Statistical Analysis of DNS Abuse in gTLDs

Prepared for the Competition, Consumer Trust, and Choice Review Team (CCT-RT)

Maciej Korczyński, Delft University of Technology, Grenoble INP - Grenoble Alps University
Maarten Wullink, SIDN Labs
Brian Aitchison, ICANN Operations and Policy Research

Community Webinar

September 2017

Agenda

- Introduction from the ICANN organization: Background of Study
- Presentation from SIDN and Delft University of Technology
- Q & A

Study Background

2009

- Mitigating Malicious Conduct: New gTLD Program Explanatory Memorandum

Question	Recommendation(s)
1) How do we ensure that bad actors do not run registries?	1. Vet registry operators
2) How do we ensure integrity and utility of registry information?	2. Require DNSSEC Deployment 3. Prohibit "wildcarding" 4. Encourage removal of "orphan glue" records
3) How do we ensure more focused efforts on combating identified abuse?	5. Require "Thick" WHOIS records 6. Centralize Zone File access 7. Document registry- and registrar-level abuse contacts and policies 8. Provide an expedited registry security request process
4) How do we provide an enhanced control framework for TLDs with intrinsic potential for malicious conduct?	9. Create a draft framework for a high security zone verification program

Study Background (cont'd)

2016

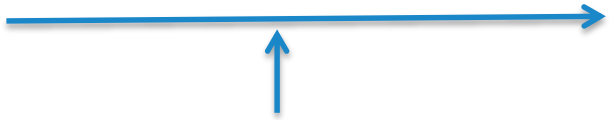
- New gTLD Program Safeguards Against DNS Abuse: Revised Report
  - Research aid to Competition, Consumer Trust, and Choice Review Team
  - How to measure effectiveness of safeguards?

Base Research Model

Explanatory Variable: DNS Expansion

Potential proxy metrics:

- Number of domain names
  - Legacy TLDs
  - New TLDs
  - Entire DNS



Intervening Variable(s) Safeguards to Mitigate DNS Abuse

What about…

- Pricing?
- Operational policies and/or practices?
- Systemic policies and/or practices?
- Cybercriminal preferences and practices?

Response Variable: DNS Abuse Rate

Potential proxy metrics:

- Spam rate
- Phishing rate
- Malware rate
- Others as relevant to the "effectiveness" objectives of the safeguards

Study Background (cont'd)

2016 -2017

- Competition, Consumer Trust, and Choice Review Team
  - Affirmation of Commitments (AoC) specified that "malicious abuse issues" be addressed in expansion of top-level domain space
  - CCT-RT mandated by AoC to examine "effectiveness of…safeguards put in place to mitigate issues involved in…the expansion [of the top-level domain space]"
  - Required comprehensive descriptive statistics as **baseline measure** of abuse rates in new compared to legacy gTLDs in order to gauge safeguard effectiveness
  - Also serves as proxy for "Trust", i.e. changes in abuse rate → changes in trust
  - CCT-RT Draft Report recommends ongoing DNS abuse measurement

Study Timeline

- RFP issued August 2016
- SIDN contracted November 2016

Study

Statistical Analysis of DNS Abuse in gTLDs (SADAG)

Consortium: SIDN and TU Delft

Requested by: Competition, Consumer Trust, and Choice Review Team

Goal

- Comprehensive statistical comparison of rates of DNS abuse in new and legacy gTLDs
  - Spam
  - Phishing
  - Malware
- Statistical analysis of potential abuse drivers

Motivation

– New Generic Top-Level Domain (gTLD) Program enabled hundreds of new generic top-level domains

Data

Blacklists

- Anti Phishing Working Group
  - Phishing URLs
- StopBadware
  - Malware URLs
- SURBL (4 blacklists)
  - Phishing domains
  - Spam domains
  - Malware domains

Data

Blacklists

- Spamhaus
  - Spam domains
- CleanMX (3 feeds)
  - Phishing URLs
  - Malware URLs
  - Defaced URLs
- Secure Domain Foundation
  - Phishing URLs
  - Malware URLs

Data

WHOIS data

- WHOIS XML API
  - All new gTLDs
  - Subset of legacy gTLDs
- DomainTools
  - Providing missing domains

Domain data

- Zone files
  - Per gTLD
  - Per day
  - 3-year period

Data

Active Web & DNS Scan

- Scanned
  - All new gTLDs
  - Sample of legacy gTLDs

Registry (ICANN)

- Sunrise periods
- Registry operators (parent companies of registry operators)

Security Metrics

- Distribution of malicious content: *
  - Number of unique domains
    - E.g. **malicious.com**

* **"Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs"**, Maciej Korczyński, Samaneh Tajalizadehkhoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten, in the *IEEE European Symposium on Security and Privacy (Euro S&P)*

Security Metrics

- Distribution of malicious content:
  - Number of unique domains
    - E.g. malicious.com
  - Number of FQDNs
    - E.g**. connect.secure.wellsfargo.malicious.com, bankofamerica.com.malicious.com, (…)**

* **"Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs"**, Maciej Korczyński, Samaneh Tajalizadehkhoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten, in the *IEEE European Symposium on Security and Privacy (Euro S&P)*

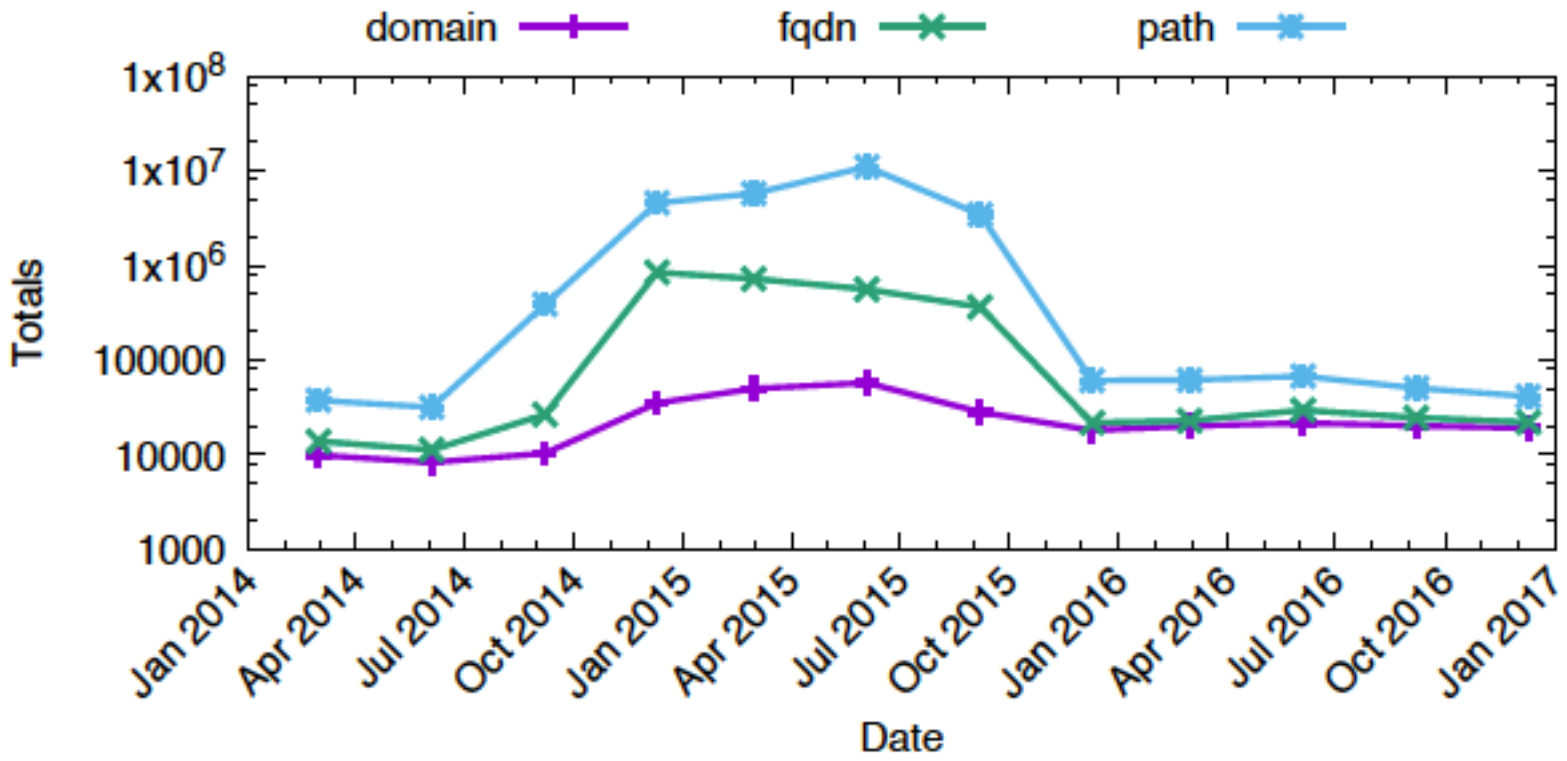
Security Metrics

- Distribution of malicious content:
  - Number of unique domains
    - E.g. malicious.com
  - Number of FQDNs
    - E.g. connect.secure.wellsfargo.malicious.com, bankofamerica.com.malicious.com, (…)
  - Number of URLs
    - E.g. **malicious.com/wp-content/file.php**, **malicious.com/wp-content/gate.php**, **(…)**

* **"Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs"**, Maciej Korczyński, Samaneh Tajalizadehkhoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten, in the *IEEE European Symposium on Security and Privacy (Euro S&P)*

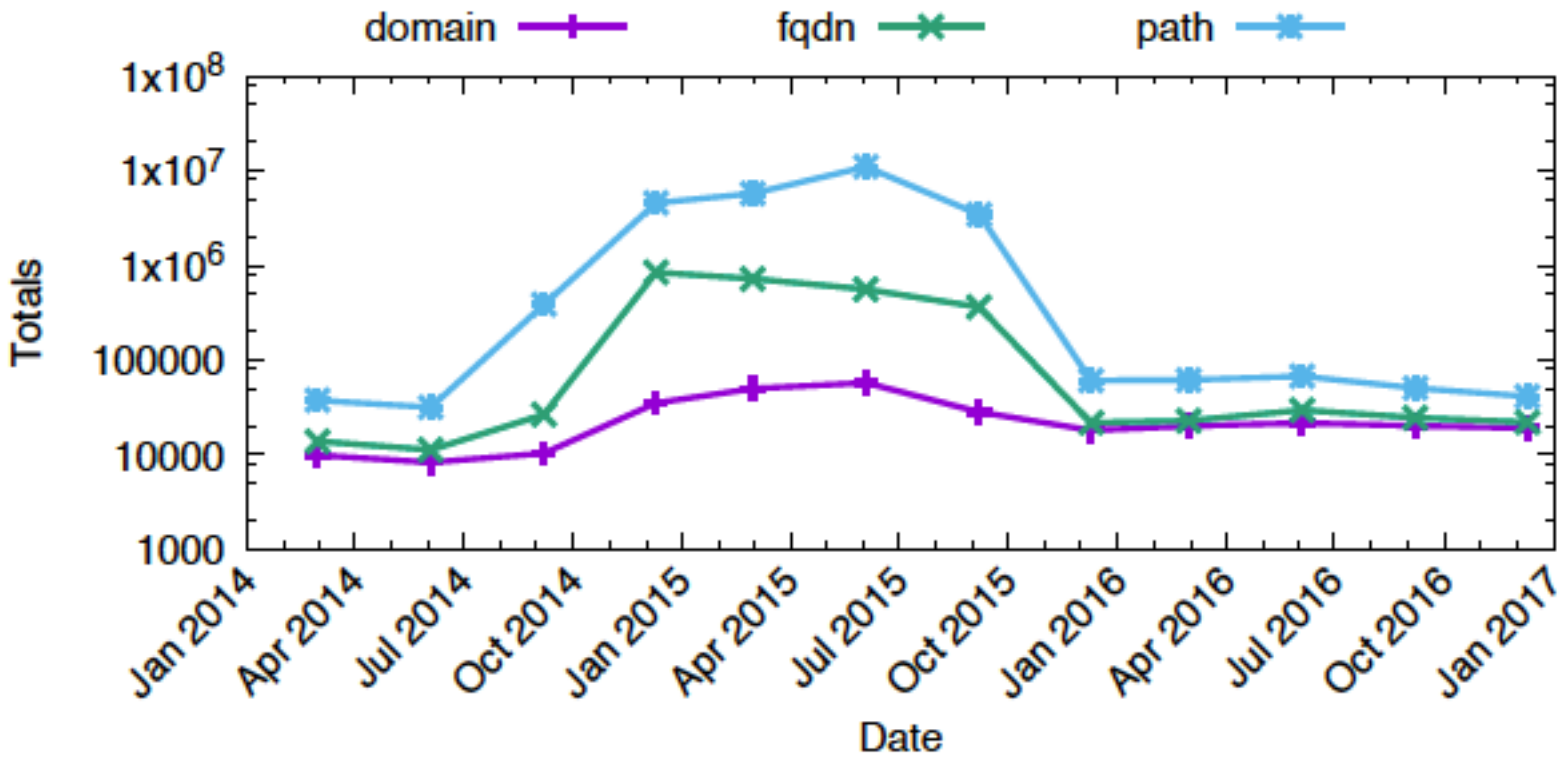
Security Metrics for gTLDs

Phishing domains, FQDNs, and URLs (APWG) per legacy gTLDs



Security Metrics for gTLDs

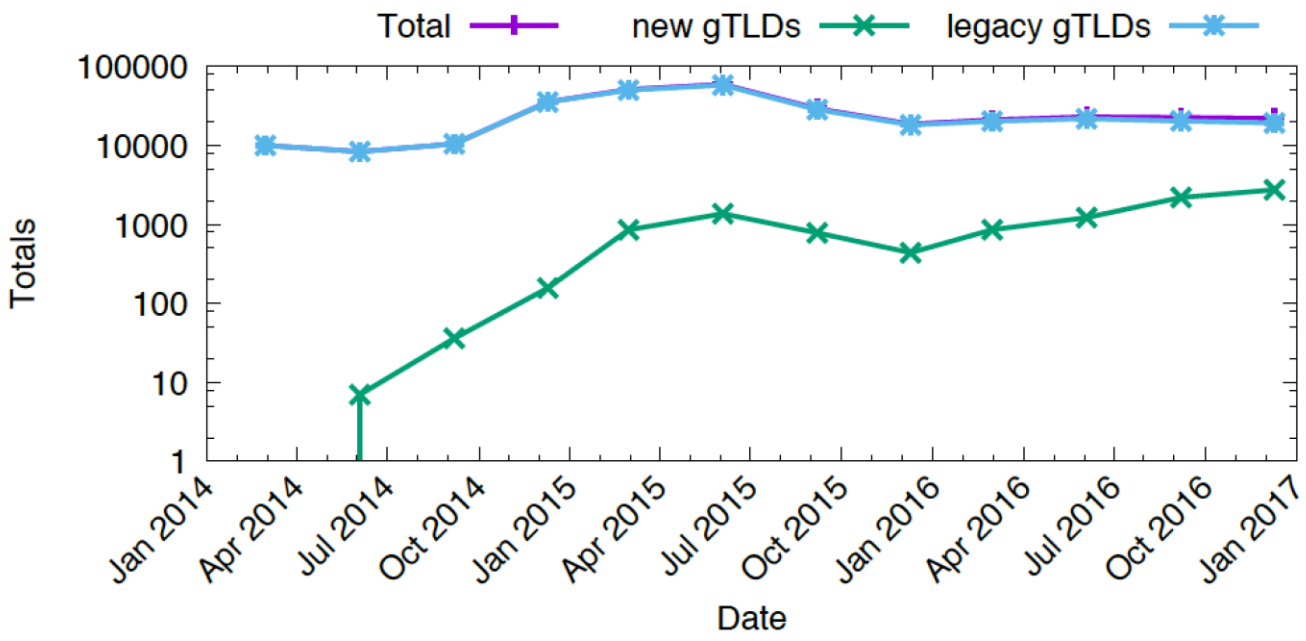
Phishing domains, FQDNs, and URLs (APWG) per legacy gTLDs



Three measures reflect attackers' profit-maximizing behavior. They abuse free legitimate services and affect the reputations of such associated services.

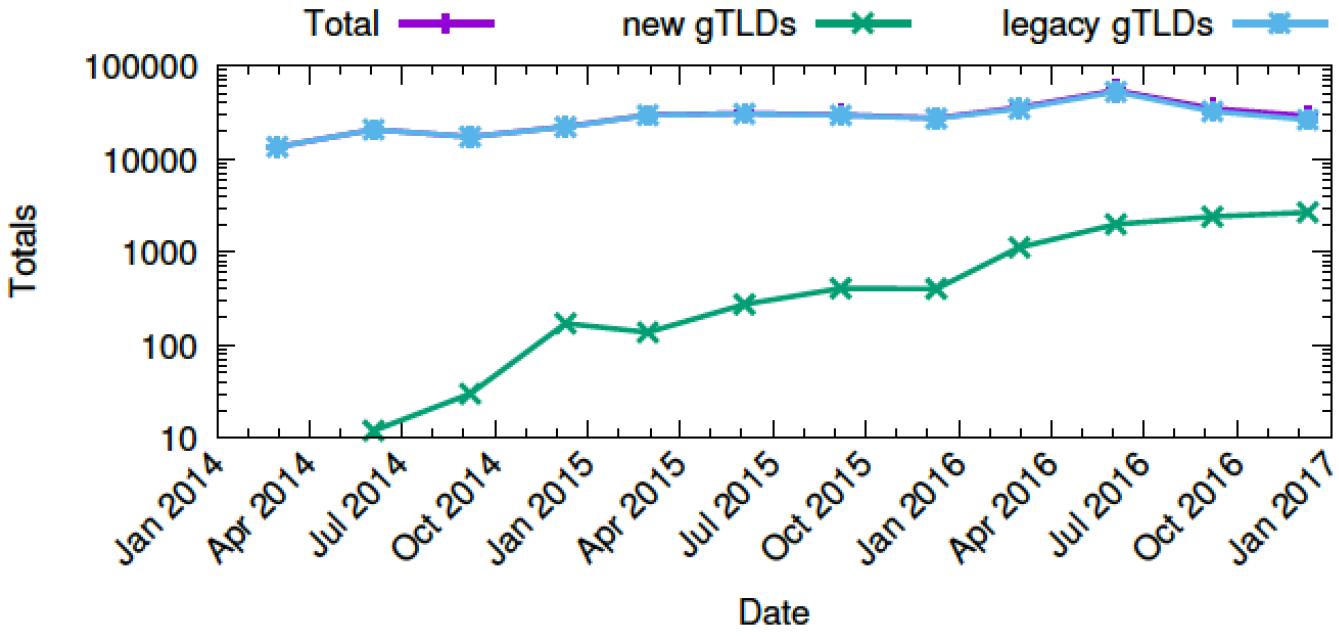
Security Metrics for gTLDs

Phishing domains (APWG) per new and legacy gTLDs



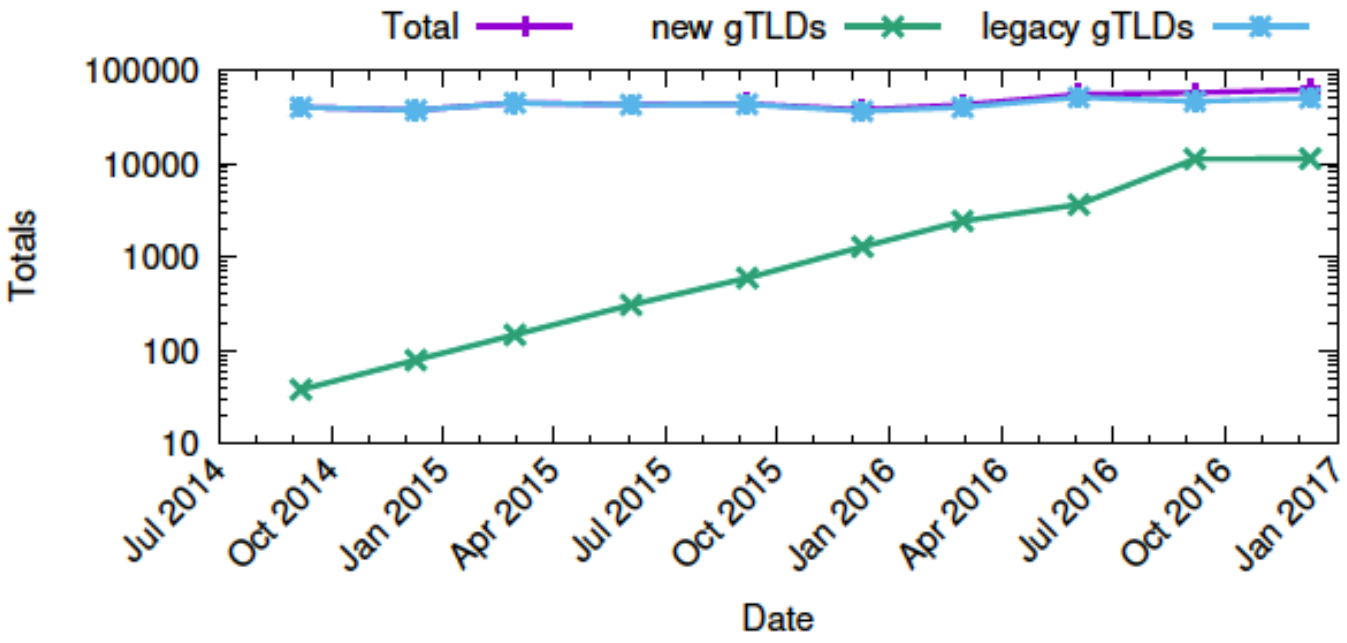
Security Metrics for gTLDs

Phishing domains (CleanMX ph) per new and legacy gTLDs



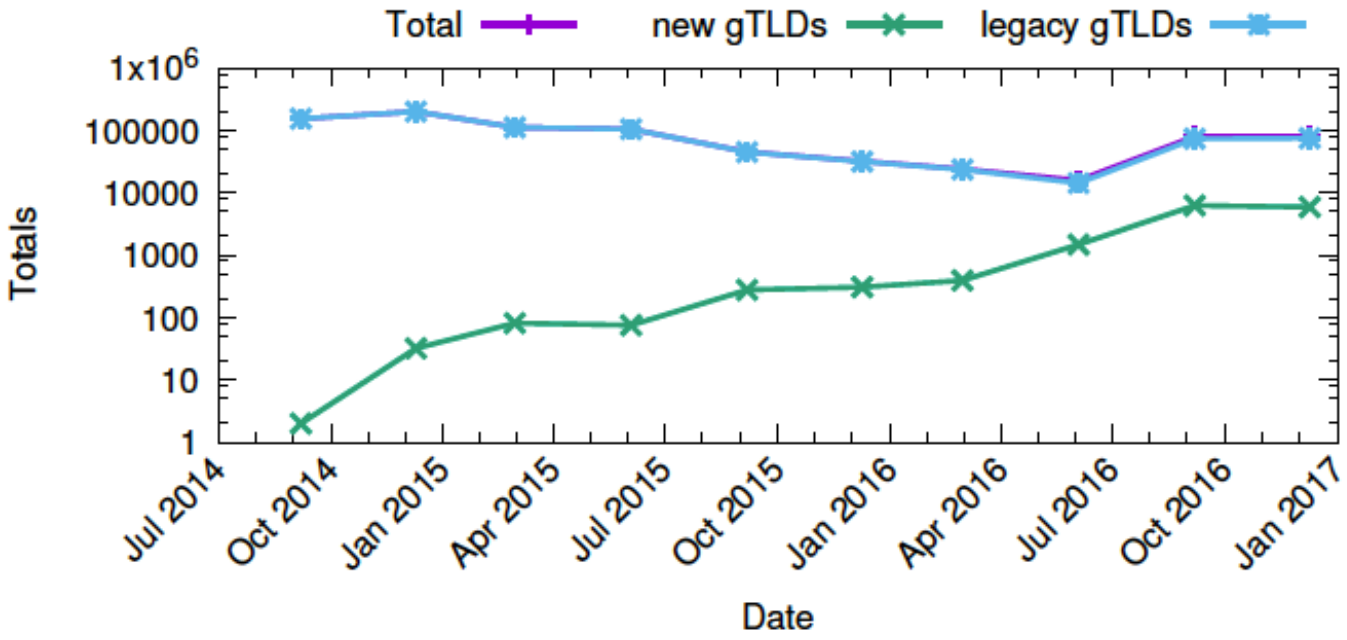
Security Metrics for gTLDs

Phishing domains (SURBL ph) per new and legacy gTLDs



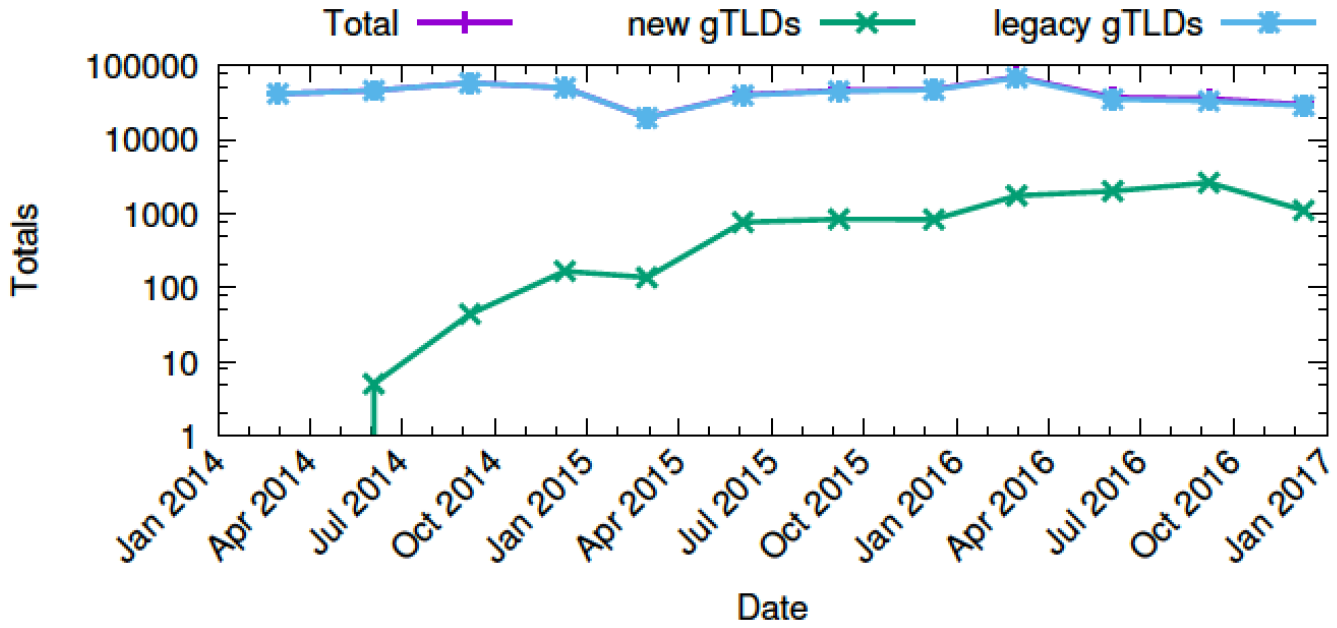
Security Metrics for gTLDs

Malware domains (SURBL mw) per new and legacy gTLDs



Security Metrics for gTLDs

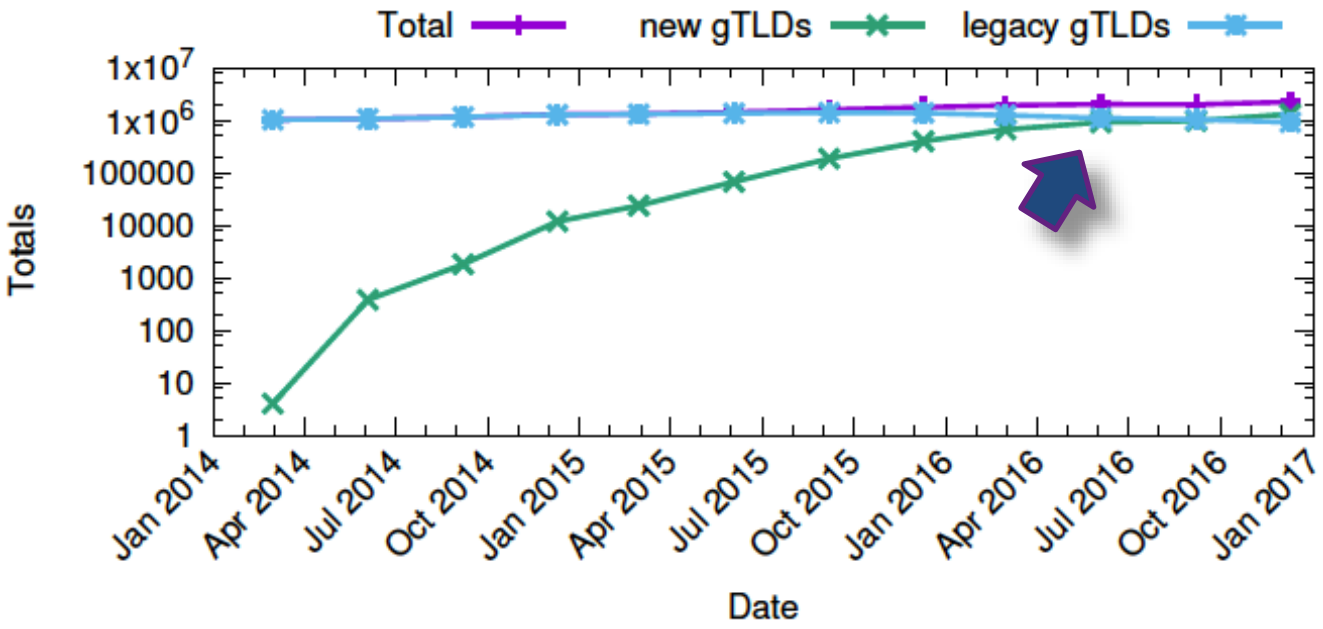
Malware domains (CleanMX mw) per new and legacy gTLDs



While the number of abused domains remains approximately constant in legacy gTLDs, we observe a clear upward trend in the absolute number of **phishing** and **malware** domains in new gTLDs.

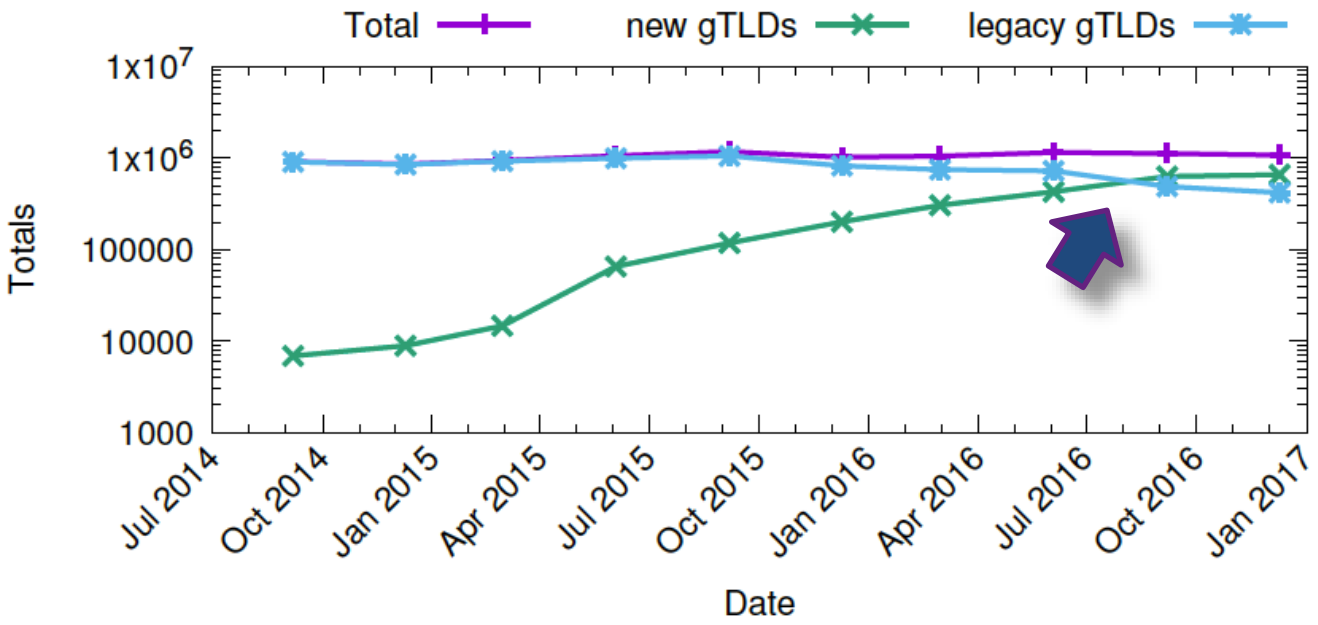
Security Metrics for gTLDs

Spam domains (Spamhaus) per new and legacy gTLDs



Security Metrics for gTLDs

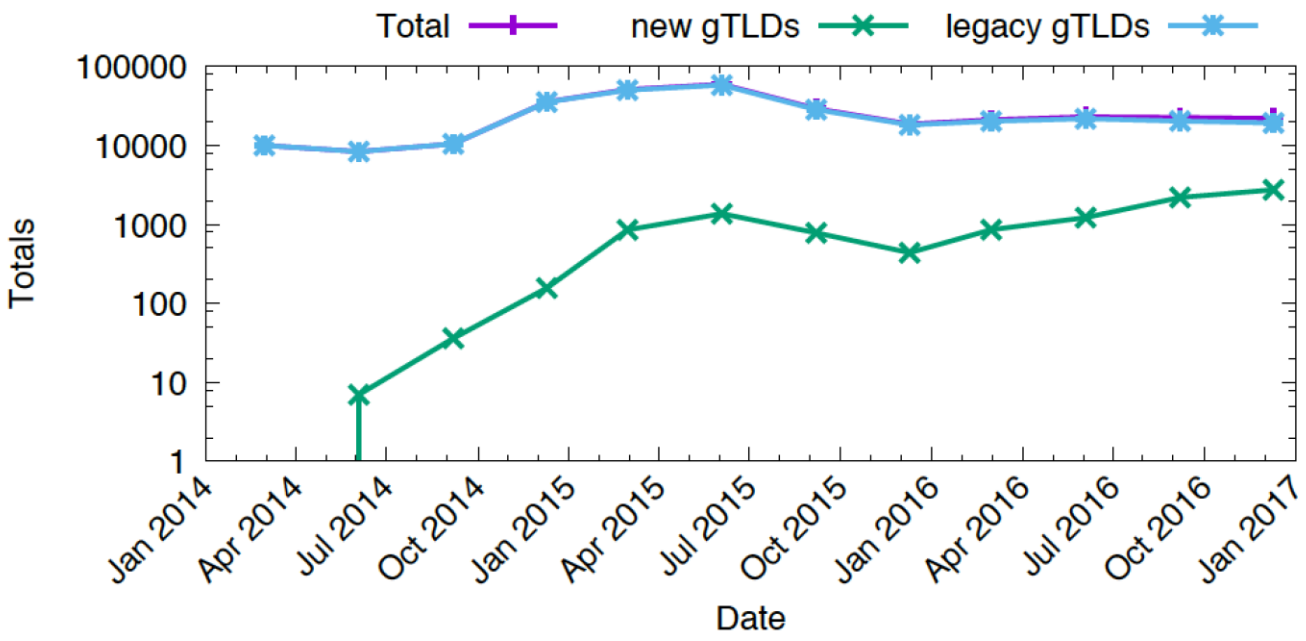
Spam domains (SURBL ws) per new and legacy gTLDs



The **absolute** number of **spam** domains in new gTLDs higher than in legacy gTLDs at the end of 2016

Security Metrics for gTLDs

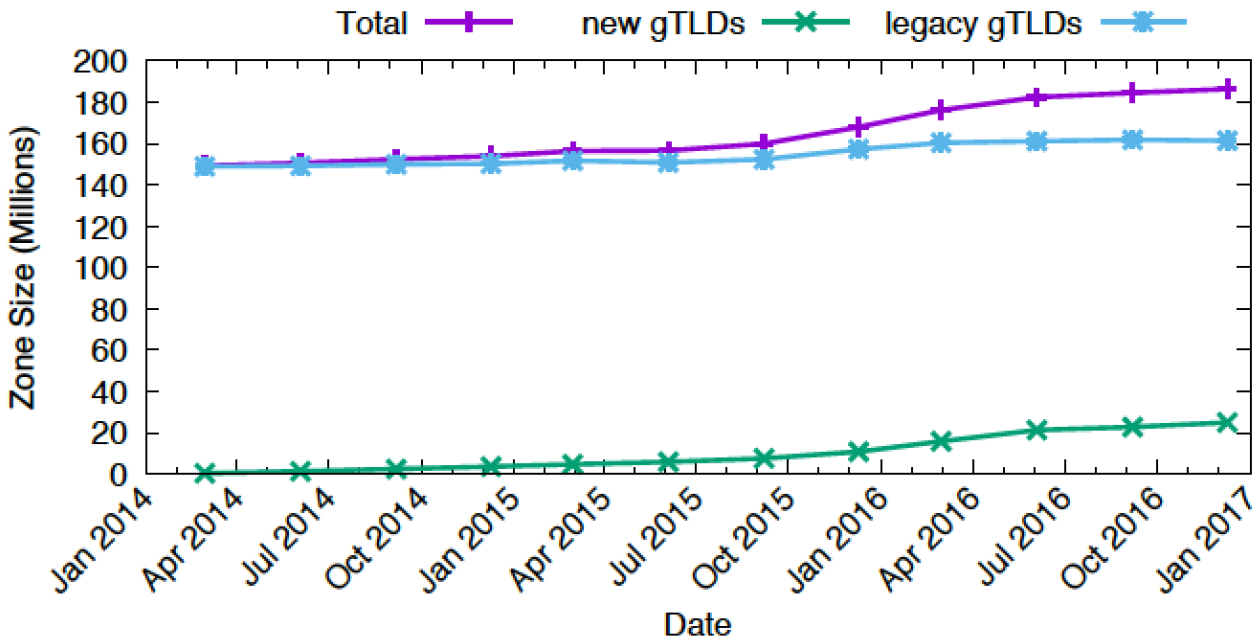
Phishing domains (APWG) per new and legacy gTLDs



Size matters!

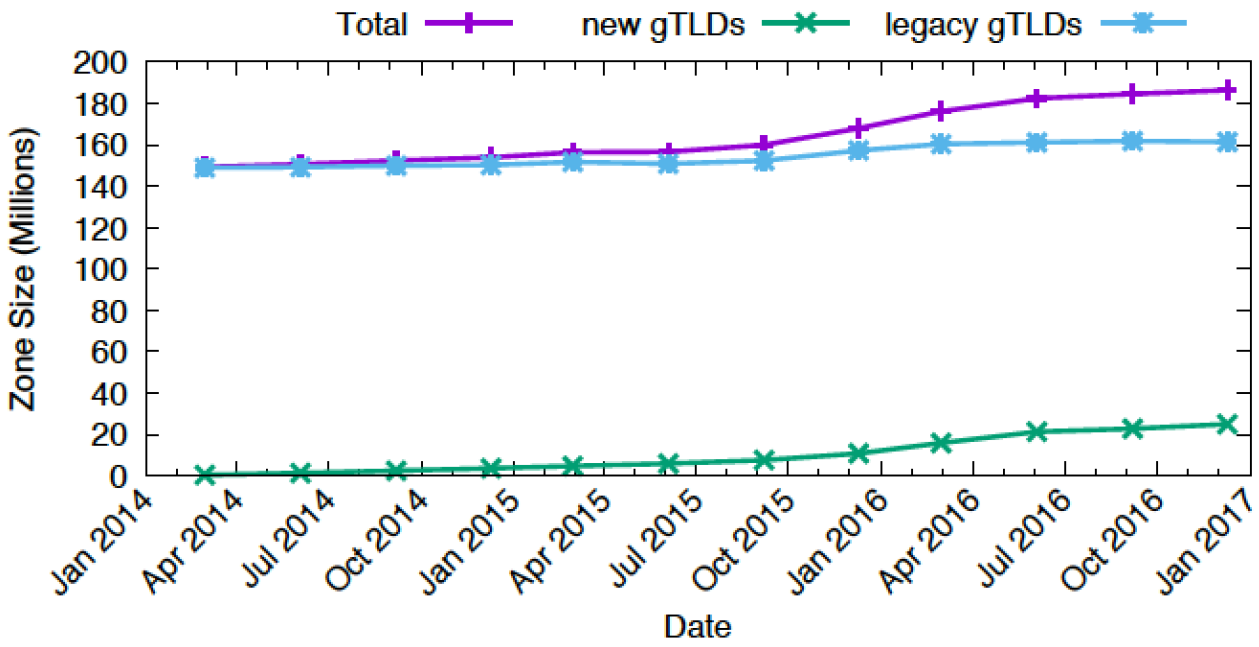
Size

– Size estimate: Number of domains in each gTLD zone file



Size

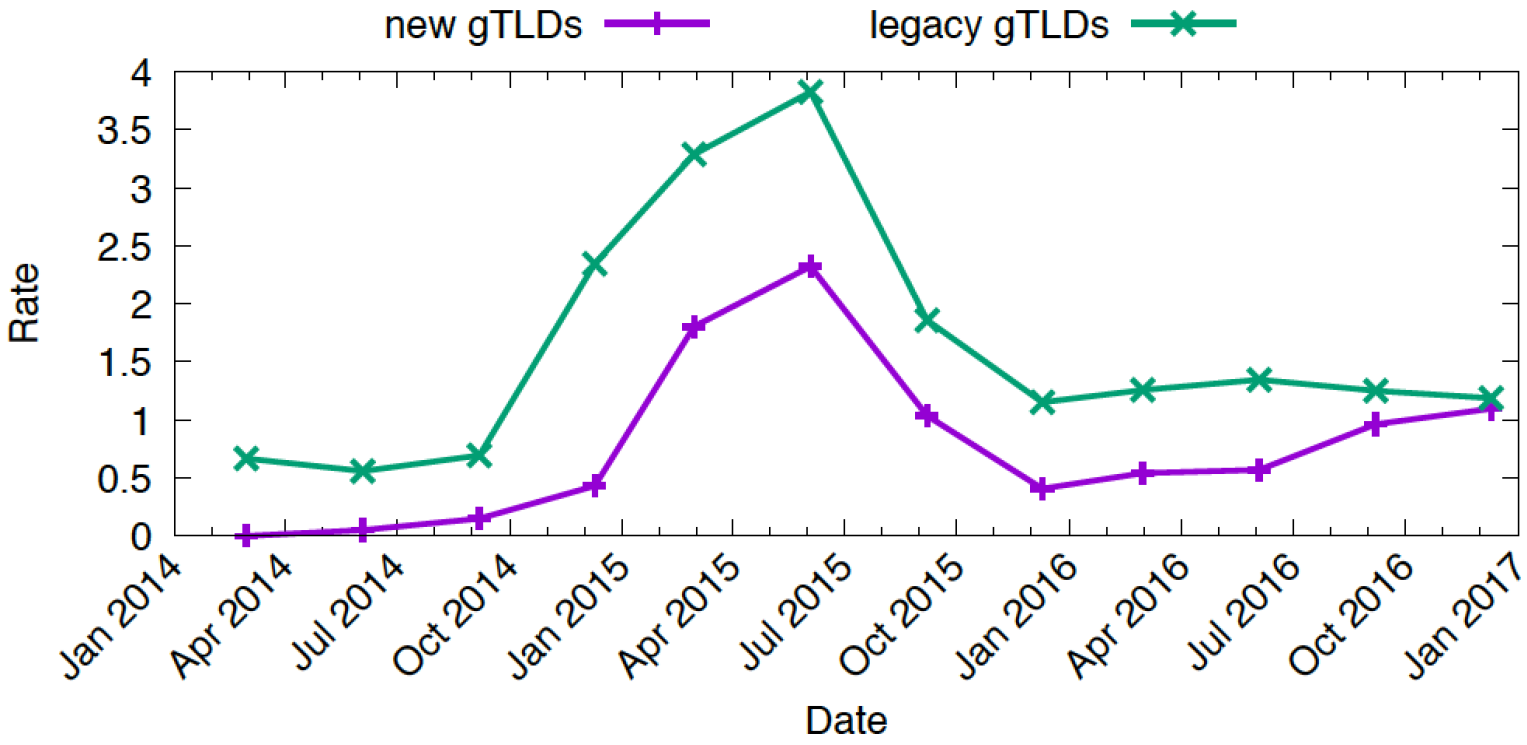
– Size estimate: Number of domains in each gTLD zone file



– Rates: (#blacklisted domains / #all domains) * 10,000

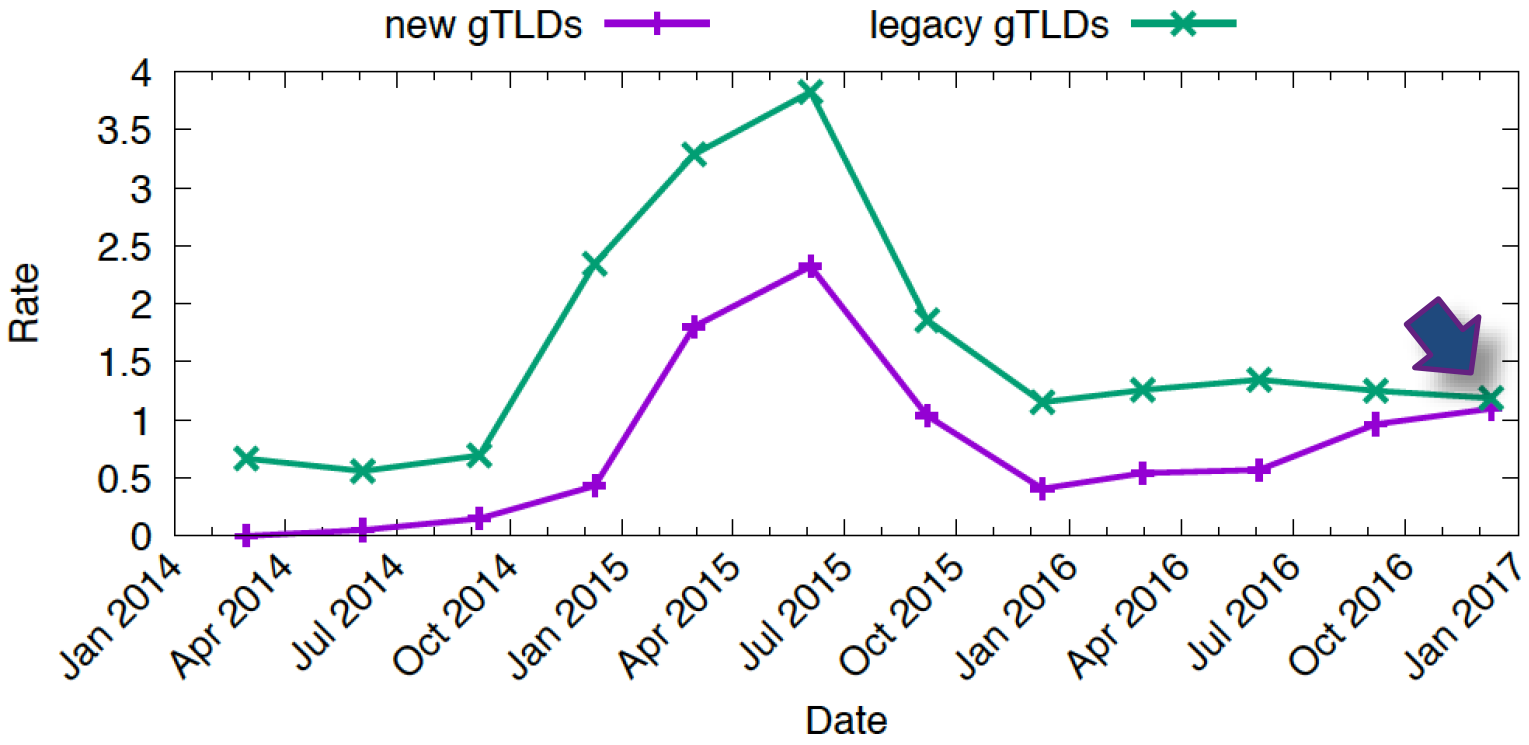
Abuse Rates

– Time series of abuse rates of **phishing** domains in legacy gTLDs and new gTLDs based on the APWG feed



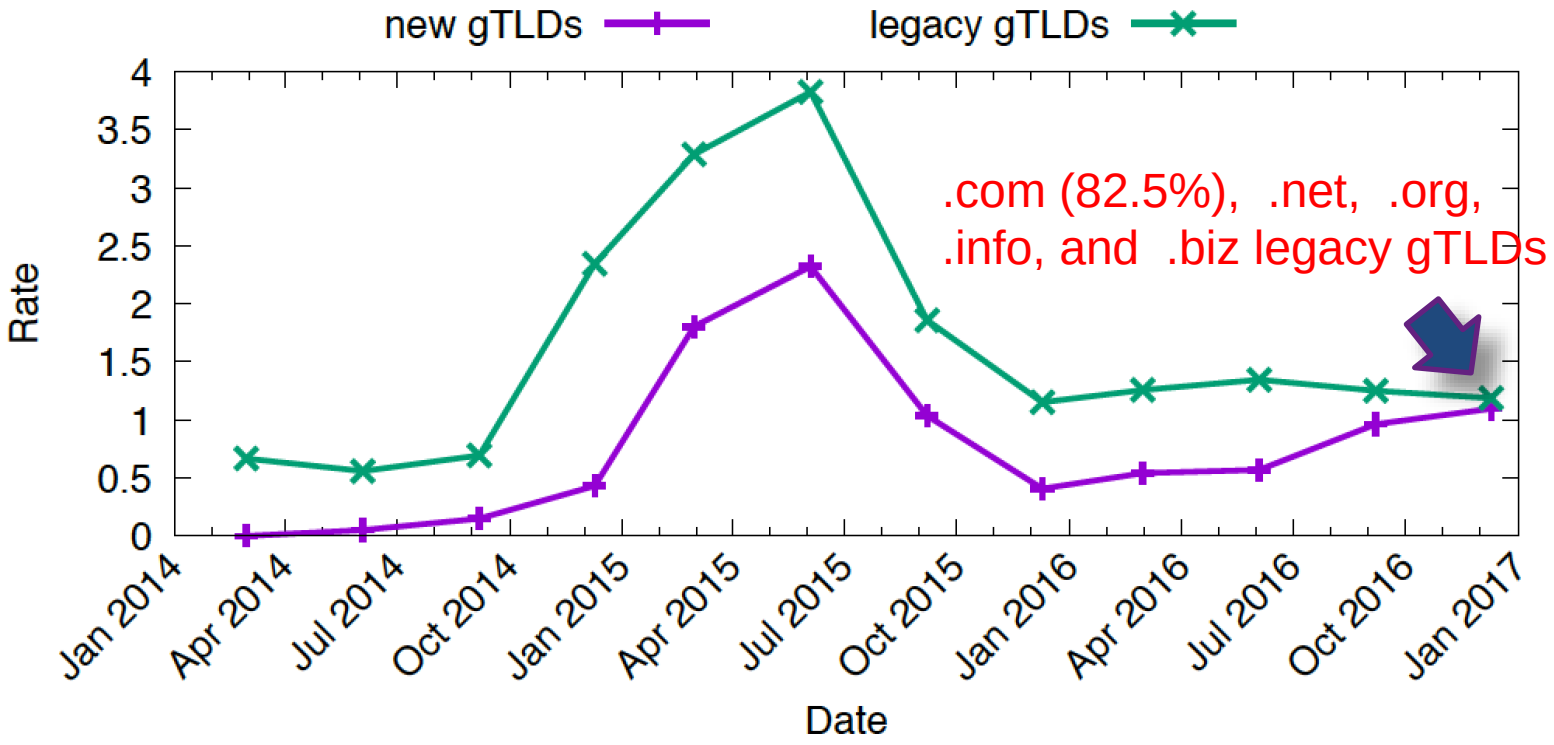
Abuse Rates

– Time series of abuse rates of **phishing** domains in legacy gTLDs and new gTLDs based on the APWG feed



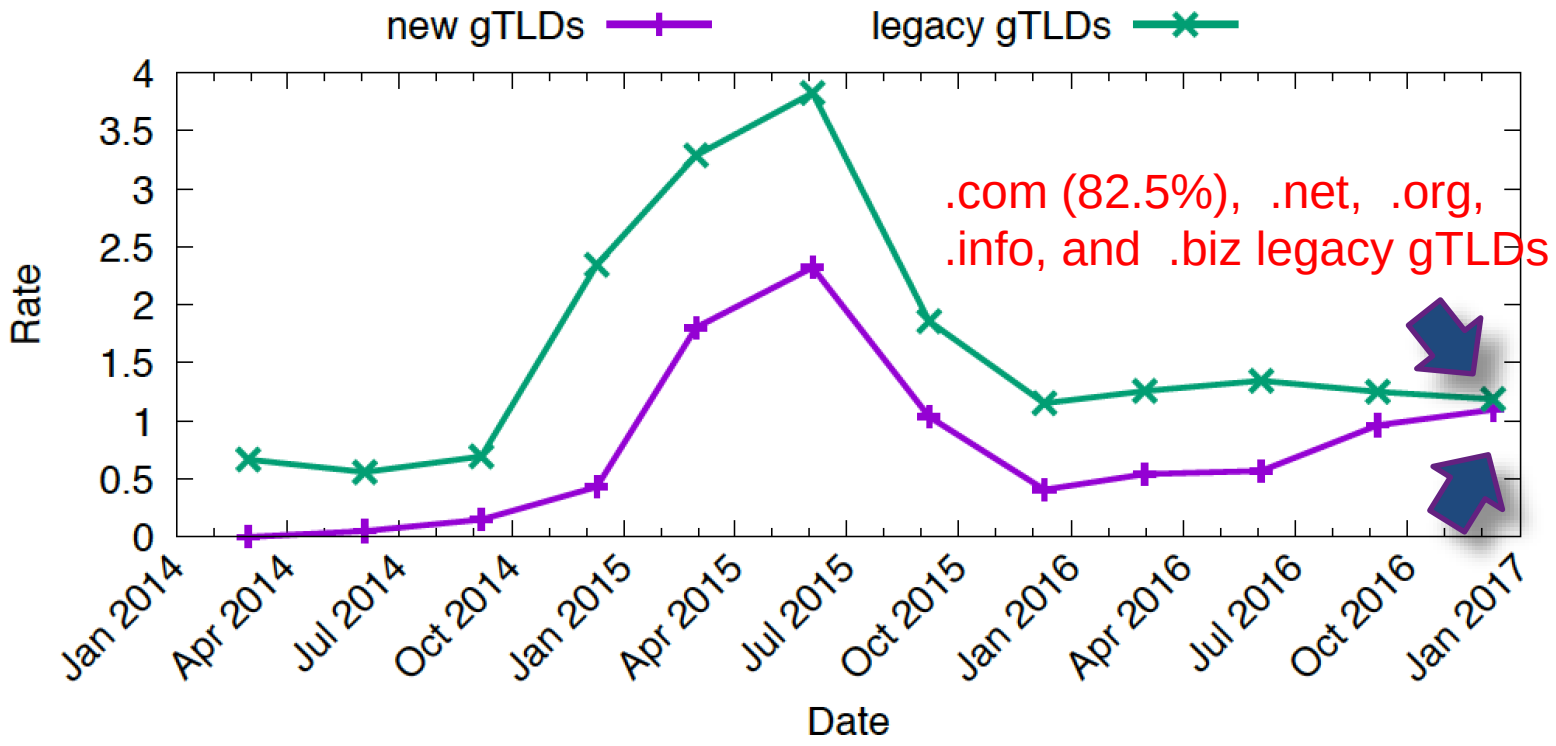
Abuse Rates

– Time series of abuse rates of **phishing** domains in legacy gTLDs and new gTLDs based on the APWG feed



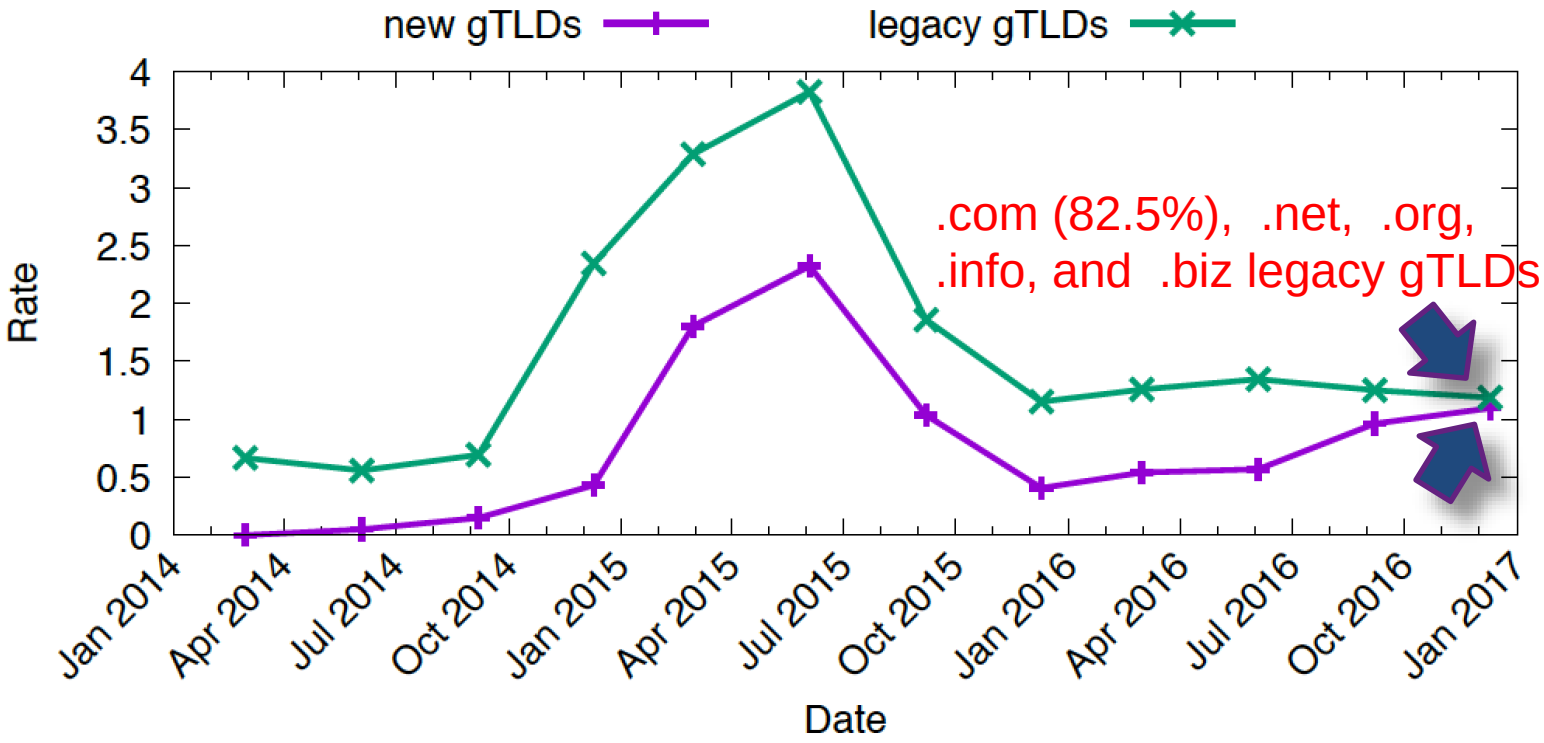
Abuse Rates

– Time series of abuse rates of **phishing** domains in legacy gTLDs and new gTLDs based on the APWG feed



Abuse Rates

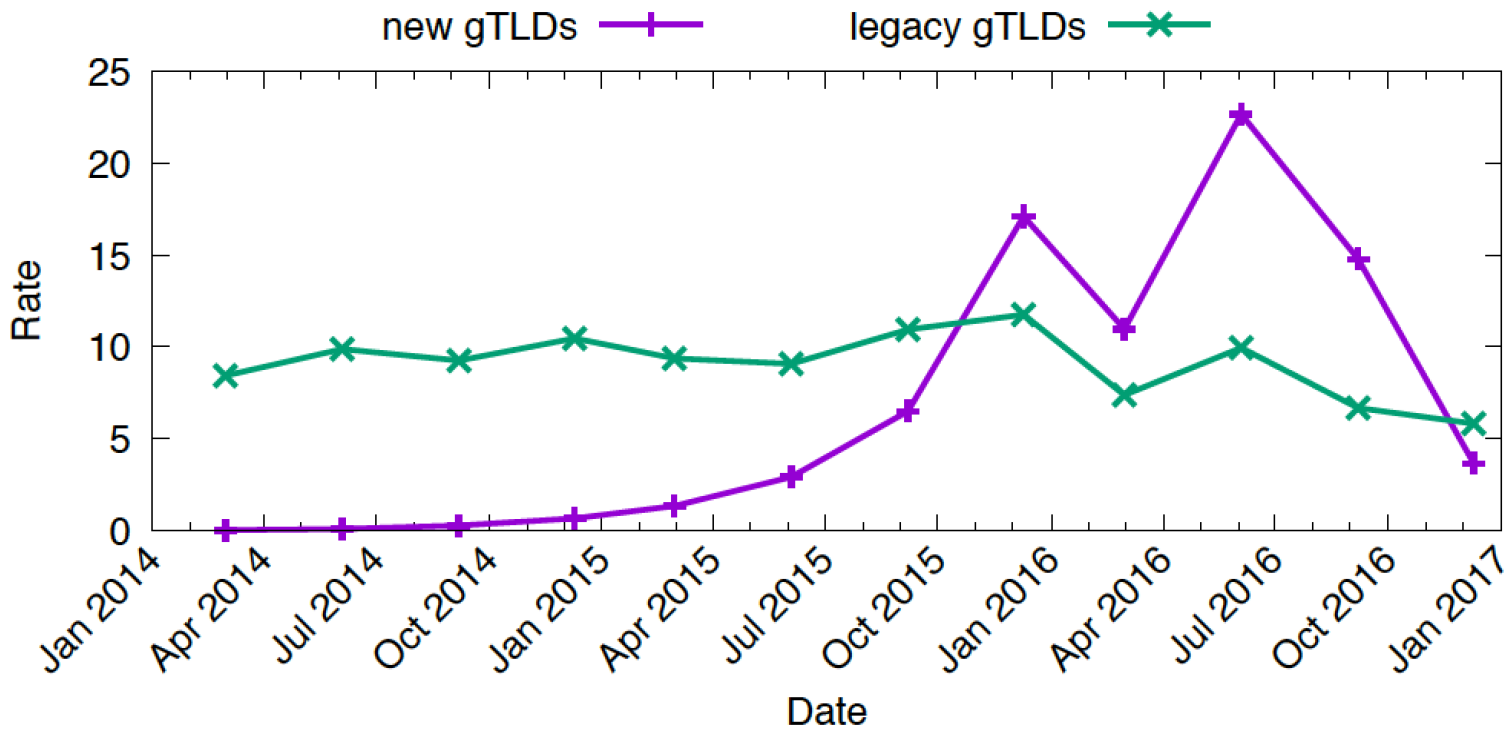
– Time series of abuse rates of **phishing** domains in legacy gTLDs and new gTLDs based on the APWG feed



Top 5 most abused new gTLDs collectively owned 58.7% of all blacklisted domains in all new gTLDs

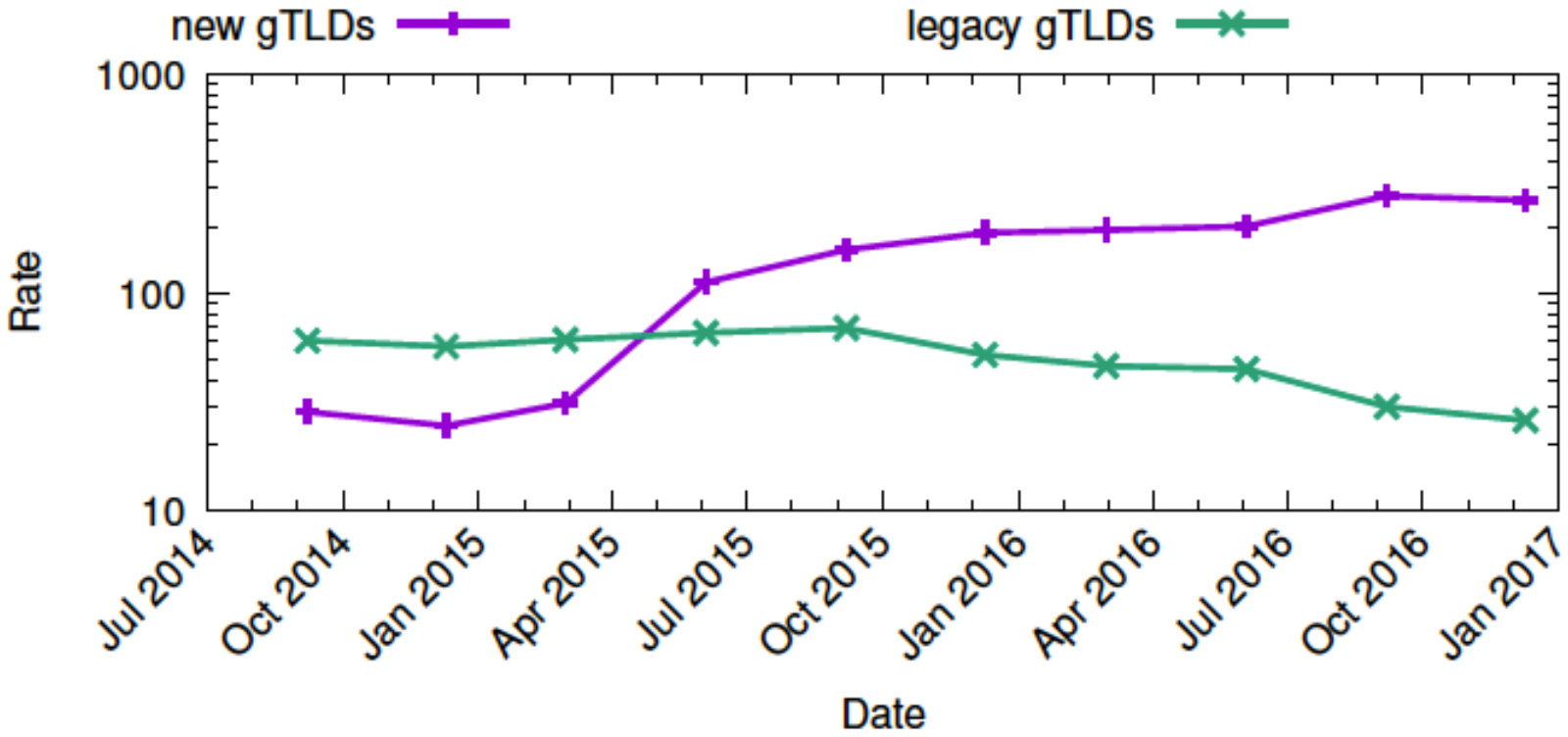
Abuse Rates

– Time series of abuse rates of **malware** domains in legacy gTLDs and new gTLDs based on the StopBadware feed



Abuse Rates

– Time series of abuse rates of **spam** domains in legacy gTLDs and new gTLDs based on the Spamhaus feed



Abuse Rates

– Top 10 new gTLDs with the highest relative concentrations of blacklisted domains for SURBL and Spamhaus datasets (4Q 2016)

Spamhaus

TLD	# Domains	Rate
SCIENCE	117,782	5,154
STREAM	18,543	4,756
STUDY	1,118	3,343
DOWNLOAD	16,399	2,016
CLICK	20,713	1,814
TOP	736,339	1,705
GDN	45,547	1,602
TRADE	23,581	1,521
REVIEW	9415	1,318
ACCOUNTANT	6,722	1,279

SURBL ws

TLD	# Domains	Rate
RACING	51,443	3,812
DOWNLOAD	21,515	2,645
ACCOUNTANT	10,543	2,007
REVIEW	12,615	1,766
GDN	49,427	1,739
FAITH	5,540	1,301
TRADE	19,330	1,247
CLICK	13,270	1,162
STREAM	4,406	1,130
DATE	1,3851	999

– Rates: (#blacklisted domains / #all domains) * 10,000

Abuse Rates

– Does the problem affect all new gTLDs?

Abuse Rates

- Does the problem affect all new gTLDs?
- No

Abuse Rates

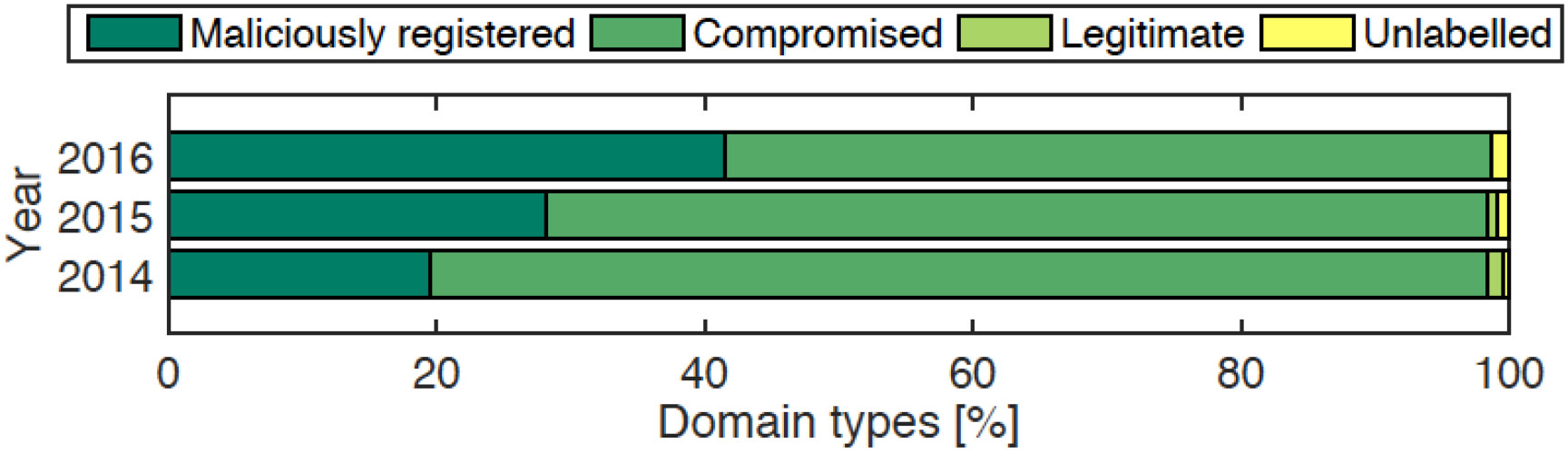
- Does the problem affect all new gTLDs?
- No
- Spamhaus and SURBL blacklists reveal that 32% and 36% of all new gTLDs available for registration did not experience a single incident in 4Q 2016.
- Spamhaus blacklisted at least 10% of all registered domains in as many as 15 new gTLDs in 4Q 2016.

Compromised and Maliciously Registered Domains

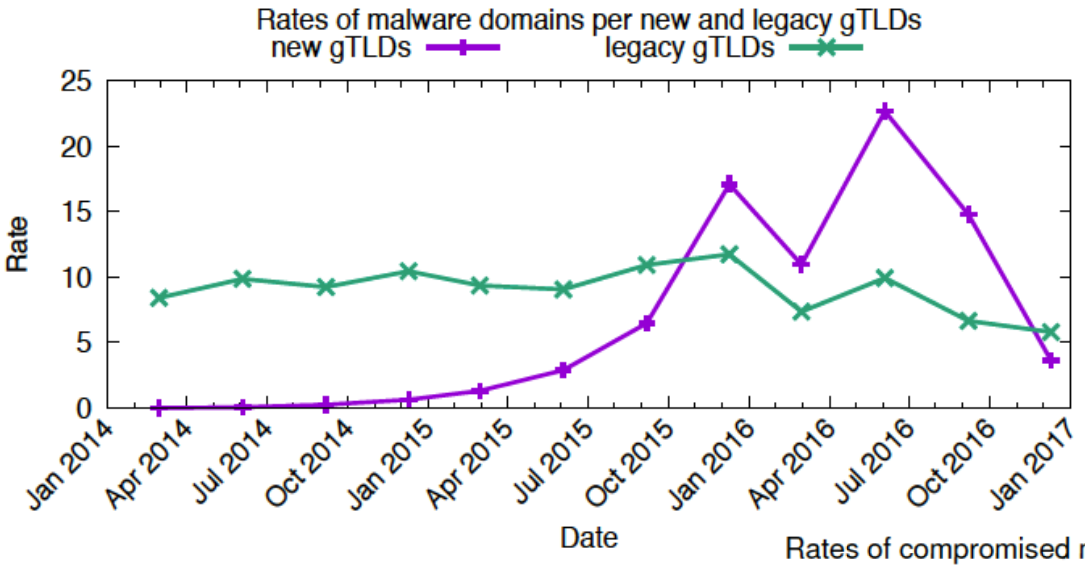
- Distinguishing between compromised and maliciously registered domains is critical because they require different mitigation actions by different intermediaries
- Three heuristics:
  - if a given domain name contains a string of a brand name, or
  - if its misspelled version, or
  - if it's involved in malicious activity within three months after creation.

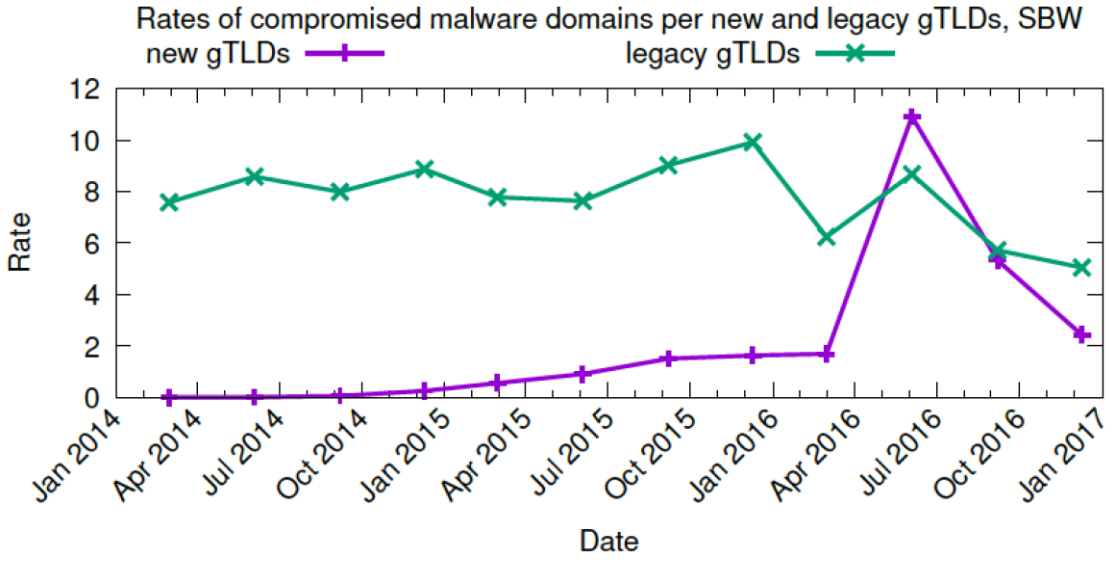
Compromised and Maliciously Registered Domains

– Distinguishing between compromised and maliciously registered domains is critical because they require different mitigation actions by different intermediaries

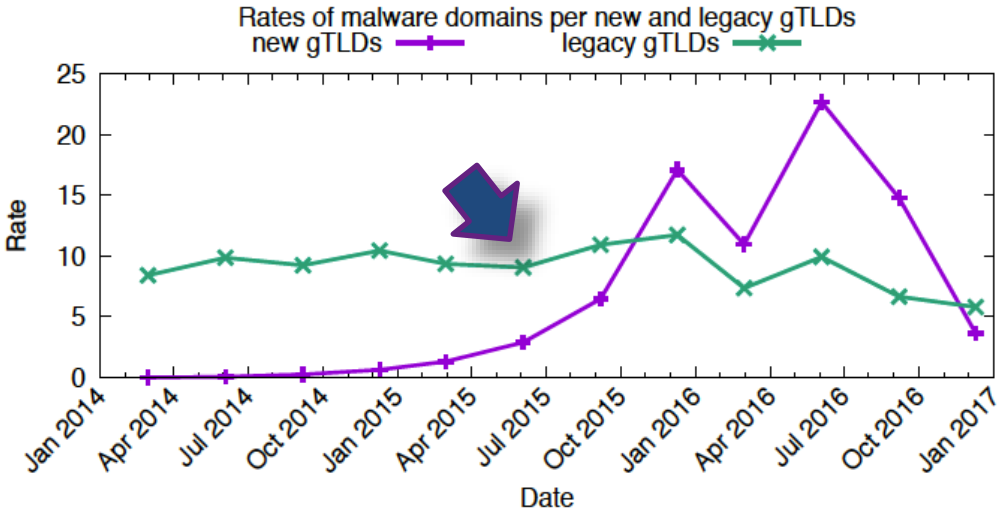


Compromised Domains

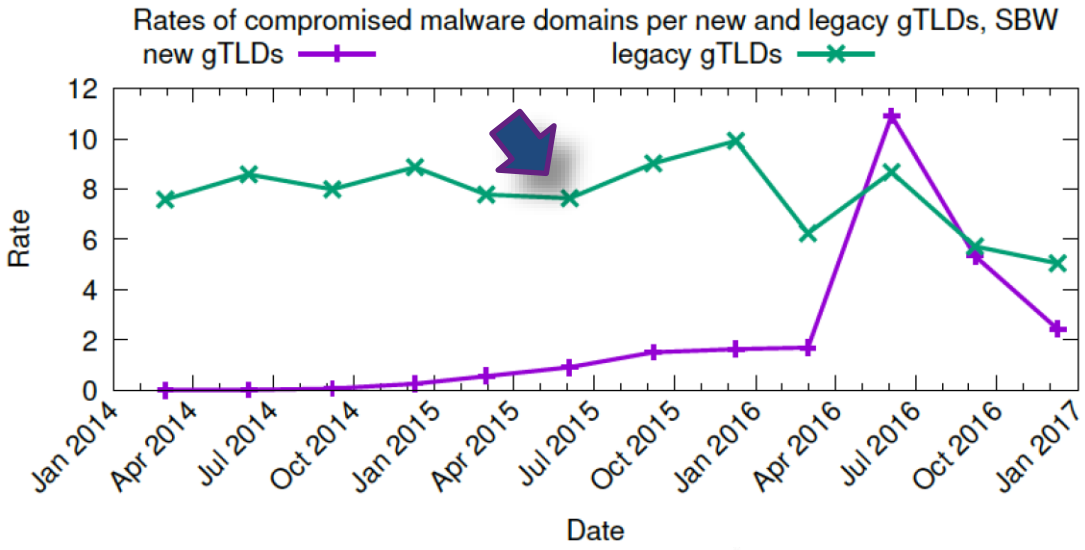




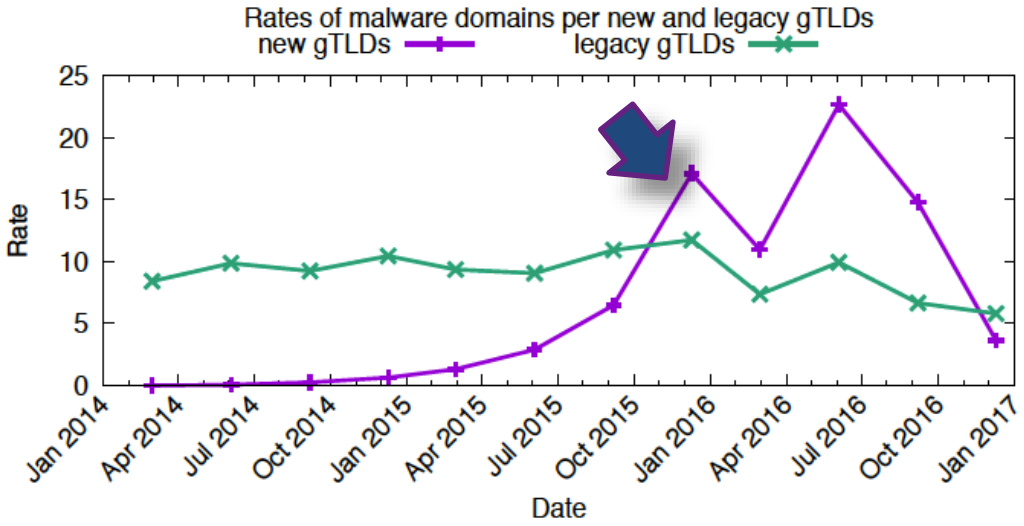
Compromised Domains



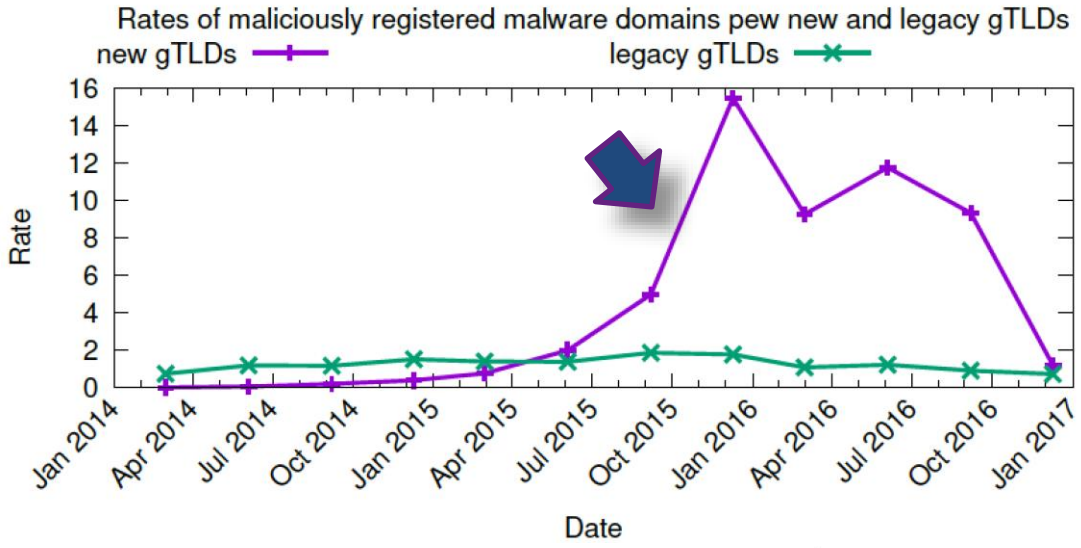
- Rates of abused domains in legacy gTLDs (StopBadware URL blacklists) are driven by compromised domains



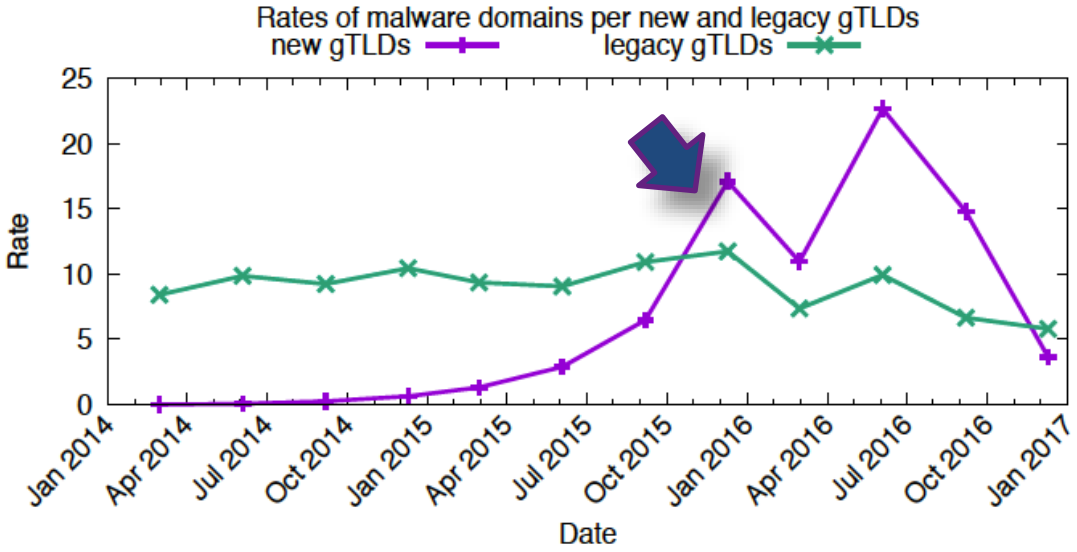
Maliciously Registered Domains



- Rates of abused domains in new gTLDs (StopBadware URL blacklist) are driven by maliciously registered domains

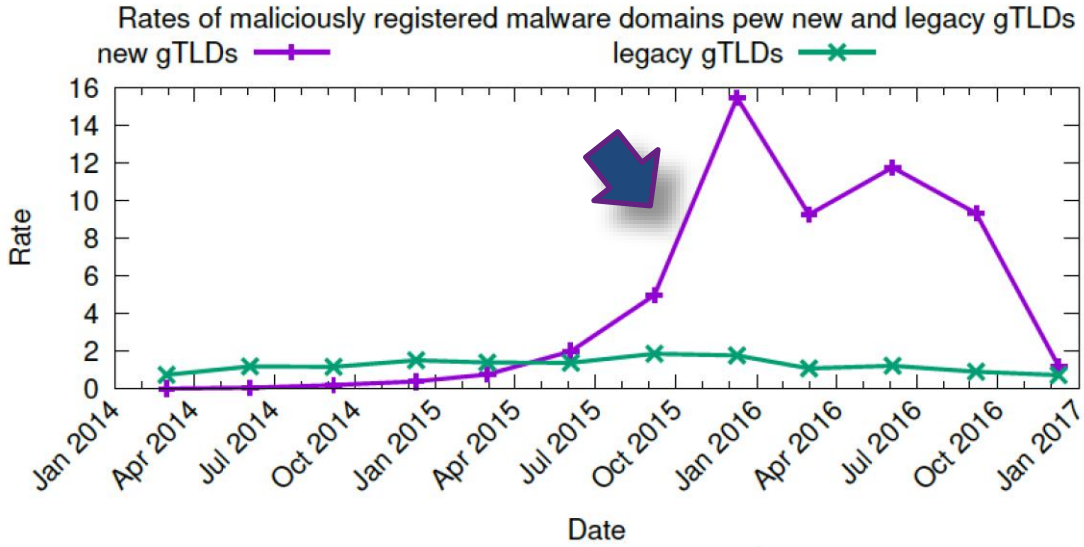


Maliciously Registered Domains



– Rates of abused domains in new gTLDs (StopBadware URL blacklist) are driven by maliciously registered domains

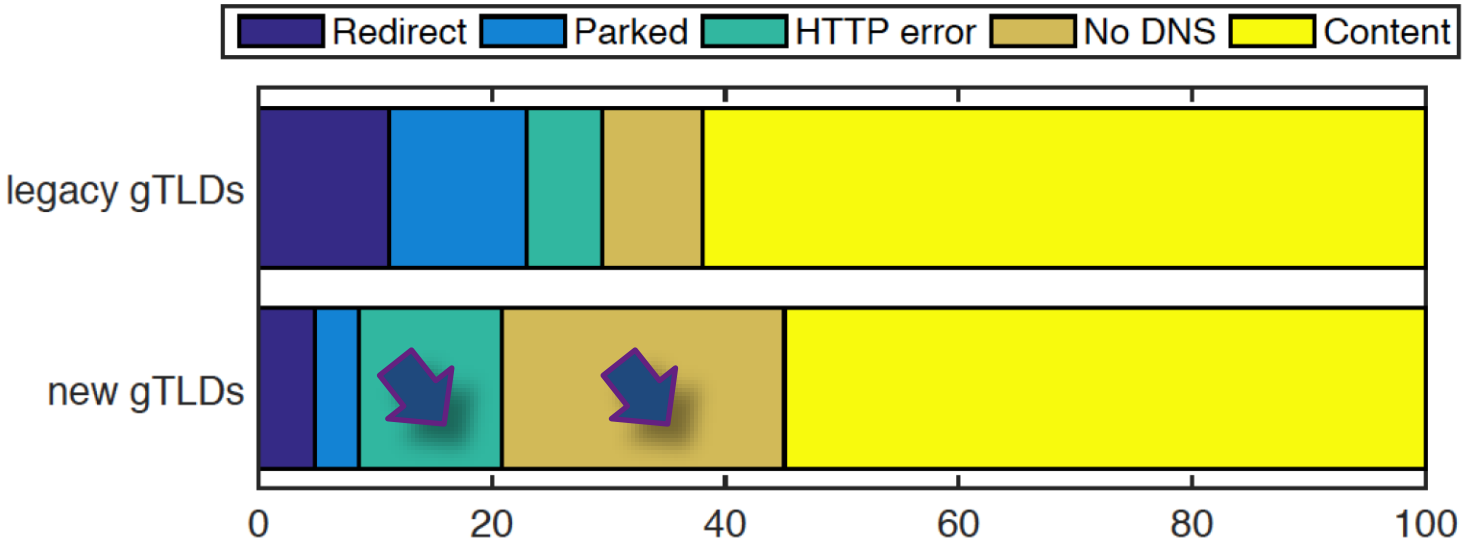
…and can be driven by single campaigns (domains registered in bulk, common patterns in domain names)



Inferential Analysis of Abuse in New gTLDs

Driver	Rationale
New gTLD size	Larger TLDs have a larger "attack surface" (compromised domains)
DNSSEC	Hypothesis: proxy for security efforts, however, miscreants could be interested in deploying DNSSEC and signing their maliciously registered domains
Parked	Domains serving content are exposed to certain types of vulnerabilities and can be hacked. However, parked domains may be used to scam users or to distribute malware
No DNS, HTTP error	Domains serving content are exposed to certain types of vulnerabilities and can be hacked
Type	Proxy for strict registration policies (registration "levels" to new gTLDs, from the least to most restricted groups: 1 generic, 2 geographic, 3 community, and 4 brand)
Registry operator (parent companies of registry operators)	Proxy for registration practices (e.g. pricing, registration in bulk, payment methods)

Inferential Analysis of Abuse in New gTLDs



"No DNS" domains account for 24.2% of all domains, whereas domains for which the websites serve an HTTP error account for another 12.2%.

Inferential Analysis of Abuse in New gTLDs

Driver	Correlation with abuse counts
New gTLD size	Very weak positive
DNSSEC	Very weak positive
Parked	Very weak positive
No DNS	Very weak negative
HTTP Error	Very weak negative
Type	Negative (statistically significant results for phishing)
Registry operator	No statistically significant results

Privacy or Proxy Services

- Why use Privacy and Proxy services
  - Protecting your personal data
  - Blocking Spam
  - Stopping unwanted solicitations
- Analyzing use of Privacy and Proxy
  - Extract list of registrants
  - keyword search using “privacy”, “proxy”, “protect” etc.
  - Manual inspection
- How many?
  - We found 570

Privacy or Proxy Services

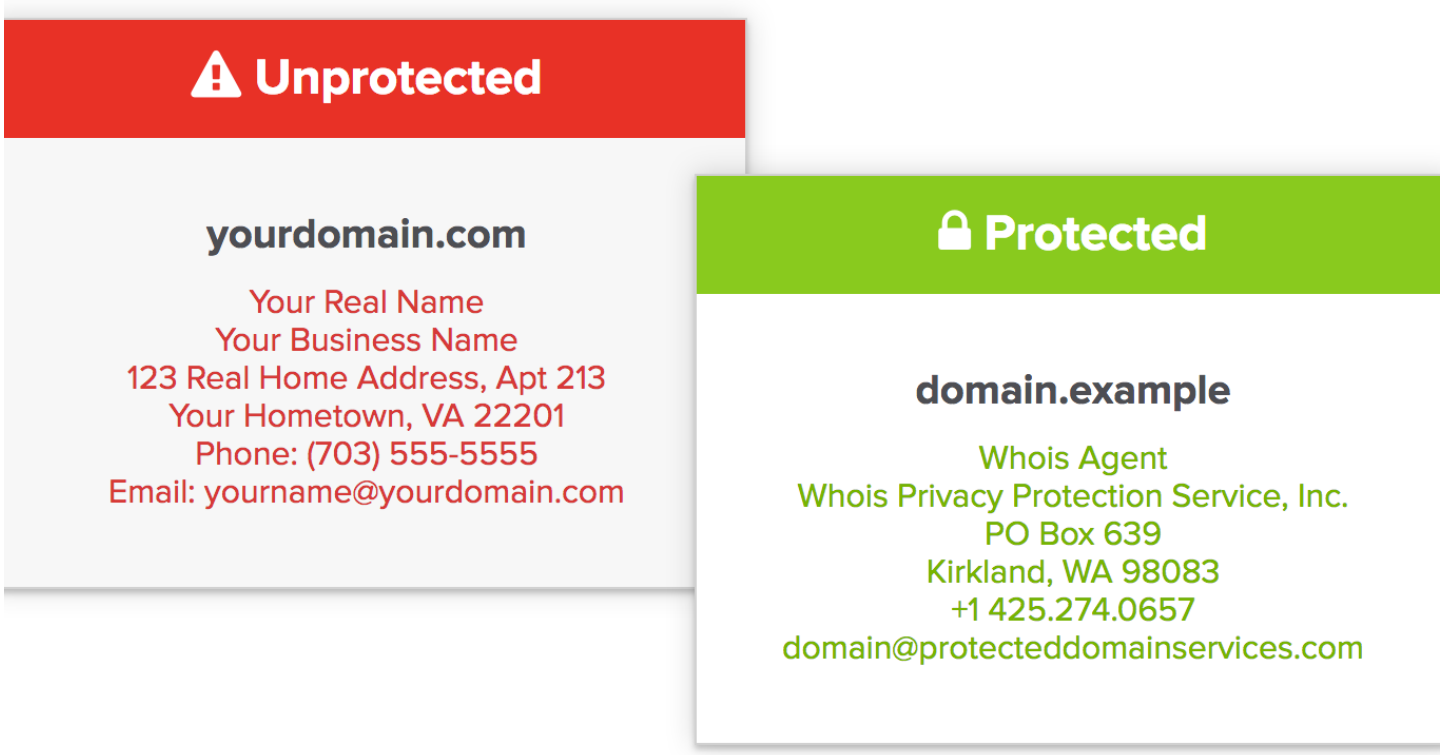
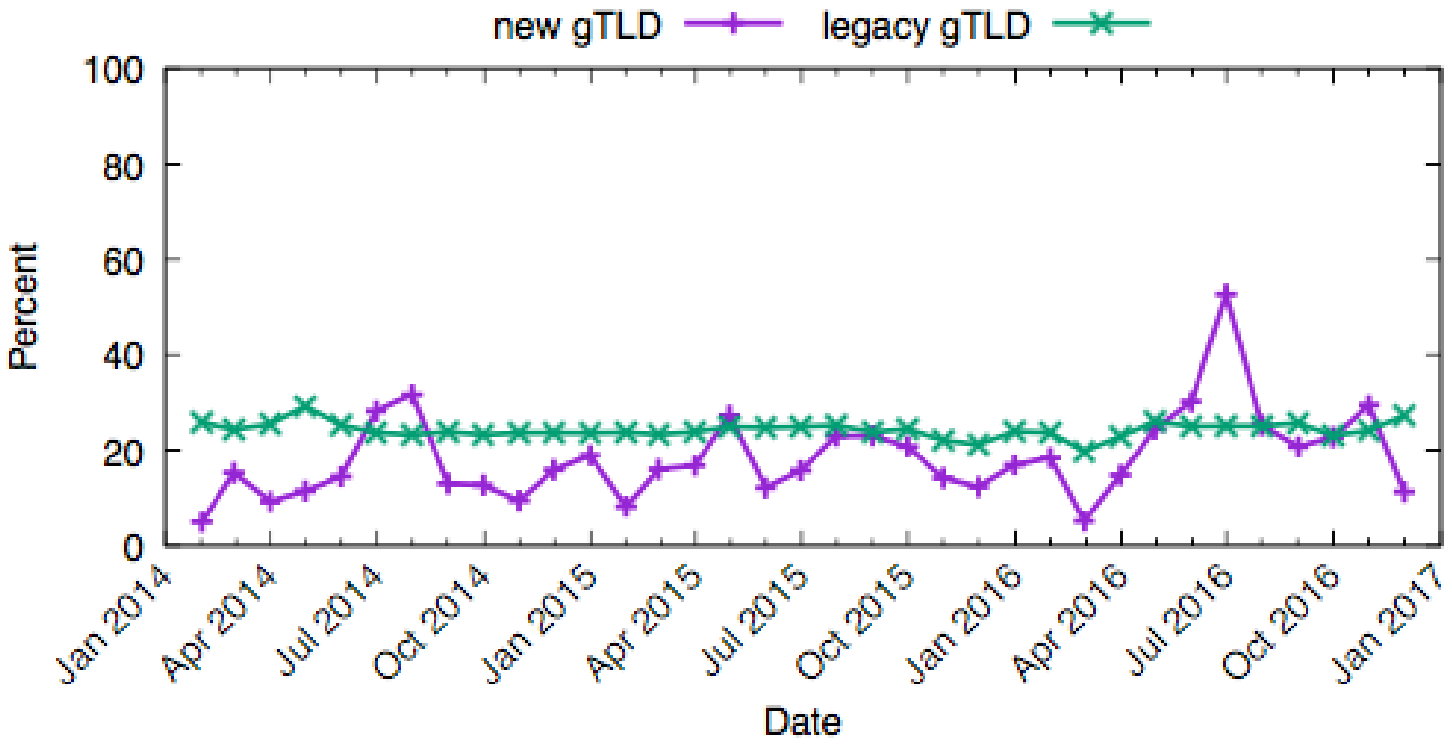


Image source: https://www.name.com/whois-privacy

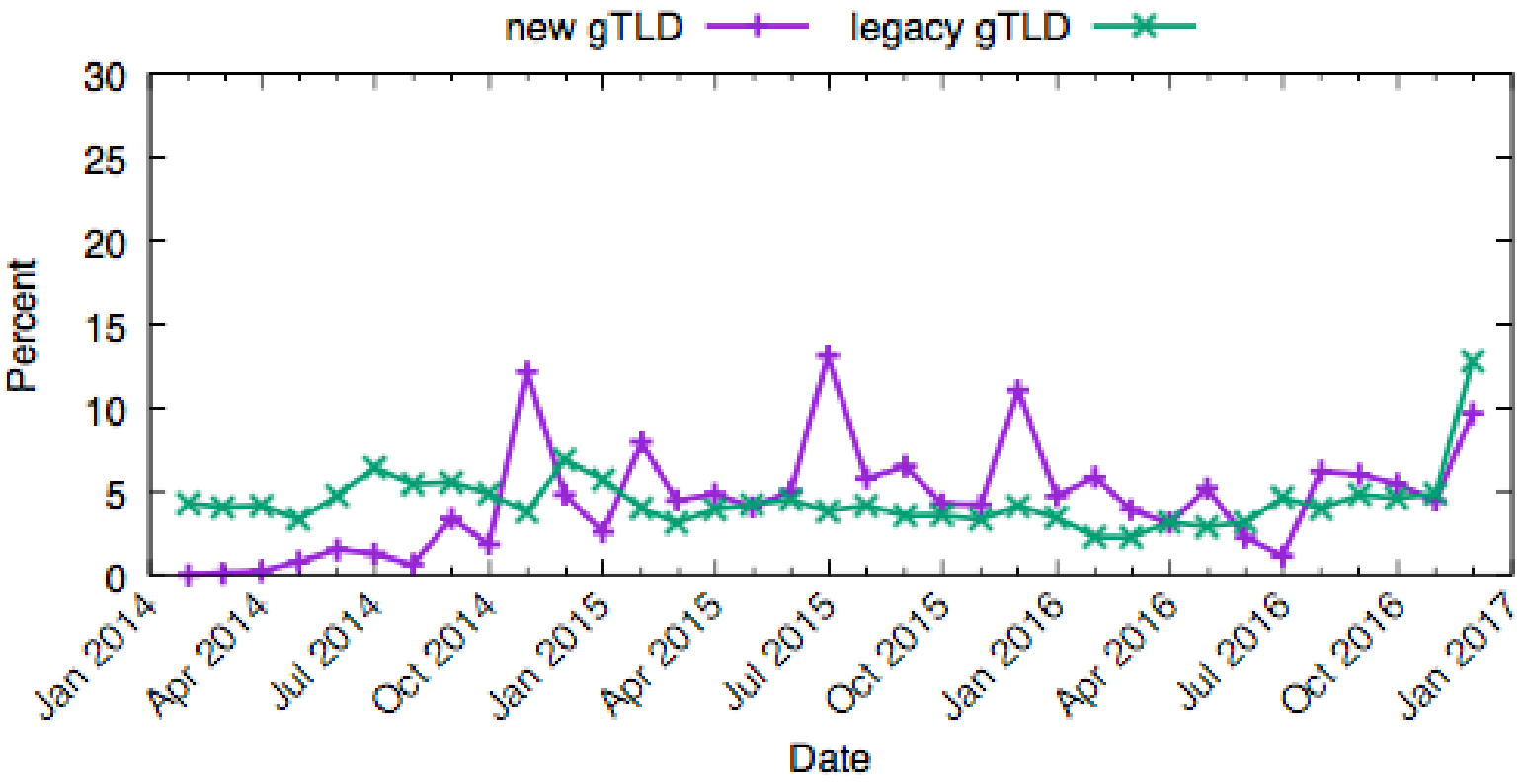
Privacy or Proxy Services

Usage for Newly Created Domains



Privacy or Proxy Services

Usage for Abusive Newly Registered Domains



Privacy or Proxy Services

- The usage of Privacy or Proxy Services by itself is not a reliable indicator of abuse.
- Usage of Privacy or Proxy Services remains higher for legacy gTLDs.

Geographical Location

- Using domain registrar location from WHOIS
  - Registrant details not reliable

- Method
  - Extract unique "registrar name" from WHOIS data.
  - Combine the registrar name with the country information for ICANN-Accredited Registrars.
  - Match remaining name variants
  - Manually lookup the country information for missing registrars

- Result
  - 5,985 registrars
  - 99.99% of domains

Geographical Location

Registrar Distribution

Country	#Registrars	share
United States	2,682	53.88
China	281	5.64
Germany	201	4.04
Canada	177	3.56
United Kingdom	160	3.21
India	144	2.89
France	116	2.33
Australia	111	2.23
Spain	105	2.11
Japan	95	1.91

Geographical Location

Domain Distribution

New	#Domains	Share	Legacy	#Domains	Share
China	8,076,776	27.92	US	152,527,872	56.72
US	6,283,269	21.72	China	24,098,150	8.96
Gibraltar	3,028,035	10.47	Germany	18,044,735	6.71
Cayman Is.	2,069,919	7.16	Canada	16,704,693	6.21
Singapore	1,870,886	6.47	India	11,135,408	4.14
Japan	1,741,228	6.02	Japan	7,935,585	2.95
India	1,323,117	4.57	Australia	6,425,896	2.39
Germany	1,105,708	3.82	France	4,988,581	1.86
Hong Kong	836,069	2.89	UK	4,511,714	1.68
France	450,371	1.56	Turkey	2,418,232	0.9

Geographical Location

SURBL Distribution

New gTLD Country	#Incidents	Percentage	Rate
Gibraltar	751,748	49.44	2482.63
Japan	295,647	19.44	976.37
China	214,332	14.1	707.83
United States	109,989	7.23	363.24
India	54,782	3.6	180.92
United Kingdom	24,955	1.64	82.41
France	20,121	1.32	66.45
United Arab Emirates	11,793	0.78	38.95
Cayman Islands	8,912	0.59	29.43
Canada	6,494	0.43	21.45

Legacy gTLD Country	#Incidents	Percentage	Rate
United States	1,985,574	47.06	130.18
Japan	1,190,409	28.21	78.05
China	319,546	7.57	20.95
India	268,812	6.37	17.62
Germany	73,185	1.73	4.8
Ireland	58,292	1.38	3.82
Canada	40,355	0.96	2.65
Australia	33,080	0.78	2.17
Turkey	32,266	0.76	2.12
Bahamas	28,918	0.69	1.9

Registrar Reputation

- Method
  - Filter out registrars designed for sinkholing domains.
  - Count number of incidents per registrar.
  - Calculate percentage of total abuse linked to registrar.

Registrar Reputation

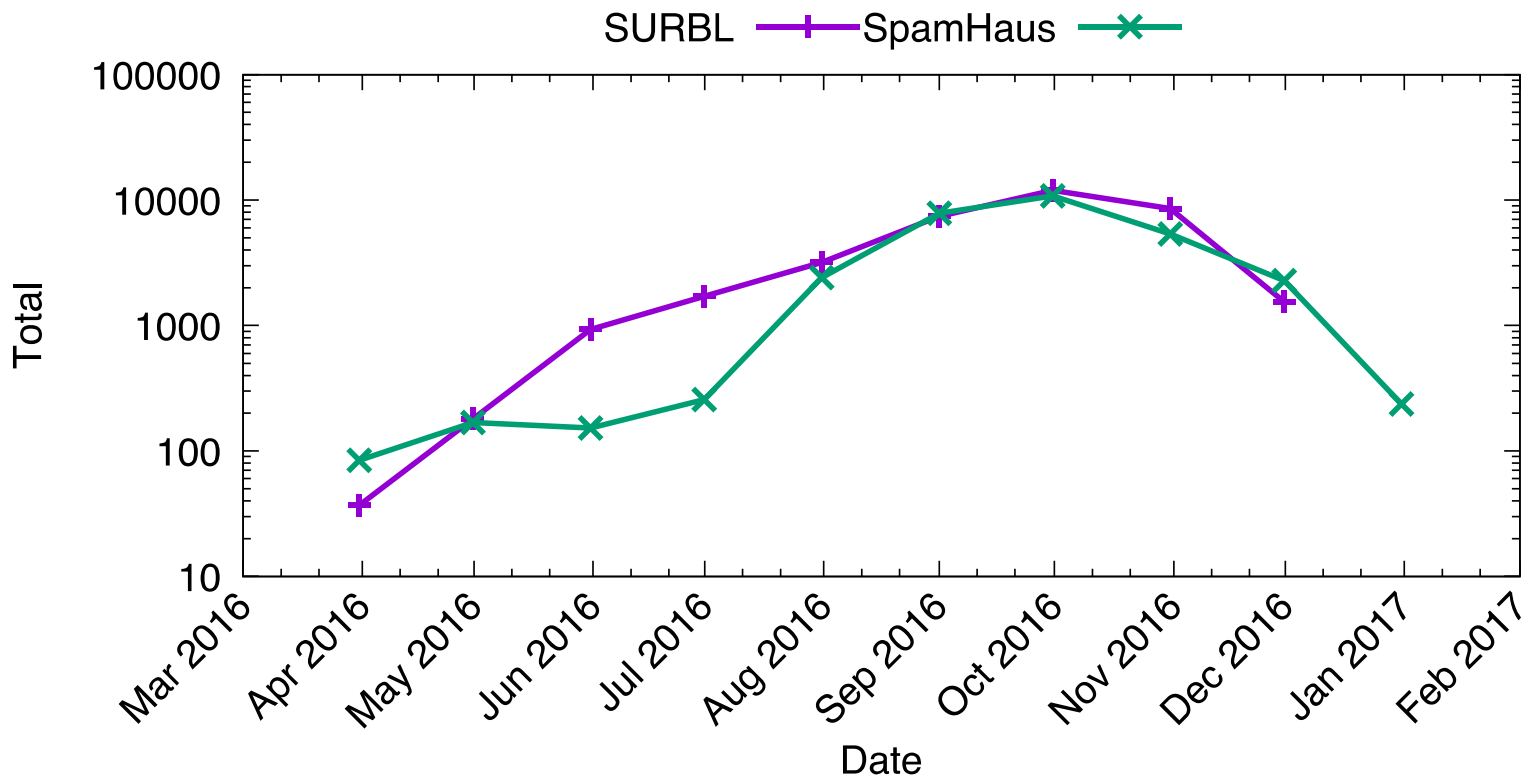
SURBL Distribution

new gTLD registrar	#Domains	#Incidents	Percent
Nanjing Imperiosus Technology	38,025	35,502	93.36
Intracom Middle East FZE	20,640	11,255	54.53
Dot Holding Inc.	153	76	49.67
Alpnames Limited	3,028,011	751,748	24.83
Todaynic.com, Inc.	329,399	69,404	21.07
Web Werks India Pvt. Ltd	785	146	18.6
GMO Internet, Inc. d/b/a Onamae.com	1,734,775	295,641	17.04
TLD Registrar Solutions Ltd.	163,988	24,700	15.06
Xiamen Nawang Technology Co., Ltd	282,925	42,089	14.88
Instra Corporation Pty Ltd.	77,642	6,200	7.99

Legacy gTLD registrar	#Domains	#Incidents	Percent
HOAPDI INC.	141	126	89.36
asia registry r2-asia (700000)	1,379	598	43.36
Nanjing Imperiosus Technology	35,309	10,834	30.68
Paknic (Private) Limited	10,525	3,083	29.29
OwnRegistrar, Inc.	22,188	5,238	23.61
Eranet International Limited	6,109	1,339	21.92
BR domain Inc. dba namegear.co	847	158	18.65
Netlynx Inc.	17,612	3,030	17.2
AFRIREGISTER S.A.	1,551	266	17.15
GMO Internet, Inc. d/b/a Onamae.com	7,306,312	1,177,886	16.12

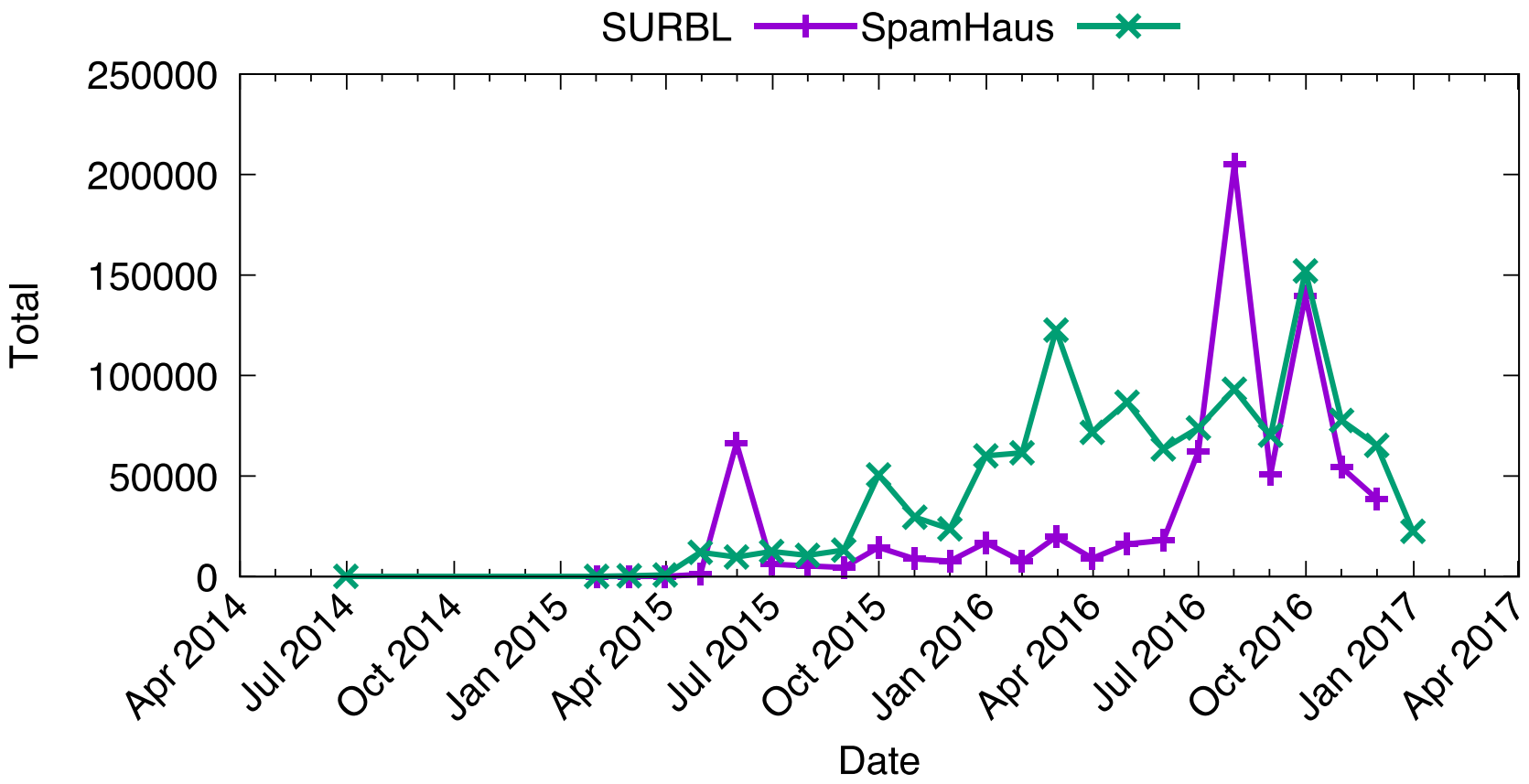
Registrar Reputation

Nanjing Imperiosus Technology Co. Ltd.



Registrar Reputation

Alpnames Ltd.



Questions?

Contact information

Maciej Korczyński
Grenoble INP - Grenoble Alps University
maciej.korczynski@univ-grenoble-alpes.fr

Maarten Wullink, SIDN Labs
maarten.wullink@sidn.nl