AVIRA
More Than Security

URLCheck
Malware and Phishing URLs aggregator

by

Sorin Mustaca <sorin.mustaca@avira.com>

for the **Virus Bulletin Conference**

vb 2008 OTTAWA

AVIRA®
More Than Security

Contents

- What is an URL aggregator ? Why do we need one ?
- Architecture
- The URL Sources
- Features
- Challenges
- Results
- Q & A

What is an URL aggregator ?

URLCheck

AVIRA
More Than Security

The same idea as an RSS Feed *aggregator:* it retrieves content (URLs) from different sources and it displays them in a central place.

Characteristics of RSS Feeds

- They must respect a standard format (XML)
- The client can *pull* data
- The server can *push* data (email, IM,html files)
- The data is displayed in a central point, usually a web portal, but also in specially built clients

Why an URL aggregator ?

AVIRA
More Than Security

Why the name „URLCheck“ ?

- it **gathers** many thousands URLs from different sources and in different ways
- it pulls content
- it receives pushed content
- it displays, manages, **checks and validates** the content in a central place

Difference from RSS feeds

- there is no standard (the URL is the only thing the sources have in common)

Why an URL aggregator ?

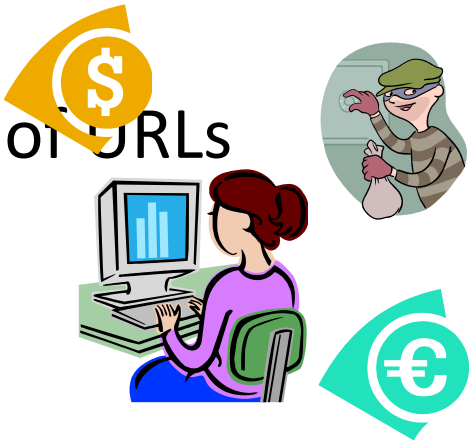
URLCheck

AVIRA
More Than Security

We‘ve built it because

- most of the threats are spreading using URLs (sent in emails, IM messages, SMS, MMS)
- there are some very good, free sources of URLs

URL



URLs point to

- Malware (viruses, trojans etc.) files
- Phishing websites
- Webshops selling fake and dangerous products
- Internet scams (nigerian scam, lottery, etc)

Why an URL aggregator ?

URLCheck

AVIRA
More Than Security

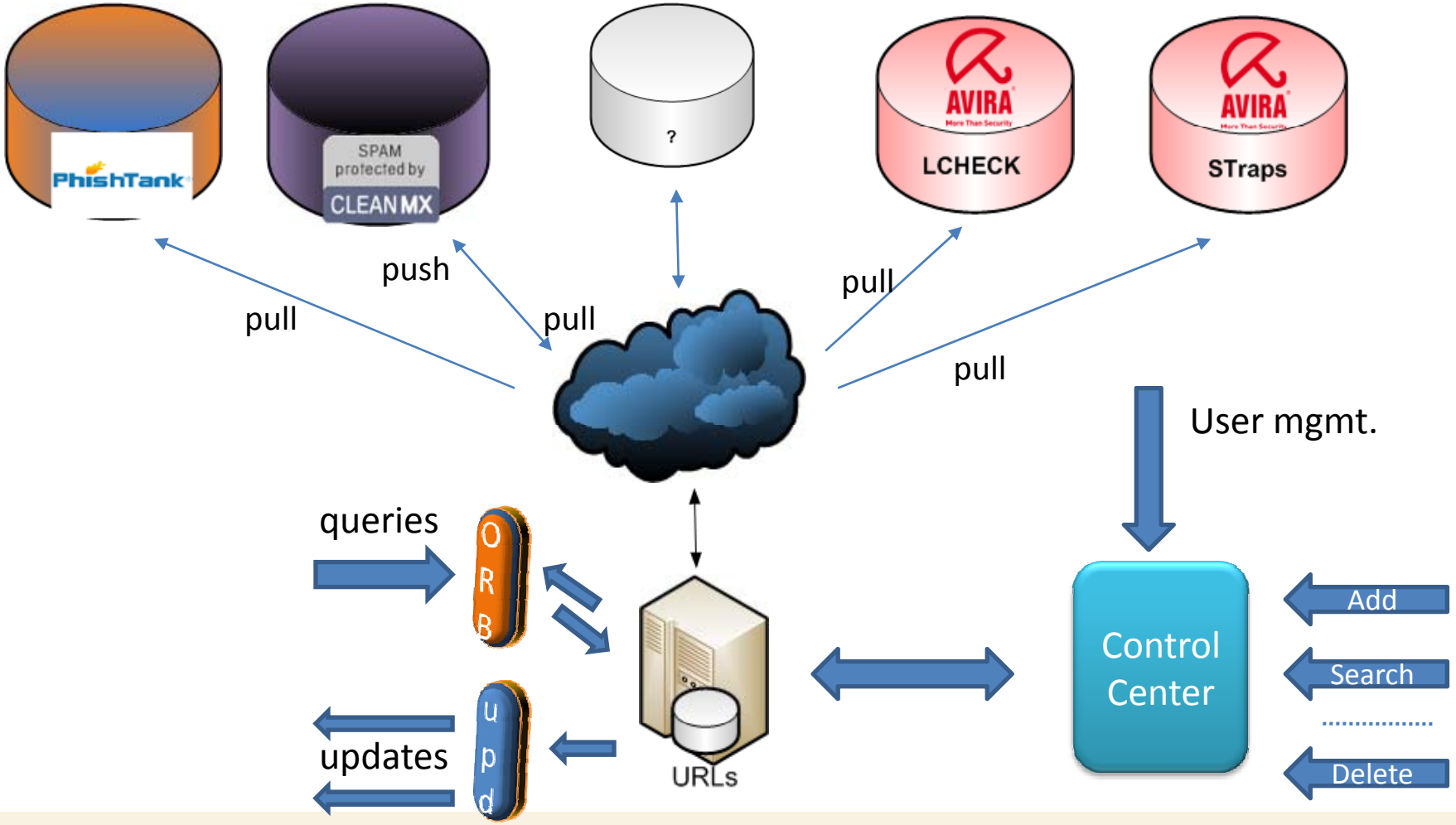
Everybody who wanted to have something
blocked or whitelisted
came to me

And I had to do that … Manually …

… A LOT OF WORK …

Architecture (simplified)

URLCheck



URL Sources - Phishtank

URLCheck

AVIRA
More Than Security

PhishTank is operated by OpenDNS, a free service that makes your Internet safer, faster, and smarter. Get started today!

PhishTank® Out of the Net, into the Tank.

Home | Add A Phish | Verify A Phish | Phish Search | Stats | Blog | FAQ | API | My Account

Join the fight against phishing

Submit suspected phishes. Track the status of your submissions.
Verify other users' submissions. Develop software with our free API.

Found a phishing site? Get started now — see if it's in the Tank:

http:// [Is it a phish?]

Recent Submissions

ID	URL	Submitted by
509414	http://abbey.disasterofasite.com/myonlineaccounts2...	tetak
509401	http://cgi1-listings4329383212372.i8.com/ws-bin/li...	PhishReporter
509392	http://www.syansan.com/news/-/paypal-secure-acces/...	PhishReporter
509379	http://mail.credinka.net/simulacion/online/	jono
509377	http://h1.ripway.com/Paypally/paypal.html	PhishReporter
509374	http://cgi1-listings4329384732382.i8.com/a-ws/well...	PhishReporter
509371	http://paypal-services.1stfreehosting.com/paypal/c...	PhishReporter
509369	http://www.webactionede.com/www[1][1].paypal.fr/ww...	PhishReporter
509368	http://www.webactionede.com/www%5B1%5D%5B1%5D.payp...	PhishReporter
509362	http://www.mihostbiz.com/	PhishReporter

URL Sources - Phishtank

URLCheck

How does it work ?

URLCheck

```
<?xml version="1.0" encoding="utf-8"?>
    <output>
        <meta>
            <generated_at>2008-09-16T06:30:00+00:00</generated_at>
            <total_entries>2475</total_entries>
        </meta>
        <entries>
            <entry>
                <url><![CDATA[http://m-maleki.com/irs.gov.cookies=5353653.html]]></url>
                <phish_id>506707</phish_id>
                <phish_detail_url><![CDATA[http://www.phishtank.com/phish_detail.php?phish_id=506707]]></phish_detail_url>
                <submission>
                    <submission_time>2008-09-16T05:21:04+00:00</submission_time>
                </submission>
                <verification>
                    <verified>yes</verified>
                    <verification_time>2008-09-16T06:28:00+00:00</verification_time>
                </verification>
                <status>
                    <online>yes</online>
                </status>
            </entry>
```

XML File with all the URLs inside :

- URL
- Phish ID
- Submission Time
- Verified or not
- Verification Time
- Online: Yes

URL Sources - Phishtank

URLCheck

AVIRA
More Than Security

Problems and Solutions

➢ False positives	✓ Report them to PT and update the XML file
➢ Invalid XML file	-- Report them to PT and update the XML file ✓ Filter the invalid chars with a special program
➢ Invalid (auto) submissions by the PT email parsing software	-- Complain
➢ Slow publication in the feed	-- PT Users do not vote in the same way , so many URLs are undecided

➢ It is NOT possible to mark an URL up (add it) or down (remove it)

URL Sources – Clean-MX

URLCheck

AVIRA
More Than Security

CLEAN MX realtime database

public access query for phishing URL
Totally watched: 413, to down: 1, to up: 5, changed ip: 20
As of 2008-09-20 23:42:59 CEST

Subscribe to the PhishWatch Mailing list, updated hourly
This database consists of Phishing URI, collected and verified since Oct 2006 some of them may be already closed, but are still recognized as fraud by firefox and opera also offending domain names are honored by SURBL.
If you detect URI'S concerning your netblock, already closed... you have made a good job, otherwise please close them as soon as possible.

Attention: all URI'S are manually verified, but not cross-checked for real phishing function in this moment you make this query.(Sites may have been closed already..)
Our automatic Phishwalker process is scheduled every hour, so you may see now a incident and this one will be resolved later on.
So please keep on sending close-feedbacks to us...

Median time to close Phishing sites: 30.8 days .... This is: **Poor, please speed up!** (Calculated for: complete database, but only for incidents since May 2008)

to look at some nice charts her are complete statistics for this database
and for our german friends some minutes on Symantec Phishing report

if you have questions, criticism, wishes or ... do not hesitate to contact us at abuse@clean-mx.de
Welcome back, would be fine to get some feedback from your site..
TIMERS: Runtime Query: 5.7332 Seconds

Line	#	Date	Closed	hours	PhishTank	ip state	response	Ip initial	ip review	Domain	
1	355801	2008-09-20 18:28:18			509362	up	alive	64.22.77.246	64.22.77.246	mihostbiz.com	http://www.mihostbiz.com/
2	355787	2008-09-20 17:05:19			509340	up	alive	68.180.151.58	68.180.151.19	cmd1safe-intl.us	http://cmd1safe-intl.us/portal/ceas.html
3	355802	2008-09-20 16:04:20			509320	up	alive	71.160.224.35	71.160.224.35	manpowersource.com	http://www.manpowersource.com/data/session/b.php
4	355803	2008-09-20 16:02:18			509317	up	alive	80.74.143.2	80.74.143.2	paplyapt.com	http://www.paplyapt.com/uk/webscr/account-update/=&i1=&bshowqi
5	355804	2008-09-20 16:00:05			509316	toggle	alive	97.100.82.84	4.238.231.94	8cfm.cc	http://ww8.associatedbank.com.8cfm.cc/web_bank/confirm.asp?font=
6	355788	2008-09-20 15:54:29	2008-09-21 00:00:18	8.1	509313	up	dead	212.89.6.11	212.89.6.11	moterosastures.com	http://www.moterosastures.com/foros/attachments/customer.html
7	355706	2008-09-20 15:39:08	2008-09-20 19:39:38	4	509308	down	dead	68.180.151.17	68.180.151.17	mailuridemailuri.com	http://mailuridemailuri.com/italia.html

URL Sources – Clean-MX

URLCheck

How does it work ?

Updates via email

From: root <root@dbserver.netpilot.net>;
Date: Freitag, 19. September 2008 16:51
To: viruswatch@lists.clean-mx.com <viruswatch@lists.clean-mx.com>;
Subject: [Viruswatch] Virus-sites with status changes: As of 2008-09-1916:15:14 CEST

Down: RIPE LV sheb.ahmad@gmail.com 78.157.141.6 http://78.157.141.6/TotalSecure2009.exe
Down: NA RIPE IT abuse@wmgitalia.it 195.225.168.186 to 195.225.168.186 lodomeni.com http://www.lodomeni.com/pindex.php
Down: NA RIPE IT abuse@wmgitalia.it 195.225.168.186 to 195.225.168.186 lodomeni.com http://www.lodomeni.com/videoPorn218hdy.exe
Up(nil): unknown_exe RIPE UA abuse@odessa.tv 88.198.8.15 78.26.179.248 to 67.228.177.143 bestdownloadsoft.com http://cdn.bestdownloadsoft.com/antiviruspcsuite.com/AntivirusPCSuite/install_sbd_en.exe
Up(nil): unknown_exe RIPE UA abuse@odessa.tv 85.17.4.6 78.26.179.248 to 67.228.177.146 bestdownloadsoft.com http://cdn.bestdownloadsoft.com/spyguardpro.com/SpyGuardPro/install_sbd_en.exe
Up(nil): unknown_exe RIPE UA abuse@odessa.tv 88.198.8.15 78.26.179.248 to 67.228.177.146 bestdownloadsoft.com http://cdn.bestdownloadsoft.com/avsystemcare.com/AVSystemCare/install_sbd_en.exe
Up(nil): unknown_html ARIN US abuse@softlayer.com 88.198.8.15 67.228.177.143 to 78.157.142.19 download-es.com http://download-es.com
Up(nil): unknown_html RIPE LV abuse@vdhost.info 67.228.177.146 67.228.177.143 to 78.157.142.19 download-drc.com http://download-drc.com
Up(nil): unknown_html RIPE UA abuse@odessa.tv 78.157.142.19 78.26.179.248 to 67.228.177.143 dwnld1.com http://dwnld1.com
Up(nil): unknown_html RIPE UA abuse@odessa.tv 85.17.4.6 78.26.179.248 to 67.228.177.146 antispywareexpert.com http://download.antispywareexpert.com
Up(nil): unknown_html RIPE UA abuse@odessa.tv 88.198.8.15 78.26.179.248 to 67.228.177.143 easydownloadsoft.com http://archive.easydownloadsoft.com
Up(nil): unknown_html RIPE UA abuse@odessa.tv 85.17.4.6 78.26.179.248 to 67.228.177.146 bestdownloadsoft.com http://cdn.bestdownloadsoft.com
Up(nil): unknown_html RIPE UA abuse@odessa.tv 88.198.8.15 78.26.179.248 to 67.228.177.146 bestdownloadsoft.com http://old.bestdownloadsoft.com

URL Sources – Clean-MX

URLCheck

AVIRA
More Than Security

Problems and Solutions

- False positives ✓ Report them and update
- Invalid updates ✓ Report them and update
- Relatively small amount of URLs
- Proprietary system - no automatic way of retrieving all URLS
- It is possible to mark an URL up (add it) or down (remove it)

URL Sources – Conclusion

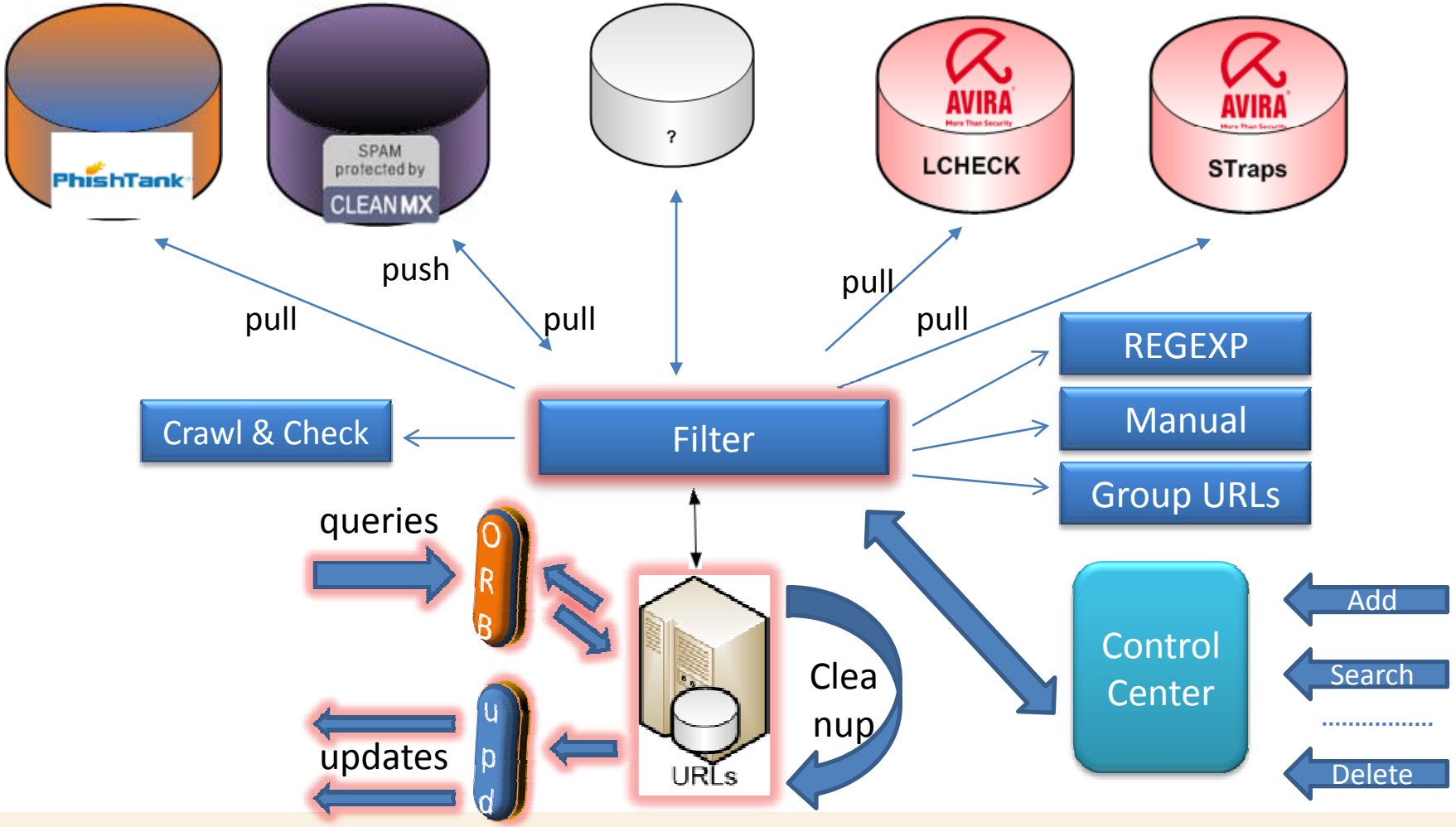
AVIRA
More Than Security

Common problems

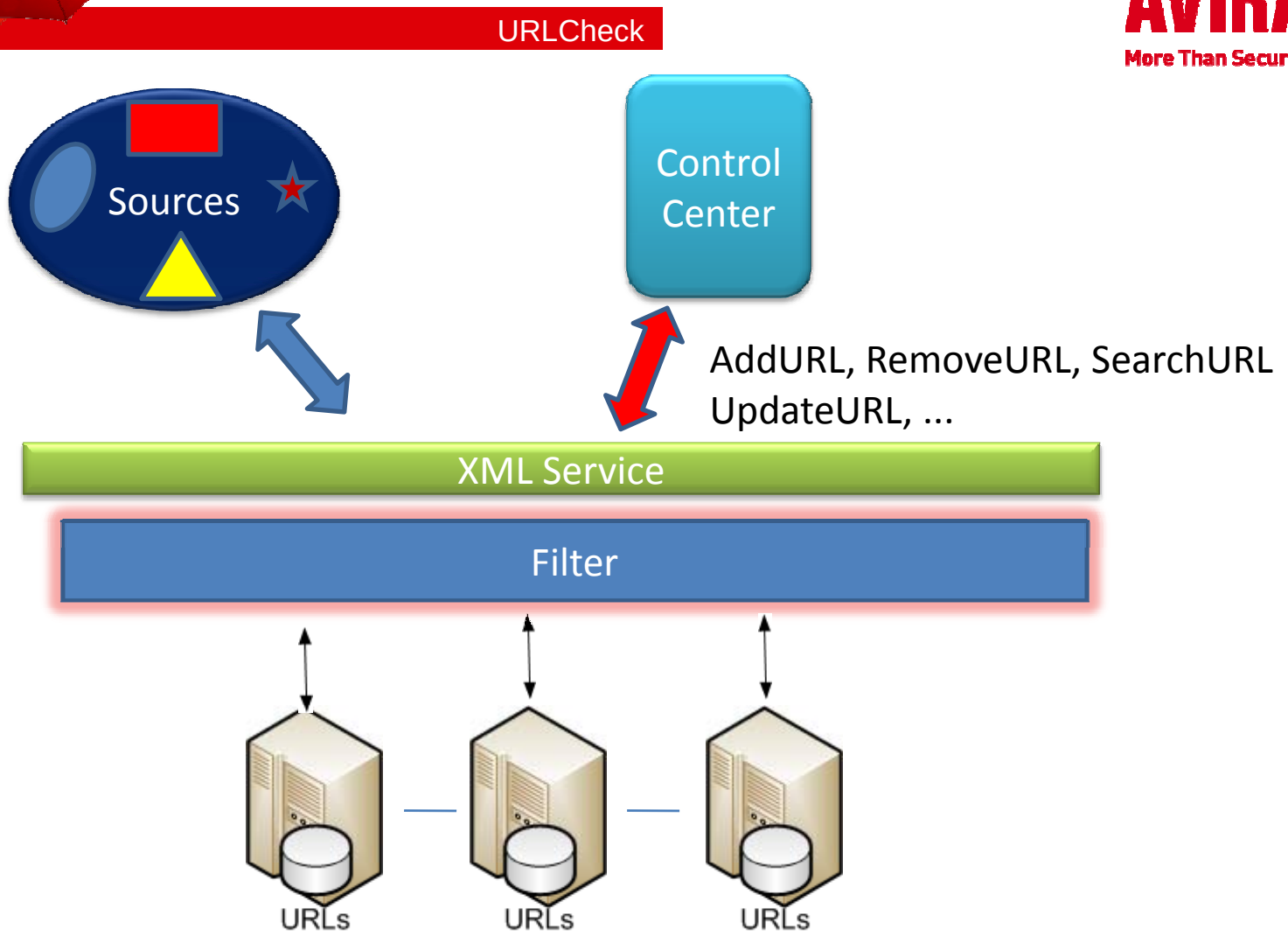
➢ False positives	✓ Whitelist (manual and automatic)
➢ Not always reliable	✓ Do not count only on one source ✓Crawl & check from the start URL
➢ (Sometimes) Slow response time on updating the status of the URLs	✓ Retry often, merge several sources in parallel

Architecture (full)

URLCheck



Architecture (future)

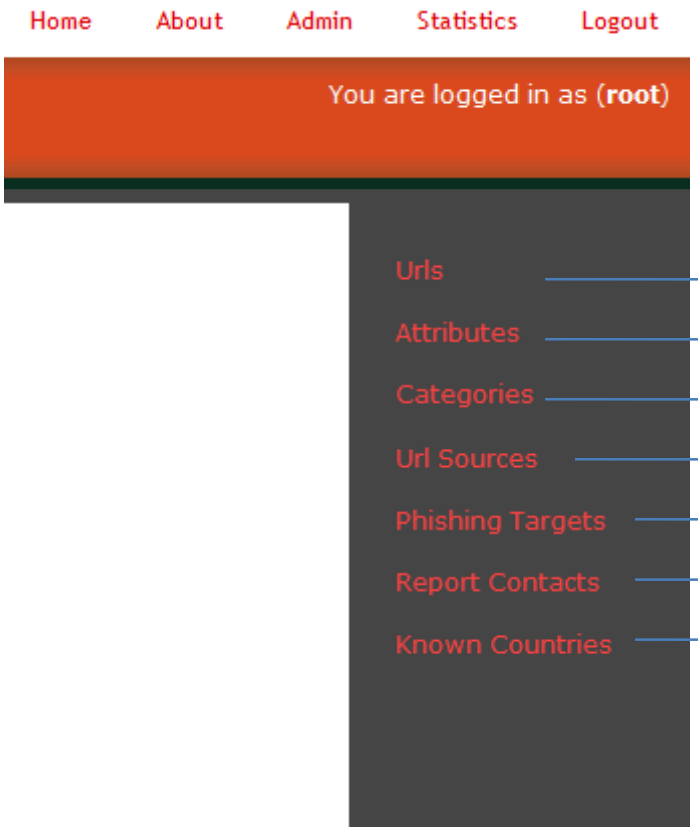


Features

URLCheck

AVIRA
More Than Security

GUI: Django framework

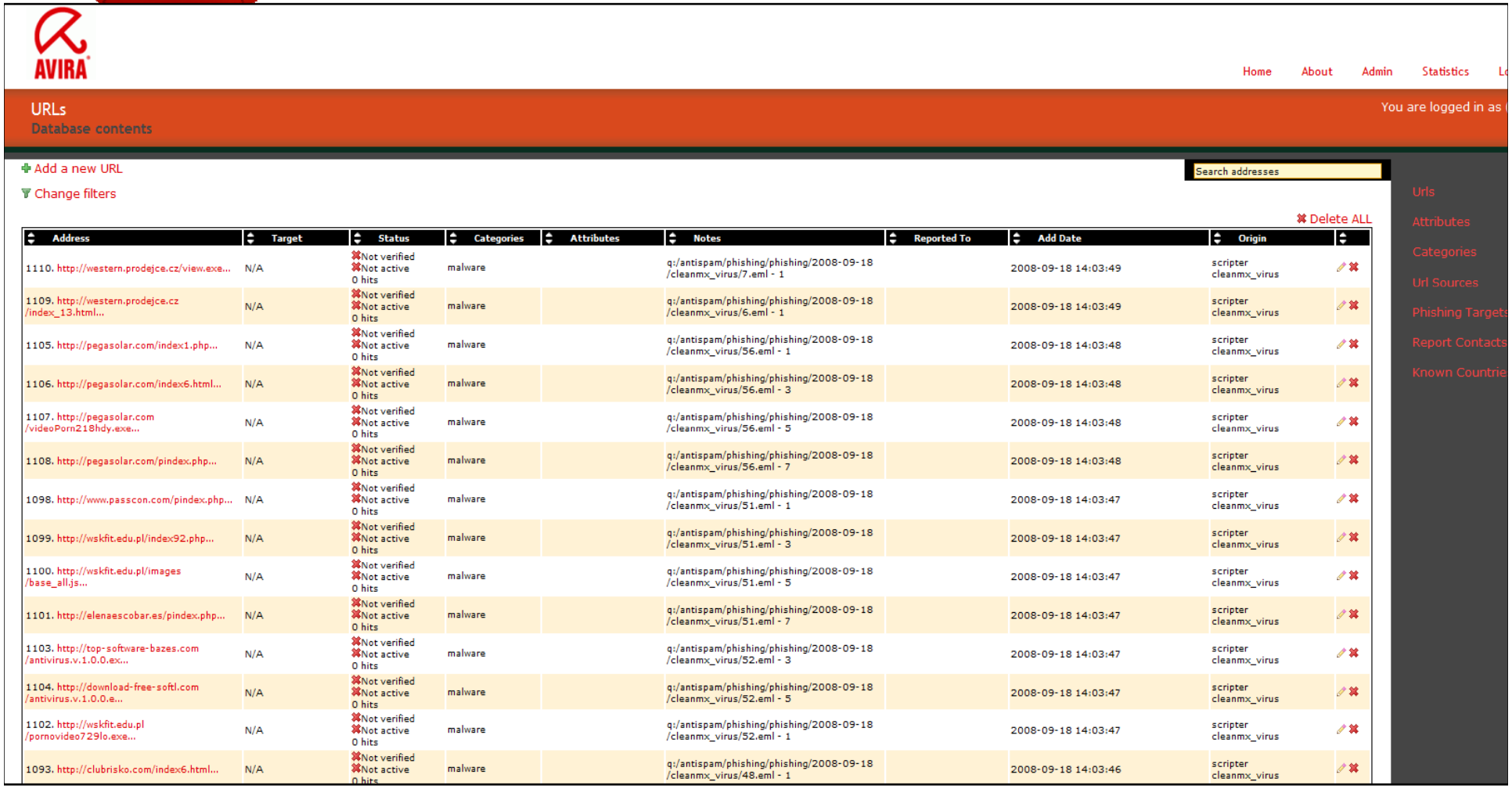


- Urls → Displays & Filters URLS
- Attributes → Wl, Bl, Outbreak,Redirector...
- Categories → Phishing, Malware, 419,Spam...
- Url Sources → Phishtank,Clean-MX...
- Phishing Targets → Name and URL of the ph. targets
- Report Contacts → Report Contacts
- Known Countries → GeoIP stuff

Features

URLCheck

AVIRA
More Than Security



Address	Target	Status	Categories	Attributes	Notes	Reported To	Add Date	Origin
1110. http://western.prodejce.cz/view.exe...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/7.eml - 1		2008-09-18 14:03:49	scripter cleanmx_virus
1109. http://western.prodejce.cz /index_13.html...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/6.eml - 1		2008-09-18 14:03:49	scripter cleanmx_virus
1105. http://pegasolar.com/index1.php...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/56.eml - 1		2008-09-18 14:03:48	scripter cleanmx_virus
1106. http://pegasolar.com/index6.html...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/56.eml - 3		2008-09-18 14:03:48	scripter cleanmx_virus
1107. http://pegasolar.com /videoPorn218hdy.exe...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/56.eml - 5		2008-09-18 14:03:48	scripter cleanmx_virus
1108. http://pegasolar.com/pindex.php...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/56.eml - 7		2008-09-18 14:03:48	scripter cleanmx_virus
1098. http://www.passcon.com/pindex.php...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/51.eml - 1		2008-09-18 14:03:47	scripter cleanmx_virus
1099. http://wskfit.edu.pl/index92.php...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/51.eml - 3		2008-09-18 14:03:47	scripter cleanmx_virus
1100. http://wskfit.edu.pl/images /base_all.js...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/51.eml - 5		2008-09-18 14:03:47	scripter cleanmx_virus
1101. http://elenaescobar.es/pindex.php...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/51.eml - 7		2008-09-18 14:03:47	scripter cleanmx_virus
1103. http://top-software-bazes.com /antivirus.v.1.0.0.ex...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/52.eml - 3		2008-09-18 14:03:47	scripter cleanmx_virus
1104. http://download-free-softl.com /antivirus.v.1.0.0.e...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/52.eml - 5		2008-09-18 14:03:47	scripter cleanmx_virus
1102. http://wskfit.edu.pl /pornovideo729lo.exe...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/52.eml - 1		2008-09-18 14:03:47	scripter cleanmx_virus
1093. http://clubrisko.com/index6.html...	N/A	Not verified Not active 0 hits	malware		q:/antispam/phishing/phishing/2008-09-18 /cleanmx_virus/48.eml - 1		2008-09-18 14:03:46	scripter cleanmx_virus

Challenges

AVIRA
More Than Security

1000+ URLs / 24h (incl. updates to the old URLs)
Refreshing of DB is now performed every hour and in the future every 15 minutes

Make sure we have a long uptime

- ➢ the server runs smoothly
  - • make sure we have decent levels of CPU, RAM, HDD usage
  - • use special parallel algorithms
  - • make a backup of the log files and DB

Challenges

URLCheck

AVIRA
More Than Security

- Reliability of the content
  - Check the URLs and remove those which are obsolete (not trivial, as described in my paper „Delivering reliable phishing protection“ published in VB Magazine in May 2008)
  - Redownload the *suspicious* files periodically and rescan them or mark them as down
- Freshness of the databases
  - Get new content as soon as it is available on the source (update often)
  - Get new sources all the time

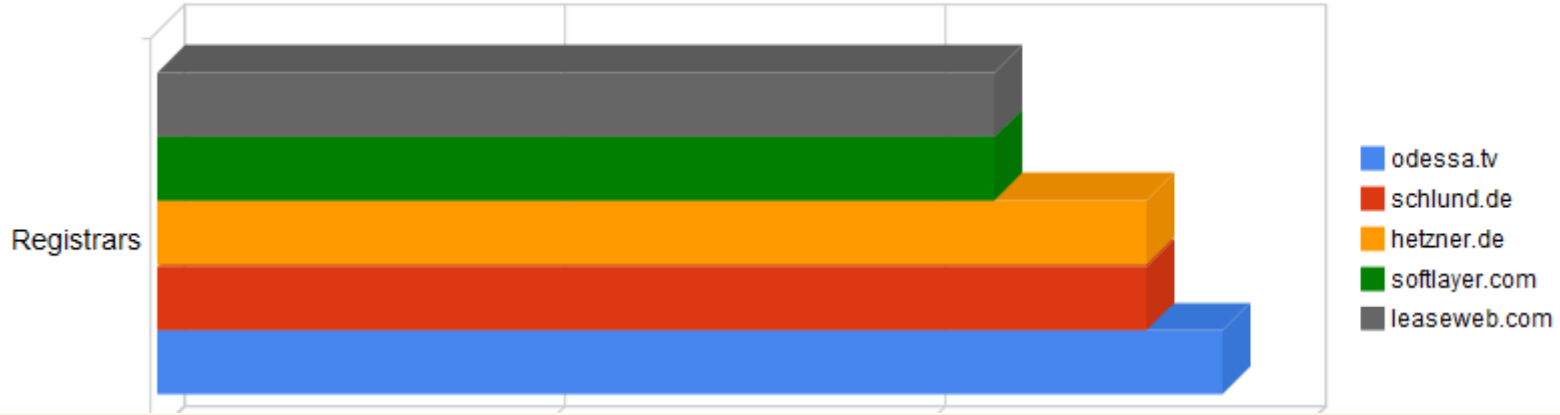
Results

URLCheck

AVIRA
More Than Security

Current world distribution of identified threats:

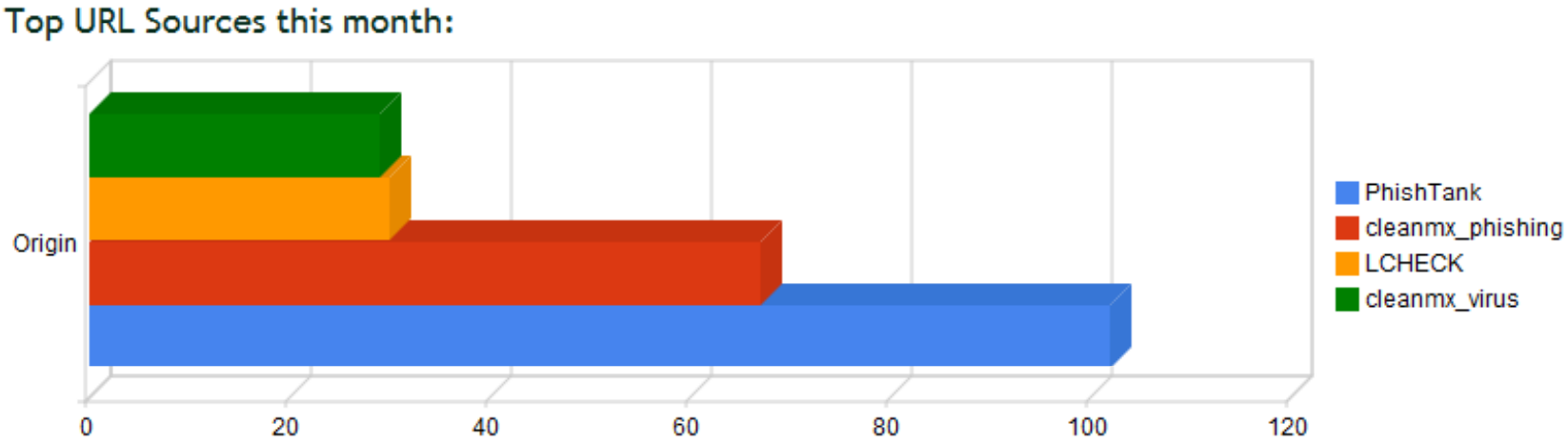
Top Registrars to whom the most Urls were reported



Results

URLCheck

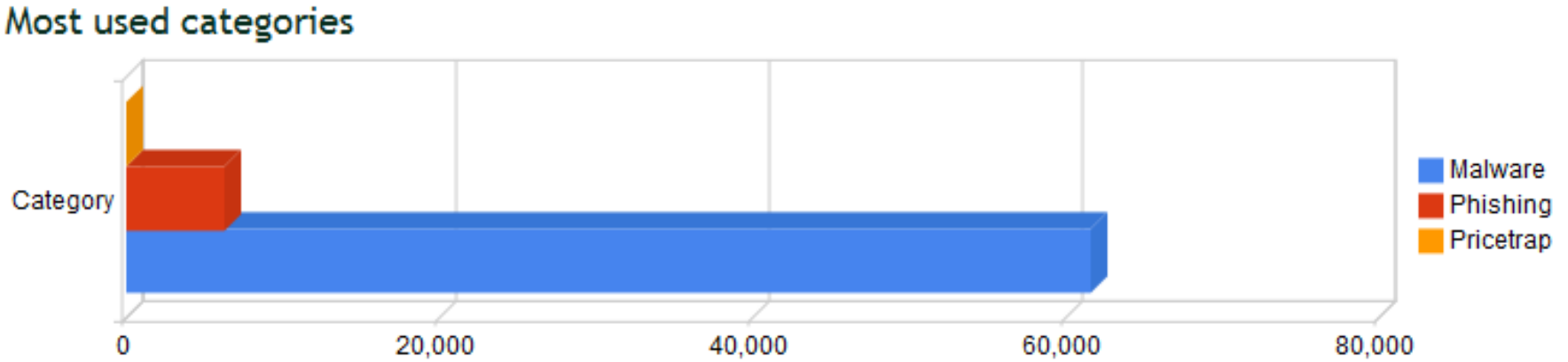
Global statistics (the system was repopulated last week)



Results

URLCheck

AVIRA
More Than Security



Every URL in the DB points to a unique file, containing not unique malware.
(there are many URLs which point to files containing the same malware)

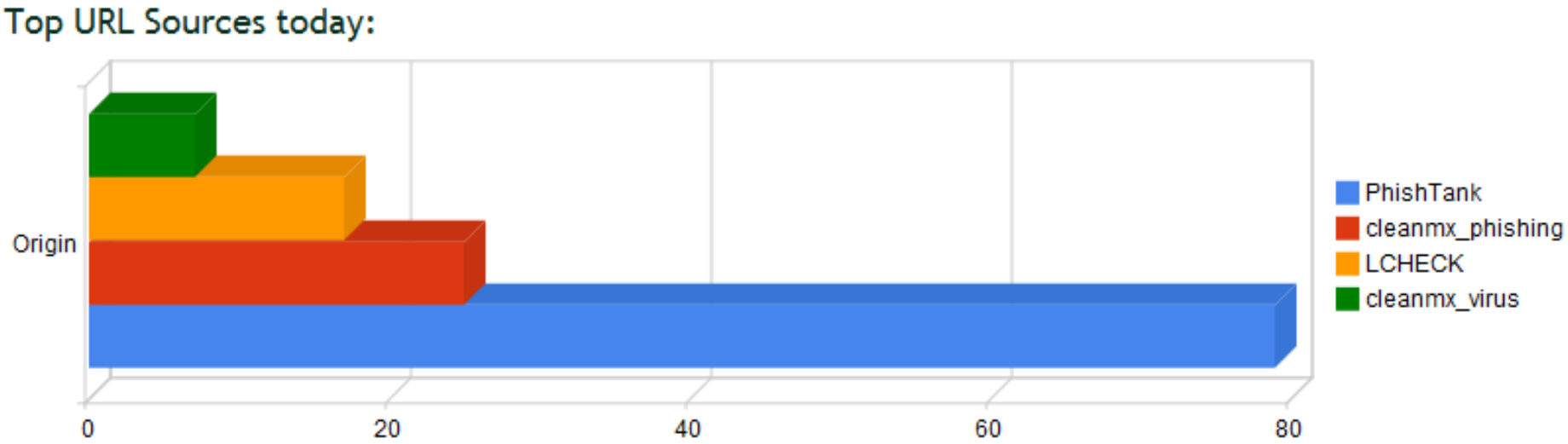
The Malware is really wide spread !

Results

URLCheck

AVIRA
More Than Security

Statistics per day ... and it was early in the morning.



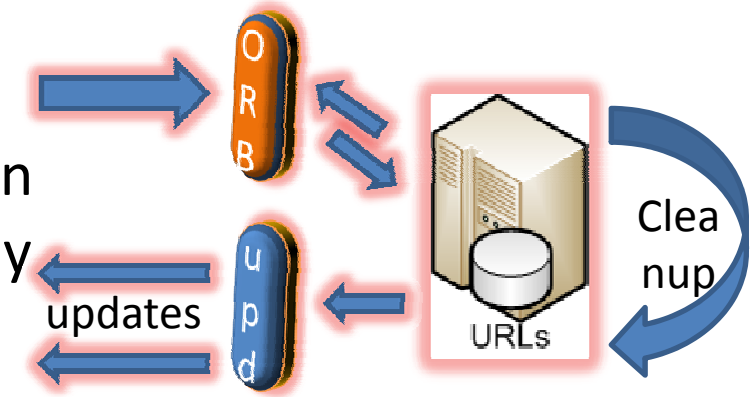
Newly unique added URLS only ...

Conclusions

URLCheck

AVIRA
More Than Security

- A lot of URLs
  - Growing with the rate of more than 100+ new unique and valid URLs / day



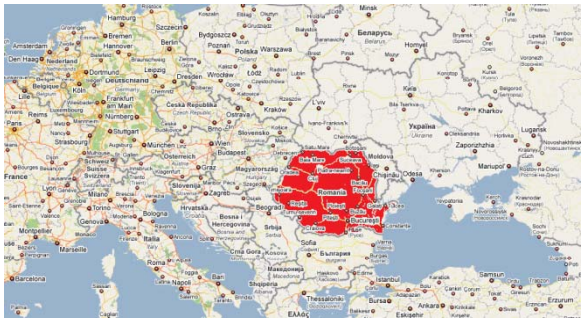
File Updates

- If we write in a single file the MD5s of 64K URLs, we have a 3 MB file (it contains also some administrative data)
- It makes sense to have
  - incremental updates, but this is not trivial considering the nature of the sources (only Clean-MX offers updates)
  - very small size of the online requests

IN THE END...

AVIRA
More Than Security

Thanks to :



Cosmin Luta
Ionut Rosoiu
Vlad Dinulescu

from Avira Romania for developing the URLCheck GUI and the software for analyzing the Spamtraps

Virus Lab

from Avira Germany for creating and maintaining LCHECK

Gerhard Recher

from CLEAN-MX

Phishtank team

Q & A

AVIRA
More Than Security

Thank you for attending !

Questions ?

If you want to discuss about an integration in URLCheck, please contact me

Sorin Mustaca
<Sorin.Mustaca@avira.com>