Using Machine-Learning to Investigate Web Campaigns at Large
Who am I?

- HITB aficionado
- Computer security geek (since 2002)
- Love research, Ph.D. (in 2011)
- “Community centric”
- Work for Trend Micro
Web Defacement
= Website Compromise + Homepage Hijacking
Evolution

“Just-for-fun” era

Geopolitical factors kick in

90s  2000  2010  Now

Digital activism

Dark propaganda
“Just-for-fun” Era

---

| owned by ssh-2 | CHILE |

uname -a;id

Linux obelix 2.6.15-1-686 #2
Mon Mar 6 15:27:08 UTC 2006
i686 GNU/Linux

uid=0(root) gid=33(www-data)
groups=33(www-data)
Digital Activism

We hack because we are fighting against bad information, all of us ain't terrorists, or stupid boyz who don't know what they're doing. We're just some guys, maybe we've grown fastly, but we're curious. Is it a crime to be interested in learning? Each of us has an own life, which can be good or bad. We don't like to live in misinformation or ignorance. We want a better world; a world where there ain't inequalities, differences, injustices: a new world! We don't want that the world will be guided by multinational companies, we don't want a world like this. Our world has to be of everyone who passes his/her own believings, religions, races, politic ideas, social conditions and everything which makes differences between people. We want this, and we're fighting for this. We are against each kind of war, against each kind of oppression, against each form of abuse of power: we're fighting and we'll fight in each place where the individual freedom is threatened. We will fight under each flag besides people we don't even know but whose idea of freedom is the same of our one. We haven't open access to the means of communication which are false, they say what they want and in the way they prefer, this is the only way to spread our ideas. Northern Ireland, Kosovo, Iran, Iraq, Afghanistan, Chile, Peru, Myanmar, Brazil, Philippines, Palestine, Pakistan, Kurdistan, Turkey, Algeria, South Africa, China, Taiwan, Kenya are places where freedom is a dream, open your eyes, open your mind, wake up, and all toghether we'll turn freedom from dream to reality. You can agree with us or not. If you agree with us let diffuse this message to make our ideas become everybody's ideas
SANTIAGO, Chile — Gen. Augusto Pinochet, who terrorized his opponents for 17 years after taking power in a bloody coup, died Sunday, putting an end to a decade of intensifying efforts to bring him to trial for human rights abuses blamed on his regime. He was 91.

Supporters saw Pinochet as a Cold War hero for overthrowing democratically elected President Salvador Allende at a time when the U.S. was working to destabilize his Marxist government and keep Chile from exporting communism in Latin America.
Coordinated campaigns:
From one to many targets

- Death statement is 2\textsuperscript{nd} of May 2011
- Campaign is 6\textsuperscript{th}-12\textsuperscript{th} of May
- Targets:

\textbf{Hello Admin!}

And Hello By all your Amircan visit..\textbf{Osama bin Laden} is not die ~ he Lives inside us .. God save you if you realy gone or still alive and be Sure will come a Millions of \textbf{Osama bin Laden} ..
The Israeli-Palestinian conflict
Data Collection
## Public Repositories

### Top 30 Users

<table>
<thead>
<tr>
<th>No</th>
<th>User Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No1</td>
<td>越南邻国宰相 [34333]</td>
<td></td>
</tr>
<tr>
<td>No2</td>
<td>Jack Riderr [19199]</td>
<td></td>
</tr>
<tr>
<td>No3</td>
<td>Mr.Kro0oz.305 [10672]</td>
<td></td>
</tr>
<tr>
<td>No4</td>
<td>AlfabetoVirtual [8907]</td>
<td></td>
</tr>
<tr>
<td>No5</td>
<td>宇少 [6559]</td>
<td></td>
</tr>
<tr>
<td>No6</td>
<td>大圣 [6486]</td>
<td></td>
</tr>
<tr>
<td>No7</td>
<td>域血 [5408]</td>
<td></td>
</tr>
<tr>
<td>No8</td>
<td>That is me [5282]</td>
<td></td>
</tr>
<tr>
<td>No9</td>
<td>惨悴 [5013]</td>
<td></td>
</tr>
<tr>
<td>No10</td>
<td>颓废 [4791]</td>
<td></td>
</tr>
<tr>
<td>No11</td>
<td>亱梦ゆ [4685]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Notifier</th>
<th>Domain</th>
<th>Country</th>
<th>PR</th>
<th>View</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-02-12</td>
<td>Tc</td>
<td><a href="http://bananabags.co.za/">http://bananabags.co.za/</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Tc</td>
<td><a href="http://bsvrenault.co.za/">http://bsvrenault.co.za/</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Tc</td>
<td><a href="http://bsgtriathlon.org/">http://bsgtriathlon.org/</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Tc</td>
<td><a href="http://broadbandondemand.co.za/">http://broadbandondemand.co.za/</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Tc</td>
<td><a href="http://bettacars.co.za/">http://bettacars.co.za/</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Tc</td>
<td><a href="http://blog.camelotkennels.co.za/">http://blog.camelotkennels.co.za/</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Al Virtual</td>
<td><a href="http://www.conseils-incendie.fr/images/j">http://www.conseils-incendie.fr/images/j</a></td>
<td>France</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Al Virtual</td>
<td><a href="http://portalmare.it/images/jdownloads/s">http://portalmare.it/images/jdownloads/s</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>Al Virtual</td>
<td><a href="http://www.fruitinfo.hu/images/jdownload">http://www.fruitinfo.hu/images/jdownload</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
<tr>
<td>2016-02-12</td>
<td>ch14</td>
<td><a href="http://0536ink.com/cm.html">http://0536ink.com/cm.html</a></td>
<td>Country</td>
<td>0</td>
<td>View</td>
</tr>
</tbody>
</table>
Others

Welcome to
MyDeface.com || Let's Show off....

Latest Verified Mirrors

<table>
<thead>
<tr>
<th>Name</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drummon</td>
<td><a href="http://upsanddowns-hordaland.no/web">http://upsanddowns-hordaland.no/web</a></td>
</tr>
<tr>
<td>Drummon</td>
<td><a href="http://bradfordscurtainssheerblinds.co.uk/">http://bradfordscurtainssheerblinds.co.uk/</a></td>
</tr>
<tr>
<td>Drummon</td>
<td><a href="http://leatherheaddistrictscouts.org.uk/">http://leatherheaddistrictscouts.org.uk/</a></td>
</tr>
<tr>
<td>Drummon</td>
<td><a href="http://leatherheaddistrictscouts.org.uk/">http://leatherheaddistrictscouts.org.uk/</a></td>
</tr>
<tr>
<td>Drummon</td>
<td><a href="http://pickwickassociation.org.uk/web">http://pickwickassociation.org.uk/web</a></td>
</tr>
<tr>
<td>Drummon</td>
<td><a href="http://rhodapartridge.co.uk/webye">http://rhodapartridge.co.uk/webye</a>...</td>
</tr>
</tbody>
</table>

Hi There!
I am so happy to announce the new Mirror-zone.org unrestricted information site, with few nice changes.

We are now at testing this script, this means the script is not stable, and you may get some errors, if you face any issues please contact us and we will be glad to work with you to in order to resolve such issues. Also, this not mean that the Gold release will have the same futures with the bugs fixes, and we will continue to add new futures to the site.

Best Regards,
Administration

Hello
mirror-zone.org followers,
Did your defaced site mirror goes to onhold?
Because your notify name are missing on your deface page
if you want your mirror goes to archive directly
then follow the instructions

Hack Mirror
@hackmirror
The place where we make you a legend
📍 The Universe
🔗 hackmirror.com
📅 Iscrizione a marzo 2012
THE site :)
## Collected Data

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Website URL</th>
<th>Acquired Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone-H</td>
<td><a href="http://www.zone-h.org">www.zone-h.org</a></td>
<td>12,303,240</td>
</tr>
<tr>
<td>Hack-CN</td>
<td><a href="http://www.hack-cn.com">www.hack-cn.com</a></td>
<td>386,705</td>
</tr>
<tr>
<td>Mirror Zone</td>
<td><a href="http://www.mirror-zone.org">www.mirror-zone.org</a></td>
<td>195,398</td>
</tr>
<tr>
<td>Hack Mirror</td>
<td><a href="http://www.hack-mirror.com">www.hack-mirror.com</a></td>
<td>68,980</td>
</tr>
<tr>
<td>MyDeface</td>
<td><a href="http://www.mydeface.com">www.mydeface.com</a></td>
<td>37,843</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>12,992,166</strong></td>
</tr>
</tbody>
</table>
# Data Trustworthiness

<table>
<thead>
<tr>
<th>Type</th>
<th>Attribute</th>
<th>Example</th>
<th>Trustworthiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata</td>
<td>URL</td>
<td><a href="http://target.gov">http://target.gov</a></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Timestamp</td>
<td>2010-01-02 15:00</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Nickname</td>
<td>Neo Hacker</td>
<td>Medium-Low</td>
</tr>
<tr>
<td></td>
<td>Webserver; Reason; Hack Mode</td>
<td>Nginx; Political; SQLi</td>
<td>Low</td>
</tr>
<tr>
<td>Raw content</td>
<td>Main page</td>
<td>HTML or TXT file</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Embeeeded resources</td>
<td>Various format</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>External resources</td>
<td>Various format</td>
<td>Medium-High</td>
</tr>
</tbody>
</table>
General Trends
Topics Over The Years

<table>
<thead>
<tr>
<th>Year</th>
<th>Most relevant topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>question, student, security, number, place</td>
</tr>
<tr>
<td>1999</td>
<td>cowboy, team, security, think</td>
</tr>
<tr>
<td>2000</td>
<td>baby, tabloid, people, provided</td>
</tr>
<tr>
<td>2001</td>
<td>lord, prime, provided, saved, better</td>
</tr>
<tr>
<td>2002</td>
<td>worry, sind, lame, care, encryption</td>
</tr>
<tr>
<td>2003</td>
<td>backup, gift, team, came, take</td>
</tr>
<tr>
<td>2004</td>
<td>best, group, micro, look, total</td>
</tr>
<tr>
<td>2005</td>
<td>normal, pope, time, familia, contact</td>
</tr>
<tr>
<td>2006</td>
<td>terror, saved, intruder, energy, user</td>
</tr>
<tr>
<td>2007</td>
<td>badger, since, high, turk, turkey</td>
</tr>
<tr>
<td>2008</td>
<td>crew, speech, warning, saved, team</td>
</tr>
<tr>
<td>2009</td>
<td>knowledge, acker, team, album, country</td>
</tr>
<tr>
<td>2010</td>
<td>posted, member, protocol, kernel, security</td>
</tr>
<tr>
<td>2011</td>
<td>contact, security, village, holding, highlander</td>
</tr>
<tr>
<td>2012</td>
<td>saved, contact, team, underground</td>
</tr>
<tr>
<td>2013</td>
<td>team, forgive, security</td>
</tr>
<tr>
<td>2014</td>
<td>eagle, crew, electronic</td>
</tr>
<tr>
<td>2015</td>
<td>clash, king, terrorism, visit, alligator</td>
</tr>
<tr>
<td>2016</td>
<td>marocain, turk, steel, anonymous, team</td>
</tr>
</tbody>
</table>

**Security Problems**

**Real World Events**
Adoption of Malicious Content
Adoption of email & Twitter handlers
Detection Engineering
Key Observations

#OpFrance

 Owned!

United Islamic Cyber Force
U Remember Us ? We Are Back Again!
#We Are Stand With Muslims in Europe#
I Challenge Your Government To Stop Us!

#OpFrance

France Coupes Website Hacked

United Islamic Cyber Force
#We Are Stand With Muslims in Europe#
We Are Not Terrorist , And Stop Calling Muslim Terrorist
Respect Existance , Or Expect Resistance
We Are Muslims, We Love Peace
I Challenge Your Government To Stop Us!
Key Observations

**Template**

- Hacked by Anonymous R4BIA Team
  
  https://www.facebook.com/R4BIAHACKERSTEAM

  Please admin or visitor I believe your whole life will change after seeing these videos
  If you are Christian see the fourth video :) you must see it

  Translation in the top of the page at your left hand
  
  What will happen to you after your death!
  Why you are exist? What is your destiny?
  You will find the solutions or answers only in Islam. Islam is the religion of equality.

**Customization**

- Hacked by Free

  Please admin or visitor I believe your whole life will change after seeing these videos

  Are you exited?
  Ask yourself
  What will happen to me after my death!
  Who am I?
  Why I exist?
  What is my destiny?
  These questions and made the whole world.
  You will find its solutions or answers only in Islam.
  Islam is the religion of equality.
Key Observations

1. Actors **cooperate** in teams
   Especially if driven by strong ideologies

2. Defacements are organized around **campaigns**

3. When a team prepares and runs a campaign, it tends to re-use a common **template** that each member can customize
Next Generation*
Defacement Explorer

(DefPloreX-NG)

(*) 1st generation presented at BH Arsenal
Deface Page Analysis

Defacement pages → De-duplication (hash based) → Dynamic analysis → Static analysis → Extracted content → Metadata (e.g., timestamp, URL, reporting source) → Elastic
Labeling & Visualization

Defacement pages
De-duplication (hash based)
Dynamic analysis
Static analysis
Extracted content
Metadata (e.g., timestamp, URL, reporting source)

Deface Page Analysis

Campaign Detection
Normalization
Clustering of Deface Pages
Feature Extraction
Re-duplication

Elastic

Labeling & Visualization
### The Indo-Pakistani conflict

#### Pre-Filtering Search

- Parameter: "Free Kashmir" AND tid_name:"in"  
- Campaigns Search

#### Choose clustering features

<table>
<thead>
<tr>
<th>Size</th>
<th>Key</th>
<th>Start</th>
<th>End</th>
<th>N. Attackers</th>
<th>N. Text Hashes</th>
<th>N. TLDs</th>
<th>N. Domains</th>
<th>Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>183</td>
<td>Details</td>
<td>2013-11-01 12:00</td>
<td>2015-04-01 12:00</td>
<td>4</td>
<td>14</td>
<td>1</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>Details</td>
<td>2011-09-01 12:00</td>
<td>2016-04-01 12:00</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>137</td>
<td>Details</td>
<td>2012-03-01 12:00</td>
<td>2015-11-01 12:00</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>Details</td>
<td>2011-06-01 12:00</td>
<td>2016-03-01 12:00</td>
<td>18</td>
<td>39</td>
<td>1</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>Details</td>
<td>2015-03-01 12:00</td>
<td>2016-05-01 12:00</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>122</td>
<td></td>
</tr>
</tbody>
</table>
Free Kashmir. Freedom is our goal. End the Occupation.

"Indian Panel Code (Act No. 45 of 1860) CHAPTER -II 18: India. India means the territory of India excluding the State of Jammu and Kashmir."

This institutionalized impunity with which the killings of civilians by military and police forces in Jammu and Kashmir continues should be a source of shame for India which propagates to be a democracy!
Implementation Details
Features Engineering

- **Format of Title**
- **Visual Features** (colors)
- **Visual Features** (images)
- **Social Features**
- **Structural Features**
Clustering

• BIRCH
  – Balanced Iterative Reducing and Clustering Hierarchies

• Do not materialize the entire distance matrix
  – Statistical values are efficient to compute
  – Quickly find the closest cluster for each new data points
Clustering

- Scalability of BIRCH vs. DBSCAN (10 runs)
Labeling

- Each cluster is represented by a succinct report
  - Time span
  - Screenshot thumbnails (by perceptual hash)
  - Name of actors and teams
  - Keywords used in campaigns (e.g. #opfrance)
  - Category of targets (e.g. news, governmental sites)
Findings
Organization of Actors

43% of actors join one team, at least

Half of joint campaigns are larger than 3

70% of campaigns do collaborate

CDF(\(x\)) = Number of (ref. legend)
Geopolitical Real-World Events

- Successfully detected
The “Charlie Hebdo” case

Hello.. Admin
I'm not Charlie
We are not Charlie
This message is clear
We are here to protect our prophet Mohamed
#OpFrance
#OpCharlie
Remember that name  Docteur Hm
The “Charlie Hebdo” case
Example of large-scale joint campaigns. While the entire Israeli-Palestinian conflict involves 12 campaigns, *opisreal* and *opsavealaqsa* represent the most aggressive and active ones.
Joint campaigns conducted by *anonymous*-affiliated groups against governmental sites.
Long-Term vs. Aggressive Campaigns

Campaign *savegaza* reacted to war events in the Gaza Strip
Example of a long-running campaign named h4ck3rsbr. The horizontal bars represent the different sub-campaigns (60) with their most targeted TLDs and teams (next slide). This is a very generic campaign, with different affiliates.
Most targeted TLDS:

- com.mx
- qc.ca
- it
- tv
- cz

- **c0d3rz (24 def.)**
- **Trustix (38 def.)**
- **Fatal Error (6 def.)**
- **Infektion Group (17 def.)**
Conclusions

- Dark propaganda
- Prevailing phenomenon
- Driven by geopolitical motivations
- Key targets, influencing sites

- Contribute to make the Internet a better world!
Conclusions

- DefPloreX-NG
- GitHub (old code, new ask):
  https://github.com/trendmicro/defplorex
- Paper*:

(*) Joint work with Federico Maggi, Ryan Flores, Lion Gu and Vincenzo Ciancaglini
Thanks! Questions?
Dr. Marco Balduzzi
@embyte, madlab.it