Göran Pestana
Incident handler and developer
Megatron

Automated Abuse Handling

by CERT-SE
Who is Megatron?
Who is Megatron?

- A system that collects and processes information about bad hosts on the Internet
  - Input are logfiles from many different sources
  - Checks if the IPs exists in the organization DB
- Developed by CERT-SE
  - Tor Johnson & Göran Pestana
- Implemented in **Java** and uses a **MySQL** DB
- In production since the end of **2009**
What can Megatron do?
What can Megatron do?

- **Day-to-day abuse handling**
- **Incident handling**
  - Running list of "bad" machines
    - Any matches in the organization database?
    - Send abuse
- **File conversion**
  - Filtering
  - Add CC, ASN, geolocation, and hostname
  - Other formats
- **Data mining for statistics**
What can Megatron do?

Typical statistics for a week @ CERT-SE:

- >10 million log lines processed
- >50 000 log records saved in db
- ~50 abuse emails sent
Design Goals

- Should work out of the box
- More than just send abuse mail
- Flexible without coding
- Easy to extend with new code
- Handle large volumes of data
- Fast, only \textit{local} lookups
- Easy to use
Input Sources

mail
rsync
http
https

RBL
Shadowserver
Sunet
Zeus
Phishtank
Input Sources

• **Current:** Shadow Server, Spamhaus, SORBS, ZeuS Tracker, DroneBL, Phish Tank, Malware Patrol, Clean MX, Blade Defender, malc0de.com, Zone-h, Turk-h, xssed.com, vs-db.info, SpamCannibal, StopForumSpam, openbl.org, AutoShun, CERT-SE Honey Net, etc.

• ~60 configurations exist so far
Automation
Automation

- Handler on duty runs megatron and have to approve every mail job
- Helper scripts
  - Download script: http/https, rsync
  - mailbox-parser: Extracts body or attachments
- "slurp"-directory: Picks config from the filename
- Notification email: Sent if high priority entries are discovered
- RT-integration
  - A ticket-id is created automagic using RT CLI
Process Steps

Fetch
Parse
Filter
Decoration
Matching
Persist
Abuse
Mail
Search
Export
Process Steps

- **Fetch files:** http/https, rsync, or mail
- **Reformat:** e.g. XML to CSV file
- **Parse:** Tokenize log row
- **Filter:** Before and after every step
- **Data decoration:** Add CC, ASN, and hostname
- **Match:** Bind a log row to an organization
- **Persist:** Store log rows and the parsed data
- **Email:** Send abuse mail
- **File export:** Export data from db using template
Configuration
Configuration

- One configuration per input type
- Compact: Global config-properties are inherited
- Parsing using `regexp` and binds result to variables
- A `regexp` is defined for every variable:

```python
parser.item.countryCode=\w{0,2}
^\w+,"logTimestamp","ipAddress",".*?",
"additionalItem_infection","$countryCode","url"
```

- A variable corresponds to a field in the DB
- Variables are used in mail and export templates
Filters

Include

Exclude
Filters

• Many injection points in the flow

• **Line Filter** *(before parsing)*
  • Regexp
  • Line number

• **Log Entry Filter** *(after parsing)*
  • Regexp for arbitrary attribute
  • Country code
  • ASN
  • Priority

• Two modes: **Inclusive** or **Exclusive**
Data Decoration

- **Fast**: Only local lookups and DNS queries
- **Types:**
  - IP → ASN (BGP data or GeoIP)
  - IP → Country Code (GeoIP)
  - IP → City, latitude, longitude (GeoIP)
  - IP → Host Name (DNS)
  - Host Name → IP (DNS)
  - Other: URL → hostname, hostname → CC
- Dumps BGP table from router and imports it to the Megatron database once a week
Matching

Binding a log record to an organization

- Governmental agencies
- Municipalities
- Health care organizations
- Critical infrastructure companies
- Etc.

Matching order:

- IP
- Domain name
- ASN
Database Model
Database Model

- Organization properties
- Log data
Database Model
Organization properties

~700 organizations right now

Diagram:
- Priority
- Organization
  - AS-Number: 0..*
  - Domain Name: 0..*
  - IP-Range: 0..*
  - name
  - email_addresses
  - country_code
  - language_code
  - description

Notes:
- Priorities range from 0 to 255.
Database Model
Log data

LogEntry

Original LogEntry
1..1

Additional Data
0..*

FreeText
0..*

log_timestamp
ip_address
hostname
port
asn
country_code
ip_address2
hostname2
port2
asn2
country_code2
ip_range_start
ip_range_end
url
Features
More Features

- XML-support: Flattens XML files
- Line splitters and merges
- File processor: diffing files
- Checks log-file hash to avoid duplicates
- Support for localized mail templates
- Quarantine IPs to avoid excessive emails
- Flexible configuration of time stamp format, e.g.
  - ISO: yyyy-MM-dd HH:mm:ss z
  - Syslog: MMM dd HH:mm:ss
Even More Features

- RSS feeds (jobs, statistics)
- Warning for old time stamps
- XML and JSON data export for charts
Where is Megatron?

Open Source:
https://download.cert.se/megatron/

Users:
- NCSC-NL
- CERT-HU
- NorCERT
- CERT-CW (Curaçao)
- One major Swedish ISP
### Infekterade datorer i Sverige

<table>
<thead>
<tr>
<th>Städer</th>
<th>IP-adresser</th>
<th>Träffar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stockholm</td>
<td>7334</td>
<td>31543</td>
</tr>
<tr>
<td>Göteborg</td>
<td>2598</td>
<td>10058</td>
</tr>
<tr>
<td>Malmö</td>
<td>1810</td>
<td>9068</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organisationstyper</th>
<th>IP-adresser</th>
<th>Träffar</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISP</td>
<td>20026</td>
<td>107958</td>
</tr>
<tr>
<td>Webbhotell</td>
<td>13067</td>
<td>63535</td>
</tr>
<tr>
<td>ISP 2</td>
<td>3093</td>
<td>6895</td>
</tr>
</tbody>
</table>

Map data ©2013 GeoBasis-DE/BKG (GZL03); Google Terms of Use; Report a map error.
Megamap

- **JavaScript** application
- Static **JSON** files generated by **Megatron**
- Two versions:
  - Public: www.cert.se
  - Intranet: IP-addresses unmasked + hostnames
- Benefits:
  - "Eye candy" (upper management **loves** it)
  - Raising public **awareness**
- It is Open Source: Download URL can be provided on request
Questions?

Thank you!