



Using Machine-Learning to Investigate Web Campaigns at Large

Dr. Federico Maggi, @phretor

Joint work with Marco Balduzzi, Ryan Flores, Lion Gu and Vincenzo Ciancaglini

Web Defacement

=

Website Compromise

+

Homepage Alteration

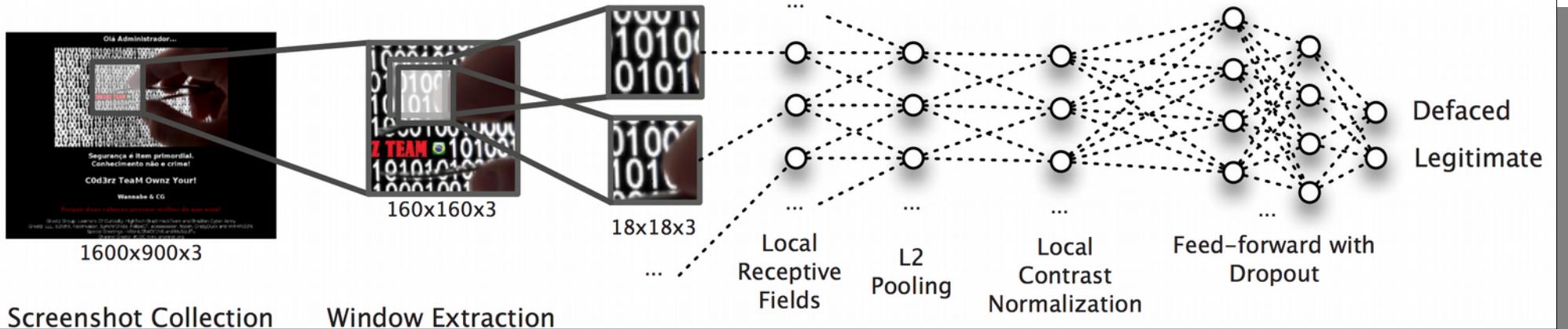
FOR IMMEDIATE RELEASE

Thursday, May 10, 2018

California Man Arrested For Hacking Websites For The Combating Terrorism Center At West Point And The New York City Comptroller

**The Defendant Committed More Than 11,000 Defacements of Various Military,
Government, and Business Websites Around the World Using the Online Pseudonym
“Alfabetovirtual”**

Geoffrey S. Berman, the United States Attorney for the Southern District of New York, and William F. Sweeney Jr., the Assistant Director-in-Charge of the New York Field Office of the Federal Bureau of Investigation (“FBI”), announced today the arrest of BILLY RIBEIRO ANDERSON, a/k/a “Anderson Albuquerque,” a/k/a “AlfabetoVirtual.” ANDERSON was charged with three separate counts of computer fraud for obtaining unauthorized access to and committing defacements of the websites for the Combating Terrorism Center at the United States Military Academy in West Point, New York (“West Point”), and the Office of the New York City Comptroller (the “NYC Comptroller”). ANDERSON was arrested earlier this



Meerkat: Detecting Website Defacements through Image-based Object Recognition

Kevin Borgolte, Christopher Kruegel, Giovanni Vigna

University of California, Santa Barbara

{kevinbo,chris,vigna}@cs.ucsb.edu

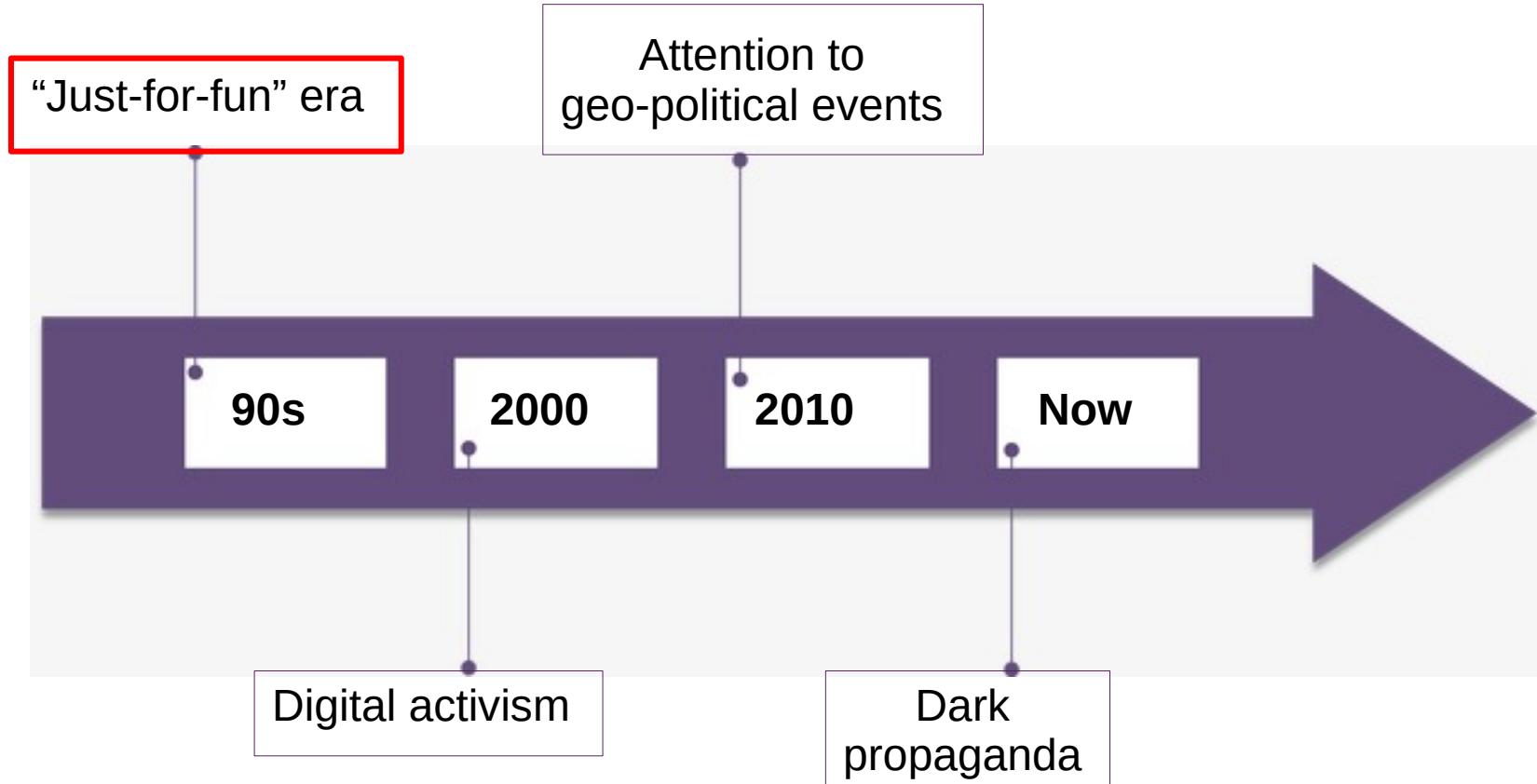
Abstract

Website defacements and website vandalism can inflict significant harm on the website owner through the loss of sales, the damage to reputation, and loss of brand recognition.

1 Introduction

The defacement and vandalism of websites is an attack that disrupts the operation of companies and organizations, tarnishes their brand, and plagues websites of all sizes, from the small organization to the website of a large

Evolution



“Just-for-fun” Era

```
=====
HACKED BY LIBERO
=====
Login : Libero      Ok !
Password: *****      Ok !
=====
Conection Ok !
=====
CFK !!!
=====
WE ARE ARGENTINA HACK TEAM
```

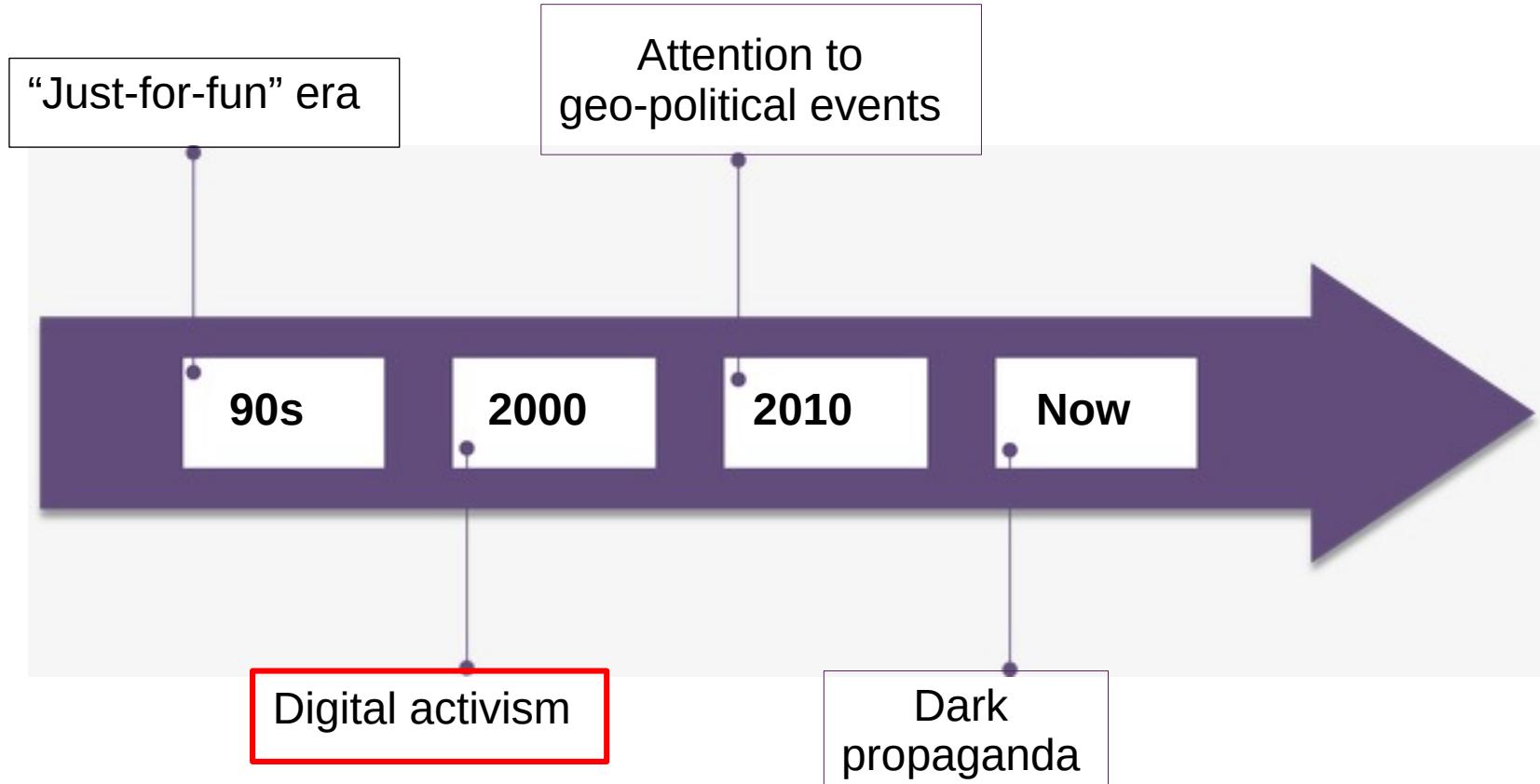
| owned by ssh-2 | CHILE |

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)

Evolution



Digital Activism

We hack because we are fighting against bad information, all of us ain't terrorists, [REDACTED] we're curious. Is it a crime to be interested in learning? [REDACTED]

[REDACTED]. Our world has to be of everyone who passes his/her own beliefings, 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,

[REDACTED] 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

Evolution



Attention to Geo-political Events

msnbc.com news services

updated 12/10/2006 11:47:52 PM ET

[Print](#) | Font: A A + -

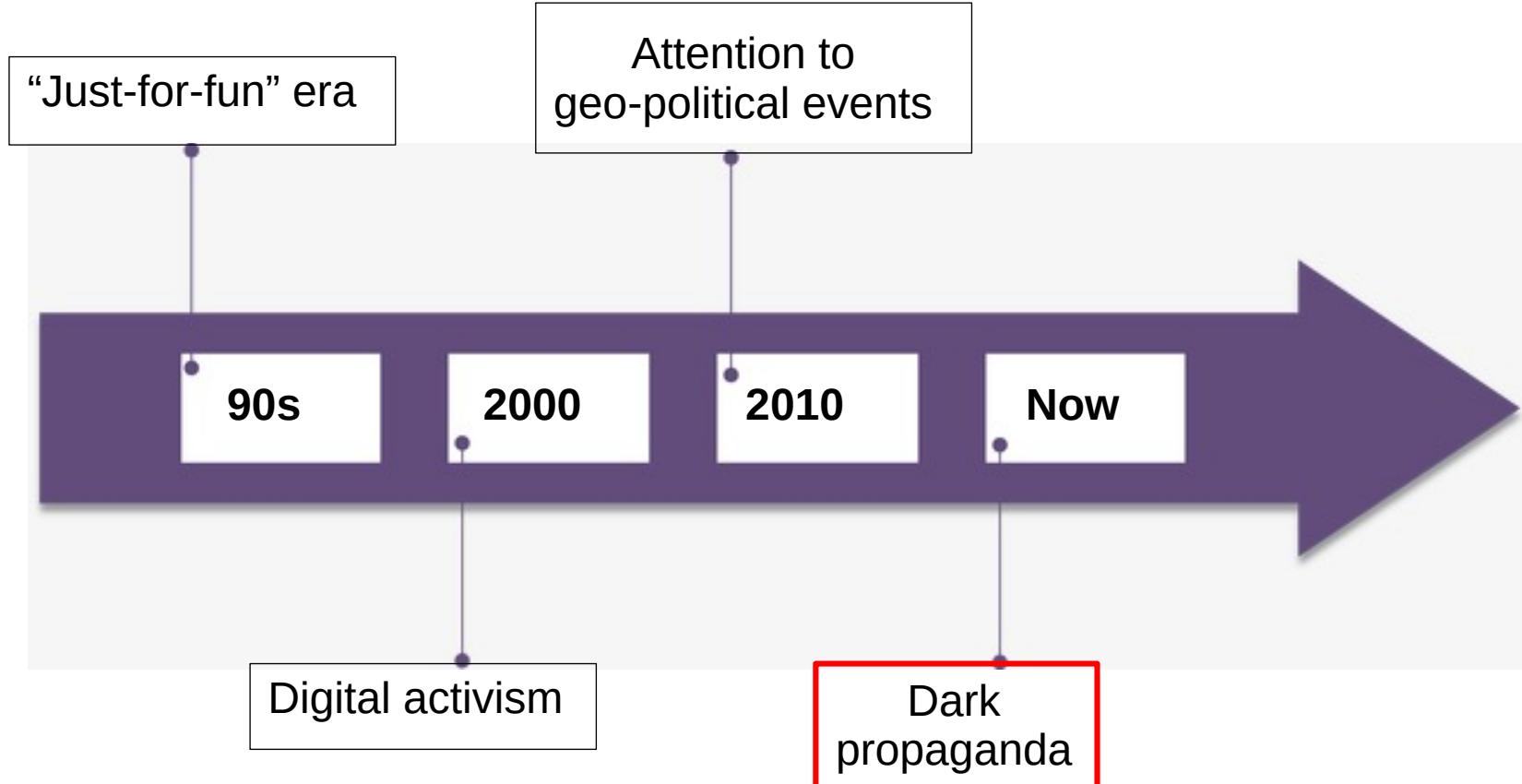
SANTIAGO, Chile — Gen. Augusto Pinochet, who terrorized his opponents for 17 years after taking power in a bloody coup, **died Sunday**, putting an end to a decade of intensifying efforts to bring him to trial for human rights abuses blamed on his regime. He was 91.

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.

Hacked by @br
R.I.P Pinochet

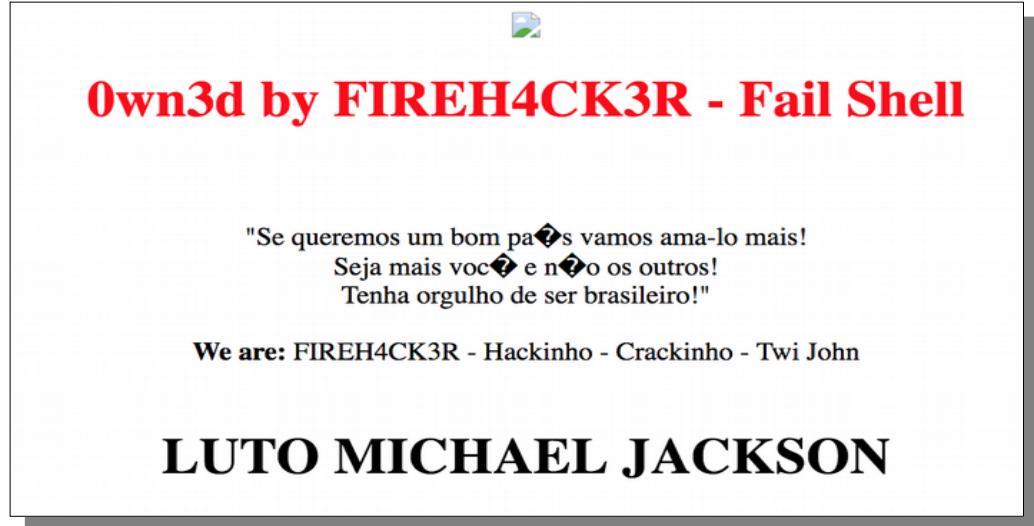
ix15

Evolution



From isolated events to coordinated campaigns

- Death statement is June 25th, 2009
- Sporadic defacements: June 28th
- Campaigns: August



That's interesting
But how do we scale?

Public Repositories



Search :

Notifier ▾

Search

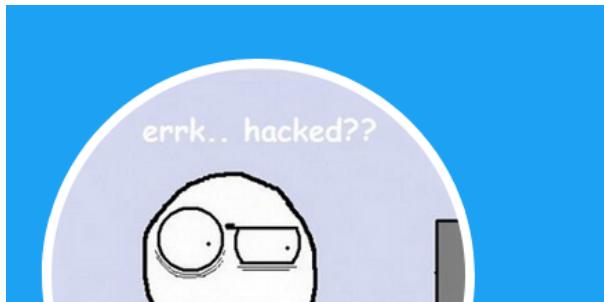
Notifier

Mass Notifier

Top100 User

TOP30 User	Date	Notifier	Domain	Country	PR	View
No1 : 越南邻国宰相 [34333]	2016-02-12	To	http://bananabags.co.za/	Country	0	View
No2 : Jack Riderr [19199]	2016-02-12	To	http://bsvrenault.co.za/	Country	0	View
No3 : Mr.KroOoz.305 [10672]	2016-02-12	To	http://bsgtriathlon.org/	Country	0	View
No4 : AlfabetoVirtual [8907]	2016-02-12	To	http://broadbandondemand.co.za/	Country	0	View
No5 : 宇少 [6559]	2016-02-12	To	http://bettacars.co.za/	Country	0	View
No6 : 大圣 [6486]	2016-02-12	To	http://blog.camelotkennels.co.za/	Country	0	View
No7 : 域血 [5408]	2016-02-12	All	Virtual	http://www.conseils-incendie.fr/images/j		0 View
No8 : That is me [5282]	2016-02-12	All	Virtual	http://portalmare.it/images/jdownloads/s	Country	0 View
No9 : 憔悴 [5013]	2016-02-12	All	Virtual	http://www.fruitinfo.hu/images/jdownload		0 View
No10 : 颓废 [4791]	2016-02-12	All	Virtual	http://www.vosges18.com/images/j		0 View

Others



errk.. hacked??

you

Hack Mirror

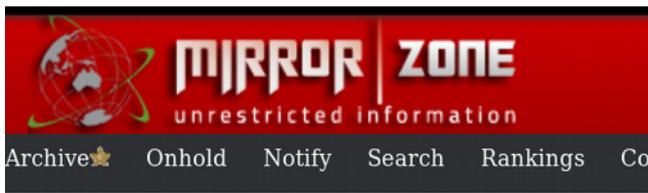
@hackmirror

The place where we make you a legend

⌚ The Universe

🔗 hackmirror.com

📅 Iscrizione a marzo 2012



MIRROR | ZONE
unrestricted information

Archive ★ Onhold Notify Search Rankings Contact

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

Welcome to

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

Latest Verified Mirrors

Name

emon	http://upsanddowns-hord
emon	http://bradfordcurtainsandbl
emon	http://leatherheaddistrictsc
emon	http://leatherheaddistrictsc
emon	http://pickwickassociation
emon	http://rhodapartridge.c

THE site :)



ZONE-H In Numbers

News: **4.738**

Admins: **4**

Registered Users: **142.949**

Early Warning subscriptions: **7312**

Digital Attacks: **13.498.756**

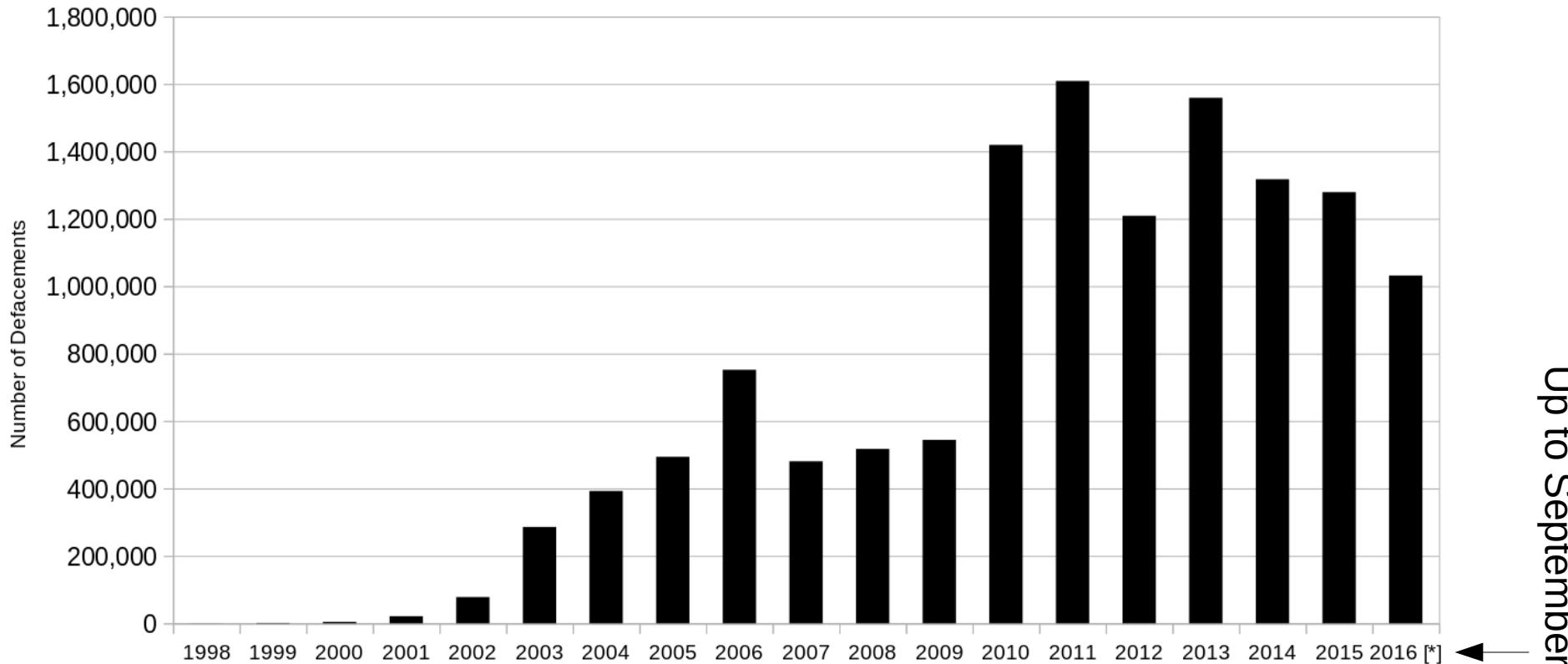
Attacks On Hold: **333.287**

Online Users: **268**

Collected Data

Source Name	Website URL	Acquired Records
Zone-H	www.zone-h.org	12,303,240
Hack-CN	www.hack-cn.com	386,705
Mirror Zone	www.mirror-zone.org	195,398
Hack Mirror	www.hack-mirror.com	68,980
MyDeface	www.mydeface.com	37,843
	TOTAL	12,992,166

Amount of Events Over Time

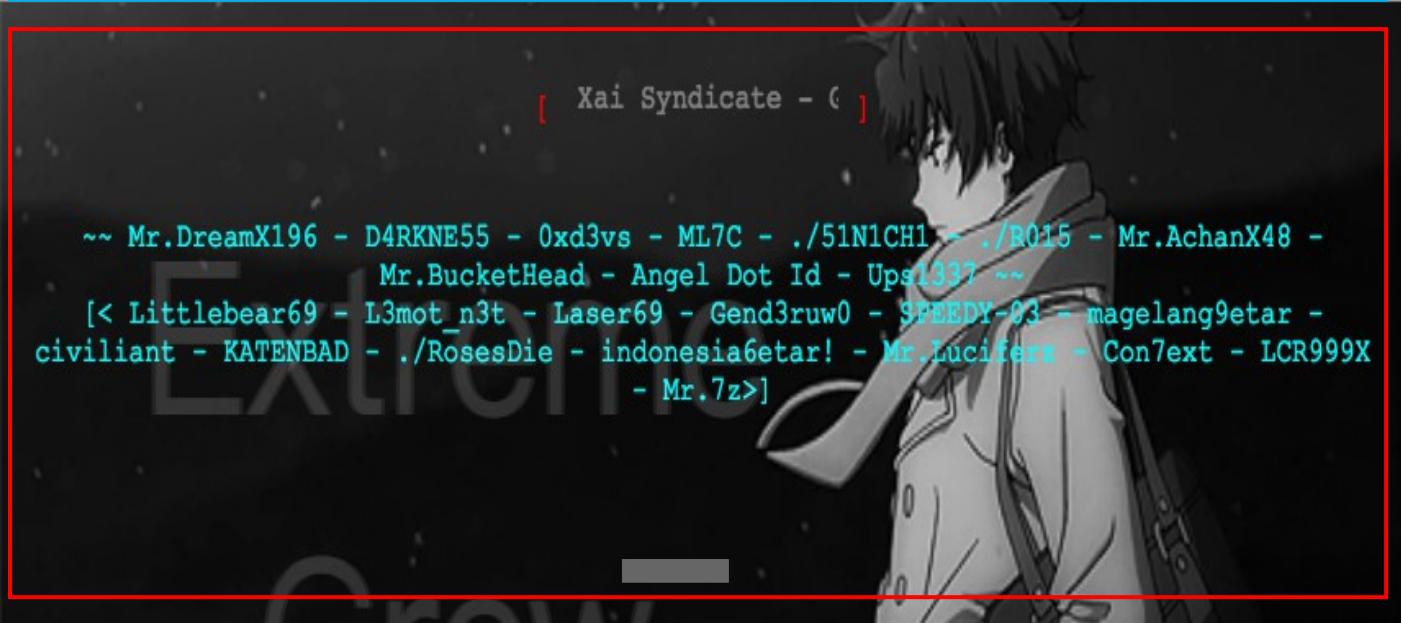


Data Format

Metadata →

Notified by: Mr.ToKeiChun69 **Domain:** http://www.redraven.co.za/files.html **IP address:** 75.119.204.64 
System: Linux **Web server:** Apache **Notifier stats:** 
This is a CACHE (mirror) page of the site when it was saved by our robot on 2018-05-24 01:28:32

Raw content →



[Xai Syndicate - C]

~~ Mr.DreamX196 - D4RKNE55 - 0xd3vs - ML7C - ./51N1CH1 - ./R015 - Mr.AchanX48 -
Mr.BucketHead - Angel Dot Id - Ups1337 ~~
[< Littlebear69 - L3mot_n3t - Laser69 - Gend3ruw0 - SPEEDY-03 - magelang9etar -
civiliant - KATENBAD - ./RosesDie - indonesia6etar! - Mr.Luciferz - Con7ext - LCR999X
- Mr.7z>]

Data Accuracy

Type	Attribute	Example	Trustworthiness
Metadata	URL	http://target.gov	High
	Timestamp	2010-01-02 15:00	Medium
	Nickname	Neo Hacker	Medium-Low
	Webserver; Reason; Hack Mode	Nginx; Political; SQLi	Low
Raw content	Main page	HTML or TXT file	High
	Embedded resources	Various format	High
	External resources	Various format	Medium-High

General Trends

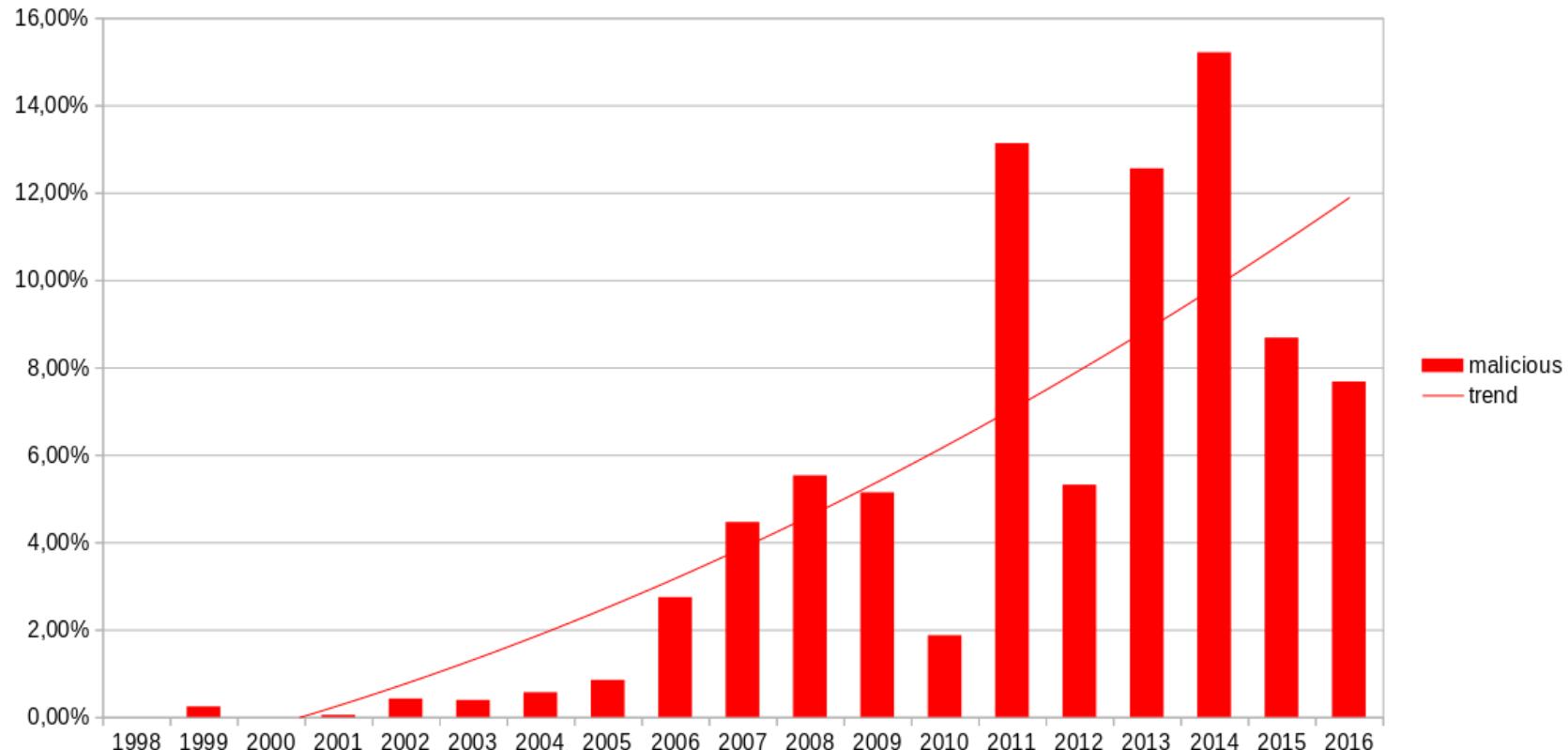
Topics

Security Problems

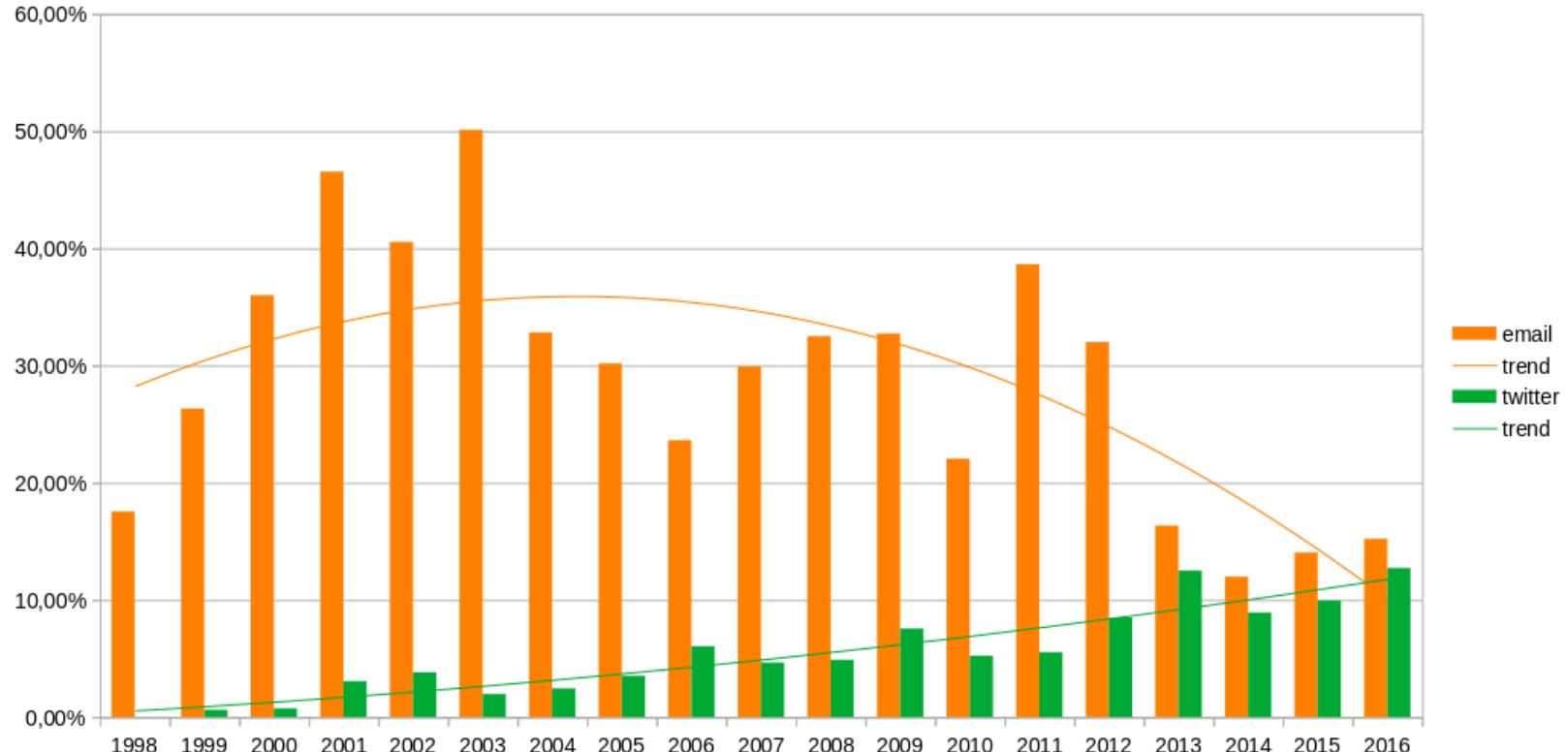
Year	Most relevant topics
1998	question, student, security , number, place
1999	cowboy, <i>team</i> , security , think
2000	baby, tabloid, people, provided
2001	lord, prime, provided, saved, better
2002	worry, sind, lame , care, encryption
2003	backup , gift, <i>team</i> , came, take
2004	best, group,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

Real World Events

Adoption of Malicious Content



Adoption of email & Twitter handlers

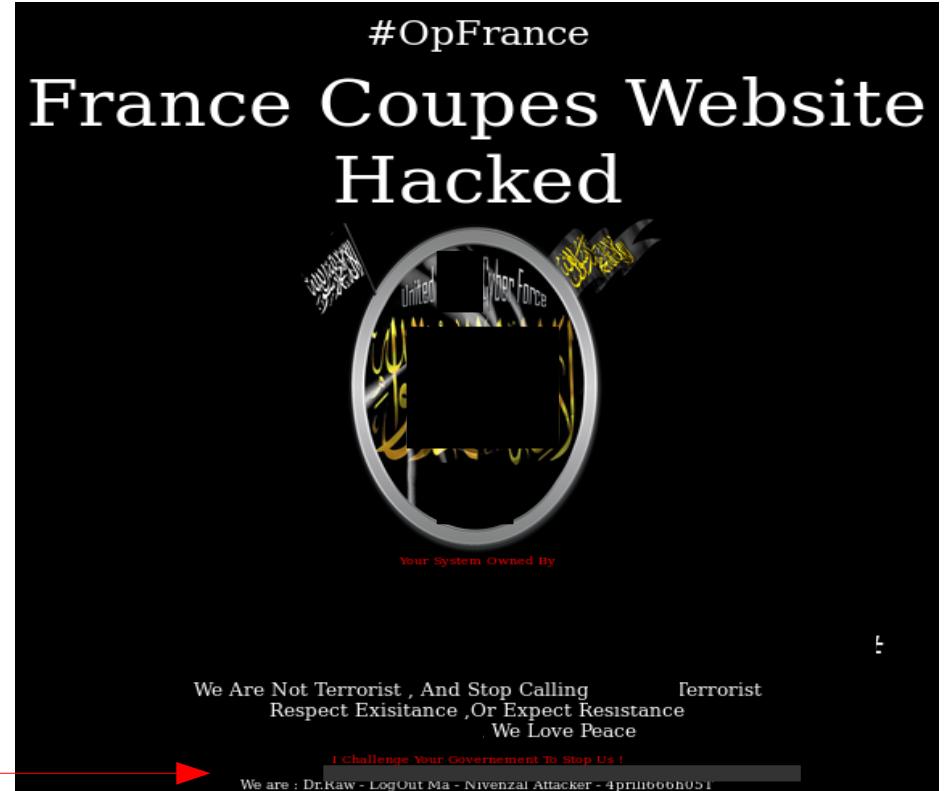


General Trends

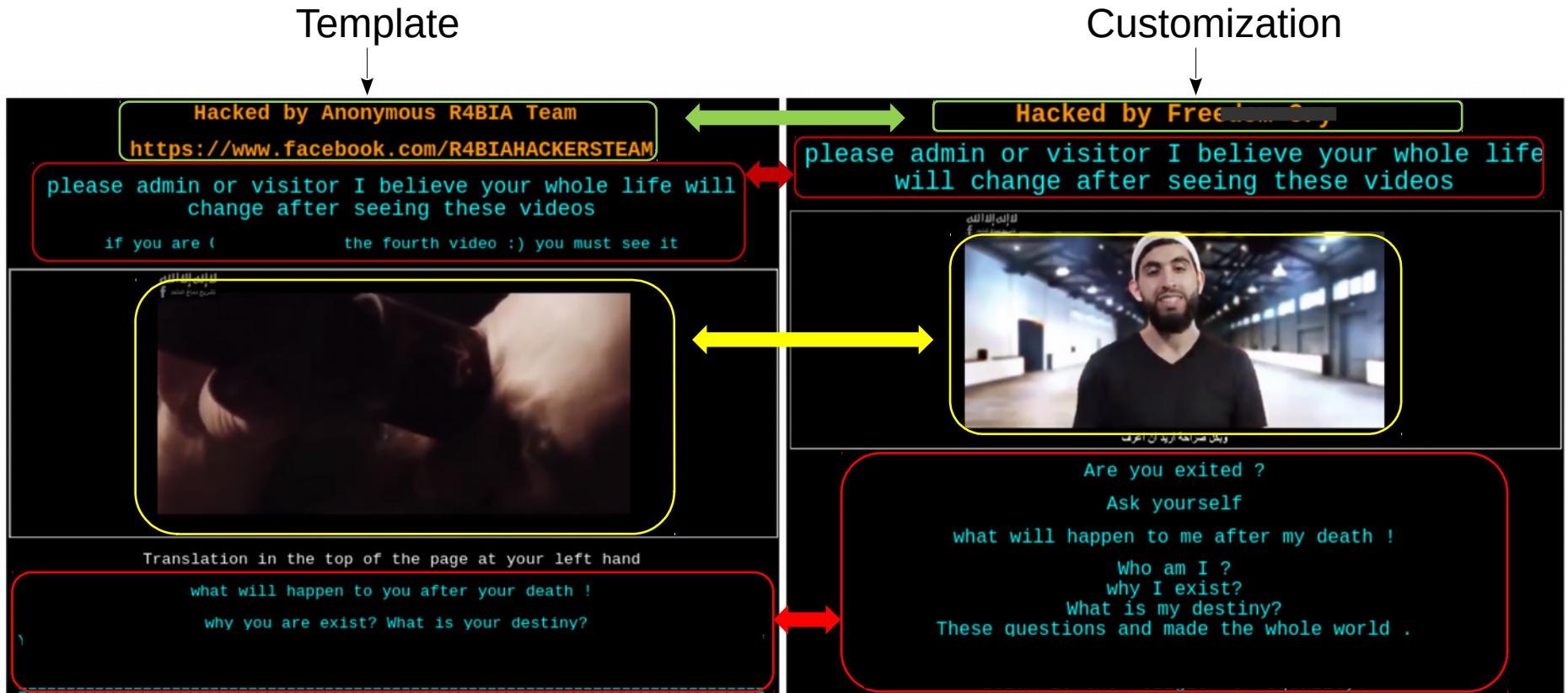


Key Observations

Teams vs. Individuals



Templates & Campaigns



Key Observations

1. Actors **cooperate** in teams

Especially if driven by **similar, strong ideologies**

2. Defacements are organized around **campaigns**

3. Teams re-use a common **template** that each member can customize

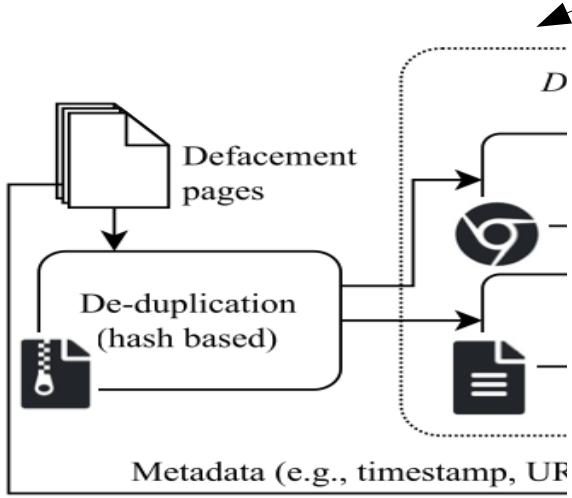
Key Observations



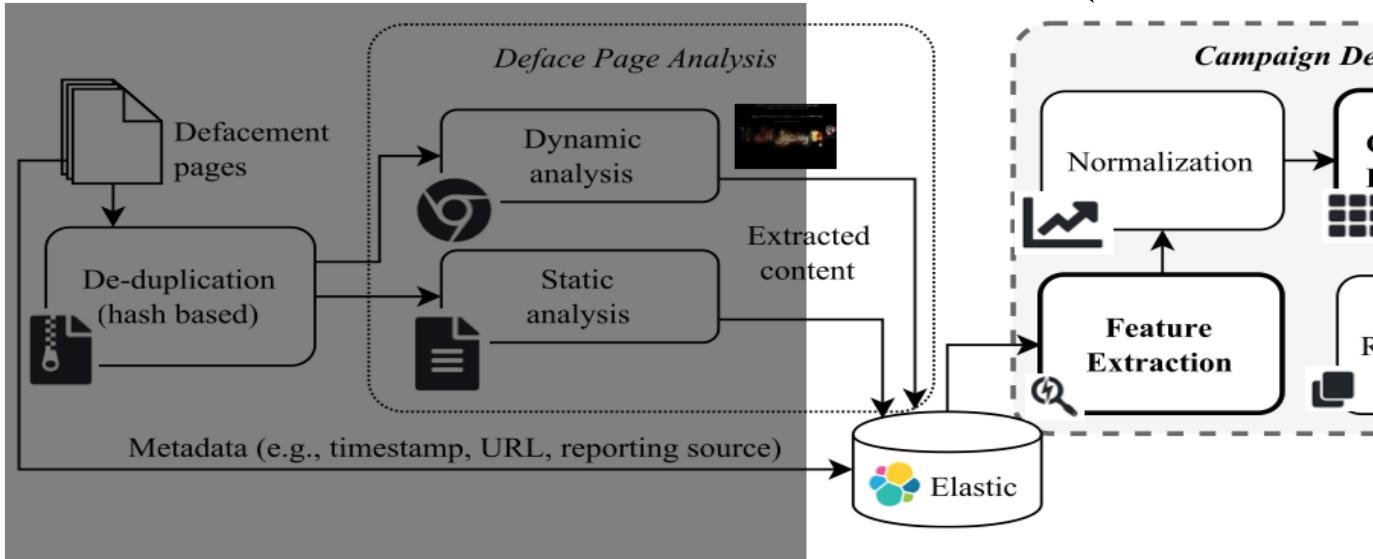
Next Generation* Defacement Explorer

(DefPloreX-NG)

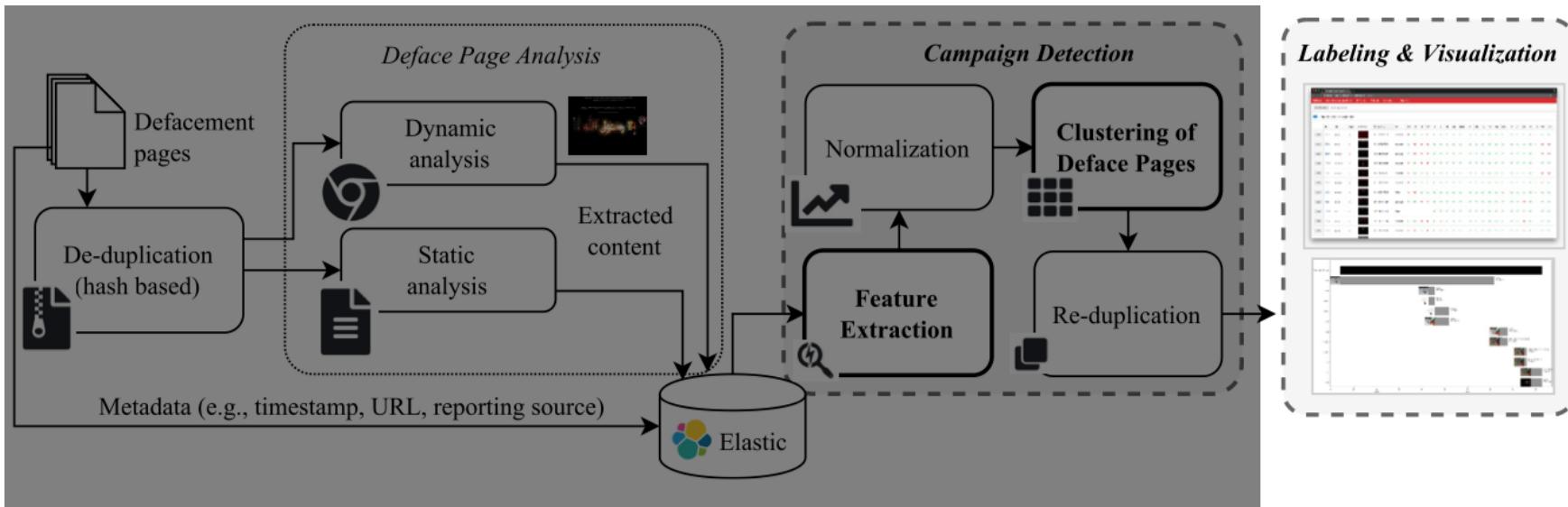
Deface Page Analysis



Campaign Detection



Labeling & Visualization



1,791 records

deepwater horizon

Pre-Filtering Search

1,267 clusters

params.cluster_size > 1

Campaigns Search

Choose clustering features

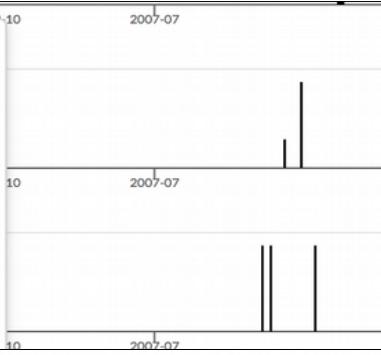
	Size	Key	Start	End	N. Attackers	N. Text Hashes	N. TLDs	N. Domains	Timeline
Details	72	European	2003-04-01 12:00	2011-08-01 12:00	2	2	3	3	
Details	56	Other	2001-05-01 12:00	2016-09-01 12:00	42	4	13	52	
Details	53	Arabic	2011-05-01 12:00	2011-05-01 12:00	1	1	3	53	
Details	31	Other	2016-08-01 12:00	2016-08-01 12:00	1	1	7	31	
Details	28	Arabic	2011-05-01 12:00	2011-05-01 12:00	1	1	3	28	

Top domains

instruction.com

ormatique.com

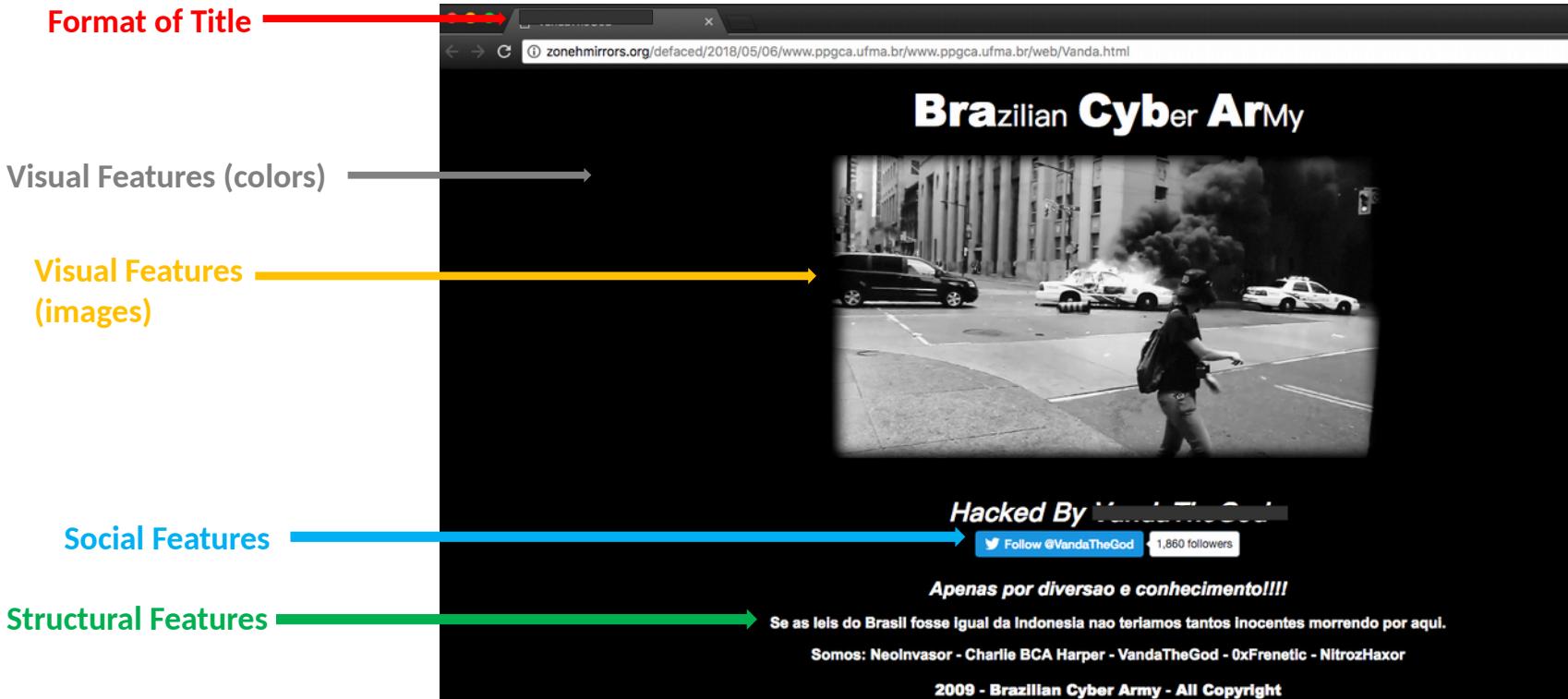
izon-micro.com



Hacked By GHoST61 - Turksec.info

Implementation Details

Feature Engineering

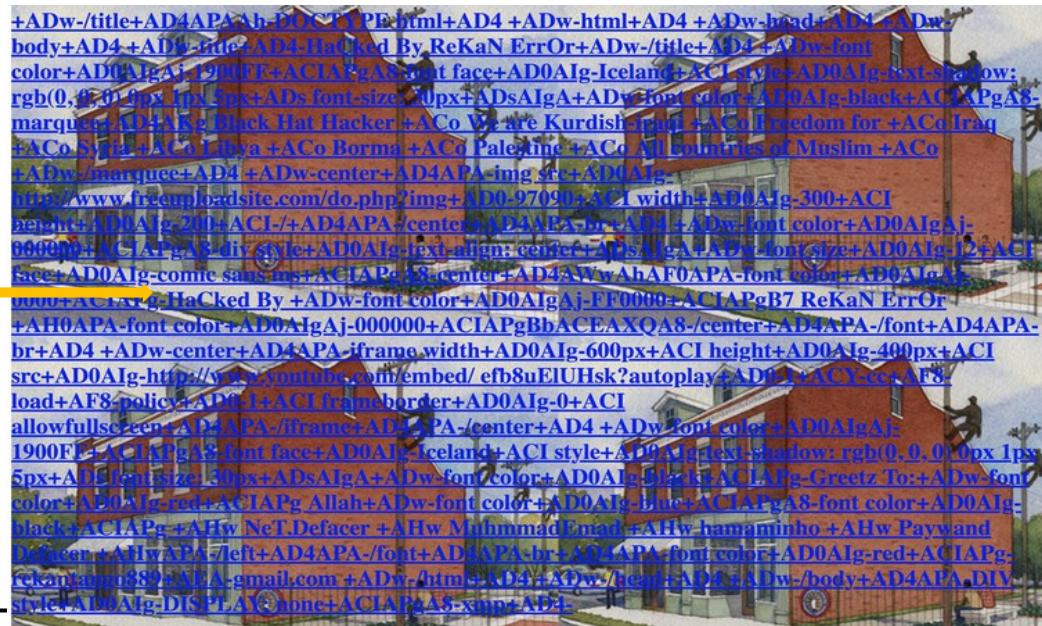


Feature Engineering



Email Addresses

Media resources
and URLs



+ADw-/title+AD4APAAb-DOCTYPE html+AD4 +ADw-html+AD4 +ADw-head+AD4 +ADw-body+AD4 +ADw-title+AD4-HaCked By ReKaN ErrOr+ADw-/title+AD4 +ADw-font-color+AD0A1gAj-1900FF+ACIAPgA8-font face+AD0A1g-Iceland+ACI style+AD0A1g-text-shadow:rgb(0, 0, 0) 0px 1px 5px+ADsA1gA+ADw-font color+AD0A1g-black+ACIAPgA8-marquee+AD4AKg Black Hat Hacker +ACo We are Kurdish-Iraqi +ACo Freedom for +ACo Iraq +ACo Syria +ACo Libya +ACo Borma +ACo Palestine +ACo All countries of Muslim +ACo +ADw/marquee+AD4 +ADw-center+AD4APA-img src+AD0A1g-<https://www.freecodecsite.com/do.php?img+AD0-97090+ACI> width+AD0A1g-300+ACI height+AD0A1g-200+ACI-/+AD4APA-/center+AD4APA-br+AD4 +ADw-font color+AD0A1gAj-090000+ACIAPgA8-div style+AD0A1g-text-align: center+ADsA1gA+ADw-font size+AD0A1g-12+ACI face+AD0A1g-comic sans ms+ACIAPgA8-center+AD4AWwAhAF0APA-font color+AD0A1g-Ak-0000+ACI fax & HaCked By +ADw-font color+AD0A1gAj-FF0000+ACIAPgB7 ReKaN ErrOr+AH0APA-font color+AD0A1gAj-000000+ACIAPgBbACEAXQA8/-center+AD4APA-/font+AD4APA-br+AD4 +ADw-center+AD4APA-iframe width+AD0A1g-600px+ACI height+AD0A1g-400px+ACI src+AD0A1g-<a href="https://www.youtube.com/embed/efb8uEIUHsk?autoplay=1+ACI-ccc+AF8-load+AF8-policy+AD0-1+ACI frameborder+AD0A1g-0+ACI allowfullscreen+AD4APA-/iframe+AD4APA-/center+AD4 +ADw-font color+AD0A1gAj-1900FF+ACIAPgA8-font face+AD0A1g-Iceland+ACI style+AD0A1g-text-shadow: rgb(0, 0, 0) 0px 1px 5px+ADsA1gA+ADw-font color+AD0A1g-black+ACIAPg-Greetz To:+ADw-font color+AD0A1g-red+ACIAPg Allah+ADw-font color+AD0A1g-blue+ACIAPgA8-font color+AD0A1g-black+ACIAPg+AHw NeT.Defacer +AHw MuhammadEmaad +AHw hamaminho +AHw PaywandDefacer +AHwAPA-/left+AD4APA-/font+AD4APA-br+AD4APA-font color+AD0A1g-red+ACIAPg-rekanlangu0389+ACIA@gmail.com +ADw-/html+AD4 +ADw-/body+AD4APA-DIV style+AD0A1g-DISPLAY:none+ACIAPgA8-xmp+AD4</p>

Summary of Features

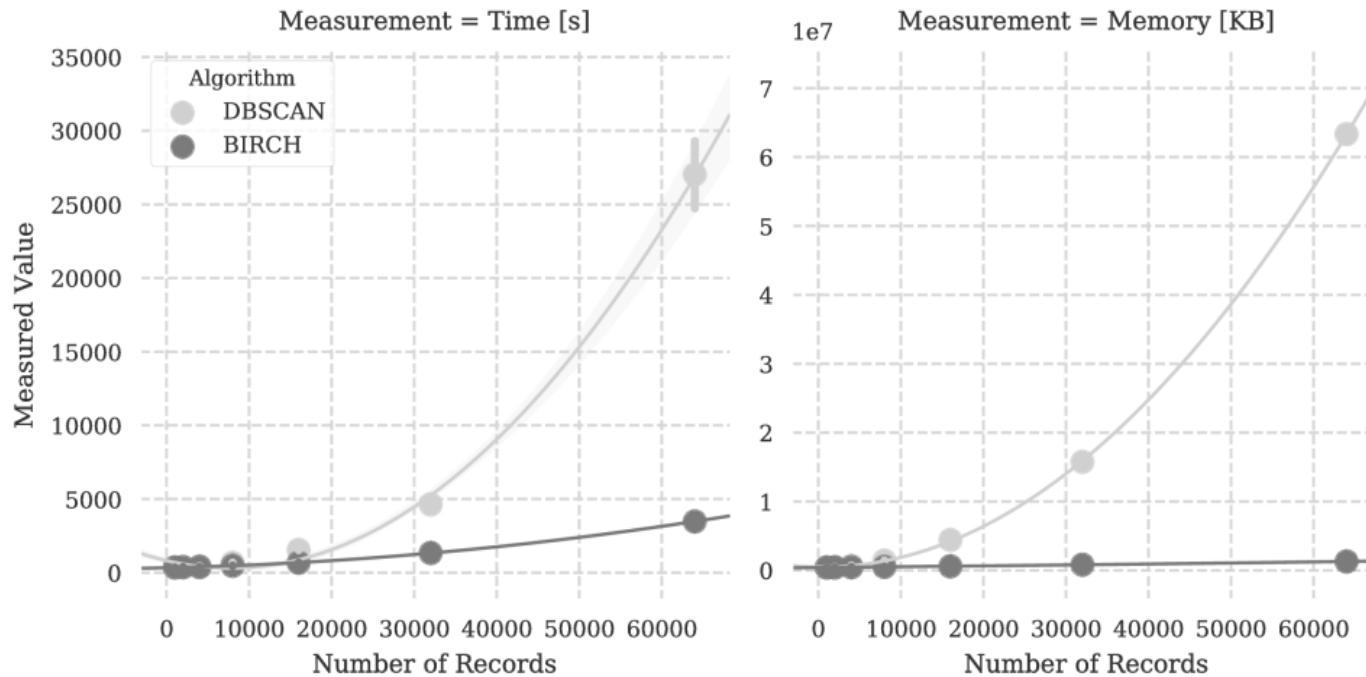
Group	Feature Name	Type and range	Description
Visual	No. of Images	integer $[0, \infty]$	Number of <code></code> tags
	Perceptual Hash	binary (64 bits)	Calculated on the north-centered 1600x900 screenshot crop
	Average Color	3 floats (RGB)	Average of the 5 most common colors in the screenshot
	No. of Sound URLs	integer $[0, \infty]$	Number of URLs pointing to sound-hosting services or files
	Type of first Sound URL	categorical	File and service type of the first (usually the only) sound URL
Structural	No. of each <code><tag></code>	7 integers $[0, \infty]$	Number of style, embed, script, meta, object, iframe, and a tags
Geographical	Encoding	categorical	Detected text encoding
	Language	categorical	Detected language (for labeling only)
Domains	External domains	real $[0, \infty]$	Ratio of links pointing to cross-origin domains
	Letters in external domains	real $[0, \infty]$	Avg. ASCII letters in the external domains string
Social	No. of online handlers	int $[0, \infty]$	Twitter @handlers, #hashtags, e-mail addresses
Title	Letters, digits, punctuation, white-spaces in title	4 real $[0, \infty]$	Ratio of the listed character classes in the page title

Campaign Detection: 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

Scaling Clustering

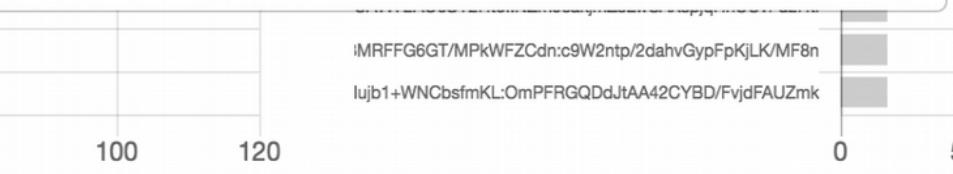
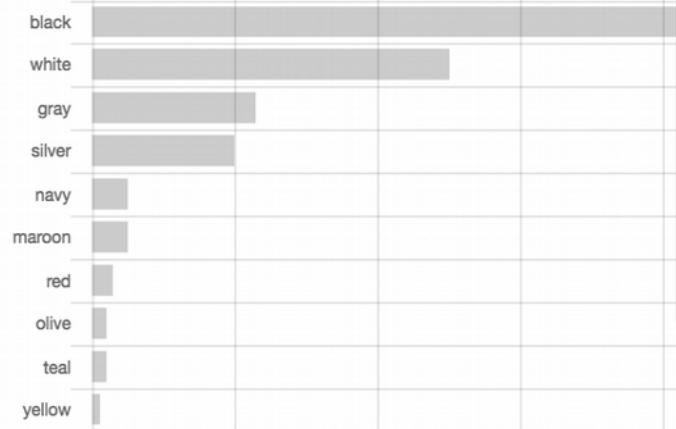
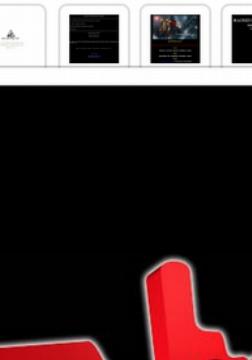
- Scalability of BIRCH vs. DBSCAN (10 runs)



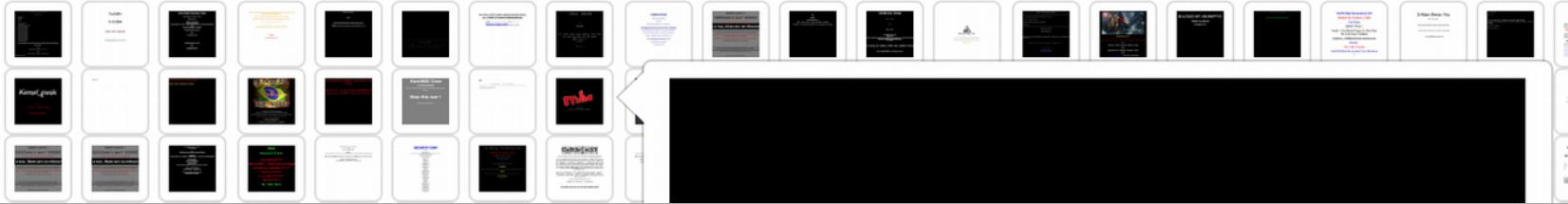
Clusters → Campaigns: Labeling

- Each cluster represented in 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)
-

Cluster 4: 217 elements (unique: 179) - 2002-03-09 05:34:56 to 2006-08-14 18:51:17



Cluster 4: 217 elements (unique: 179) - 2002-03-09 05:34:56 to 2006-08-14 18:51:17



Cluster 4

portfolio (0.524)

nicholas (0.355)

steel (0.355)

close (0.254) mohan (0.186)

your (0.531)

cyber71 (0.214)

lucif42 (0.214)

5pid32 (0.214)

that (0.212)

kurdish (0.236)

saved (0.236)

kä¼rd (0.236)

tä¼rk (0.236)

turkish (0.236)

maroon

red

olive

teal

yellow

0 20 40 60 80 100 120

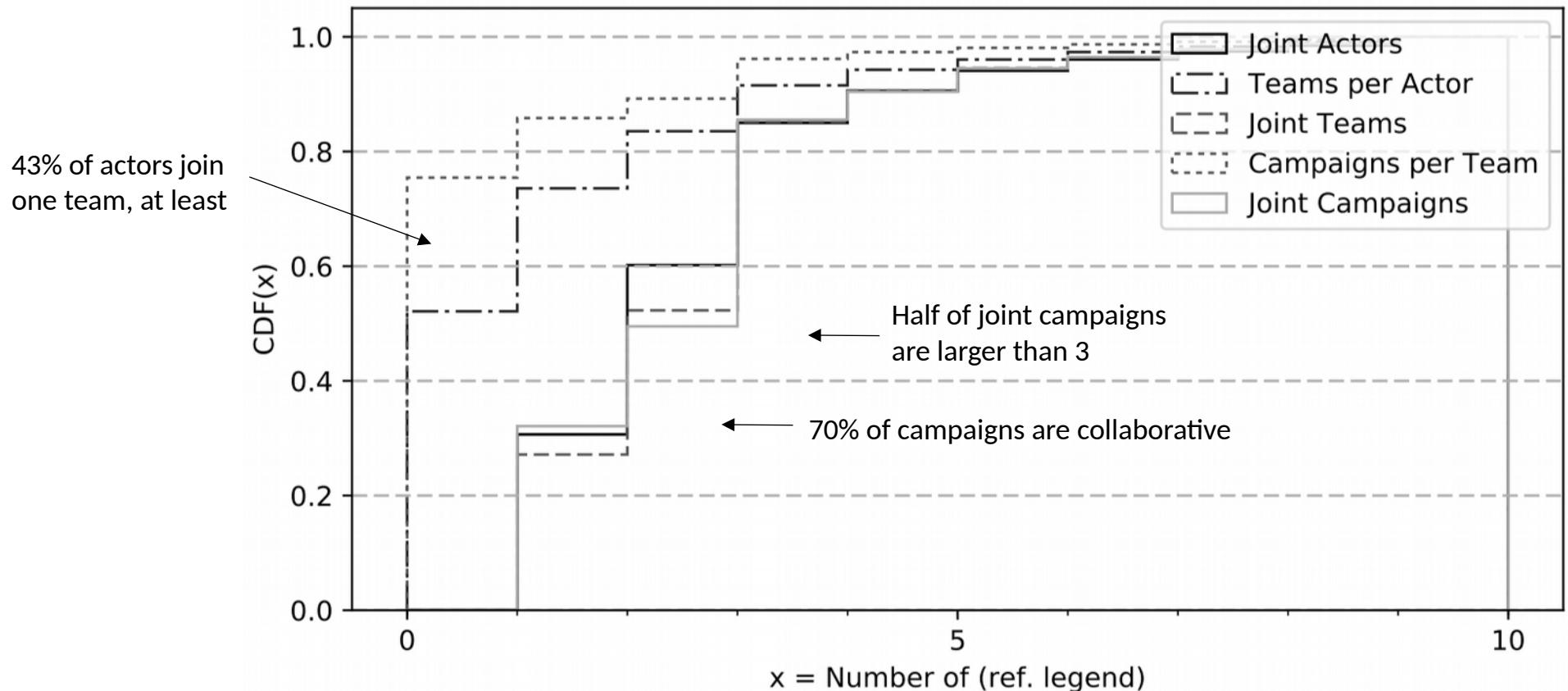
0 5

iMRFFG6GT/MPkWFZCdn:c9W2ntp/2dahvGypFpKjLK/MF8n

lujb1+WNCbsfmKL:OmPFRGQDdJtAA42CYBD/FvjdFAUZmk

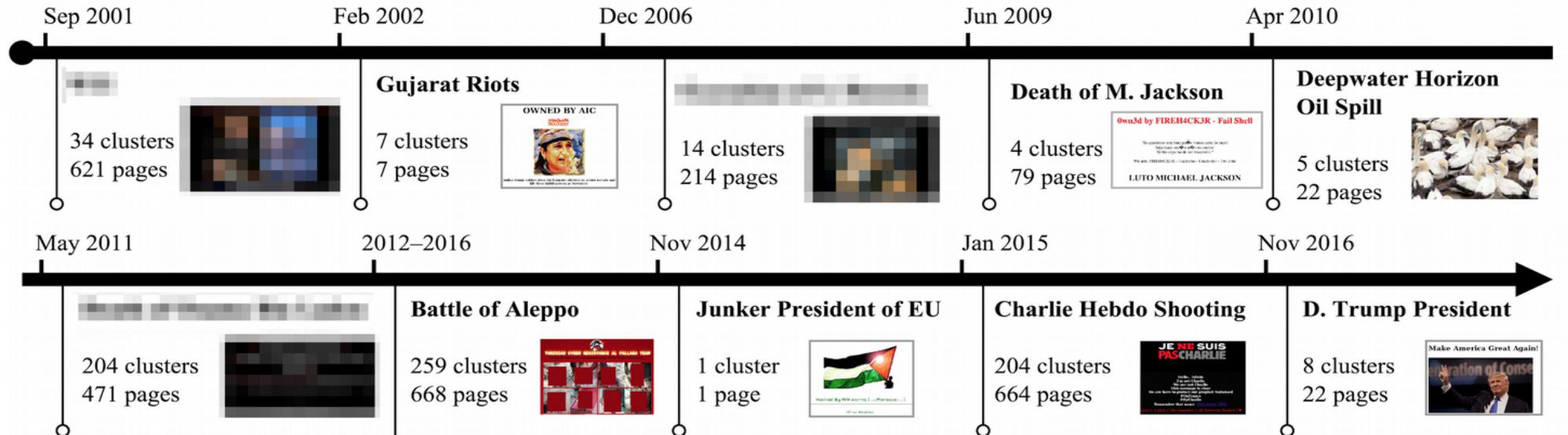
Findings

Organization of Actors



Geopolitical Real-World Events

Successfully detected



The “Charlie Hebdo” case



The “Charlie Hebdo” case



Campaigns

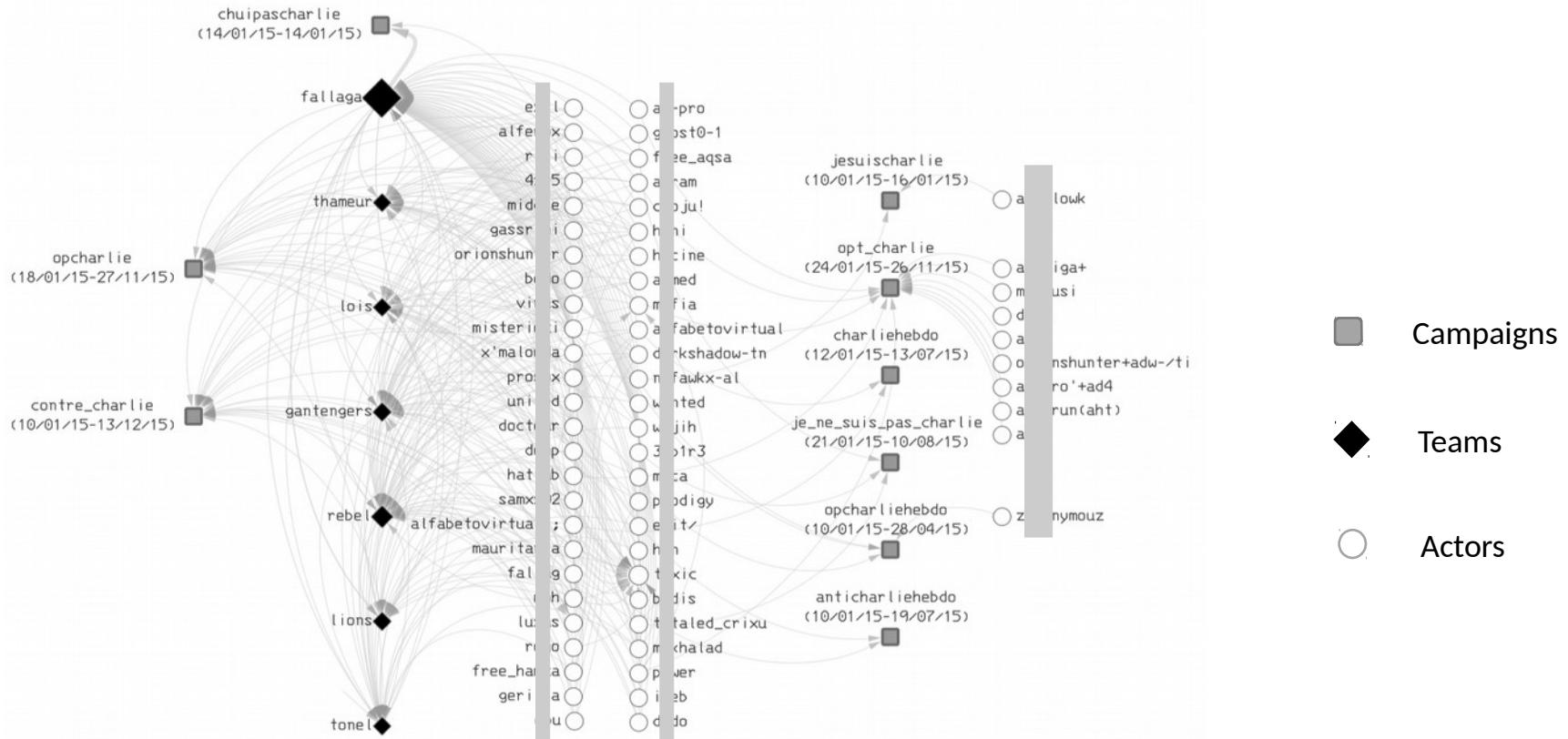


Teams

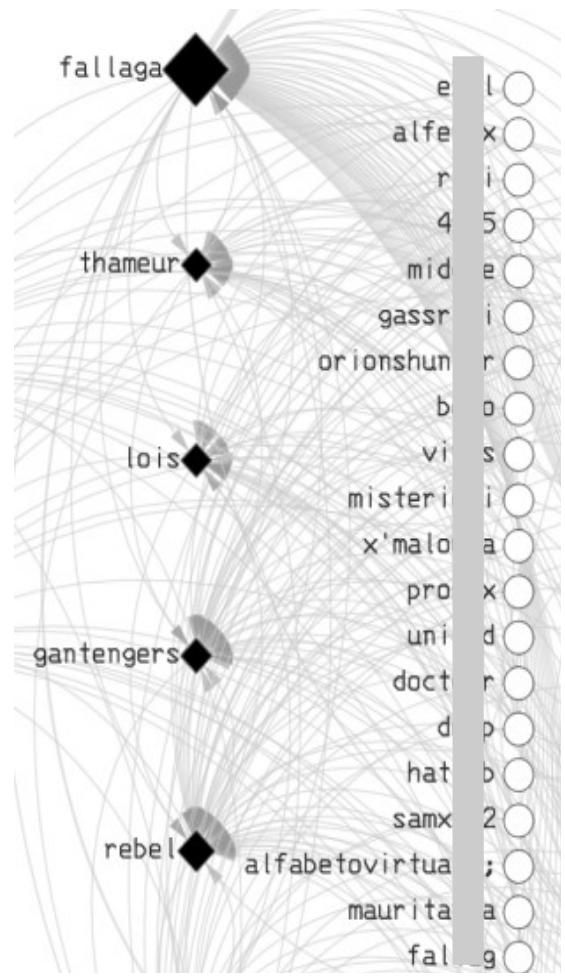


Actors

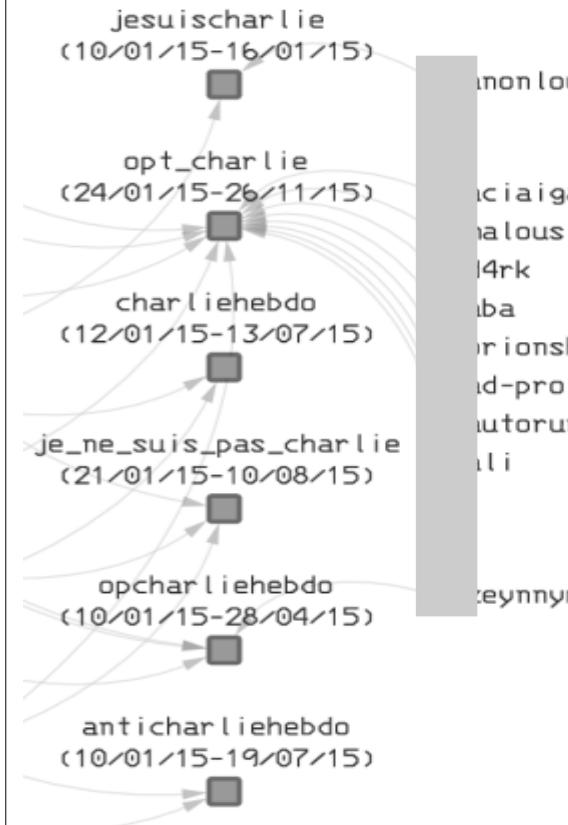
The “Charlie Hebdo” case



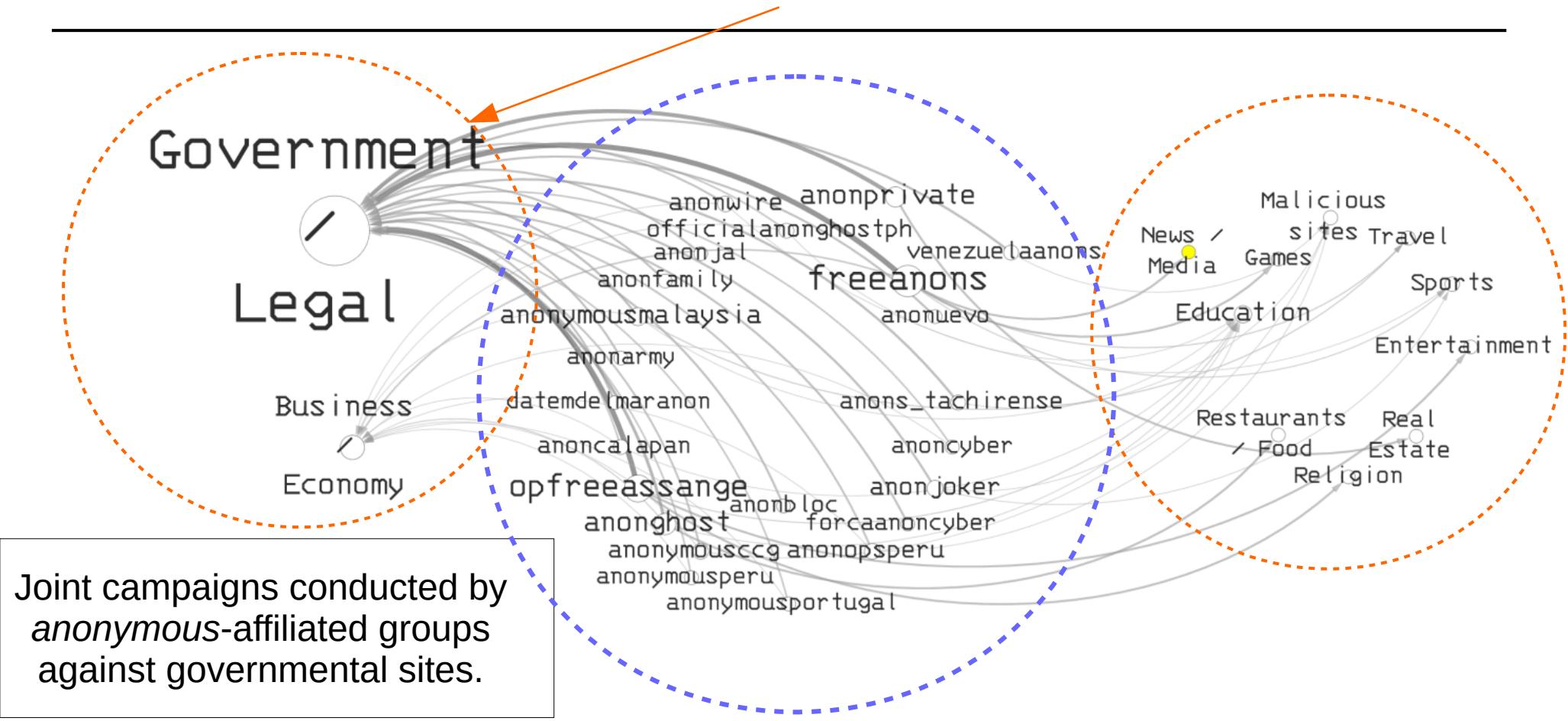
Affiliated actors



Lone wolves



Anonymous Ops. Target Gov. Sites

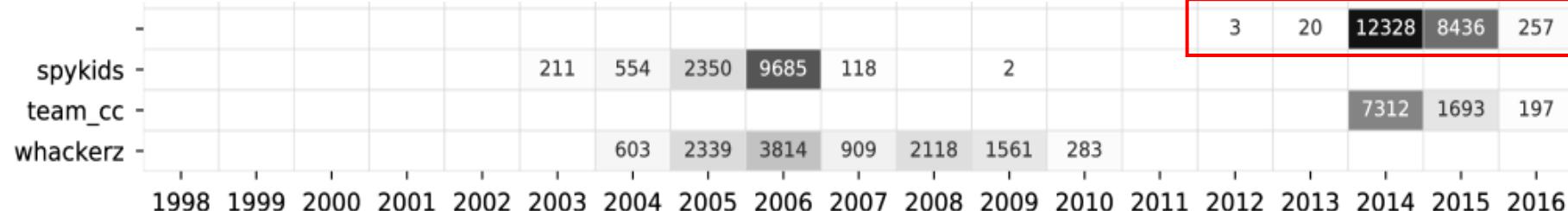


Long-Term vs. Aggressive Campaigns

Long-Term vs. Aggressive Campaigns

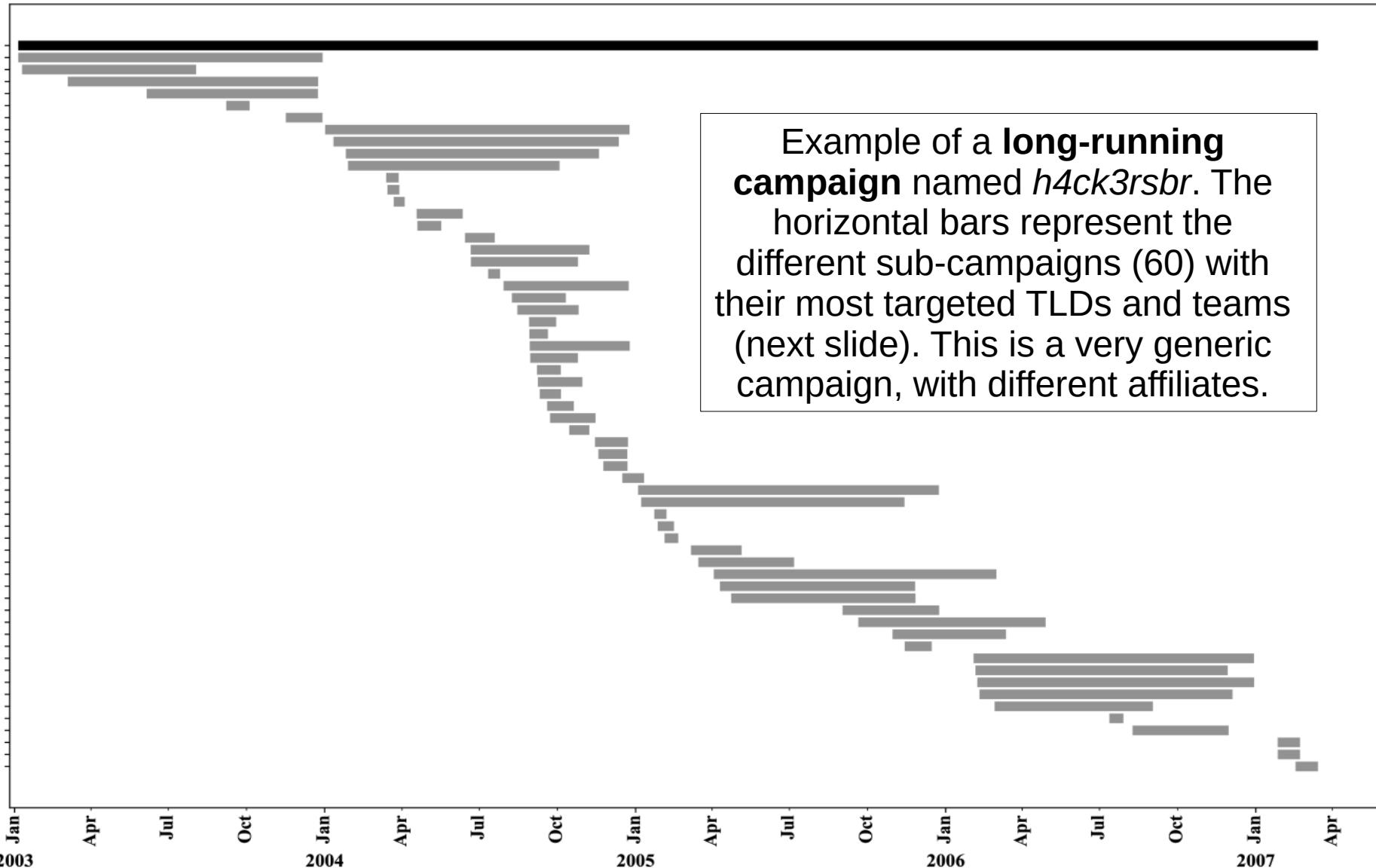
palestine -				1					1			29	2	98	902	1629	1124	
powhack -				238	557	133	27		28					1	1			
r00t -	1	3	18	4	18				5			138	117	10	217	189	375	
redhack -						69	1973	538	3731	279	22	636	17	22	1	49	6	
rodape -										1	1	14	13	11	24	69	29	
samarindahack -				8	96	49	40	104	1	9	2			498	4			
syshack -				2	2	2	42	2	11	28				67		1	1	
turkish -						2		6		101	7	6	120	5	27	2	31	16
venezuela -			1			1	18	65	62					2		54	1	

Spot, aggressive campaign in reaction to war events



Campaign: h4ck3rsbr

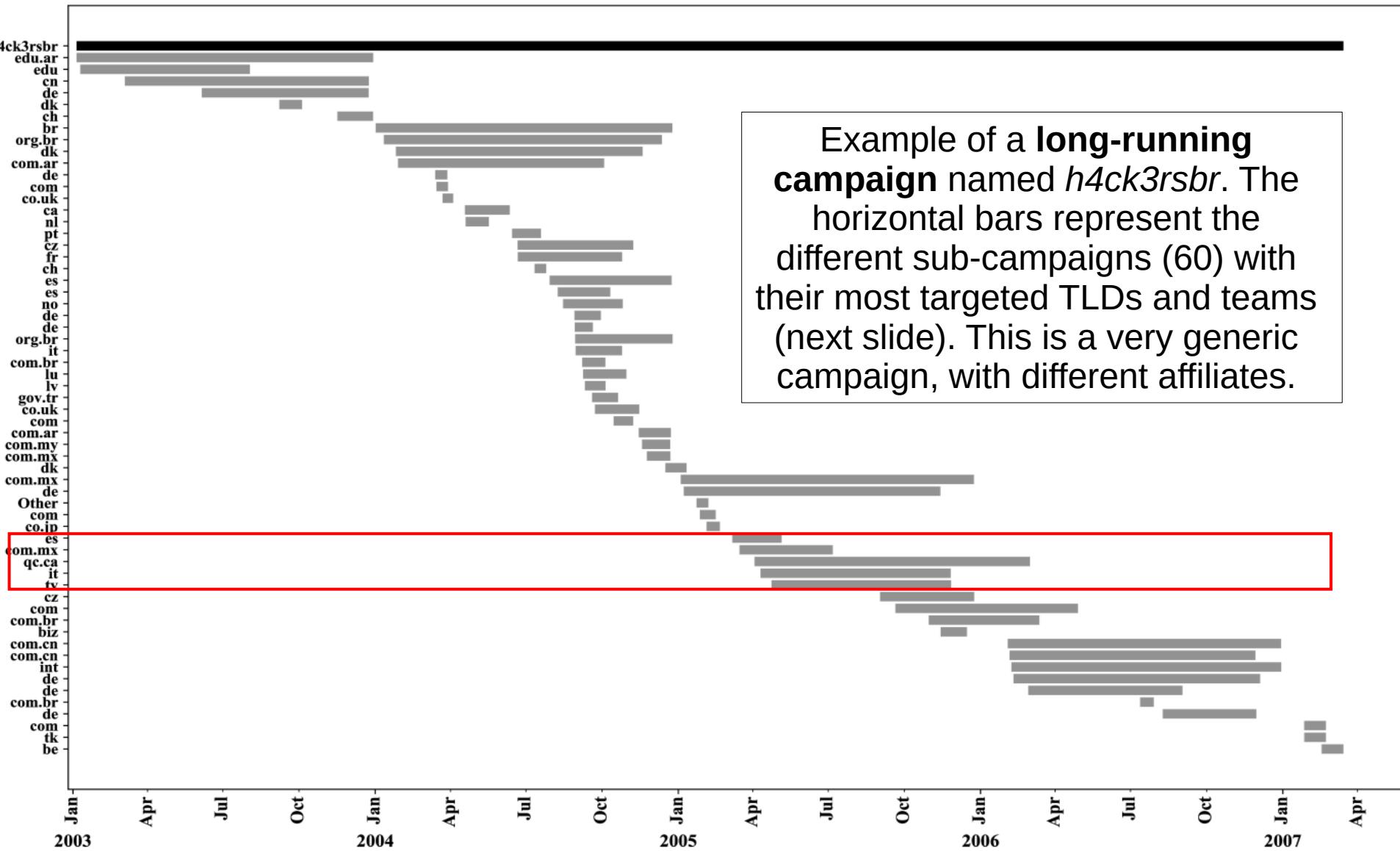
Most targeted TLDs



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.

Campaign: h4ck3rsbr

Most targeted TLDs



Most targeted TLDs

com.mx



qc.ca



c0d3rz (24 def.)

it



Trustix (38 def.)

tv



Fatal Error (6 def.)

cz



Infektion Group (17 def.)

Oct

Conclusions

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

References

- DefPloreX-NG
- GitHub:
<https://github.com/trendmicro/defplorex>
- Paper*:
<https://documents.trendmicro.com/assets/wp/wp-web-defacement-campaigns-uncovered-gaining-insights-from-deface-pages-using-defplorex-ng.pdf>

(*) Joint work with Marco Balduzzi, Ryan Flores, Lion Gu and Vincenzo Ciancaglini



Using Machine-Learning to Investigate Web Campaigns at Large

Dr. Federico Maggi, @phretor

Joint work with Marco Balduzzi, Ryan Flores, Lion Gu and Vincenzo Ciancaglini