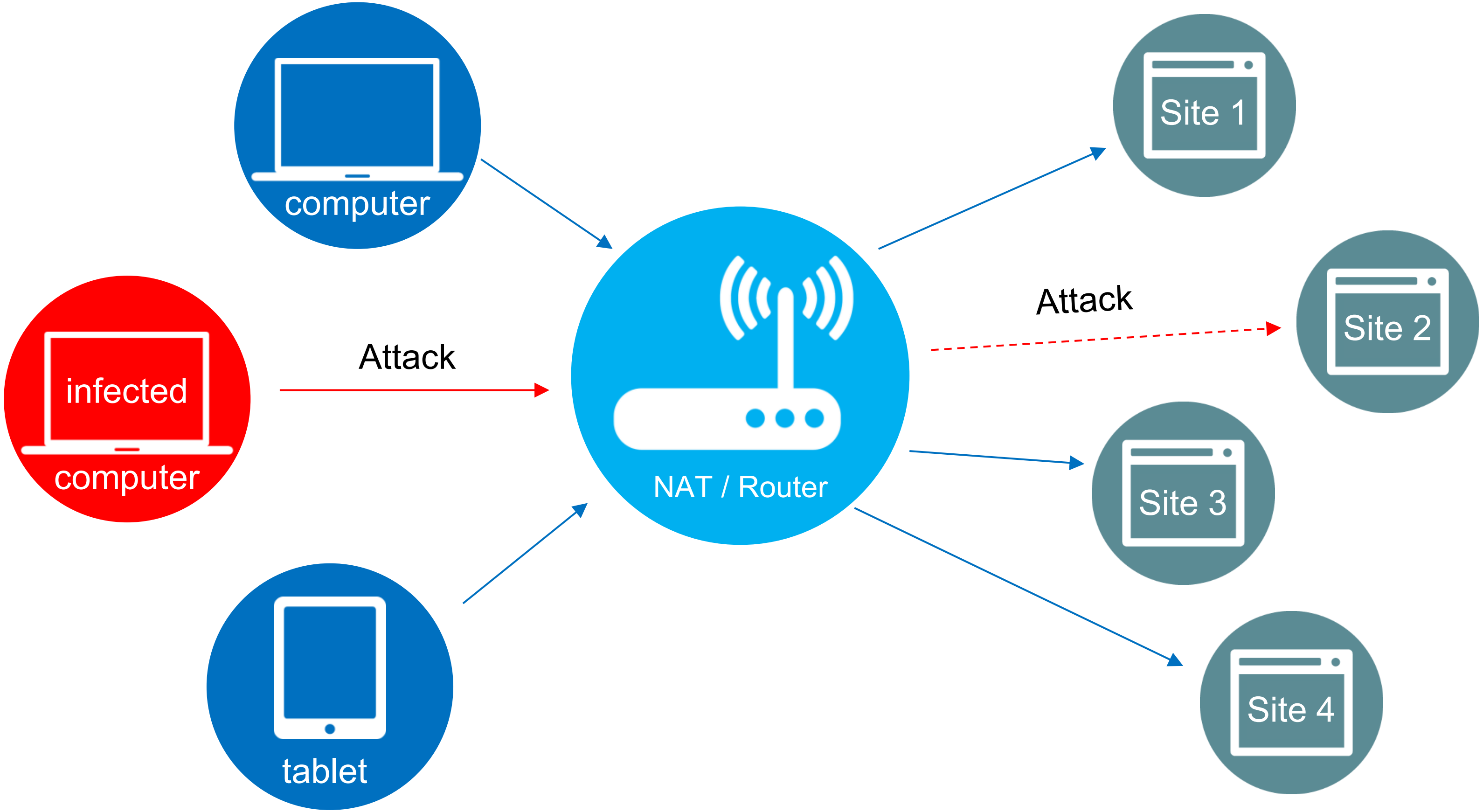
Please make yourself comfortable and have a look around.

I'll be right here if you need any assistance. :)

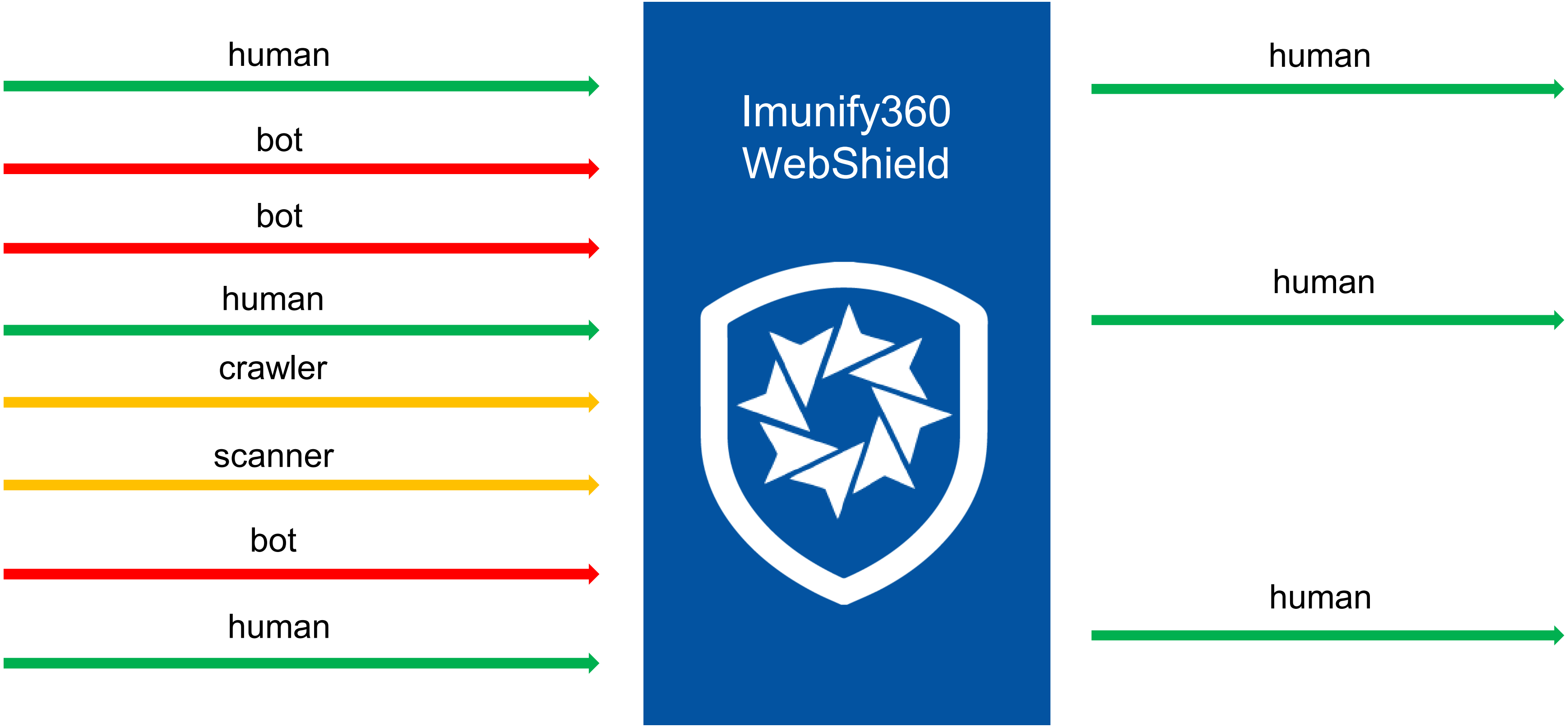
Write a reply...

Captcha / DHCP / NAT

We are CloudLinux



Next Stage: bot filtering



human

bot

bot

human

crawler

scanner

bot

human

Imunify360 WebShield

human

human

human

Machine Learning

- Feedback loop / captcha
  - Captcha bypass
  - Still important
- Look for anomalies & trends
- Next stage → client

What is next

- NAT
- WebShield
  - NAT
  - Client tagging (cookie)
  - Proof of JS
  - mod_sec
  - Passive client fingerprinting
    - Detect impersonators
  - IP & URI fingerprinting / correlation

Thank you!
Questions?