PharmaGuard
Automatic Identification of Illegal Search-Indexed Online Pharmacies

Igino Corona*, Matteo Contini*, Davide Ariu*, Giorgio Giacinto*, Fabio Roli*, Michael Lund+, Giorgio Marinelli+

*DIEE University of Cagliari, ITALY,
+Danish Institute of Fire and Security Technology, DENMARK

CYBERSEC 2015 Gdynia, Poland, June 24th
This research has been funded with support from the European Commission (Grant Agreement HOME/2012/ISEC/AG/4000004360)

Buster of ILLEGAL contents spread by malicious computer networks

Thanks to my co-authors!

Matteo Contini
Davide Ariu
Giorgio Giacinto
Fabio Roli
Michael Lund
Giorgio Marinelli
Throughout the history, technological innovations have changed the power balance between attacker and defender.

[...] Understanding how advances in technology affect security - for better or for worse - is important to building secure systems that stand the test of time

Bruce Schneier - Beyond Fear, 2003
From **Physical** to **Virtual** worlds

- Physical World
- Virtual World

Technology

Threats

Opportunities

http://pralab.diee.unica.it
From Pharmacies to online Pharmacies

Physical World

Virtual World

http://pralab.diee.unica.it
Online Pharmacies

BE AWARE: Today, most of online pharmacies are ILLEGITIMATE (black)!

Online Pharmacies

- ILLEGITIMATE: 94.3%
- POTENTIALLY LEGITIMATE: 5.1%
- LEGITIMATE: 0.6%

Data Source: LegitScript - https://www.legitscript.com
(Illegal) Online Pharmacies

- Easy way for cyber-criminals to make money without qualms about other people’s health
- May sell any kind of drug - no need of medical prescription
- Sold products expose people to severe health threats that may even lead to death

International efforts to fight this threat

- **June 2013** - FDA seized 1,700 illegal online pharmacies
- **May 2014** - Interpol coordinated an international operation, in more than 100 countries, to disrupt more than 11,800 illegal pharmacies
  - seized pharmaceuticals worth more than **32 million US dollars**
Illegal online pharmacies - Key issues

Cyber-criminals can make money from illegal online pharmacies with little effort and low risk

• many countries miss a clear legislation in this regard

• laws across different countries are typically heterogeneous

• law enforcement authorities are often ill-equipped to find and investigate them

• new victim users may easily reach such sites through simple queries on web search engines

• notably, the search giant Google has been even prosecuted for illegal profits from those activities

http://pralab.diee.unica.it
Our Proposal: **PharmaGuard**

- **Objective**: *automatically detect illegal online pharmacies* advertised throughout the web
- We focus on web sites *indexed by web search engines*
  - reachable by million users through simple queries
- Aimed at assisting law-enforcement toward their early identification, blacklisting and shutdown.

- **PharmaGuard** is the equivalent of a *virtual police dog trained to automatically “smell” illegal pharmacies*
Our Proposal: **PharmaGuard**

- **Objective:** automatically detect illegal online pharmacies indexed by web search engines
User Emulation

- Web pages may be generated through JavaScript or Flash code, Cascade Style Sheets (CSS), HTML frames
- Content might not be visible without using a real browser
- **PharmaGuard** navigates through the suspicious web pages through a **Browser Automation Framework**

![Diagram showing the process of PharmaGuard navigating through suspicious web pages using Selenium and Firefox]
Illegal Pharmacy Detector

Is online Pharmacy?

Pharma vs Other Classifier

queries: urls, images, text

web search engines

discard

Pharma vs Pharma Classifier

Is illegitimate?

yes

User Interface (Law Enforcement)

feedback

no: low priority

yes: high priority

Is online Pharmacy?

web page

http://pralab.diee.unica.it
Pharma vs Other/Pharma Classifiers

- Text-based content analysis
- Lightweight, fast, effective
  - web search engines need to see text to index such pages
  - cyber criminals are thus incentivized to insert relevant textual content
- **Feature selection**: TF-IDF Technique
  - we firstly strip all stop words from within each webpage $p$
  - Each term (word) $t$ within page $p$ receives a weight $w_{t,p}$, computed as follows:
    \[
    w_{t,p} = \begin{cases} 
    (1 + \log(tf_{t,p})) \cdot \log\left(\frac{N}{pf_t}\right) & \text{if } tf_{t,p} > 0 \\
    0 & \text{otherwise}
    \end{cases}
    \]
  - $tf_{t,p}$: number of times term $t$ appears in page $p$
  - $pf_t$: number of pages that $t$ occurs in
  - $N$: number of webpages in the collection.
  - For each webpage $p$ we then reshape its weights $w_{t,p}$ using cosine ($l_2$) normalization
    \[
    \sum_{t \in p} w_{t,p}^2 = 1.
    \]
- Classification: linear classification algorithms
  - fast, and suitable for the high dimensionality of the feature space
Experimental Evaluation

• Evaluation objectives
  • accuracy when discriminating illegitimate online pharmacies from other kind of websites, including legitimate online pharmacies
  • learning time and throughput
  • complementarity of the approach with respect to state-of-the-art tools
  • more insights into the characteristics of illegal online pharmacies in the wild, as well as of the threat posed by drugs sold online

• Dataset (more than 1200 manually-validated webpages)
  • Built through a bootstrap phase: starting from few known-as-illegitimate online pharmacies as ‘seeds’
  • L: set of 172 legitimate online pharmacies
    • mainly built using LegitScript
  • I: set of 446 illegitimate online pharmacies
  • O: set of 647 other webpages
Detection Accuracy

- **5 runs** (random splits)
  - Training: 70% (parameter optimization: 3-fold cross-validation)
  - Test: 30%

- **Tested Linear classifiers:** Ridge Regression, **Stochastic Gradient Descent**, Passive-Aggressive, K- nearest Neighbors Vote, Support Vector, Nearest Centroid and Naive Bayes

### TABLE I

**Accuracy of the detector of illegal online pharmacies.**

<table>
<thead>
<tr>
<th>Class</th>
<th>FPR for TPR=100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>1.64% (± 0.6%)</td>
</tr>
<tr>
<td>Legitimate</td>
<td>-</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>0.59% (± 0.57%)</td>
</tr>
<tr>
<td>Other</td>
<td>0.41% (± 0.38%)</td>
</tr>
<tr>
<td>Legitimate</td>
<td>1.09% (± 1.2%)</td>
</tr>
</tbody>
</table>

http://pralab.diee.unica.it
Learning and Detection time

- **Machine**: CPU Intel Core i7-3630QM, 4GB of RAM, Hard Drive: 5400 RPM. Operating System: Ubuntu 14.04 LTS

- **Single-thread**
  - Average Parsing time per webpage:
    - 0.5 seconds (throughput: 172,800 webpages per day)
  - Average Learning time (whole dataset)
    - Pharma vs Other: 40 seconds
    - Pharma vs Pharma: 18 seconds
  - Average Classification time:
    - Negligible: 5 milli seconds
Comparison with State-Of-the-Art

• Almost all publicly available tools/blacklists:
  • DNS-BH, DShield, Feodo Tracker, Google SafeBrowsing, Malc0de, Malwarebytes hpHosts, Malwareed, MalwareDomainList, OpenPhish, PhishTank, Spam404, Spamhaus DBL, SURBL, Yandex SafeBrowsing, Zeus Tracker
  • thanks to Guido Mureddu @ DIEE

• Only 5.24% of the 446 malicious websites detected by means of PharmaGuard were also listed in such blacklists

• Only 0.06% detected by Google Safebrowsing
  • default protection for many popular web browsers such as Firefox, Safari and Chrome

Almost no protection from Google Safe Browsing
User Interface

Pharma vs Other

Pharma vs Pharma
# Top 10 Autonomous Systems

<table>
<thead>
<tr>
<th>Illegitimate</th>
<th></th>
<th></th>
<th>Legitimate</th>
<th></th>
<th></th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share</td>
<td>Autonomous System</td>
<td>Organization</td>
<td>Share</td>
<td>Autonomous System</td>
<td>Organization</td>
<td></td>
</tr>
<tr>
<td>7.39%</td>
<td>AS8972</td>
<td>Intergenia AG</td>
<td>17.44%</td>
<td>AS33070, AS19994, AS27357</td>
<td>Rackspace Hosting</td>
<td></td>
</tr>
<tr>
<td>6.05%</td>
<td>AS16265</td>
<td>LeaseWeb B.V.</td>
<td>9.88%</td>
<td>AS26496</td>
<td>GoDaddy.com, LLC</td>
<td></td>
</tr>
<tr>
<td>5.15%</td>
<td>AS13335</td>
<td>CloudFlare, Inc.</td>
<td>9.30%</td>
<td>AS32577</td>
<td>The Kroger Co.</td>
<td></td>
</tr>
<tr>
<td>4.04%</td>
<td>AS24940</td>
<td>Hetzner Online AG</td>
<td>5.81%</td>
<td>AS3525</td>
<td>Albertson’s Inc.</td>
<td></td>
</tr>
<tr>
<td>3.81%</td>
<td>AS45839</td>
<td>Piradius NET</td>
<td>4.65%</td>
<td>AS22527</td>
<td>Safeway, Inc.</td>
<td></td>
</tr>
<tr>
<td>3.36%</td>
<td>AS16276</td>
<td>OVH SAS</td>
<td>2.90%</td>
<td>AS13626</td>
<td>Walgreens Co</td>
<td></td>
</tr>
<tr>
<td>3.14%</td>
<td>AS27956</td>
<td>Cyber Cast International</td>
<td>2.33%</td>
<td>AS16509, AS14618</td>
<td>Amazon.com, Inc.</td>
<td></td>
</tr>
<tr>
<td>2.69%</td>
<td>AS29073</td>
<td>Ecatel Network</td>
<td>1.74%</td>
<td>AS19271</td>
<td>Peak 10</td>
<td></td>
</tr>
<tr>
<td>2.24%</td>
<td>AS59642</td>
<td>Uab Duomenu Centras</td>
<td>1.74%</td>
<td>AS46606</td>
<td>Unified Layer</td>
<td></td>
</tr>
<tr>
<td>2.24%</td>
<td>AS20860</td>
<td>Iomar</td>
<td>1.74%</td>
<td>AS36351</td>
<td>SoftLayer Technologies</td>
<td></td>
</tr>
</tbody>
</table>

TABLE II. Top 10 autonomous systems and related organizations, source of either illegitimate or legitimate online pharmacies.
### Table IV. Top 20 Popular Domain Names According to Alexa, Associated to Either Illegitimate or Legitimate Online Pharmacies. The Table Also Shows the Ranking of the Domain Names on Two Well-Known Illegal Pharmacy Advertisement Sites, Namely, MuscleGurus.com and Eroids.com (From 0 to 100).

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Alexa Rank</th>
<th>Muscle Gurus Rank</th>
<th>Eroids Rank</th>
<th>Domain Name</th>
<th>Alexa Rank</th>
<th>Muscle Gurus Rank</th>
<th>Eroids Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1napsgear.org</td>
<td>88755</td>
<td>91.51</td>
<td>96.34</td>
<td>walmart.com</td>
<td>186</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>qualityvet.tblog.com</td>
<td>94315</td>
<td>-</td>
<td>-</td>
<td>target.com</td>
<td>370</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>steroid-seller.com</td>
<td>226578</td>
<td>0.0</td>
<td>0.0</td>
<td>walgreens.com</td>
<td>1423</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pharmaoffshore.com</td>
<td>261819</td>
<td>-</td>
<td>0.0</td>
<td>cvs.com</td>
<td>1931</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>unitedpharmacies.com</td>
<td>292003</td>
<td>-</td>
<td>0.0</td>
<td>drugstore.com</td>
<td>4677</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>alphamuscle.co</td>
<td>315631</td>
<td>99.12</td>
<td>87.56</td>
<td>safeway.com</td>
<td>6812</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>body-building-steroids.com</td>
<td>392873</td>
<td>0.0</td>
<td>-</td>
<td>kroger.com</td>
<td>7926</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>anabolicshop.ro</td>
<td>412654</td>
<td>0.0</td>
<td>-</td>
<td>shoprite.com</td>
<td>9160</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>steroidsone.eu</td>
<td>418820</td>
<td>0.0</td>
<td>-</td>
<td>drsfostersmith.com</td>
<td>16517</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>swoleoil.com</td>
<td>445293</td>
<td>-</td>
<td>96.36</td>
<td>1800petmeds.com</td>
<td>17051</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uk-anabolics.biz</td>
<td>470204</td>
<td>96.56</td>
<td>88.0</td>
<td>pet360.com</td>
<td>23104</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>originalsteroid.com</td>
<td>482693</td>
<td>97.6</td>
<td>-</td>
<td>medcohealth.com</td>
<td>25211</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>samento.eu</td>
<td>490764</td>
<td>0.0</td>
<td>-</td>
<td>vons.com</td>
<td>29300</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>purchasepeptides.com</td>
<td>615302</td>
<td>-</td>
<td>-</td>
<td>express-scripts.com</td>
<td>29926</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>anabolics.ro</td>
<td>663547</td>
<td>0.0</td>
<td>-</td>
<td>goodrx.com</td>
<td>37888</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>roidbazaar.com</td>
<td>674012</td>
<td>97.19</td>
<td>-</td>
<td>wegmans.com</td>
<td>44893</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>eurobolic.com</td>
<td>697948</td>
<td>0.0</td>
<td>-</td>
<td>caremark.com</td>
<td>45077</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>101fitnesspharma.com</td>
<td>711583</td>
<td>0.0</td>
<td>89.22</td>
<td>fredmeyer.com</td>
<td>45713</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>getroids.net</td>
<td>760984</td>
<td>0.0</td>
<td>86.88</td>
<td>petfooddirect.com</td>
<td>50949</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>alvgear.com</td>
<td>766986</td>
<td>0.0</td>
<td>86.67</td>
<td>ralphs.com</td>
<td>52477</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Top illegally-sold Prescription Drugs

• **Top drugs sold online** (found in this research): Peptides, Sustanon, Stanozolol, Prolinix Enanthate, Trenbolone, Clenbuterol, Human Growth Hormone (HGH)
  • Keywords extracted automatically from the learned Pharma vs Pharma Classifier

• **Severe adverse effects**: Strokes and heart attacks, irreversible tardive dyskinesia, irreversible clitomegaly, oligospermia, urinary obstruction, priapism, edema, death
  • These effects can be verified by looking for the above substances into the reputable website [http://www.drugs.com](http://www.drugs.com)

• **Concrete and persistent health threat**
**PharmaGuard** conclusions

- Novel architecture that can **automatically discover illegal online pharmacies**
  - advertised throughout the web
  - indexed by popular web search engines

- **Accurate and fast** learning & detection engine

- Substantial **complement for current blacklists** (focused analysis)

- Allows **law enforcement** operators to **focus only on** webpages that are most likely related to **illegal online pharmacies**

- Our results confirm that **illegal online pharmacies are a concrete, threatening problem**
Thank you! Questions?