Security Crawler

Scanning 20K+ IP addresses at a rate of knots

Michael Markevich
Opera Software

Our Scale

- Software products and services for 350 million users
- Datacenters in Europe, Asia, America and Africa
- Thousands of servers and network devices
- Vast external perimeter and several intranet namespaces
- Multiple teams making deployments and introducing changes all the time

Requirements

- Perform a quick assessment of network services and firewall rules
- Flexibly run extended and custom checks on demand
- Monitor changes in the infrastructure and detect issues
- Integrate with inventory and other internal systems

Market Research

- A well-known commercial scanner: expensive and slow (15 minutes per host)
- Nmap: even `-sC -p 1-65535 -T4 x.x.x.x/20` takes ages to complete
- Masscan: better speed, but less reliable and this is only a port scanner
- None of the tested solutions provide a convenient way to store scan results

So … let’s build something ourselves! :-)

Key Concepts

- An assessment is split into multiple tasks, at least one task per scanned host
- Tasks are managed by the task queue manager
- Tasks are executed in parallel when possible
- Scanning itself is performed by any of the existing open-source tools
- An output of one task can be used as an input for another

Workflow

- An assessment is created through web UI and pushed into the queue as separate tasks
- Agents consume tasks, perform scans and upload results
- If configured, additional tasks can be run based on the scan data
- A host inventory and scan reports are stored in the database for future reference and comparison

Technology Stack

- **Masscan** as a default port scanner
- **Nmap** as a vulnerability and script scanner for open ports
- **Beanstalkd** as a queue manager
- **Ruby** scripts for scanning agent daemons and plugins
- **Ruby on Rails** + **PostgreSQL** for a web application with UI and REST API
- **GCP**, **OpenStack**, bare **KVM**, and physical servers are supported

Performance Summary

- A full port scan for a host: **2-5 minutes**
- A /24 subnet (254 addresses): **20-30 minutes**
- A /22 subnet (1022 addresses): **40-50 minutes**
- A full external network perimeter scan (20K addresses): **7-8 hours**

… and it can be even faster if we use more agents and bandwidth

Performance Optimization

- A task is selected by a closest agent (based on geolocation)
- The agent performs a full port scan with Masscan to detect open ports
- The agent scans open ports with Nmap and saves results to our database
- Only one target is scanned at a time by any agent process and scanner and can be terminated by timeout
- Extra scans are performed in a background based on agents' availability
- Some TCP stack tweaks are applied to agent nodes

Current Limitations

- All IP addresses in the inventory should be unique
- Limited IPv6 support (due to GCP constraints)

Other Highlights

- When scanning from GCP, we use “disposable” and cheap preemptive instances in an auto-scaling group, which manages the load automatically
- Masscan doesn’t work well with certain virtual network configurations, so we had to develop a custom TCP scanner that uses asynchronous IO
- Any security tool can be added as a plugin (e.g. to fetch Shodan data, brute-force dictionary passwords, or perform TLS security checks)

Competitors and Similar Work

- Scantron project
- Various scripts to run Nmap in parallel and collect results (e.g. nmapthrottle)
- Cloud services like Shodan or Censys

Questions?

https://security.opera.com

https://jobs.opera.com