Exploring the ecosystem of malicious domain registrations in the .eu TLD

Lieven Desmet – OWASP BeNeLux Day 2017 – Tilburg, NL

Lieven.Desmet@cs.kuleuven.be – @lieven_desmet

DistriNet EURid

Joint research between KU Leuven and EURid

- EURid:
  - Dirk Jumpertz
  - Peter Janssen
  - Marc Van Wesemael

- DistriNet, KU Leuven:
  - Thomas Vissers
  - Jan Spooren
  - Pieter Agten
  - Frank Piessens
  - Wouter Joosen
  - Lieven Desmet

Overview

- Research Context
- Domain name registrations in .eu
- Longitudinal campaign analysis
- Towards automatic campaign identification
- Towards pro-active detection and prevention
- Conclusion

Research context

Malicious use of domain names

- Domain names are often abused by cyber criminals
  - Spam, botnet C&C infrastructure, phishing, malware, …
- To avoid blacklisting, malicious actors often deploy a hit-and-run strategy
  - Fast flux in domain names
  - Single shot: 60% are only active for 1 day after registration [Hao et al][1]

[1] Hao et al. “Understanding the Domain Registration Behavior of Spammers” IMC 2013

Research hypothesis:

“Malicious actors register domains in bulk, and do so for longer periods of time.”

Research question

- “Can we identify such bulk behavior based on commonalities between individual registrations?”

- Long-term goal of this research:
  - Understand the malicious domain registration ecosystem in order to detect and prevent malicious registrations.

Domain name registrations in .eu

Domain name registrations in the .eu TLD

- **.eu** – 7th largest ccTLD (European Economic Area)
  - ~3.8 million domain names

- Dataset used in this research:
  - 824,121 new registrations over 14 months (Apr 2015 – May 2016)
  - 20,870 registrations end up on blacklists (2.5%)

Available registration data

- Basic registration information
  - domain name, datetime of registration, and registrar
- Contact information of the registrant
  - company name, name, language, email address, phone, fax, as well as postal address
- Name server information
  - Name servers and/or glue records

Dataset enrichments

- Maliciousness of a domain name
  - Spamhaus DBL
  - SURBL multi list
  - Google Safe Browsing
- Geolocation information of name servers
  - MaxMind GeoLite2 Free database

Longitudinal campaign analysis

Concept of a “registration campaign”

- Set of registrations with malicious intent
- Most probably linked to the same actor
- Running over a longer period of time

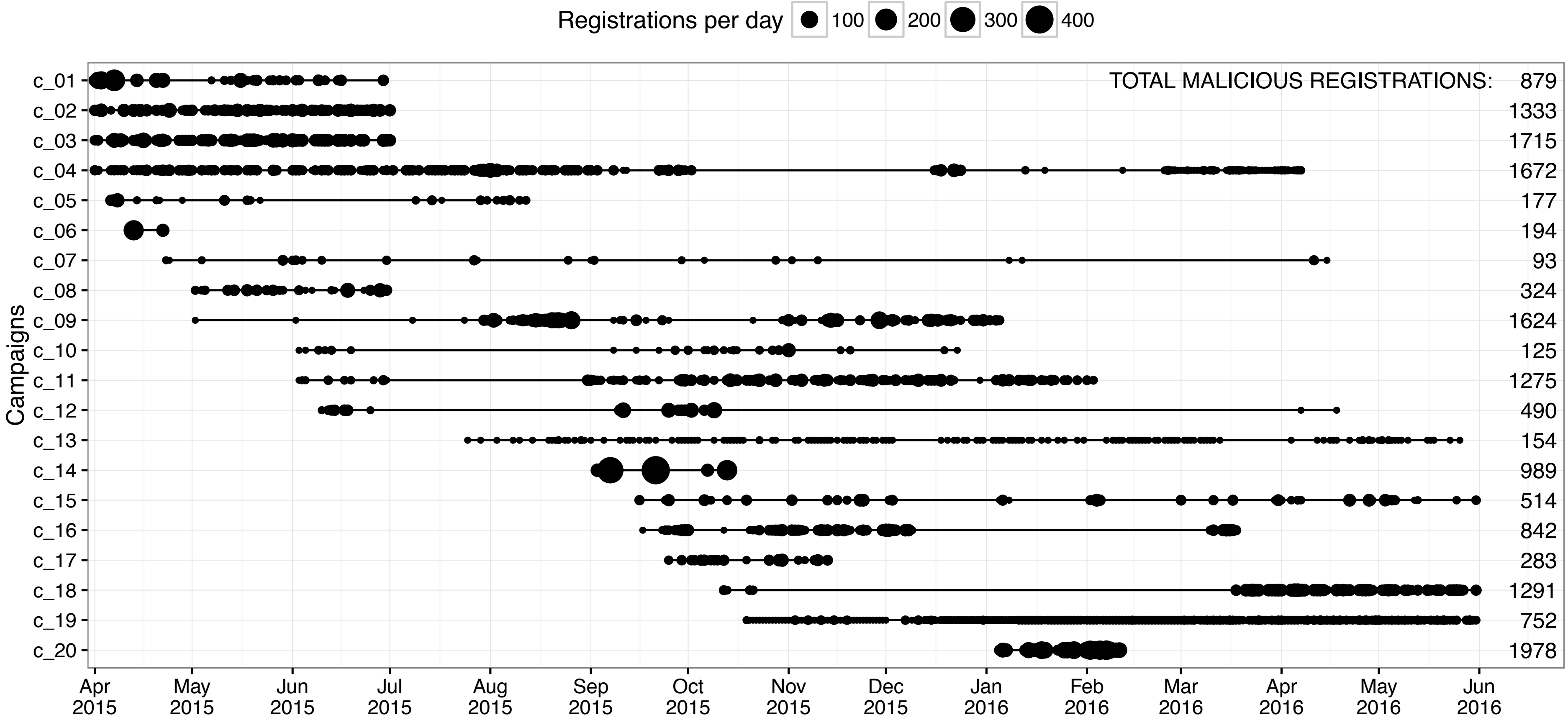
- Our approximation: Manually selected based on common characteristics in the registration details

Example campaign (c_11)

- Fixed email domain
  - j***n.com
- Multiple fake registrant details
  - Combinations of
    2 email accounts,
    3 phone numbers,
    2 street addresses
- 4 registrars used back-to-back

- **8 months active**
  (Jun 3, 2015 – Feb 3, 2016)
- **1,275 blacklisted registrations**

Activity of identified campaigns

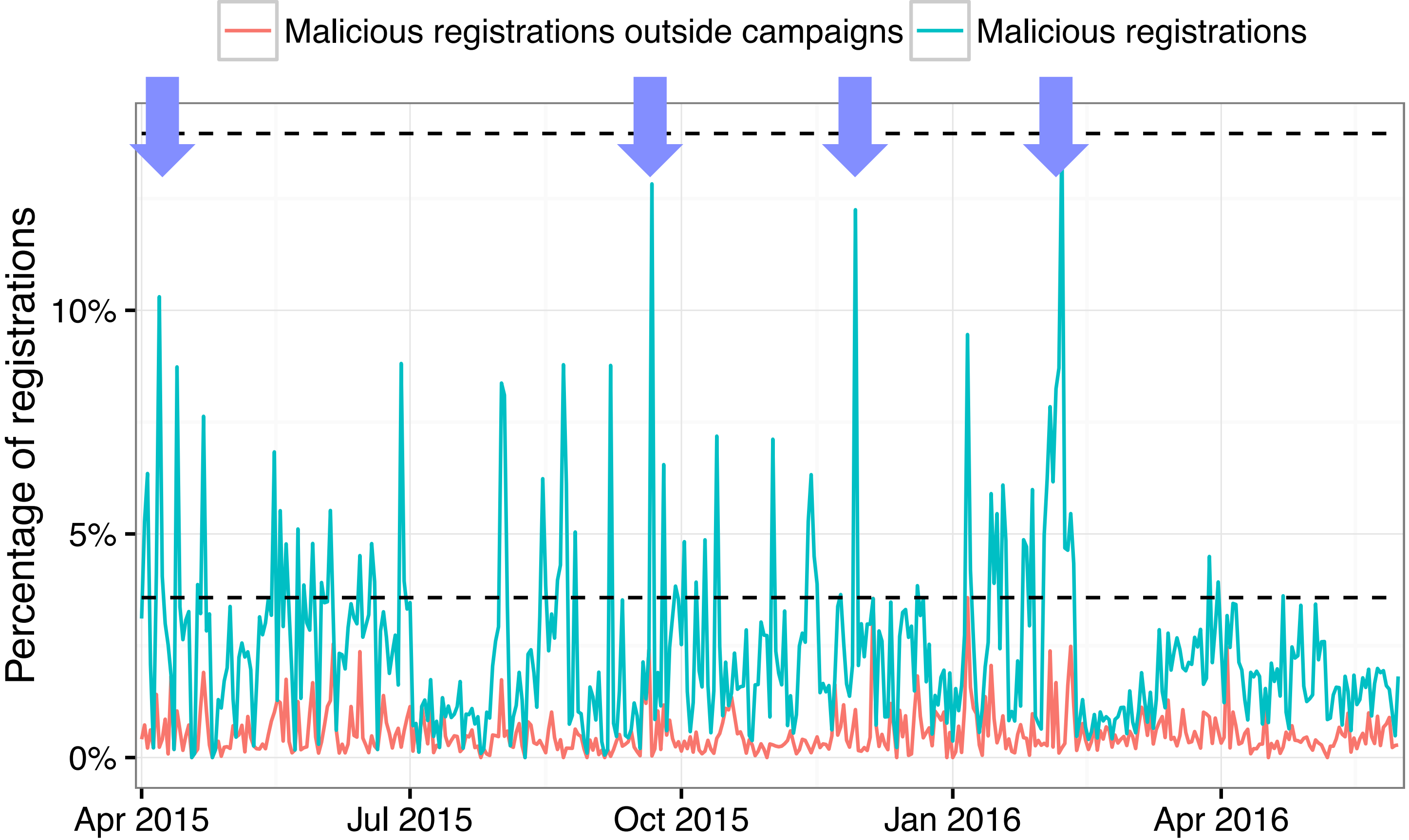


Campaign identification process

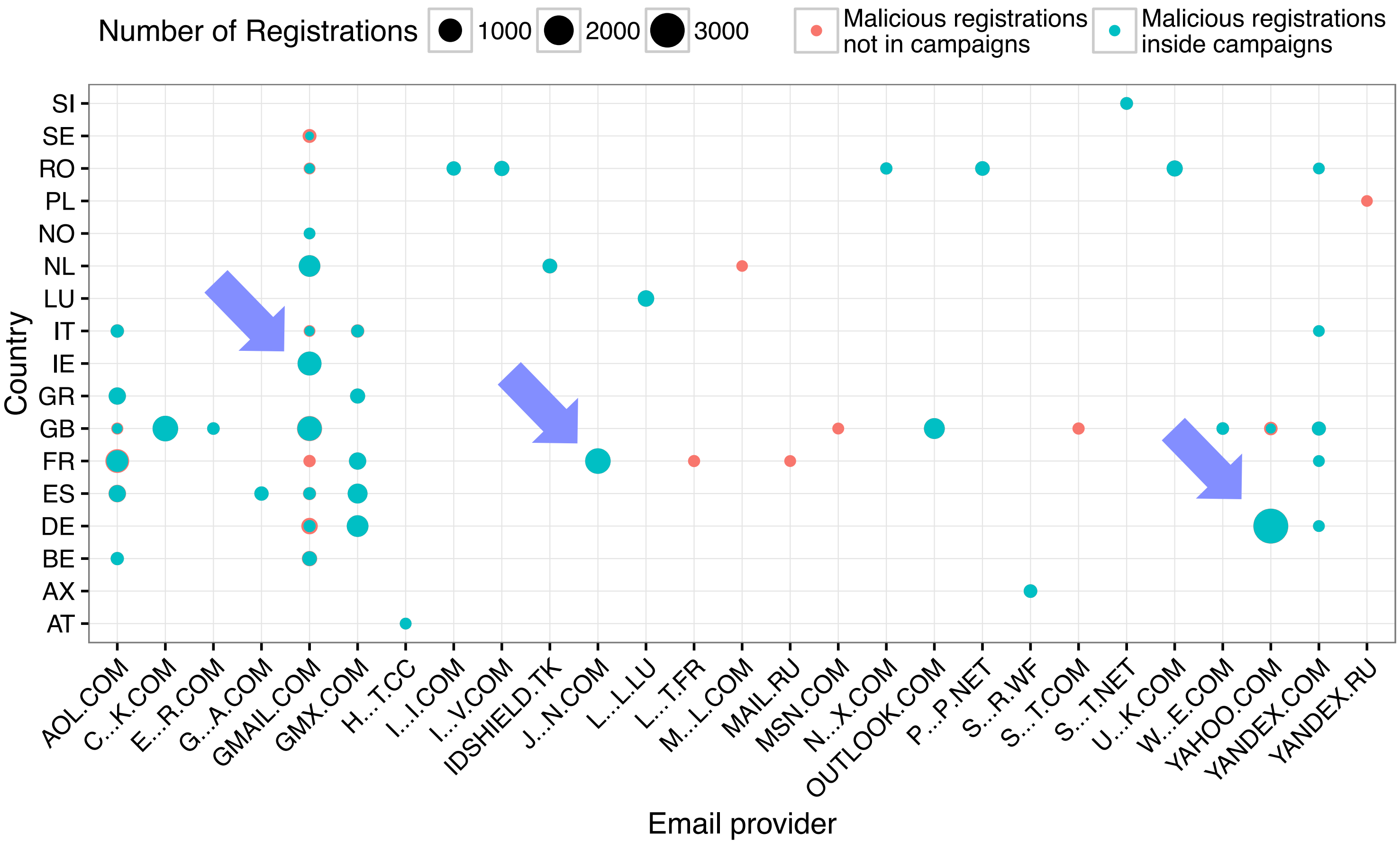
Manual campaign identification process

- Start from maliciously flagged registrations
- Identify:
  - days with high number of malicious registrations
  - most reused registrations details (email address, phone, street, …)
  - recognizable patterns in registration details (e.g. ….202@mymail.com)
  - frequent combinations of two independent registration details
- Apply selection criteria over benign and malicious registrations

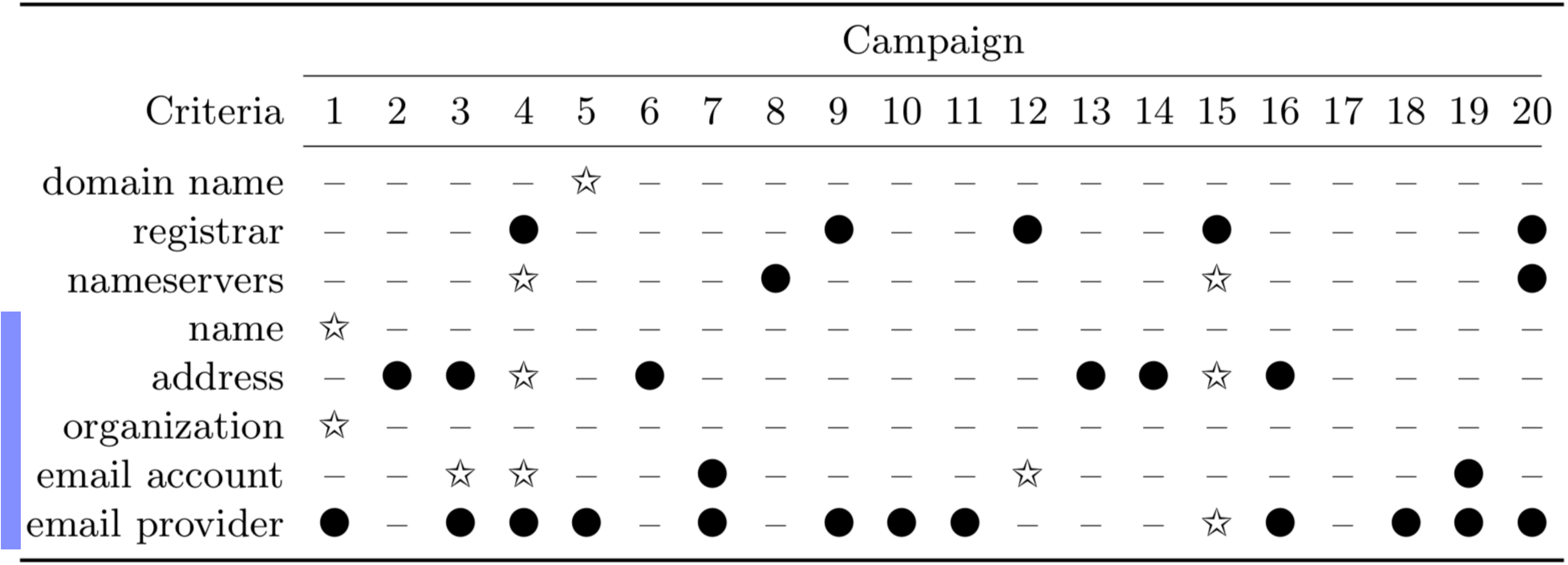
a) Days with high number of malicious registrations



b) Frequent combinations of registration details



Campaign selection criteria



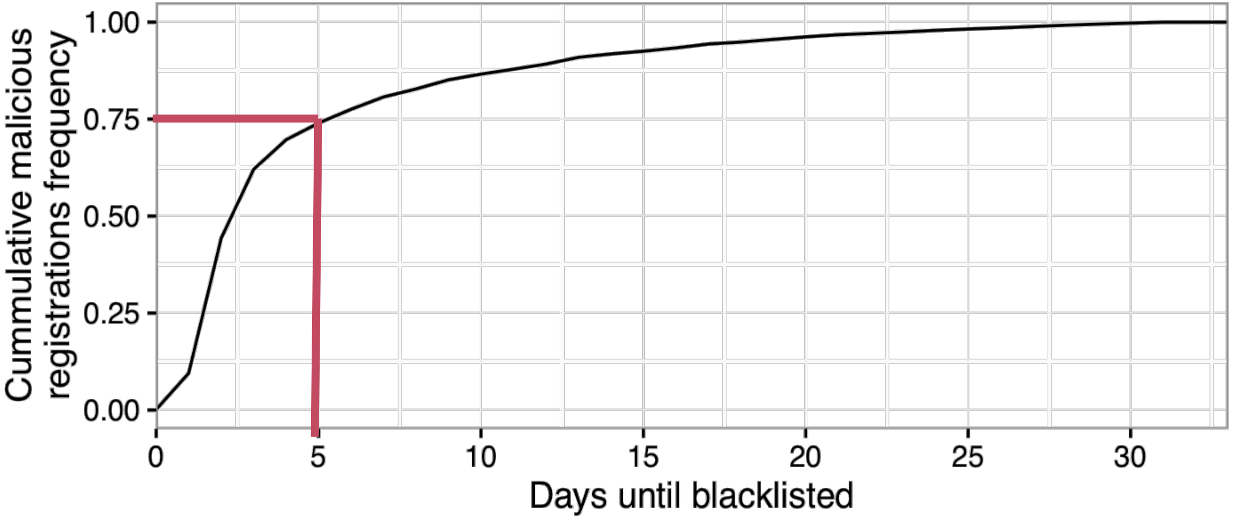
		Campaign																			
	Criteria	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	domain name	–	–	–	–	☆	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	registrar	–	–	–	●	–	–	–	–	●	–	–	●	–	–	●	–	–	–	–	●
	nameservers	–	–	–	☆	–	–	–	●	–	–	–	–	–	–	☆	–	–	–	–	●
Registrant	name	☆	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	address	–	●	●	☆	–	●	–	–	–	–	–	–	●	●	☆	●	–	–	–	–
	organization	☆	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	email account	–	–	☆	☆	–	–	●	–	–	–	–	☆	–	–	–	–	–	–	●	–
	email provider	●	–	●	●	●	–	●	–	●	●	●	–	–	–	☆	●	–	●	●	●

● represents a string match, and ☆ a regular expression pattern

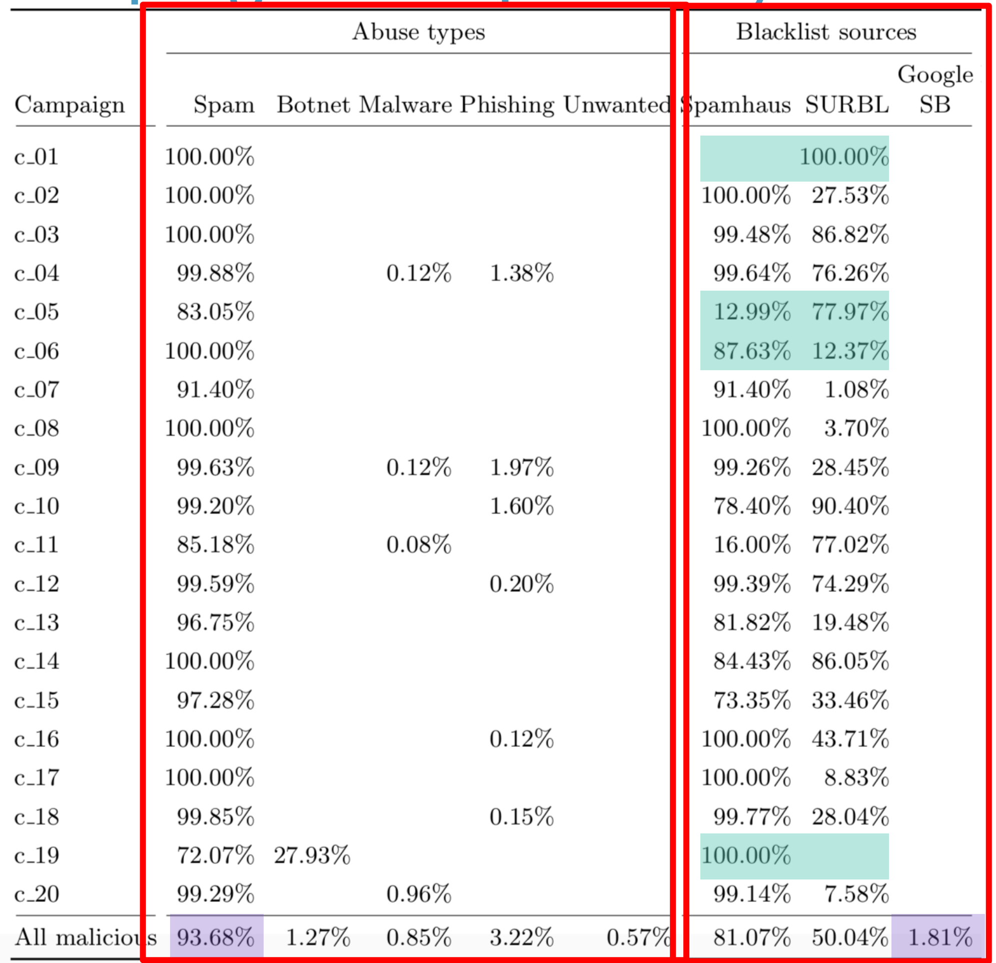
Insights in malicious domain registration

Insight 1: Hit-and-run strategies

- Small window of opportunity:
  - Domain rendered useless once blacklisted
  - 73% is blacklisted 5 days after registration, 98% after 30 days

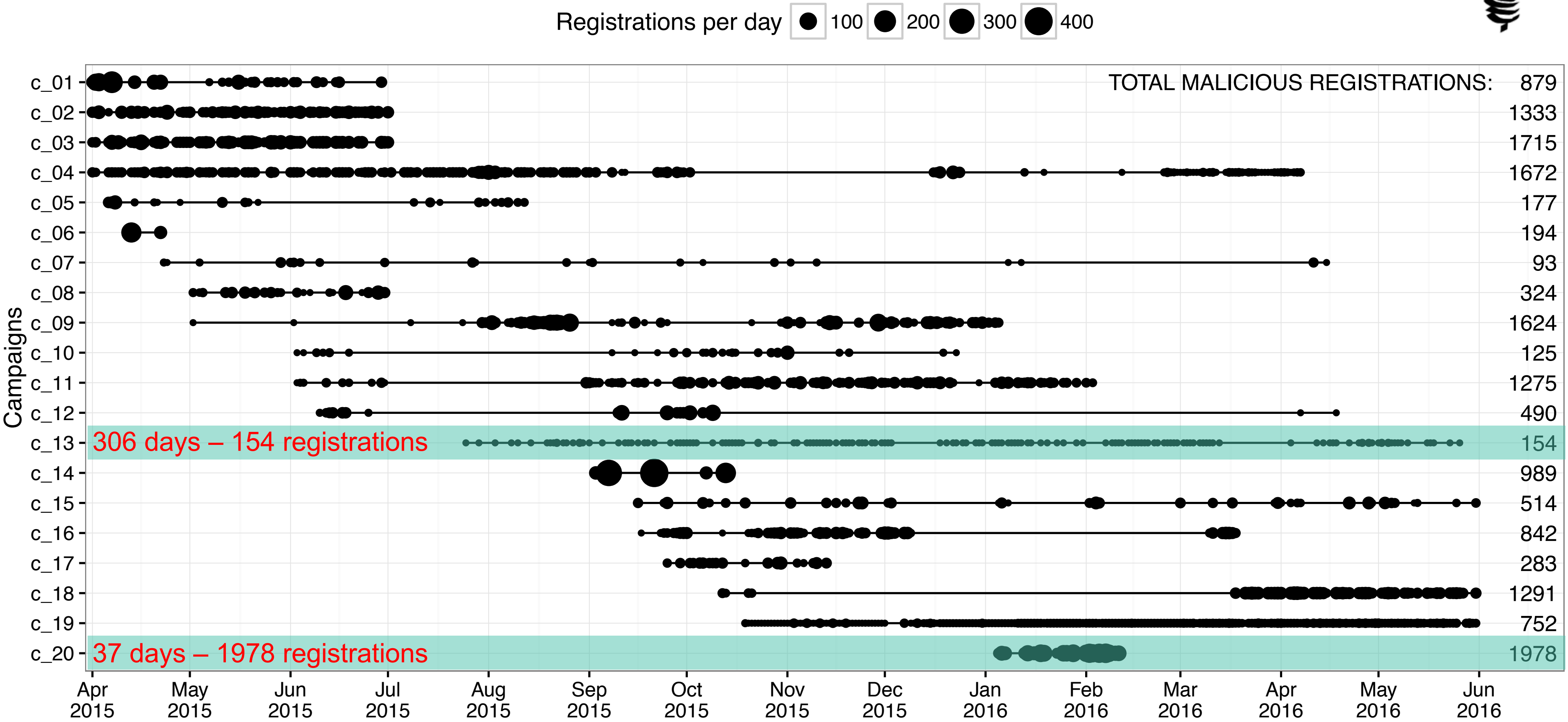


Insight 2: Campaigns are primarily linked to spam



	Abuse types					Blacklist sources		
Campaign	Spam	Botnet	Malware	Phishing	Unwanted	Spamhaus	SURBL	Google SB
c_01	100.00%						100.00%	
c_02	100.00%					100.00%	27.53%	
c_03	100.00%					99.48%	86.82%	
c_04	99.88%		0.12%	1.38%		99.64%	76.26%	
c_05	83.05%					12.99%	77.97%	
c_06	100.00%					87.63%	12.37%	
c_07	91.40%					91.40%	1.08%	
c_08	100.00%					100.00%	3.70%	
c_09	99.63%		0.12%	1.97%		99.26%	28.45%	
c_10	99.20%			1.60%		78.40%	90.40%	
c_11	85.18%		0.08%			16.00%	77.02%	
c_12	99.59%			0.20%		99.39%	74.29%	
c_13	96.75%					81.82%	19.48%	
c_14	100.00%					84.43%	86.05%	
c_15	97.28%					73.35%	33.46%	
c_16	100.00%			0.12%		100.00%	43.71%	
c_17	100.00%					100.00%	8.83%	
c_18	99.85%			0.15%		99.77%	28.04%	
c_19	72.07%	27.93%				100.00%		
c_20	99.29%		0.96%			99.14%	7.58%	
All malicious	93.68%	1.27%	0.85%	3.22%	0.57%	81.07%	50.04%	1.81%

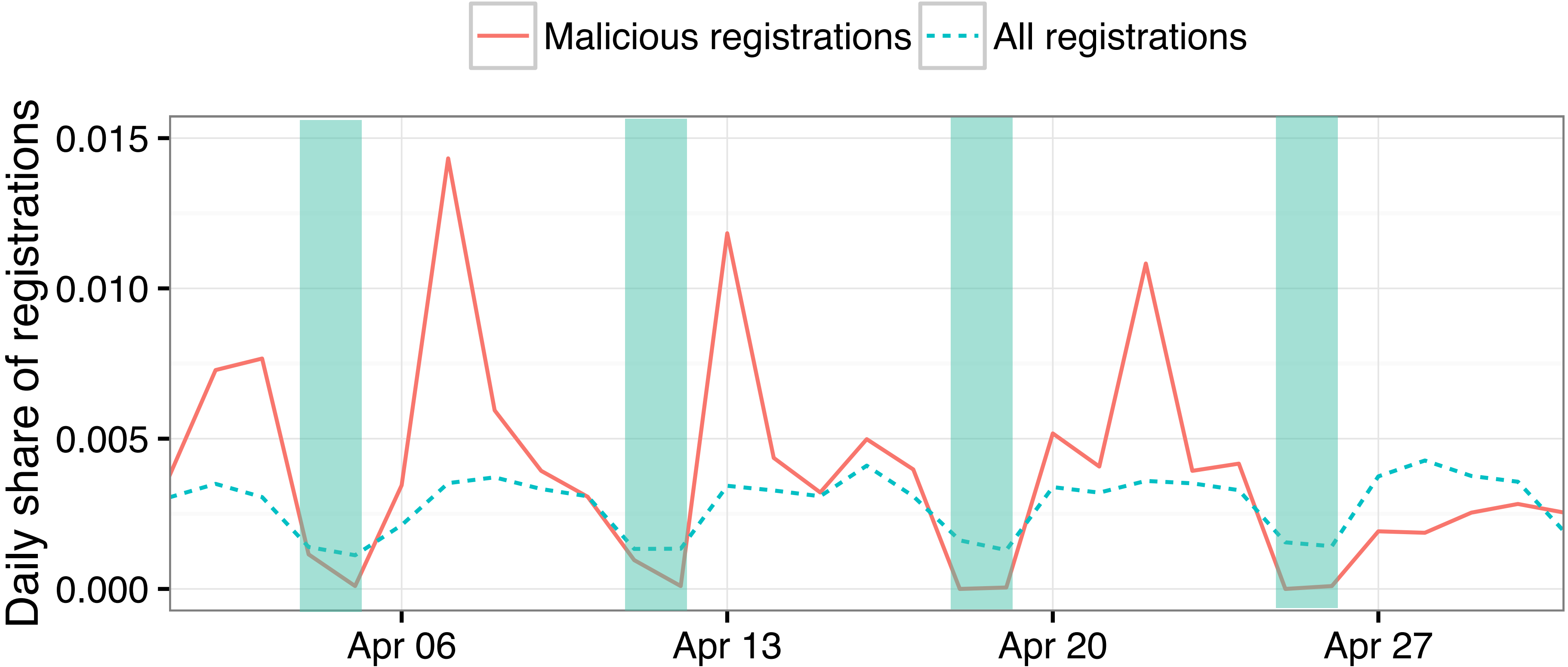
Insight 3: Variety in intensity and duration



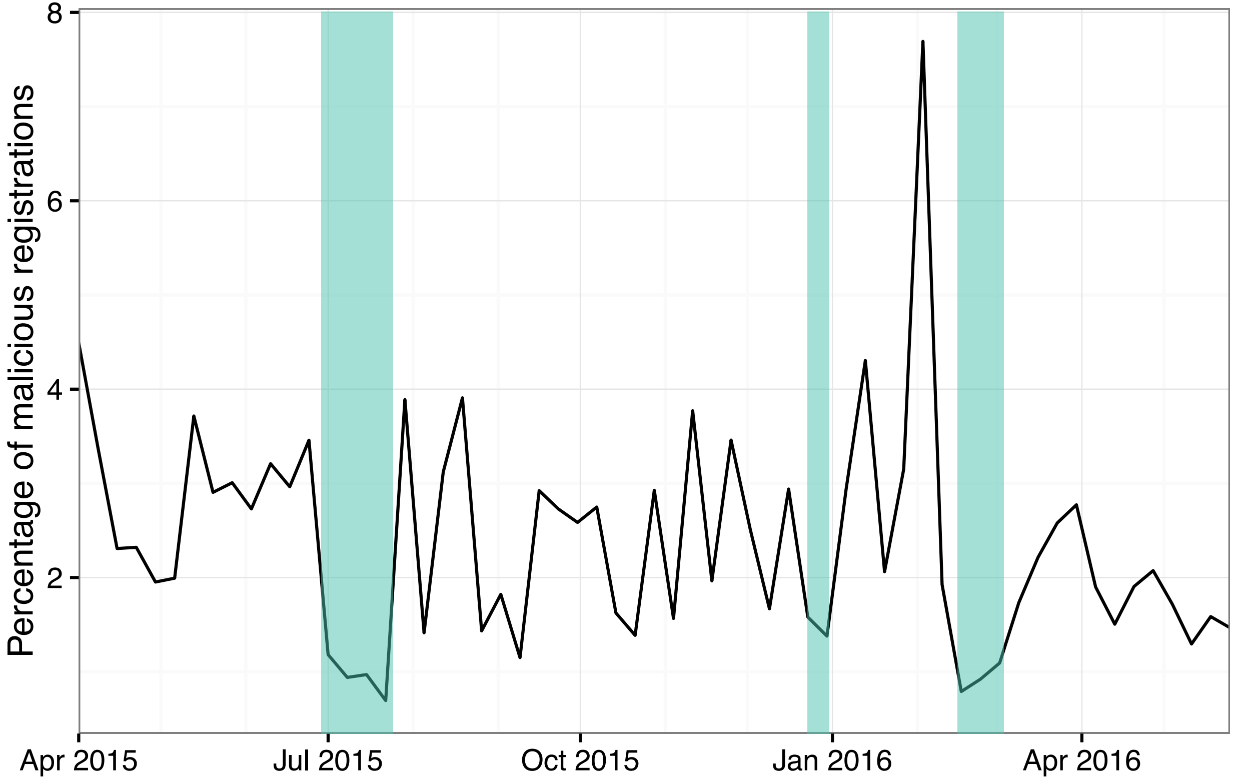
306 days – 154 registrations

37 days – 1978 registrations

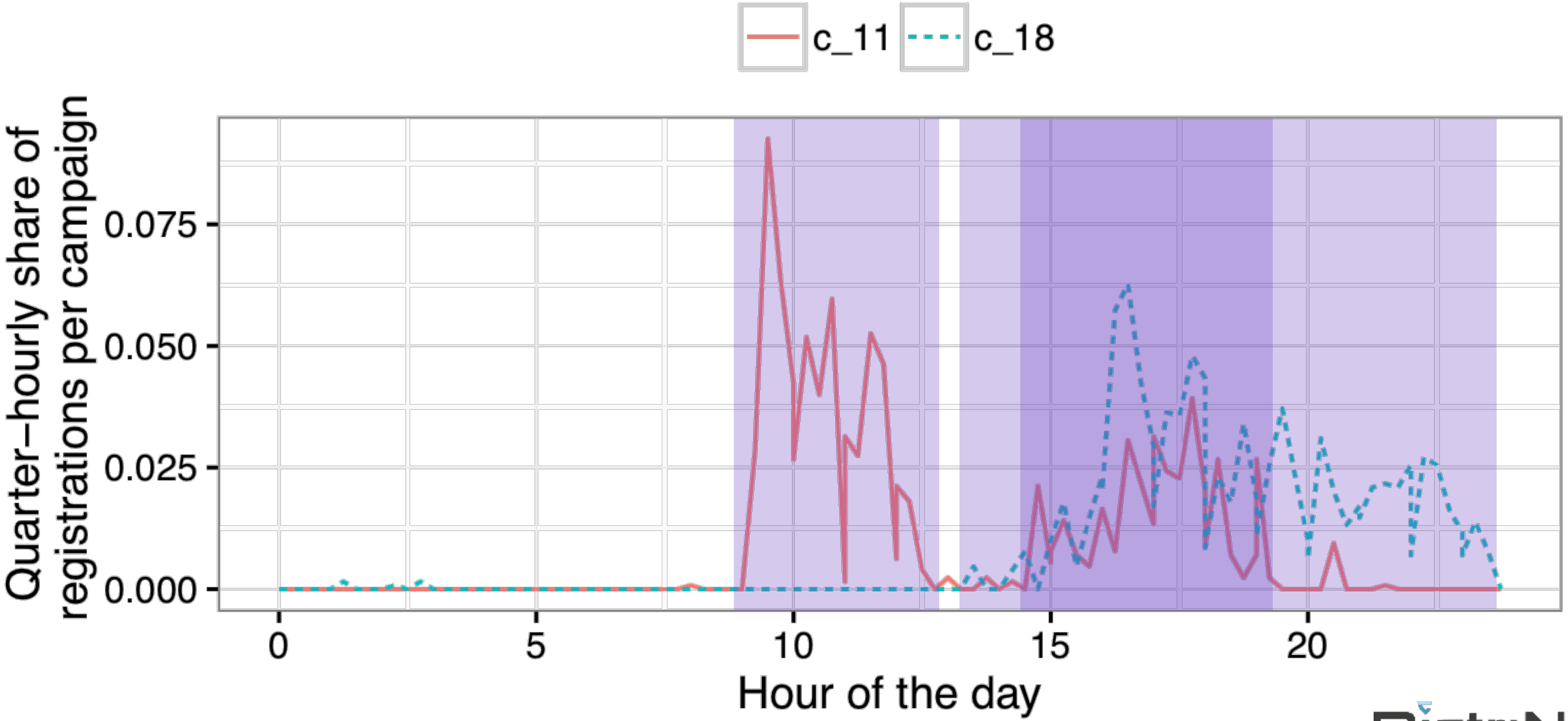
Insight 4: Some campaigns align with regular business activity patterns (1)



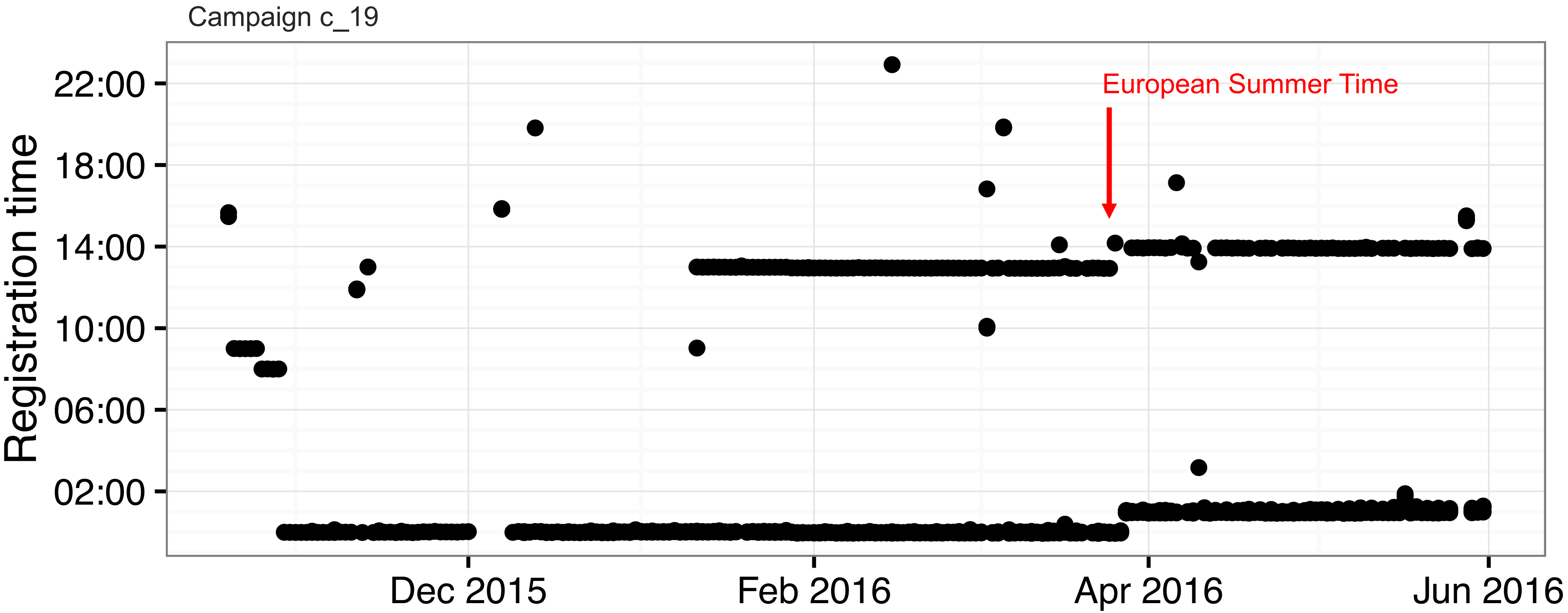
Insight 4: Some campaigns align with regular business activity patterns (2)



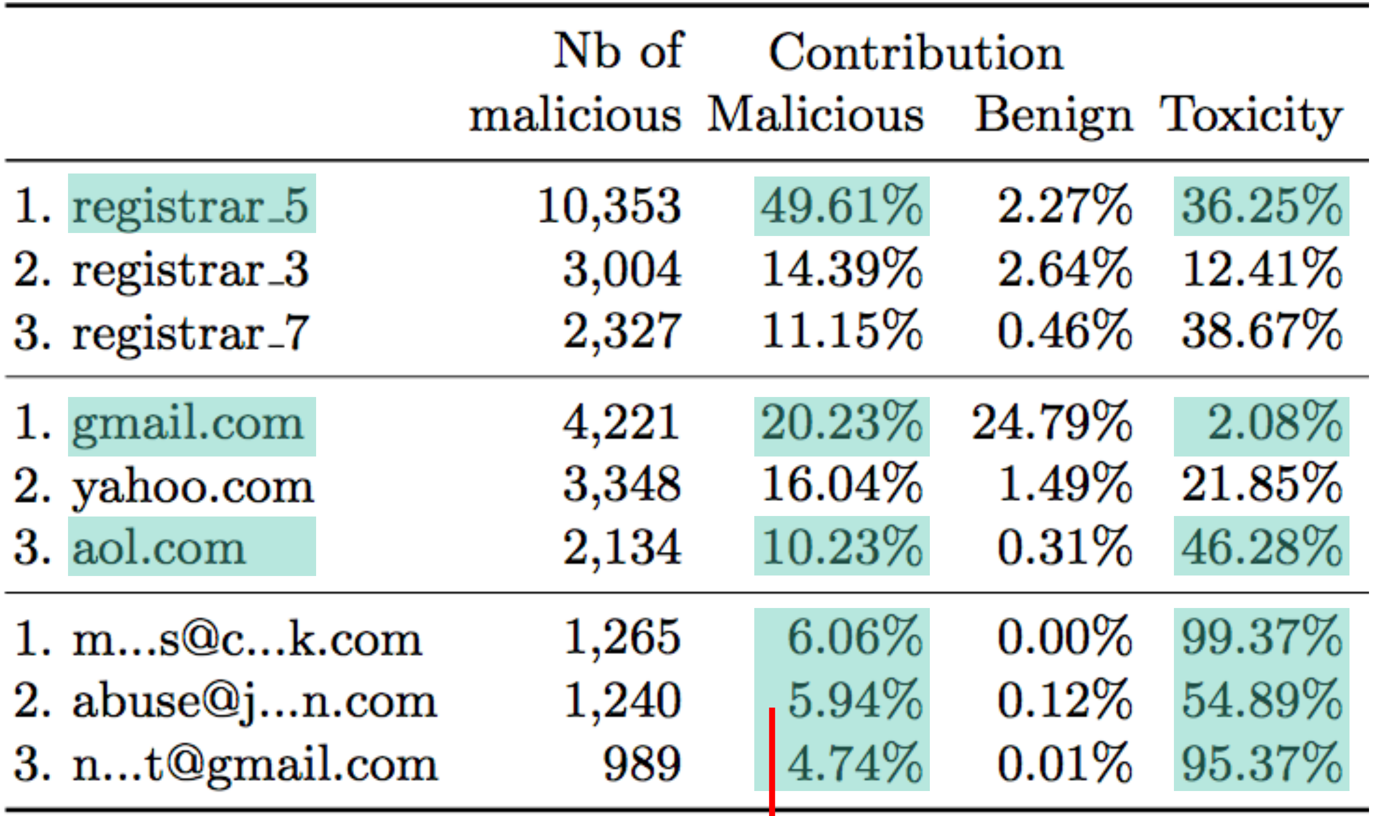
Insight 4: Some campaigns align with regular business activity patterns (3)



Insight 5: Some campaigns are fully automated



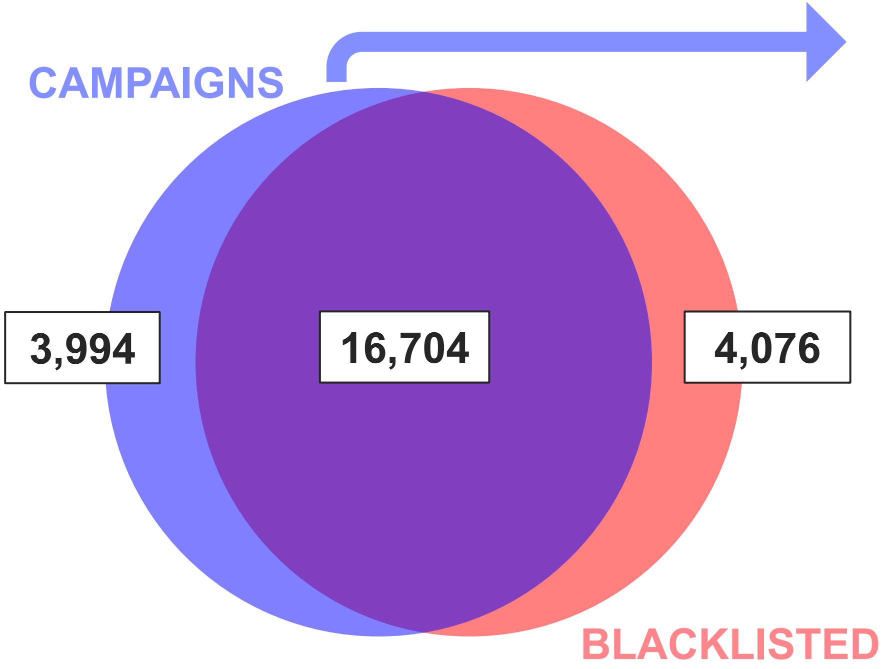
Insight 6: Top facilitators for malicious registrations



	Nb of	Contribution		
	malicious	Malicious	Benign	Toxicity
1. registrar_5	10,353	49.61%	2.27%	36.25%
2. registrar_3	3,004	14.39%	2.64%	12.41%
3. registrar_7	2,327	11.15%	0.46%	38.67%
1. gmail.com	4,221	20.23%	24.79%	2.08%
2. yahoo.com	3,348	16.04%	1.49%	21.85%
3. aol.com	2,134	10.23%	0.31%	46.28%
1. m...s@c...k.com	1,265	6.06%	0.00%	99.37%
2. abuse@j...n.com	1,240	5.94%	0.12%	54.89%
3. n...t@gmail.com	989	4.74%	0.01%	95.37%

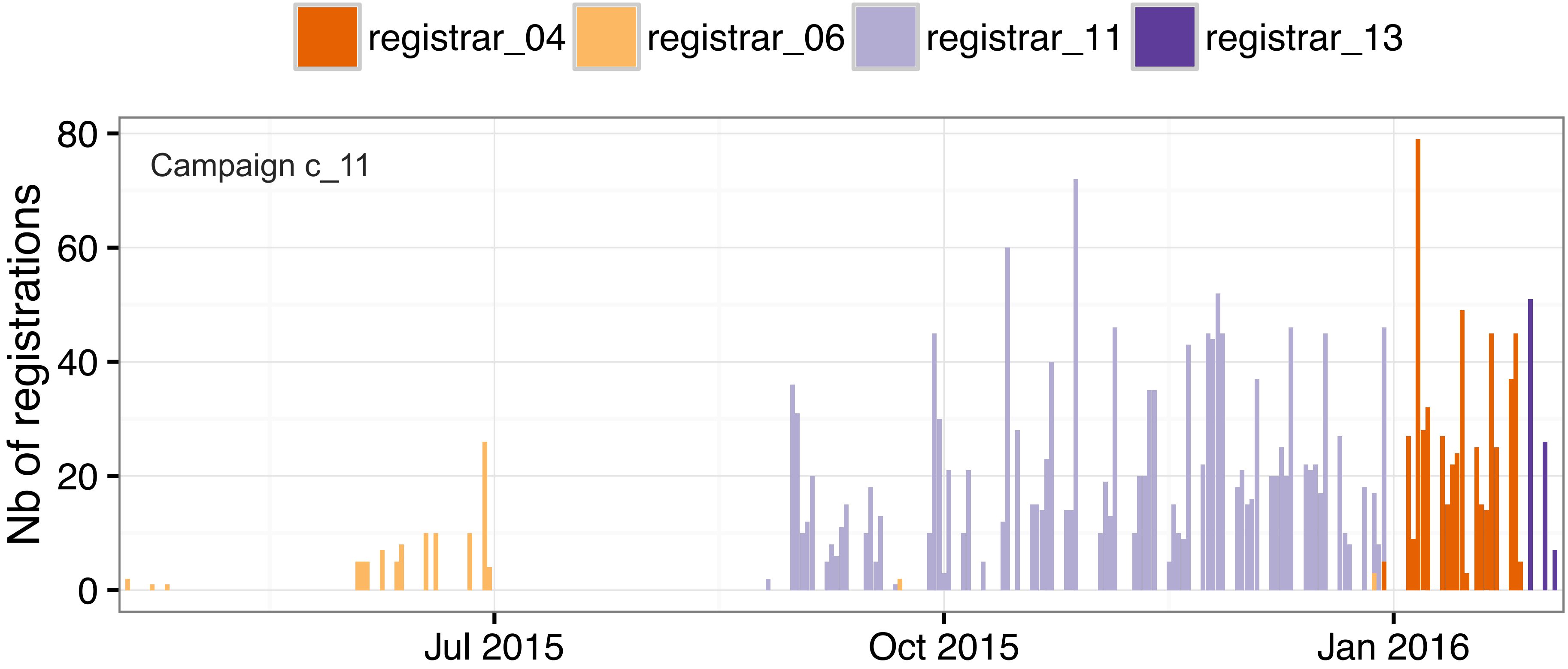
~ 17% of all registrations

Insight 7: Campaigns vs blacklists

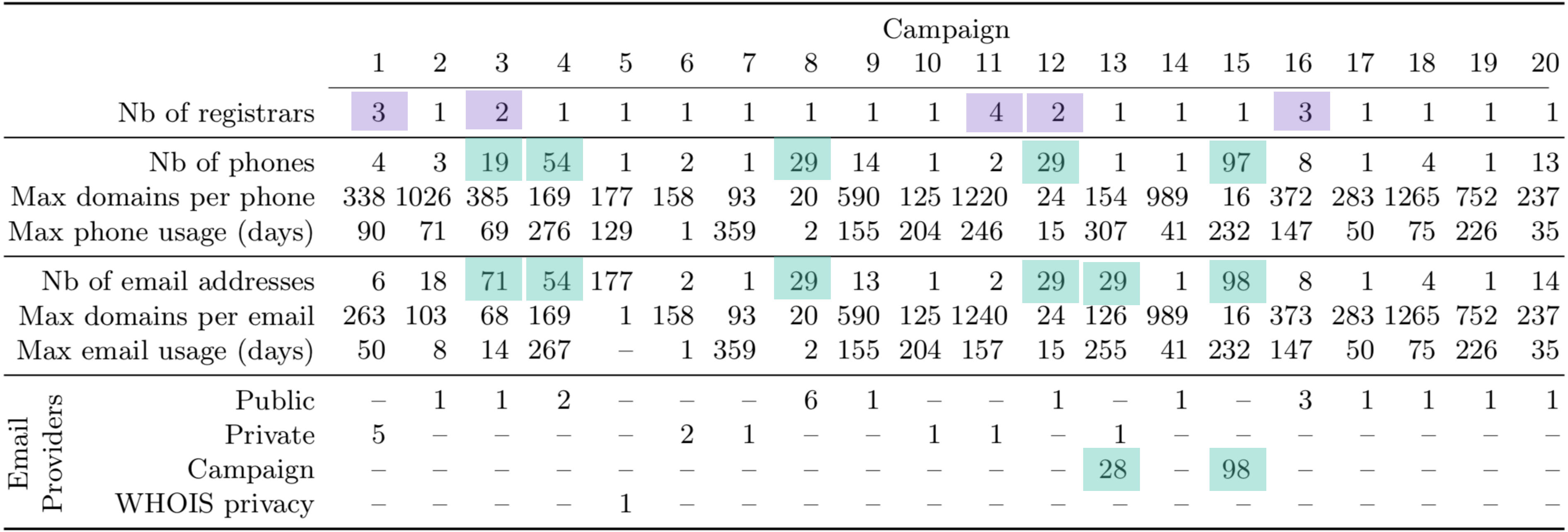


- Manual analysis of non-blacklisted domains
- Result: < 1% false positives
- About 20% extra on top of existing blacklists

Insight 8: Adaptive campaign strategies



Insight 8: Adaptive campaign strategies (2)



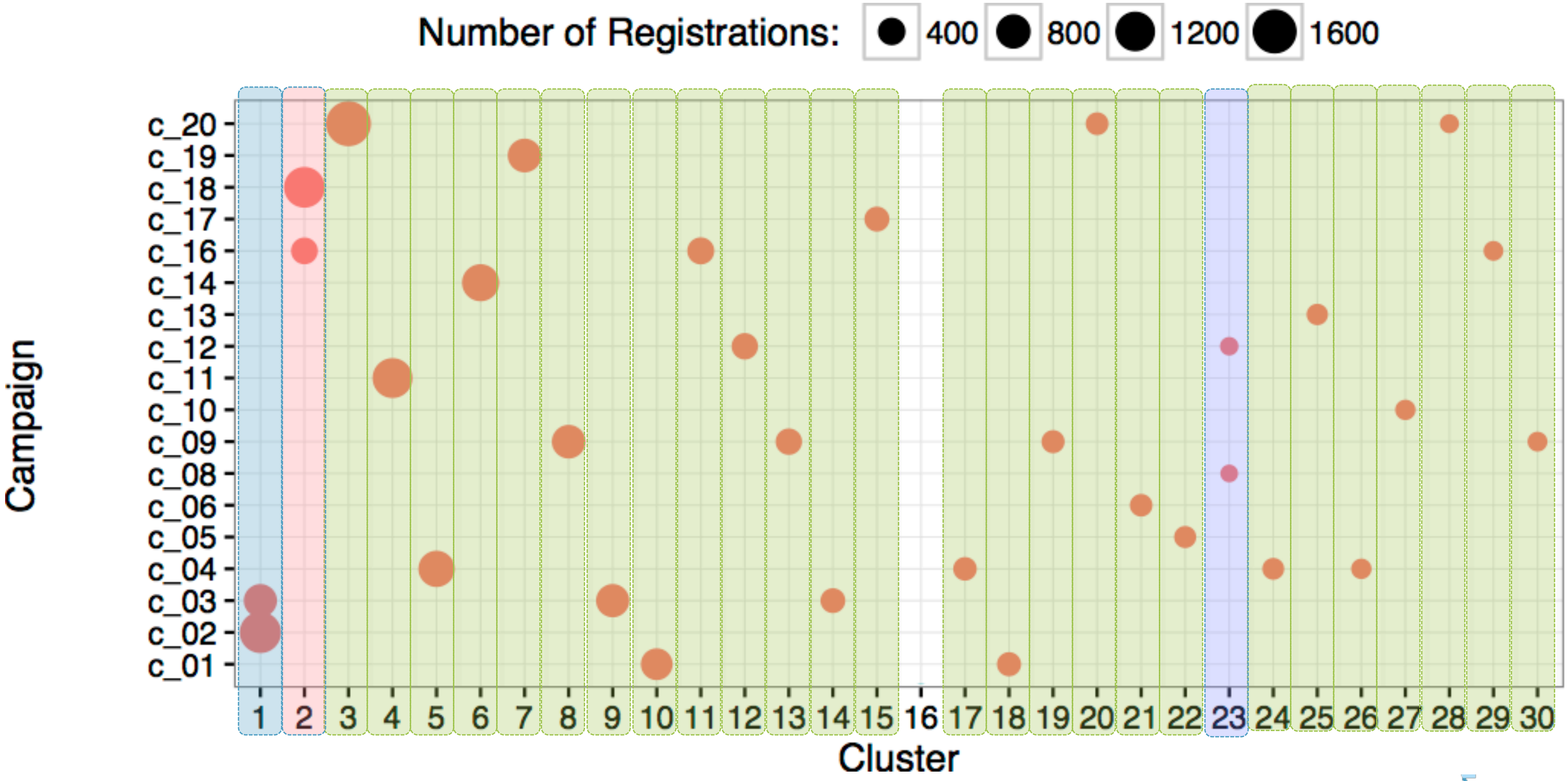
		Campaign																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Nb of registrars	3	1	2	1	1	1	1	1	1	1	4	2	1	1	1	3	1	1	1	1
	Nb of phones	4	3	19	54	1	2	1	29	14	1	2	29	1	1	97	8	1	4	1	13
	Max domains per phone	338	1026	385	169	177	158	93	20	590	125	1220	24	154	989	16	372	283	1265	752	237
	Max phone usage (days)	90	71	69	276	129	1	359	2	155	204	246	15	307	41	232	147	50	75	226	35
	Nb of email addresses	6	18	71	54	177	2	1	29	13	1	2	29	29	1	98	8	1	4	1	14
	Max domains per email	263	103	68	169	1	158	93	20	590	125	1240	24	126	989	16	373	283	1265	752	237
	Max email usage (days)	50	8	14	267	–	1	359	2	155	204	157	15	255	41	232	147	50	75	226	35
Email Providers	Public	–	1	1	2	–	–	–	6	1	–	–	1	–	1	–	3	1	1	1	1
	Private	5	–	–	–	–	2	1	–	–	1	1	–	1	–	–	–	–	–	–	–
	Campaign	–	–	–	–	–	–	–	–	–	–	–	–	28	–	98	–	–	–	–	–
	WHOIS privacy	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

Towards automatic campaign identification

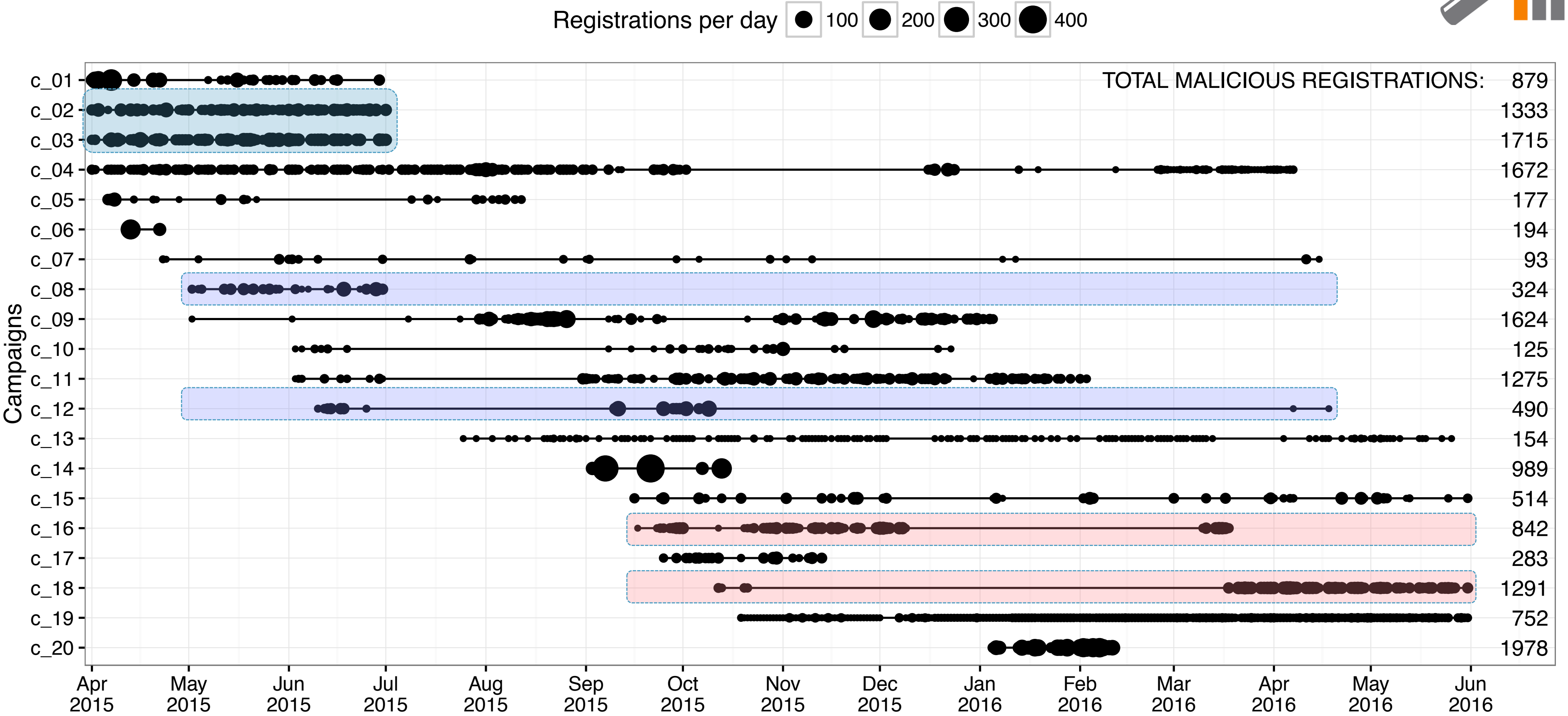
Campaign validation: clustering algorithm

- Machine learning technique to group registrations based on similarities between registration details
  - Agglomerative clustering of blacklisted registrations
  - Iteratively merge two closest clusters
- 30 largest (of 432) clusters represent 92% of campaign registrations

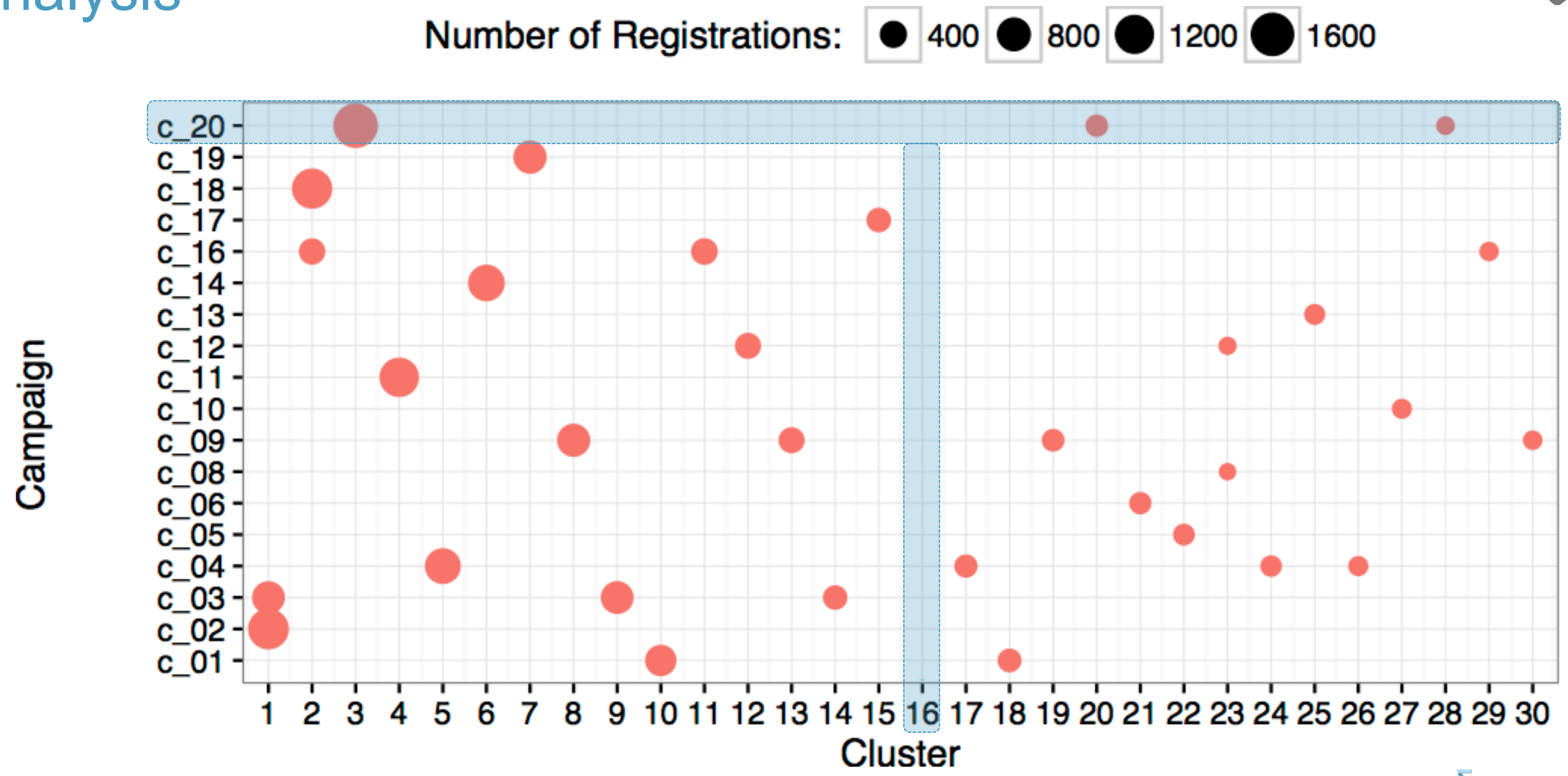
Cluster - campaign mapping



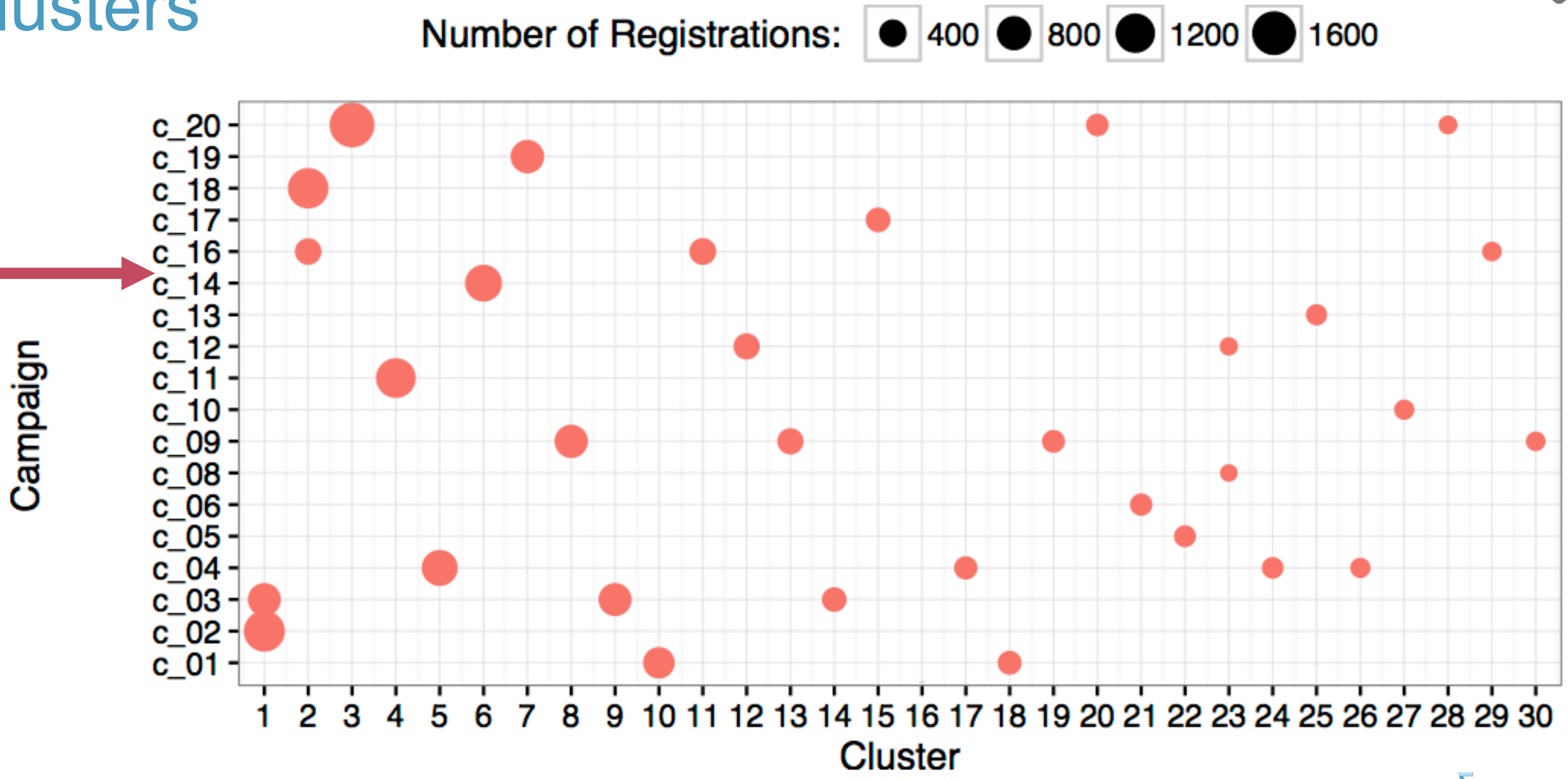
Finding 1: Some campaigns are linked to each other



Finding 2: Some registrations were missed during campaign analysis

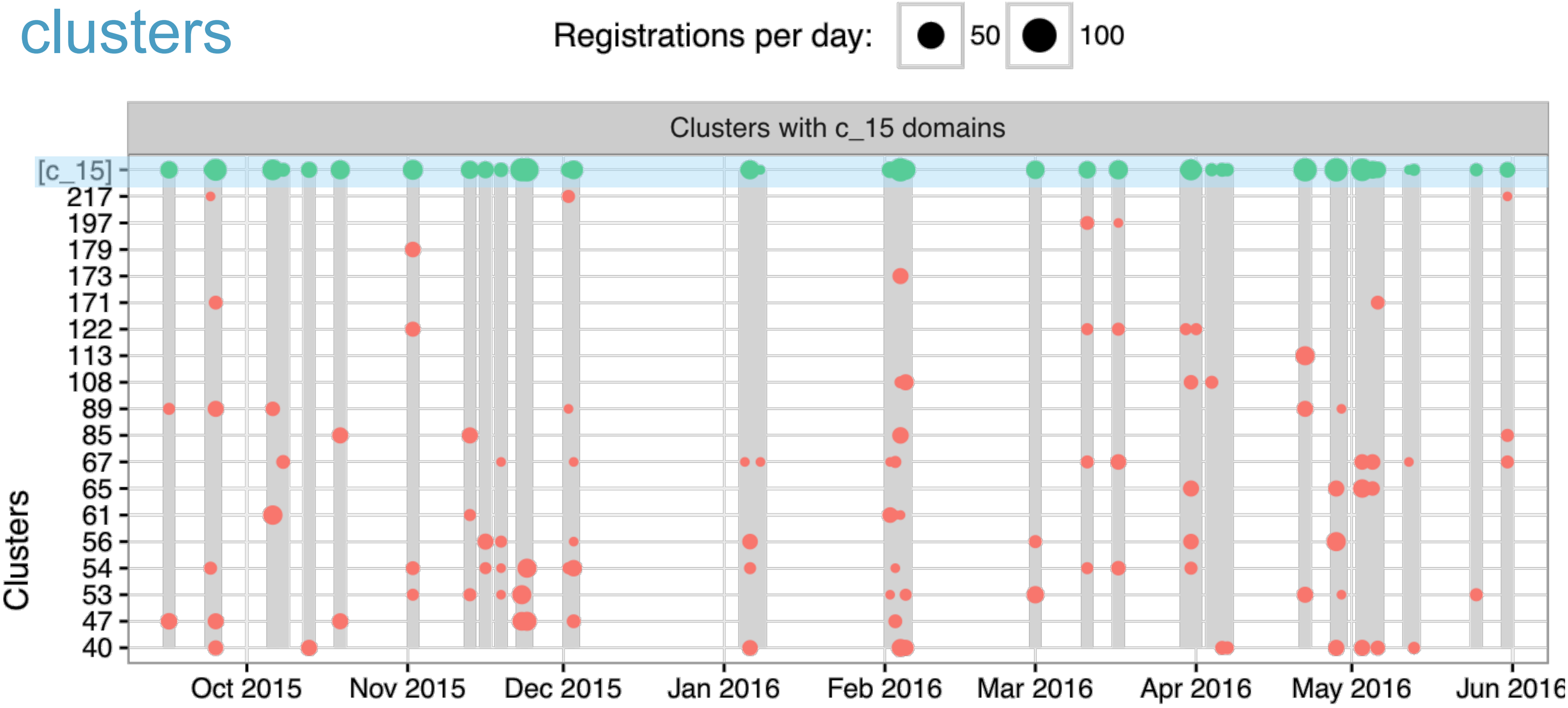


Finding 3: Advanced campaigns are not part of large clusters



15

Finding 3: Advanced campaigns are not part of large clusters



Example of an advanced campaign (c_15)

- Campaign c_15 is much more advanced
  - 514 domains registrations during 258 days
  - 98 registrants generated by Laravel Faker tool
  - Domain names consist out of 2-3 Dutch words
  - Dutch words are reused across registrants
  - Batches of 8, 16, 24 or 32 registrations
- Hard to automatically detect this type of patterns

Towards pro-active detection and prevention

“Given the commonalities between registrations in long-running campaigns, can newly registered domains with malicious intent be detected or prevented?”

Pro-active detection and prevention

- Based on previously-registered domain names, prediction models are trained:
  - Similarity-based agglomerative clustering
  - Reputation-based classification
- Early results:
  - About 60% of the malicious domain name registrations can proactively be detected and/or prevented at registration time
- Currently being deployed as part of EURid’s Trust & Security program

Conclusion

Campaign analysis on 14 months of registration data

- Hit-and-run strategies
- Some long-running campaigns
- Variety in intensity, duration and complexity/adaptiveness
- Alignment with business activity
- Top 3 facilitators have huge footprint
- Campaign analysis can strengthen existing blacklists

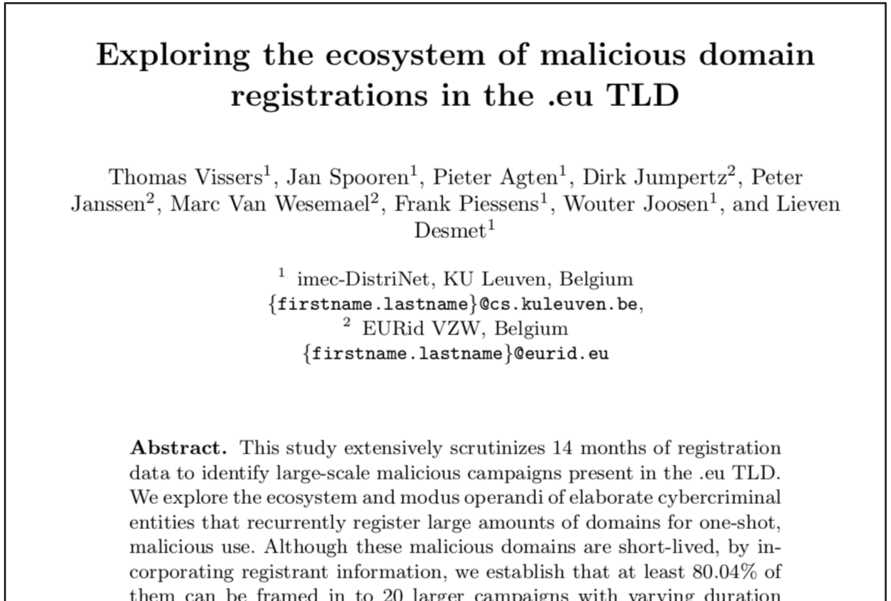
Towards …

- Automatic campaign identification
  - Validation of manual analysis process
  - Nice interplay between manual and automatic analysis

- Pro-active detection and prevention
  - Early results look promising
  - More to come within next 6 months!

Interested in more?

- Thomas Vissers, Jan Spooren, Pieter Agten, Dirk Jumpertz, Peter Janssen, Marc Van Wesemael, Frank Piessens, Wouter Joosen, Lieven Desmet, **Exploring the ecosystem of malicious domain registrations in the .eu TLD**, Research in Attacks, Intrusions, and Defenses, (RAID 2017), Atlanta, USA, September 18-20, 2017



Exploring the ecosystem of malicious domain registrations in the .eu TLD

Thomas Vissers[1], Jan Spooren[1], Pieter Agten[1], Dirk Jumpertz[2], Peter Janssen[2], Marc Van Wesemael[2], Frank Piessens[1], Wouter Joosen[1], and Lieven Desmet[1]

[1] imec-DistriNet, KU Leuven, Belgium
{firstname.lastname}@cs.kuleuven.be,
[2] EURid VZW, Belgium
{firstname.lastname}@eurid.eu

Abstract. This study extensively scrutinizes 14 months of registration data to identify large-scale malicious campaigns present in the .eu TLD. We explore the ecosystem and modus operandi of elaborate cybercriminal entities that recurrently register large amounts of domains for one-shot, malicious use. Although these malicious domains are short-lived, by incorporating registrant information, we establish that at least 80.04% of them can be framed in to 20 larger campaigns with varying duration

Final version: https://doi.org/10.1007/978-3-319-66332-6_21

Exploring the ecosystem of malicious domain registrations in the .eu TLD

Lieven Desmet – OWASP BeNeLux Day 2017 – Tilburg, NL

Lieven.Desmet@cs.kuleuven.be – @lieven_desmet