Project

Statistical Analysis of DNS Abuse in gTLDs (SADAG)

Consortium: SIDN and TU Delft

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

Goal

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

Motivation

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

Data Providers

Blacklists

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

Data Providers

Blacklists

- Spamhaus
  - Spam domains

- CleanMX (3 feeds)
  - Phishing URLs
  - Malware URLs
  - Defaced URLs

Data Providers

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

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)*

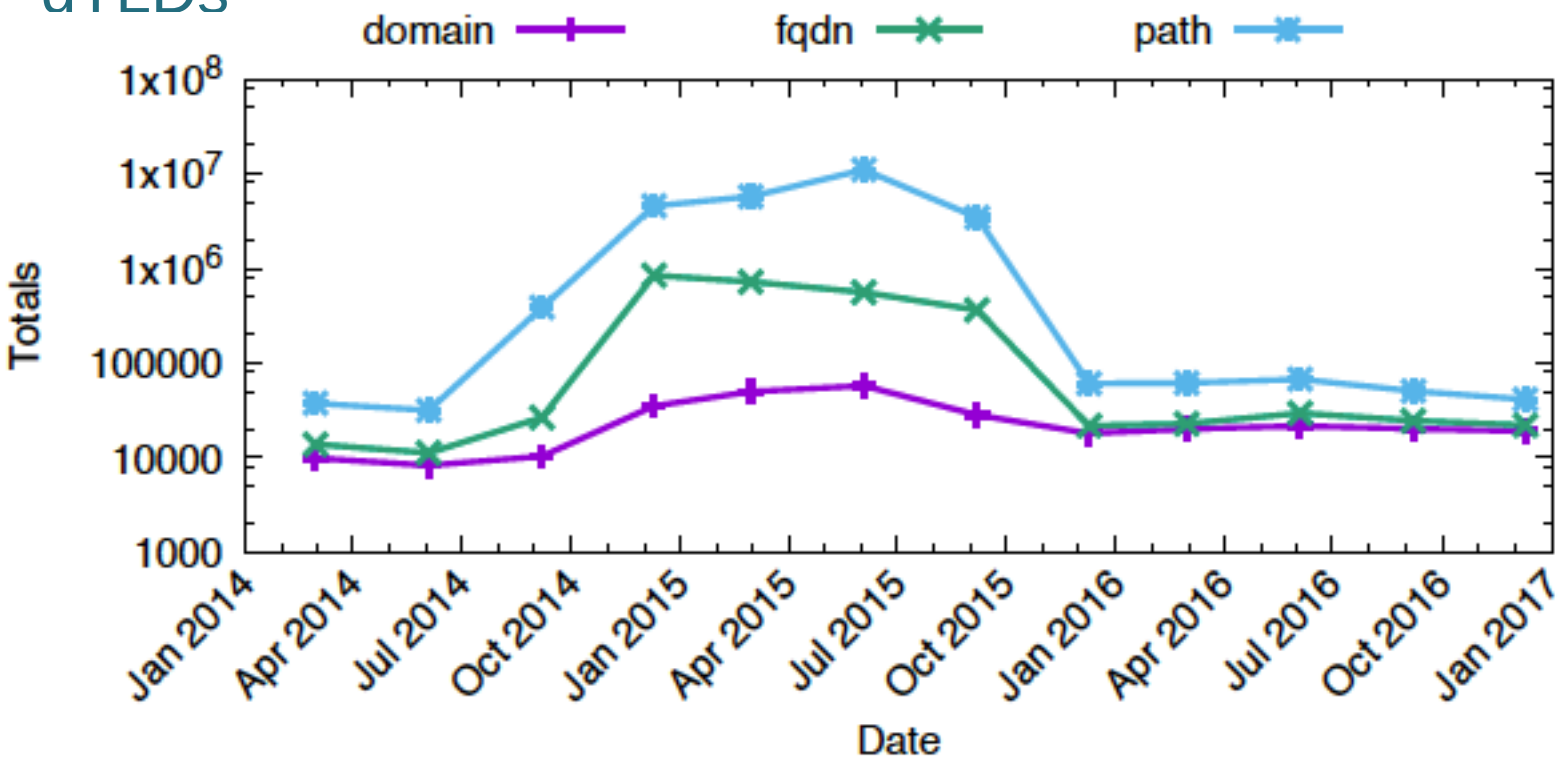
Security metrics

- 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)*

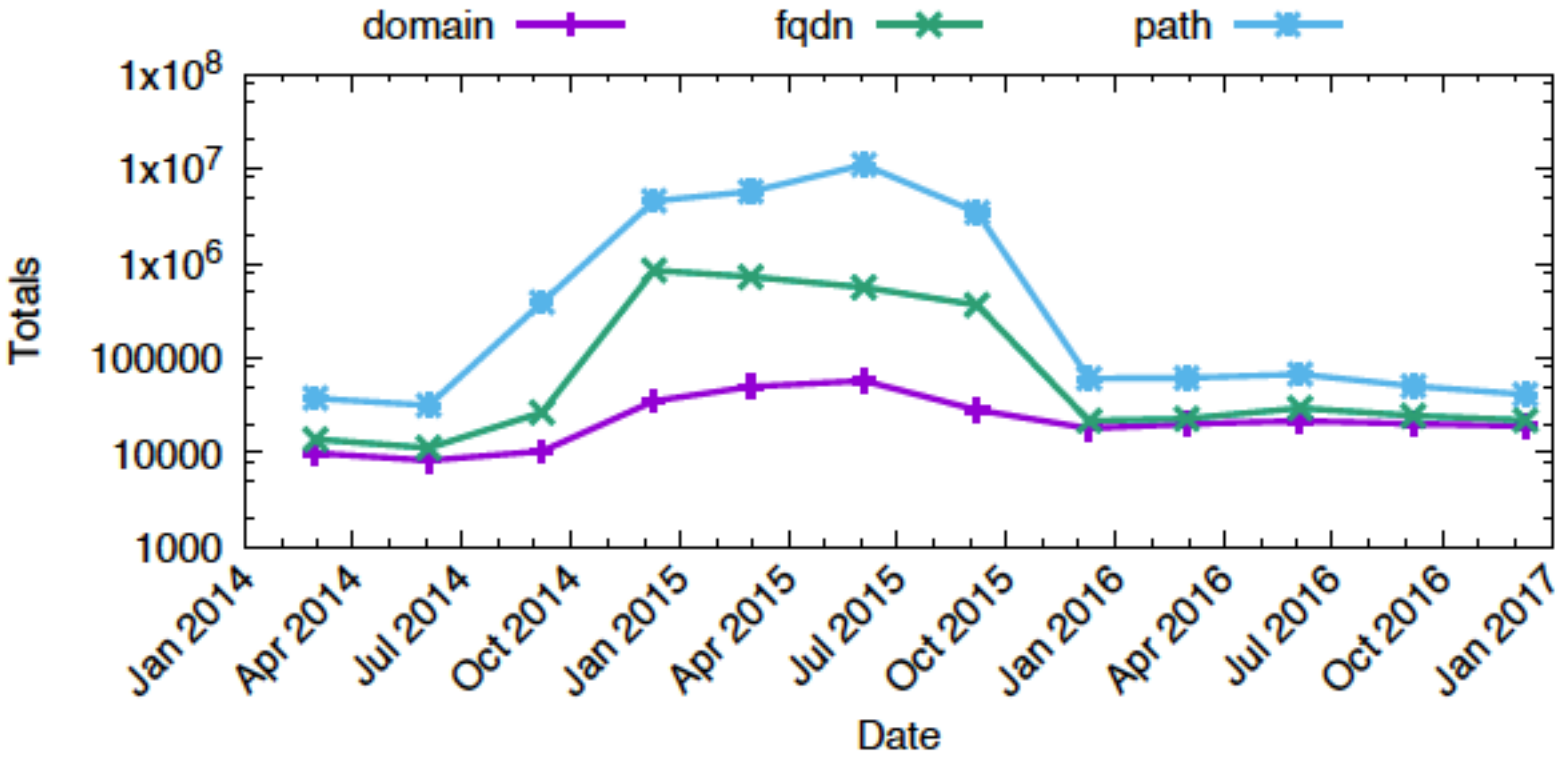
Security metrics for gTLDs

Phishing domains, FQDNs, and URLs (APWG) per lega
αTLDs



Security metrics for gTLDs

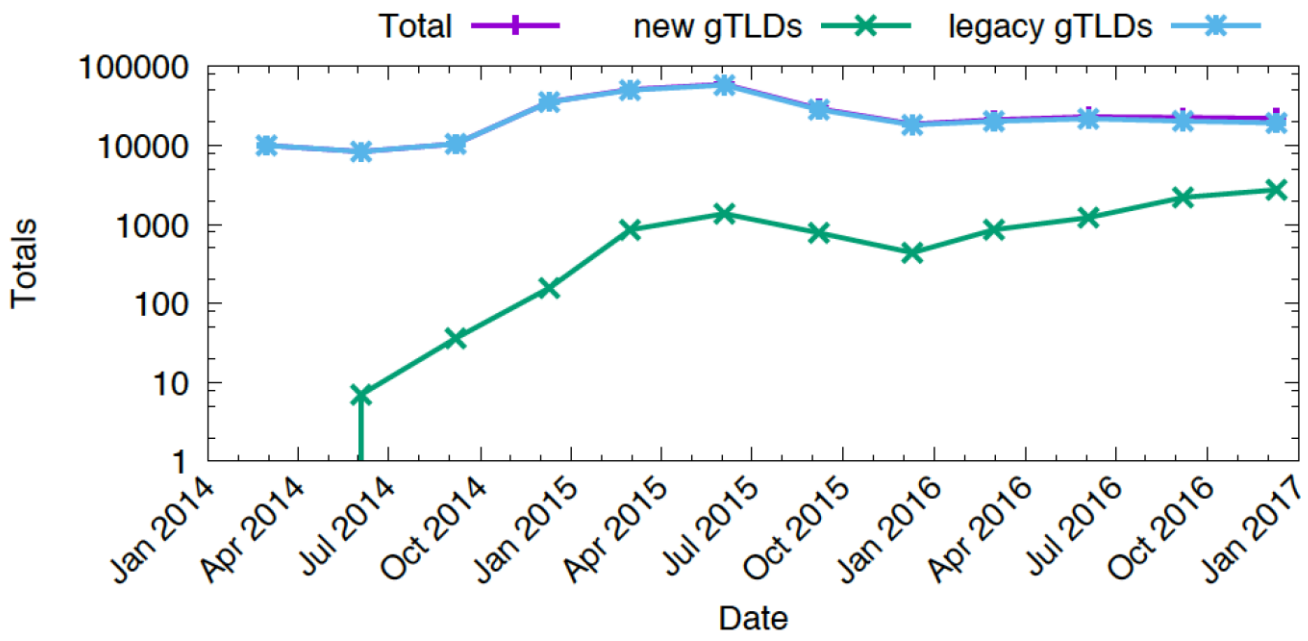
Phishing domains, FQDNs, and URLs (APWG) per legacy gTLDs



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

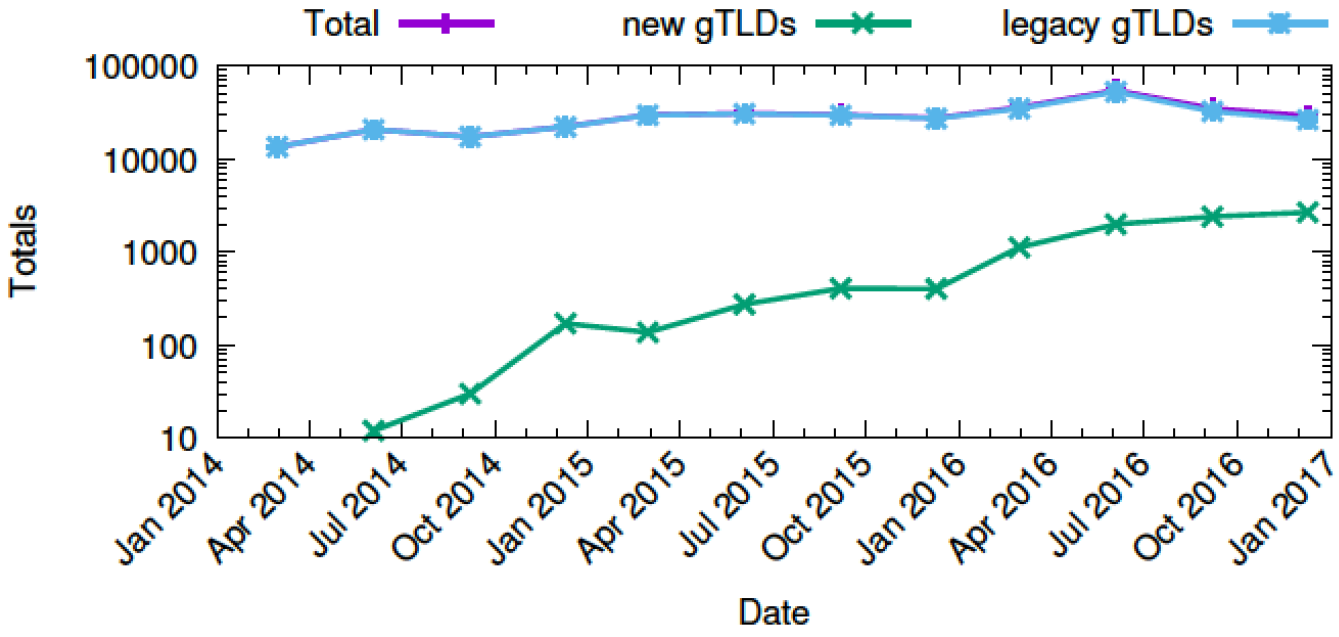
Security metrics for gTLDs

Phishing domains (APWG) per new and legacy gTLDs



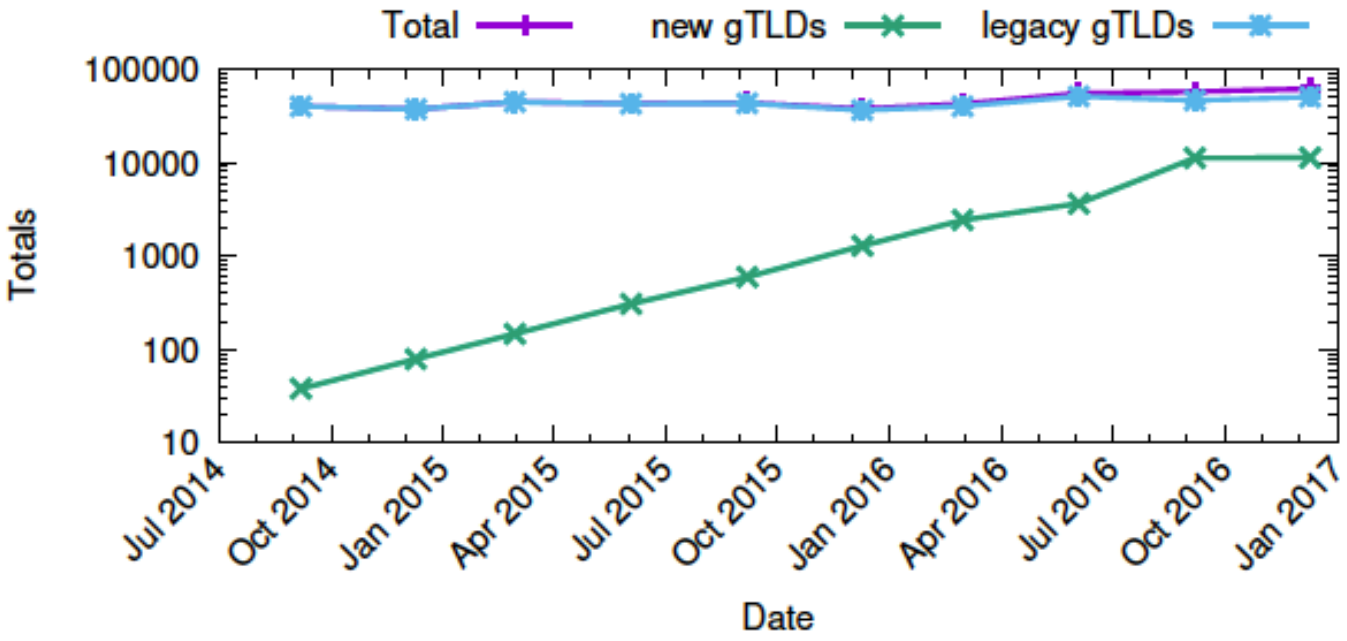
Security metrics for gTLDs

Phishing domains (CleanMX ph) per new and legacy gTLD



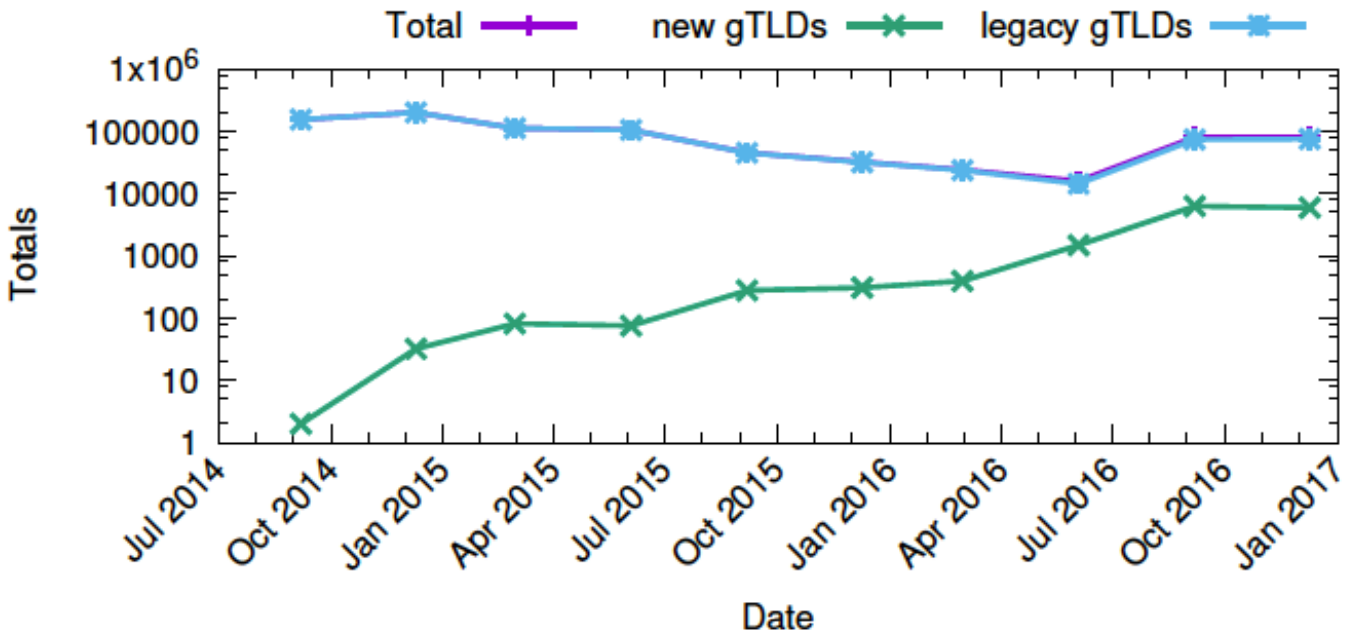
Security metrics for gTLDs

Phishing domains (SURBL ph) per new and legacy gTLDs



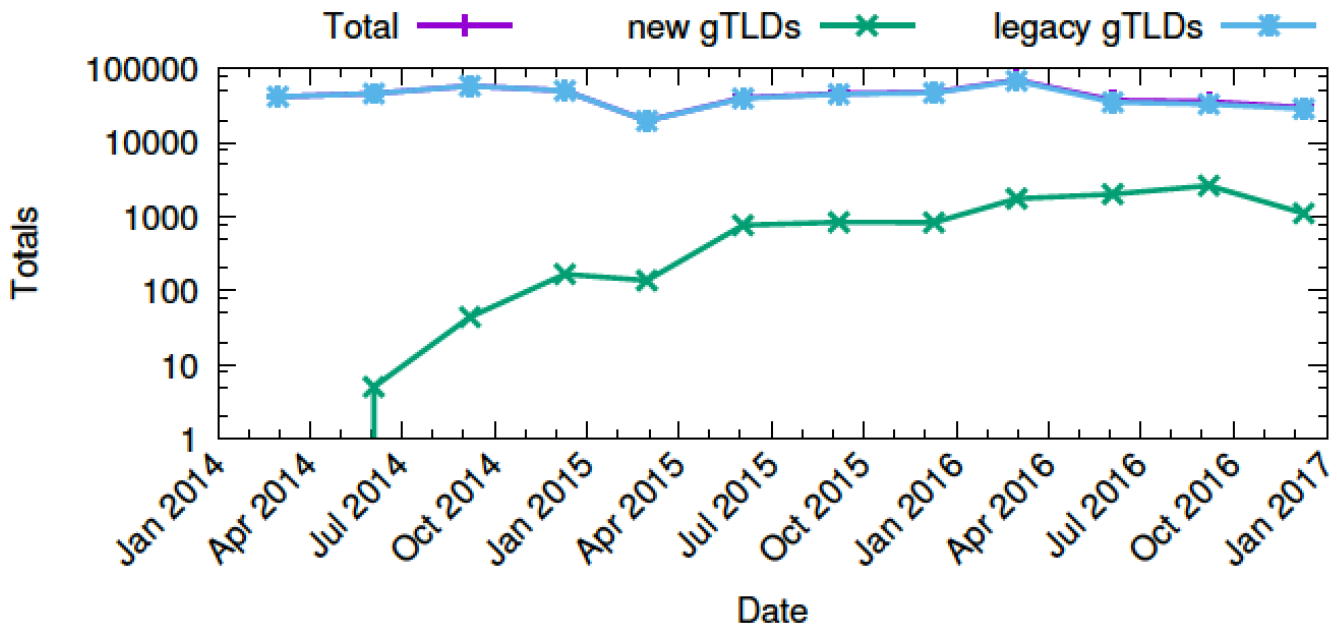
Security metrics for gTLDs

Malware domains (SURBL mw) per new and legacy gTLD



Security metrics for gTLDs

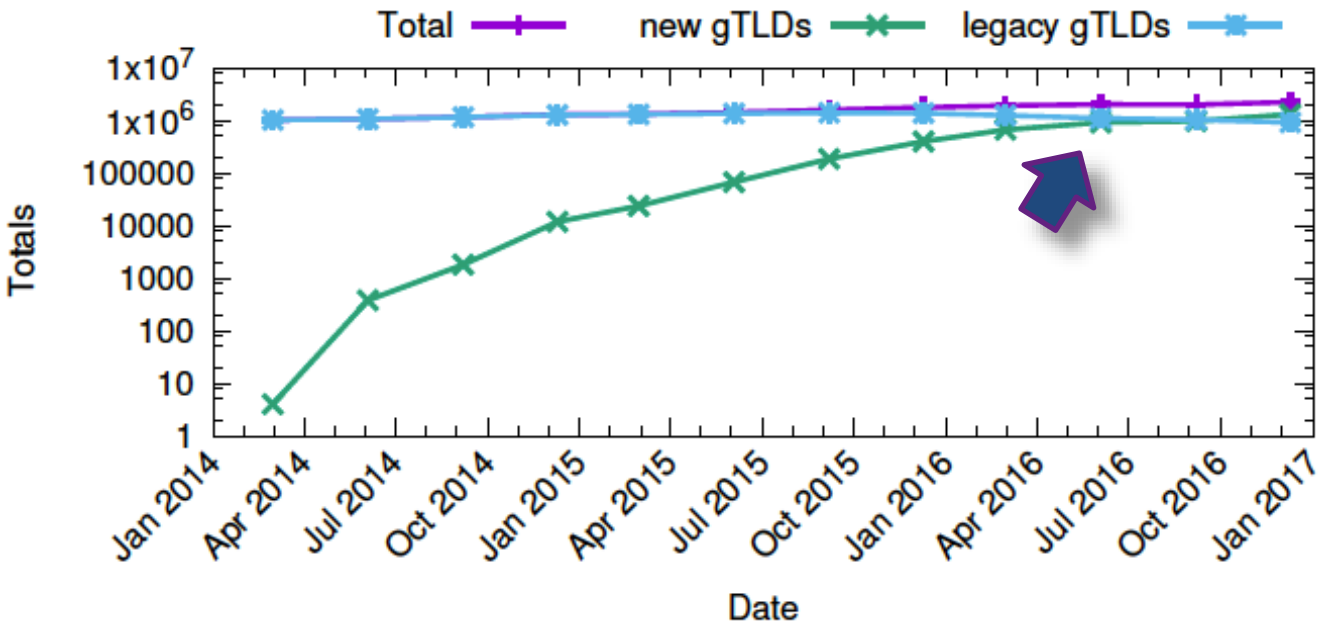
Malware domains (CleanMX mw) per new and legacy gTLDs



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.

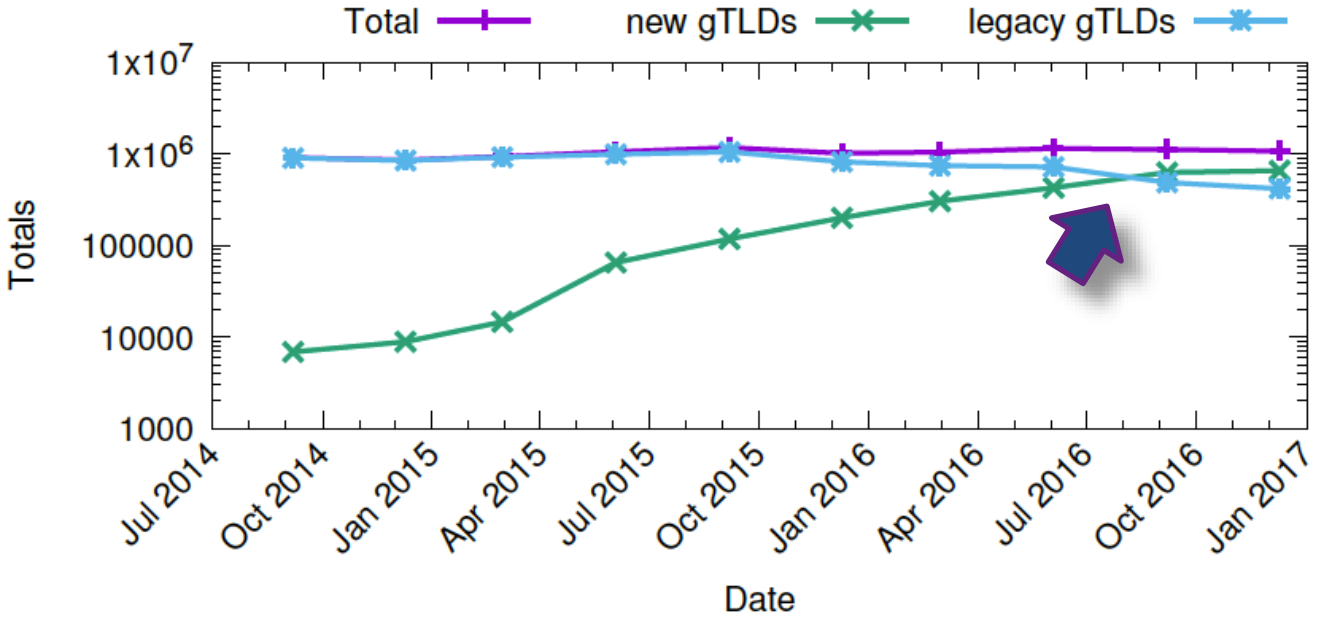
Security metrics for gTLDs

Spam domains (Spamhaus) per new and legacy gTLDs



Security metrics for gTLDs

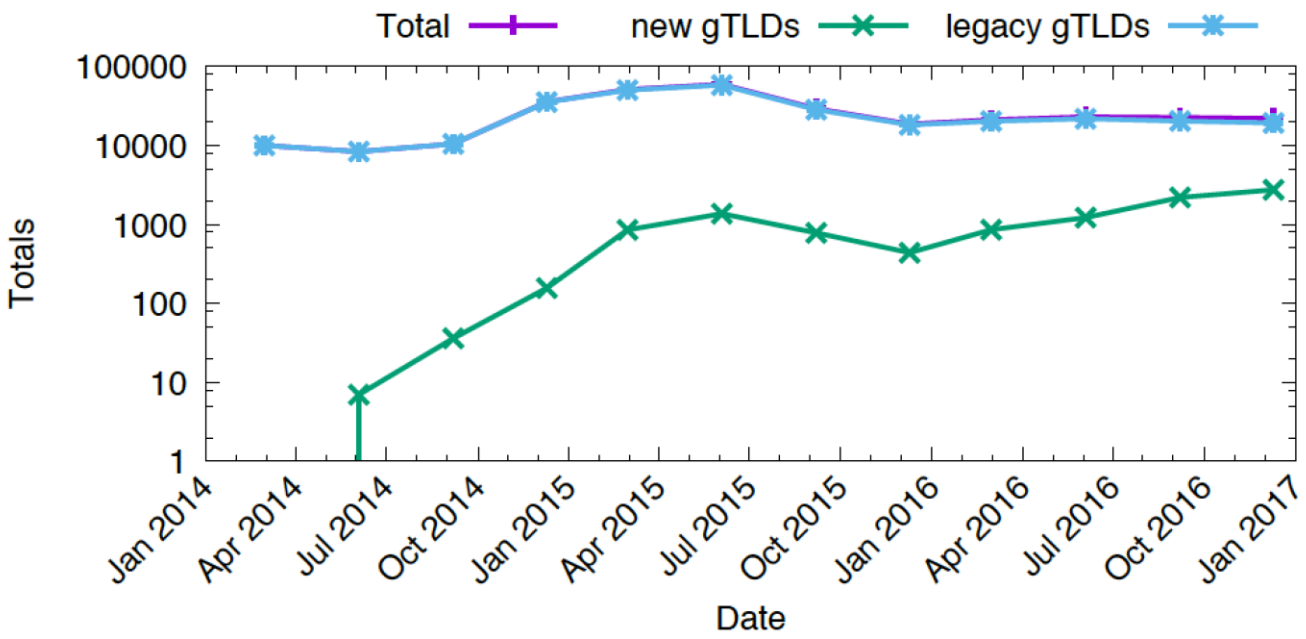
Spam domains (SURBL ws) per new and legacy gTLDs



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

Security metrics for gTLDs

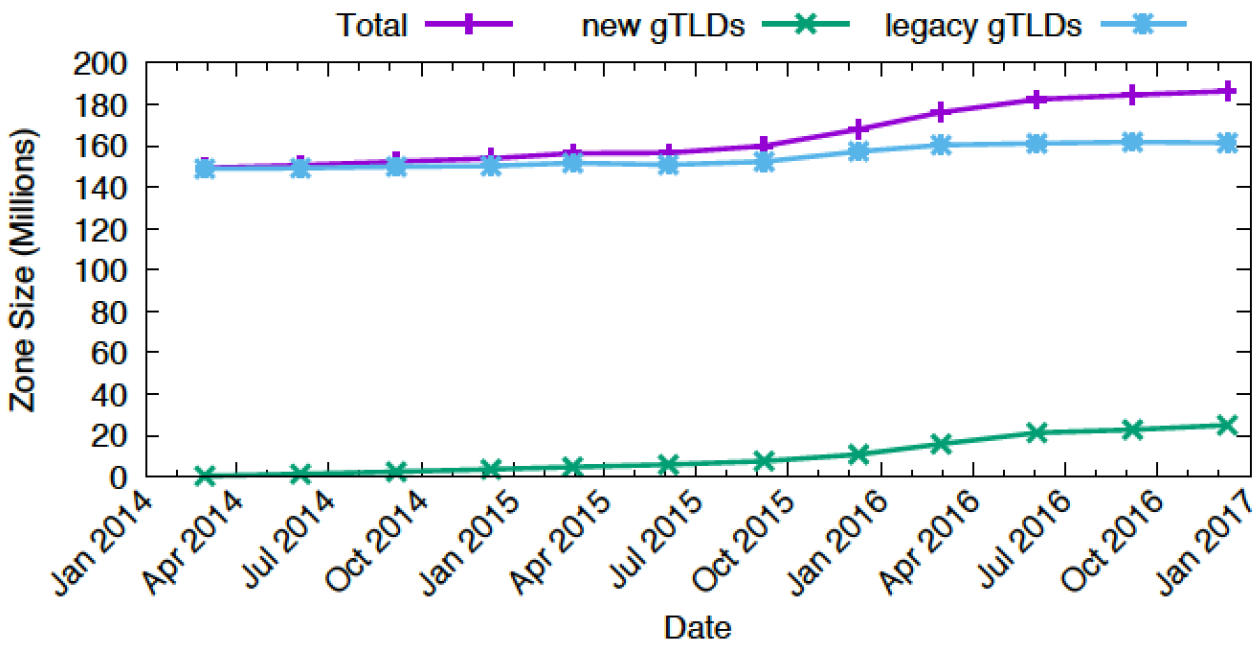
Phishing domains (APWG) per new and legacy gTLDs



– Size matters!

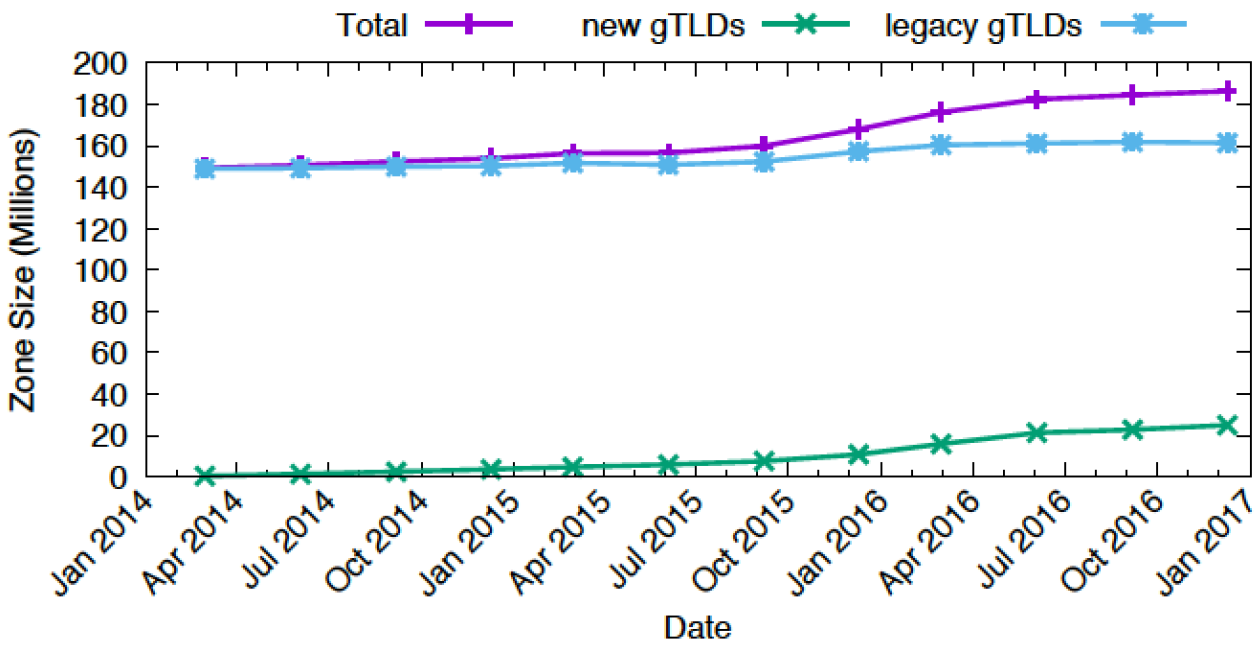
Size

– Size estimate: Number of 2nd–level domains in each gTLD zone file



Size

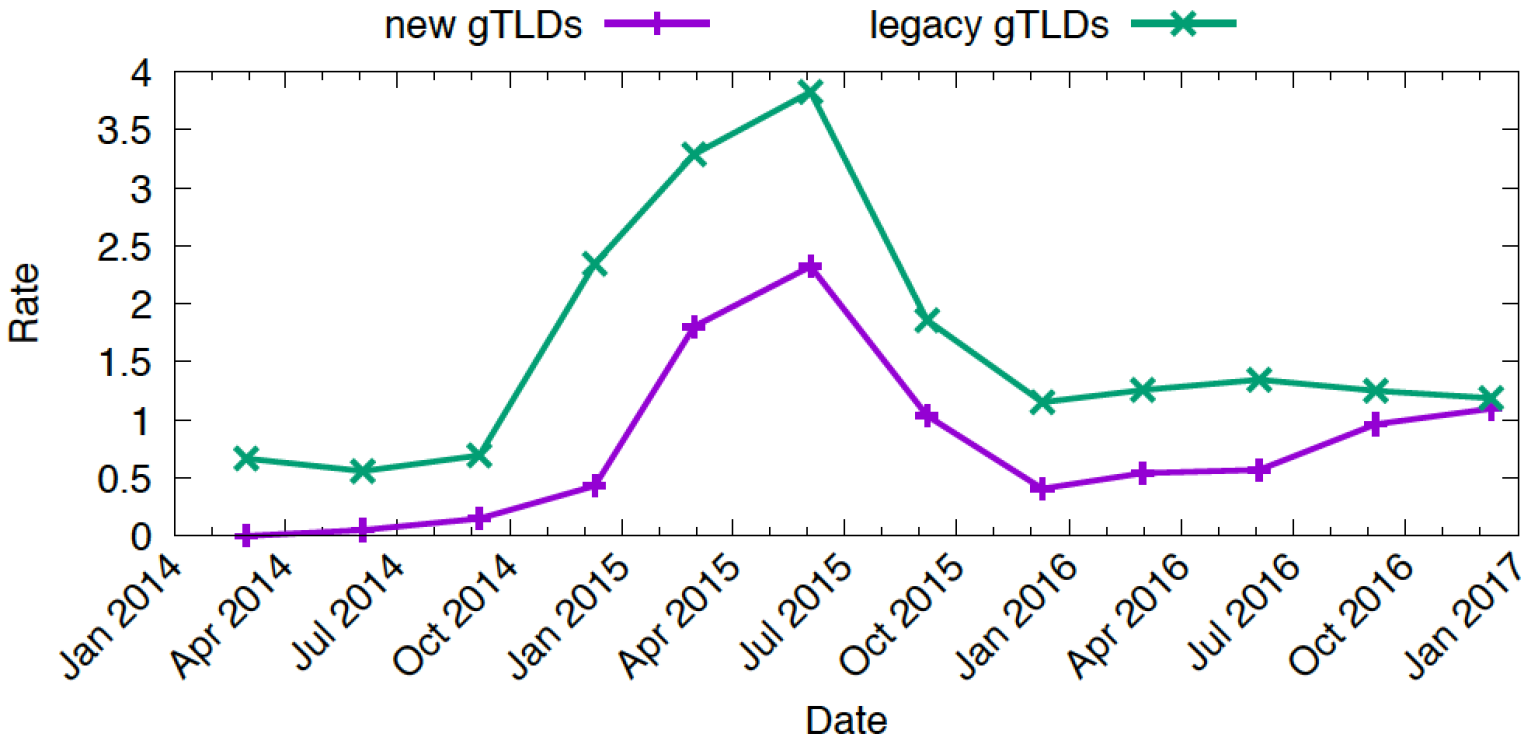
– Size estimate: Number of 2nd–level domains in each gTLD zone file



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

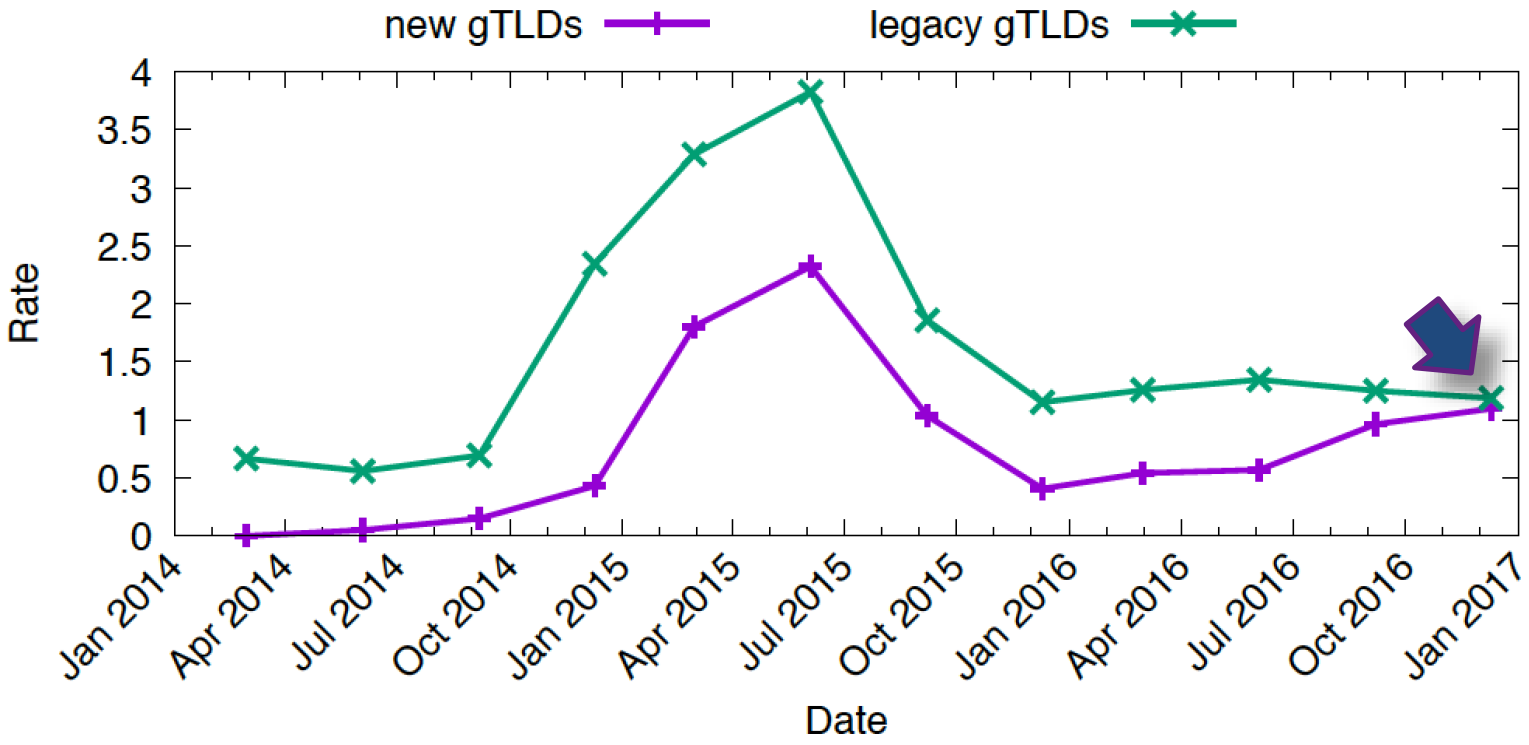
Abuse rates

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



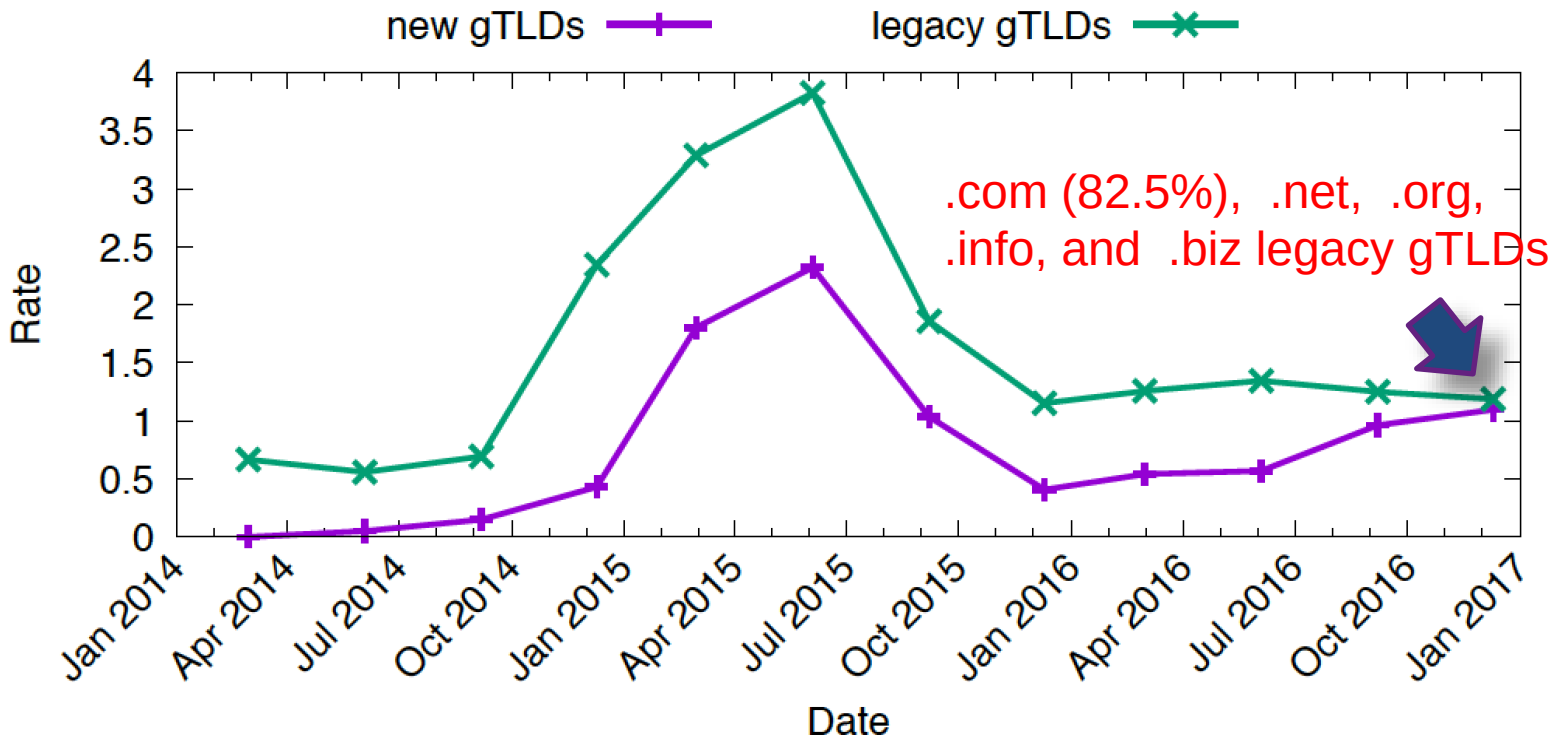
Abuse rates

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



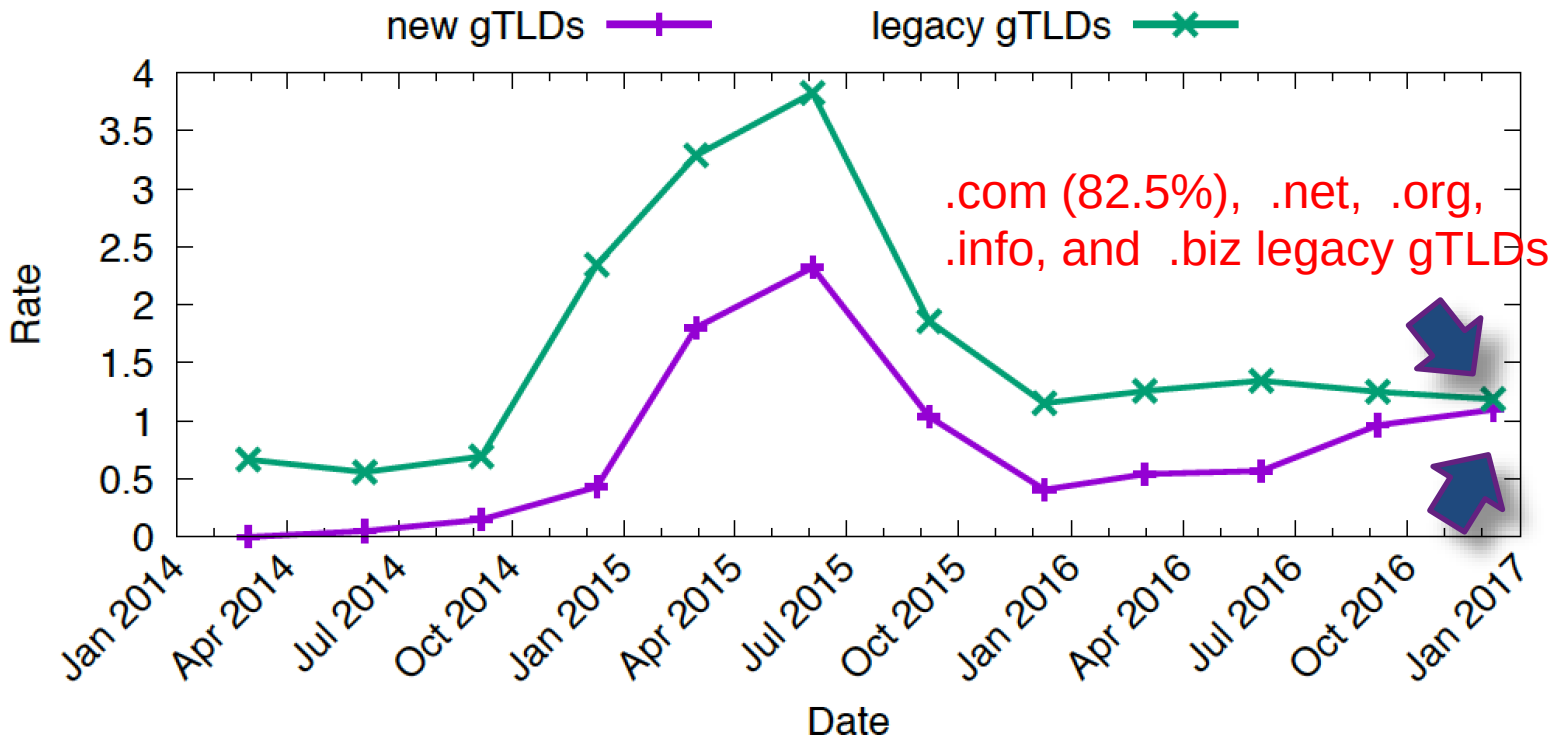
Abuse rates

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



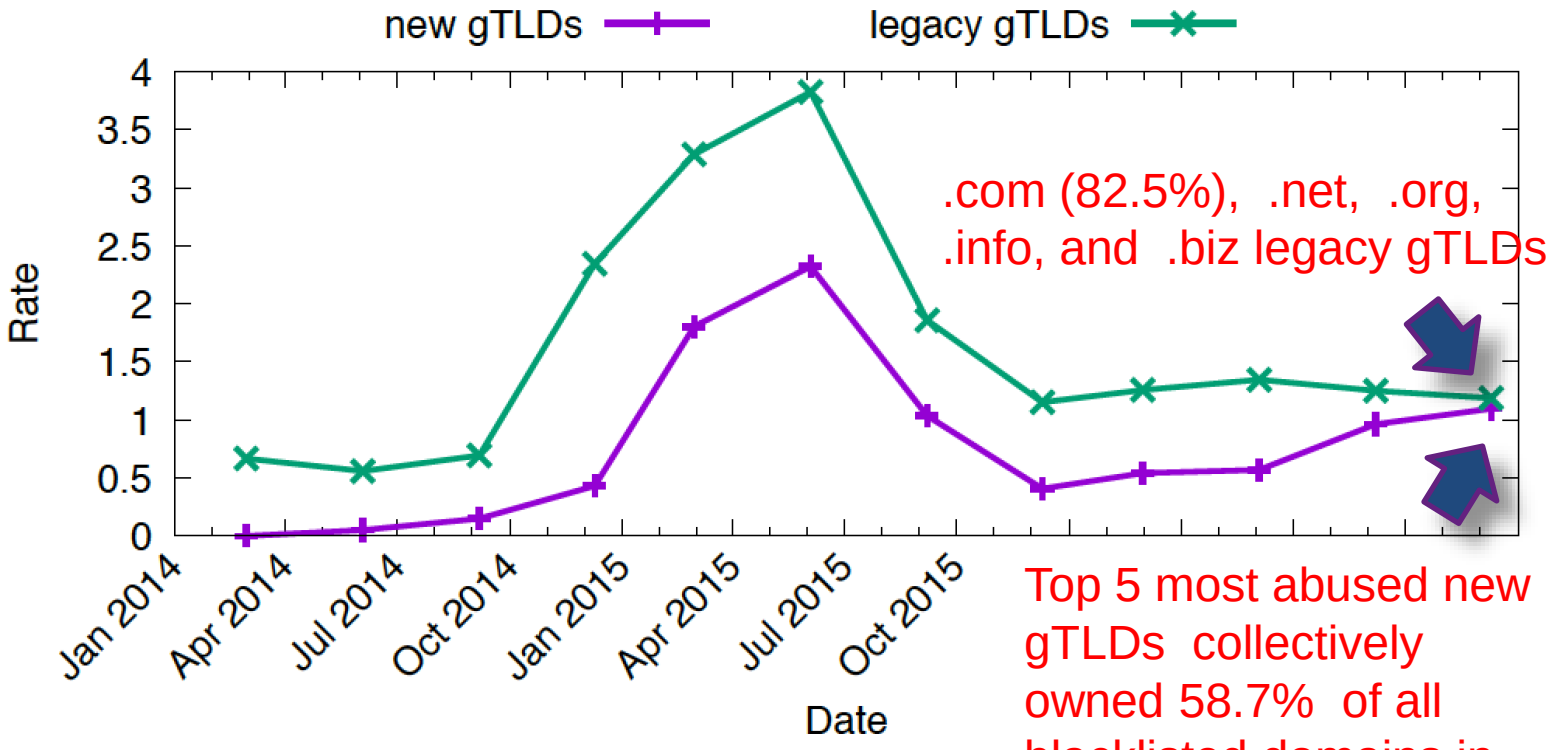
Abuse rates

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



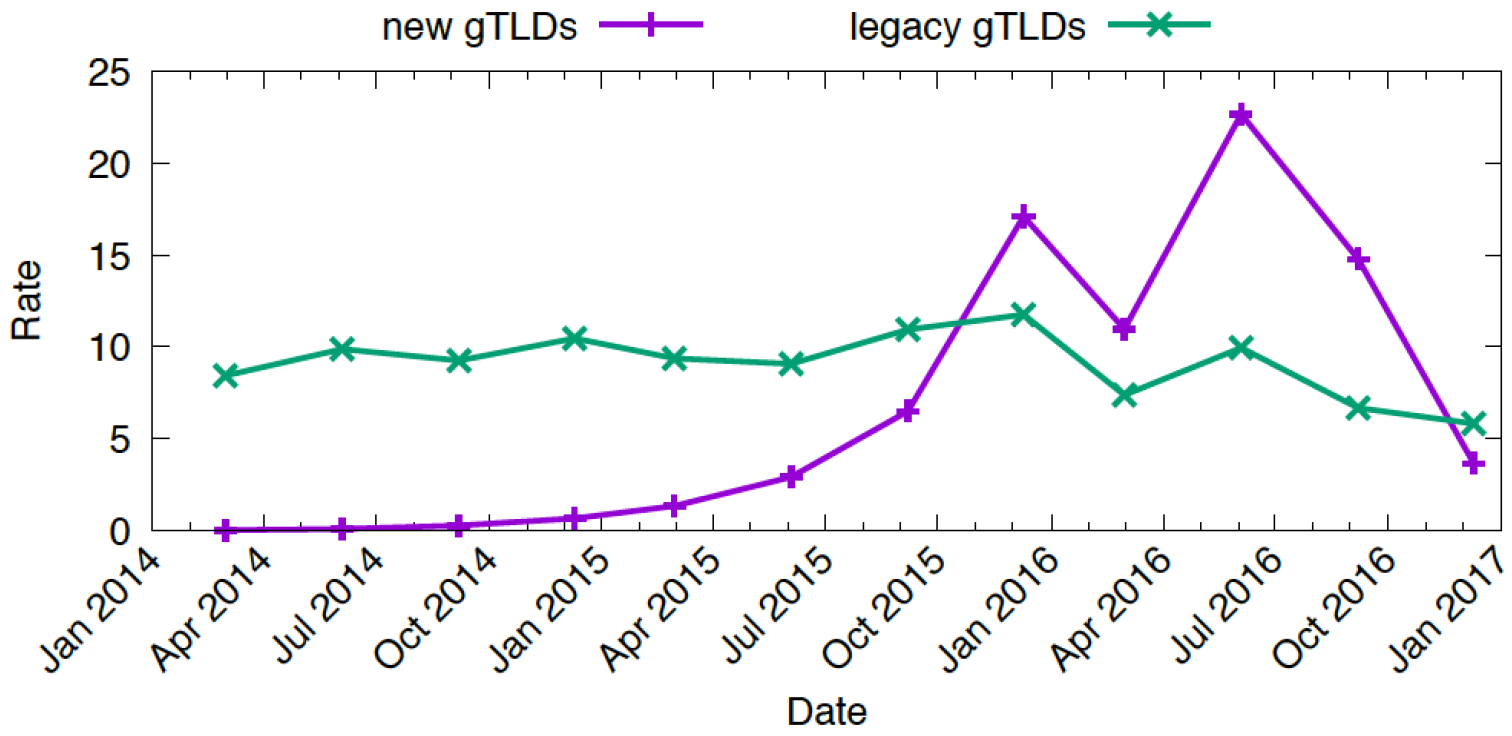
Abuse rates

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



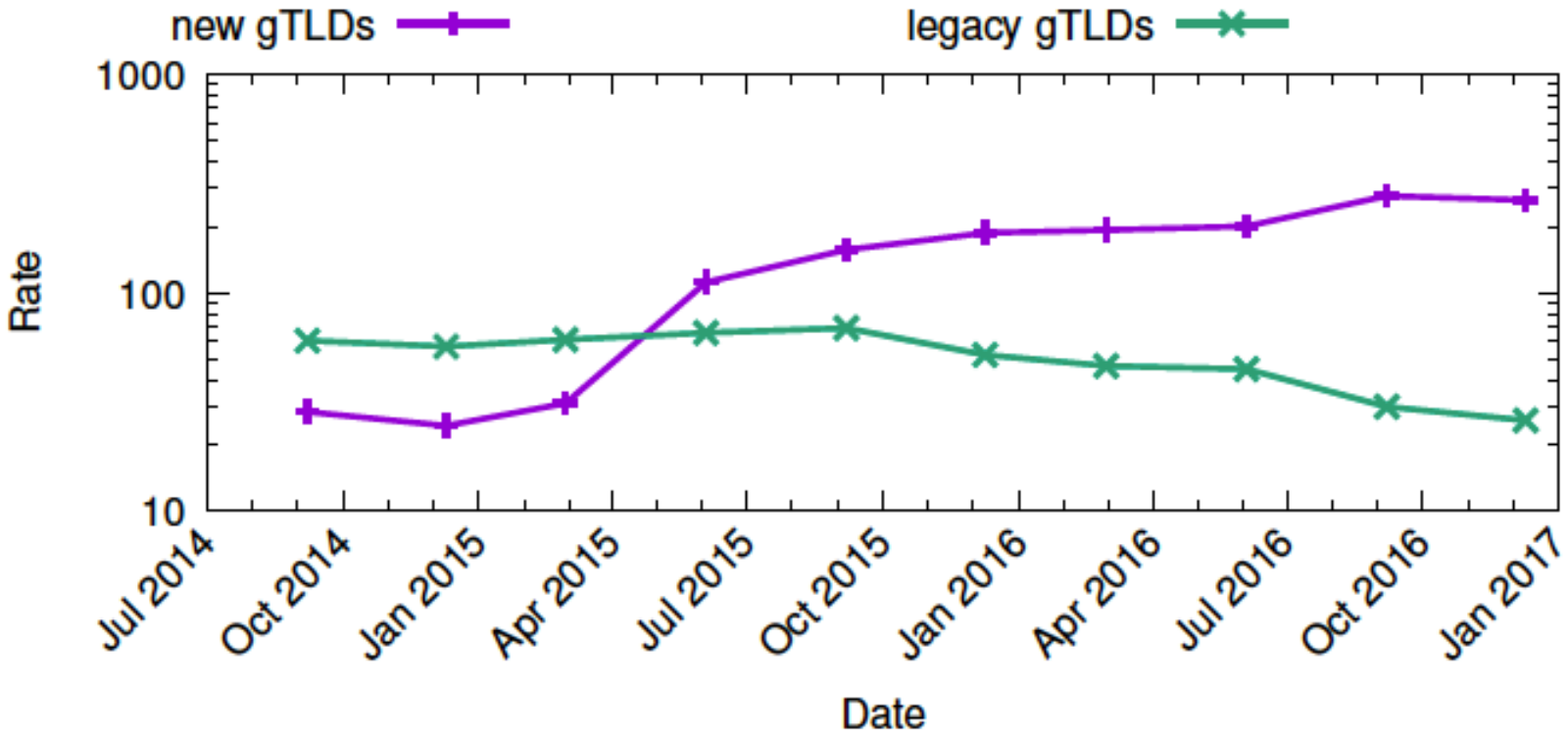
Abuse rates

– 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

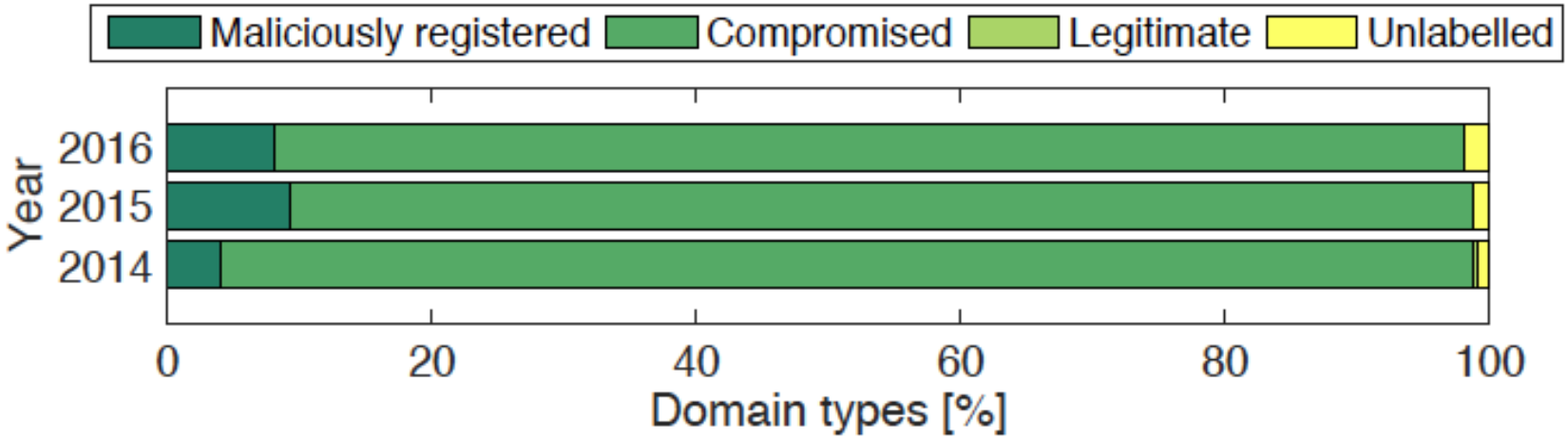


Compromised and maliciously registered domains

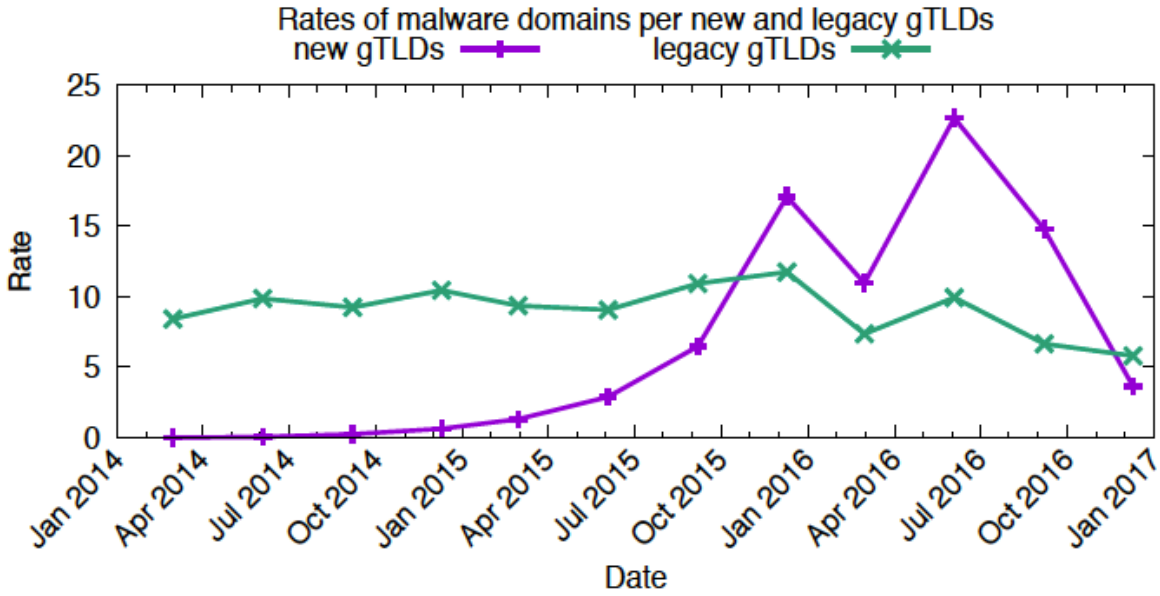
- Distinguishing between compromised and maliciously registered domains is critical because they require different mitigation actions by different intermediaries
- Assumption: maliciously registered domains are involved in a criminal activity within a short time after the registration.
- Other heuristics: if a given domain name contains a string of a brand name or its misspelled version indicating malicious registration, URLs indicating compromised content management systems, etc.

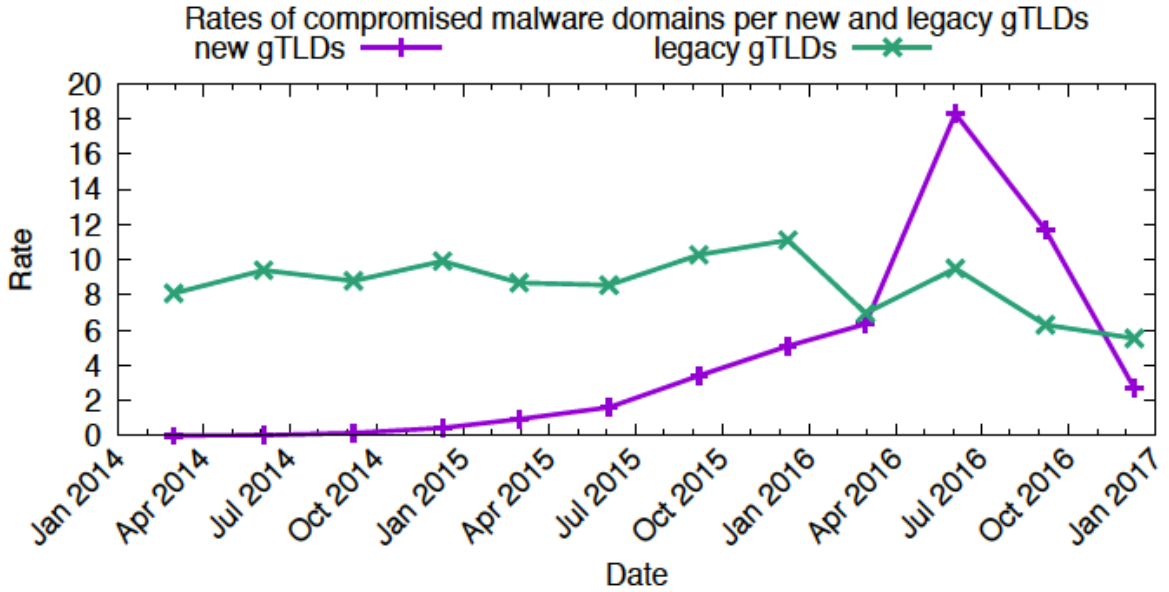
Compromised and maliciously registered domains

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

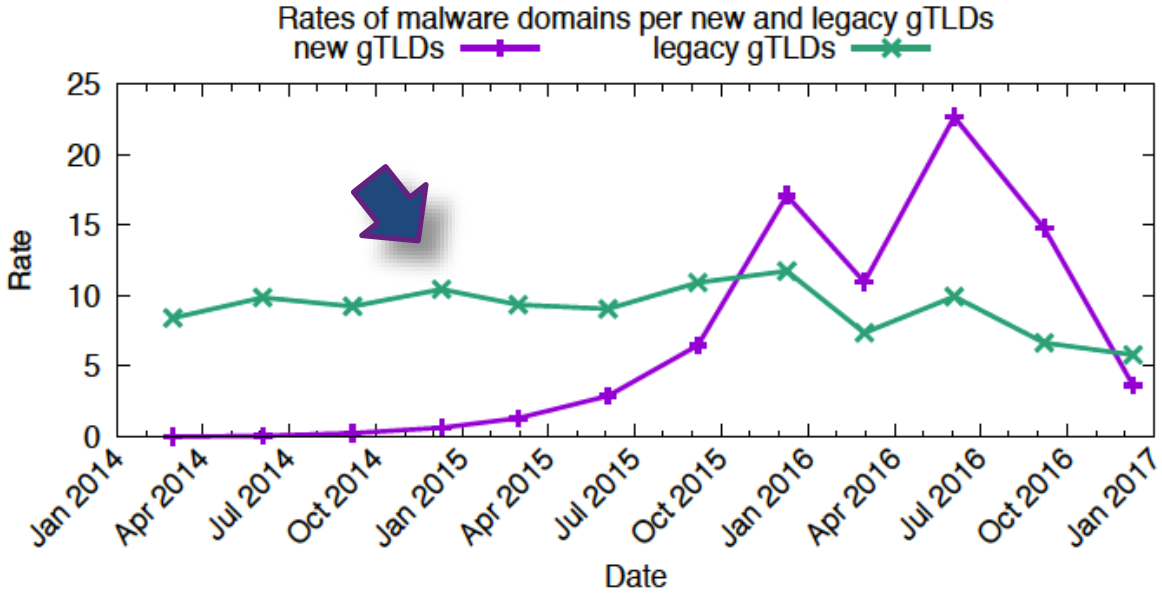


Compromised domains

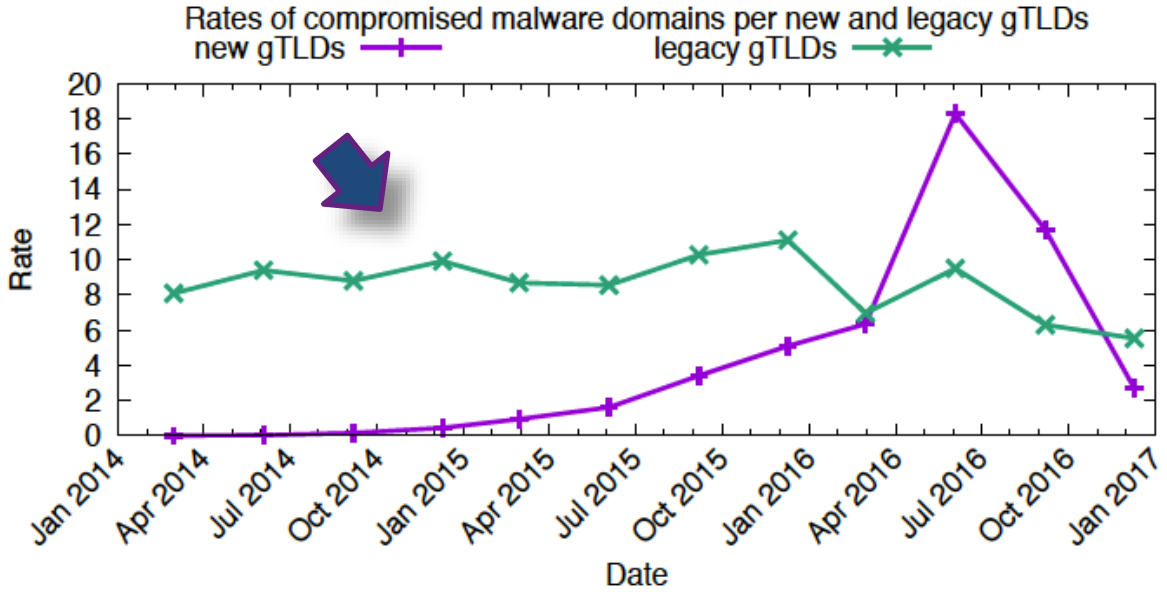




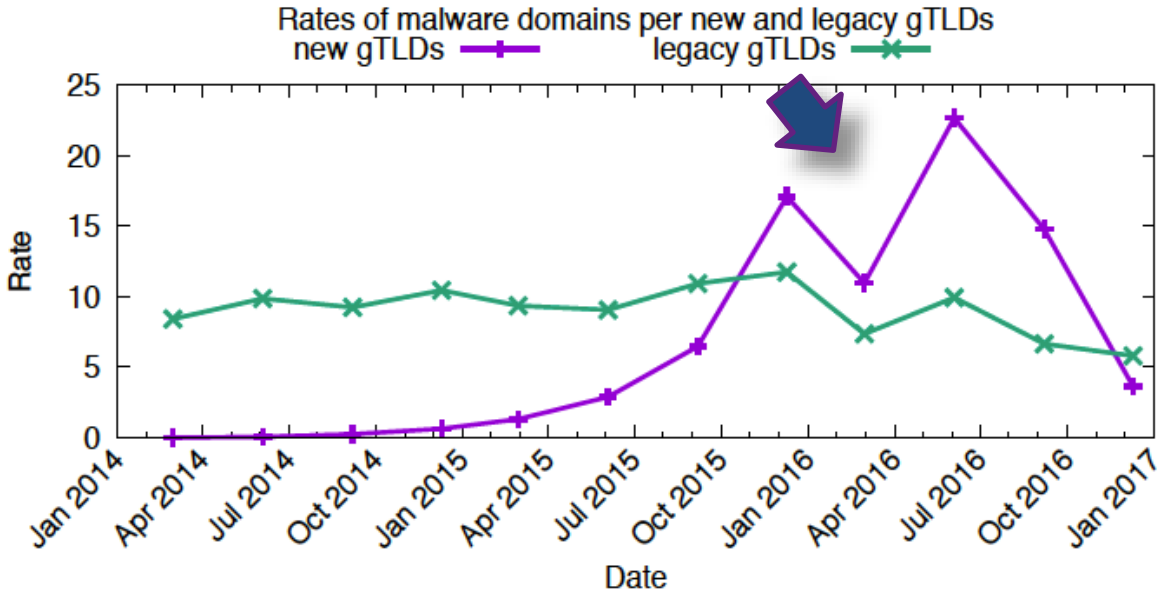
Compromised domains



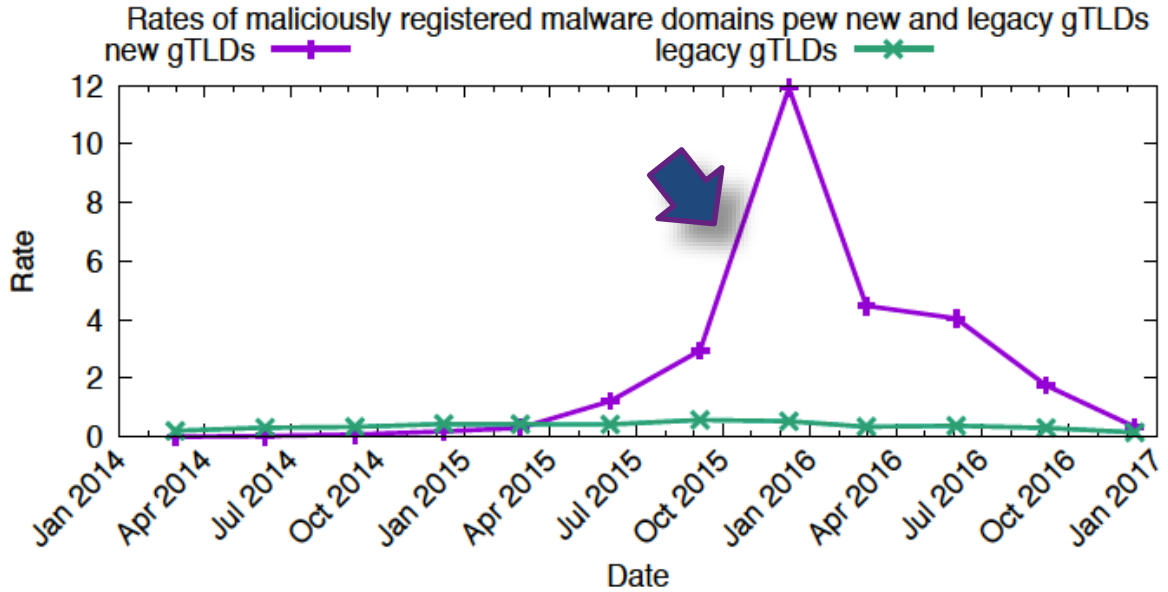
- Rates of abused domains in legacy gTLDs (StopBadware URL blacklists) are driven by compromised domains



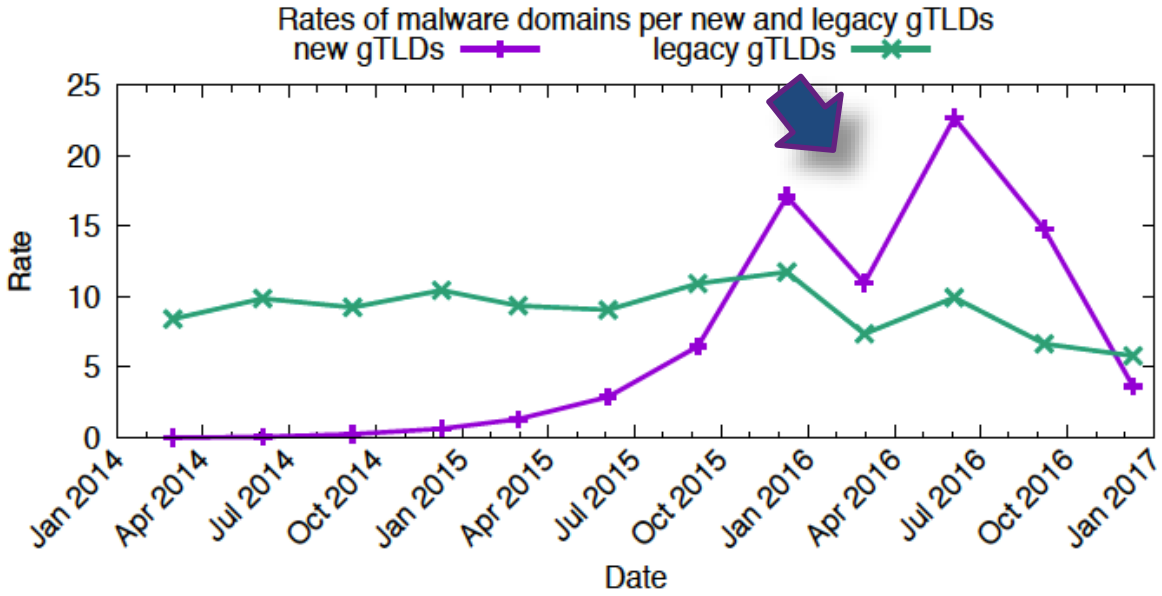
Maliciously registered domains



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

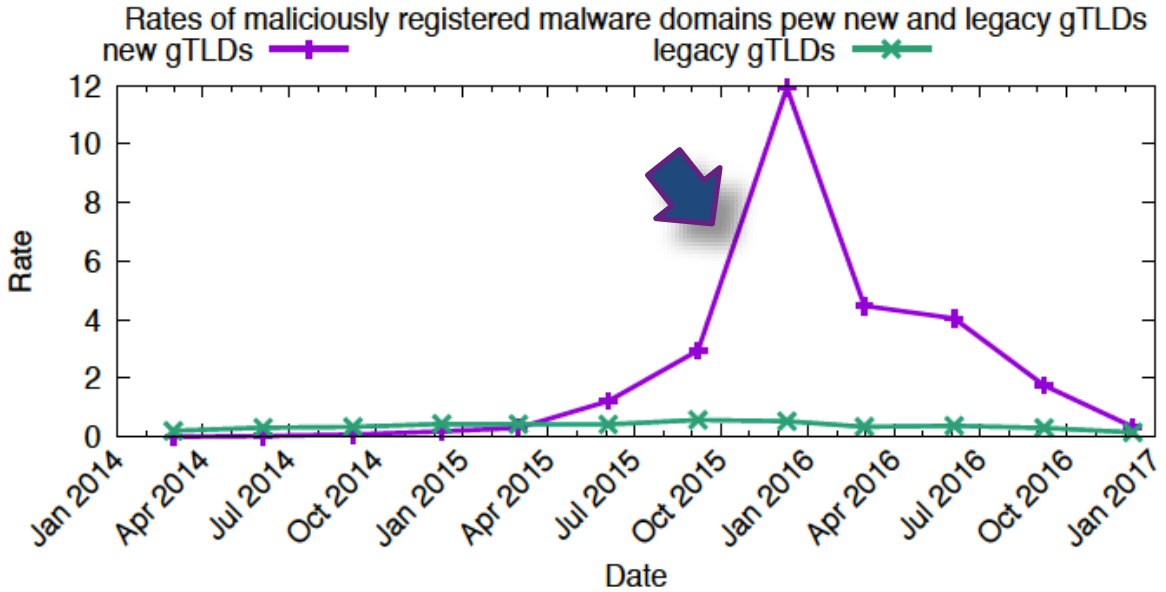


Maliciously registered domains



– 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)



Privacy or Proxy Services

- Why use PP services
  - Protecting your personal data
  - Blocking Spam
  - Stopping unwanted solicitations
- Analyzing use of PPs'es
  - Extract list of registrants
  - keyword search using "privacy", "proxy", "protect" etc
  - Manual inspection
- How many?
  - We found 570

Privacy or Proxy Services

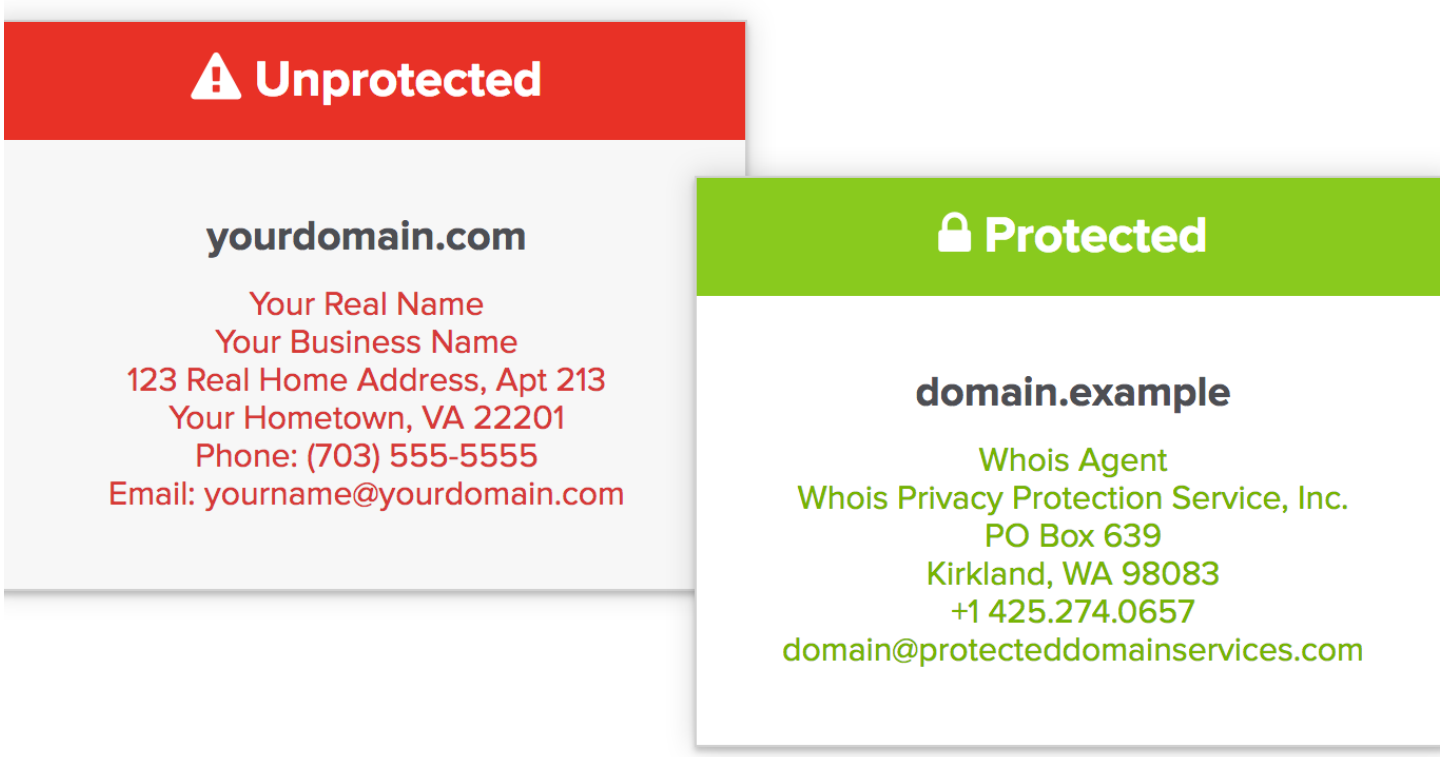
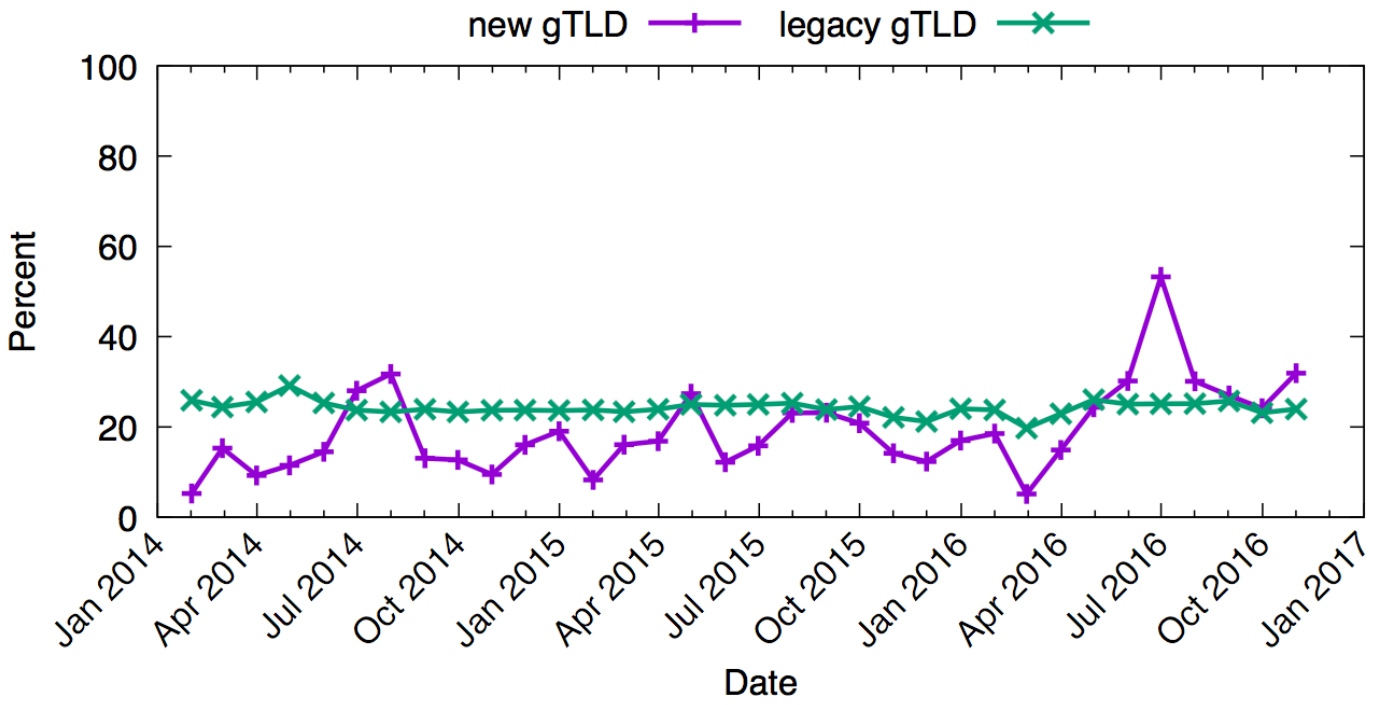


Image source: https://www.name.com/whois-privacy

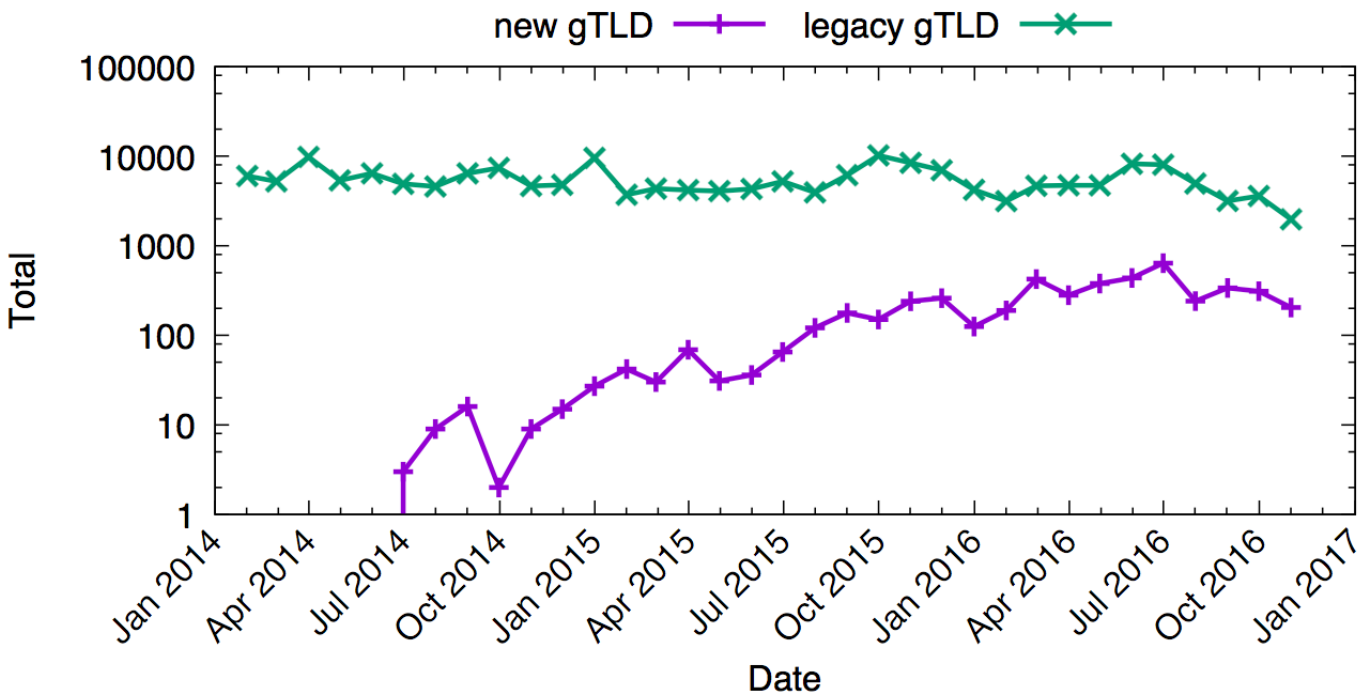
Privacy or Proxy Services

Usage for newly created domains per month



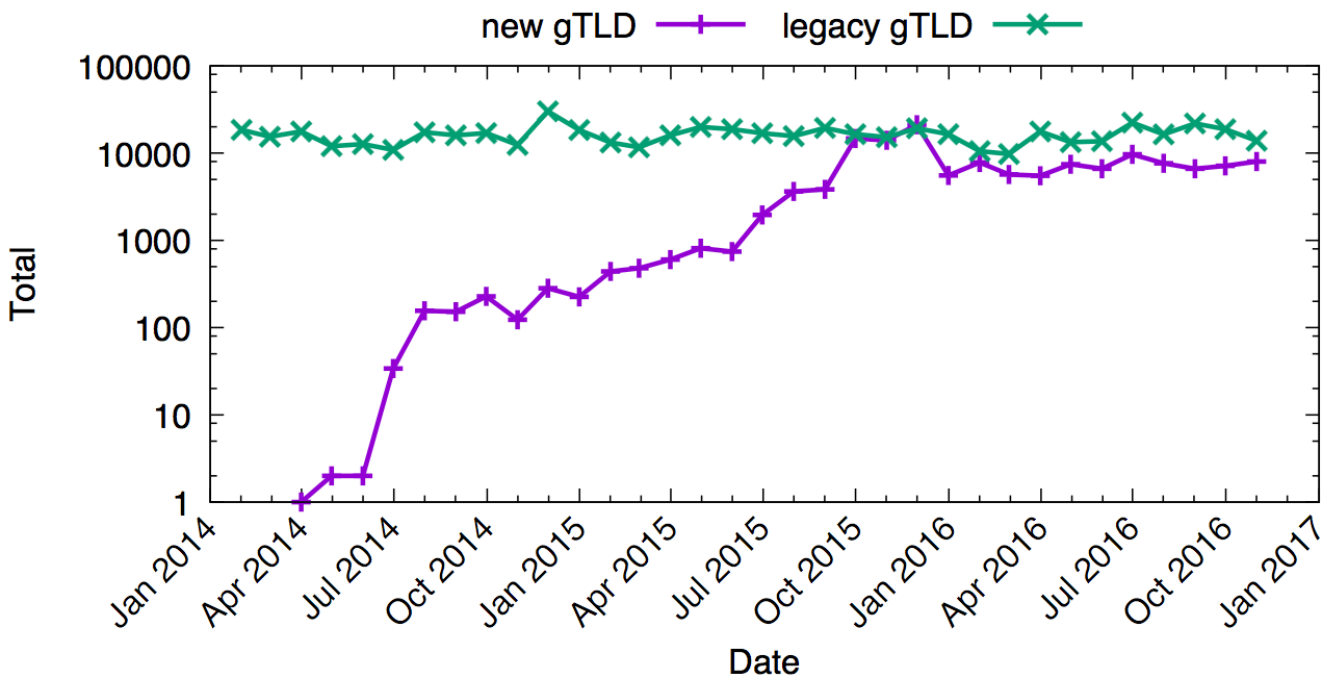
Privacy or Proxy Services

StopBadware



Privacy or Proxy Services

Spamhaus



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	7,832,264	28.57	USA	145,652,390	58.81
USA	6,114,944	22.31	China	22,409,117	9.05
Gibraltar	2,603,236	9.5	Germany	16,574,944	6.69
Cayman Islands	1,959,580	7.15	Canada	14,198,455	5.73
Singapore	1,700,985	6.2	India	9,509,405	3.84
Japan	1,667,079	6.08	Japan	6,400,530	2.58
India	1,274,622	4.65	Australia	5,950,392	2.4
Germany	1,056,541	3.85	France	4,573,133	1.85
Hong Kong	815,039	2.97	UK	3,670,192	1.48
Canada	422,834	1.54	Turkey	2,216,396	0.89

Geographical Location

SURBL distribution

new gTLD Country	#Incidents	percentage	rate
Gibraltar	585,839	47.4	2233.07
Japan	249,426	20.18	950.75
China	201,869	16.33	769.47
United States	87,139	7.05	332.15
India	45,059	3.65	171.75
United Kingdom	19,775		
United Arab Emirates	11,746		
Canada	6,110		
France	6,073		
Australia	5,852		

Legacy gTLD country	#Incidents	percentage	rate
United States	1,893,528	47.87	124.27
Japan	1,074,165	27.15	70.49
China	312,560	7.9	20.51
India	243,127	6.15	15.96
Germany	66,075	1.67	4.34
Ireland	58,226	1.47	3.82
Canada	37,861	0.96	2.48
Turkey	32,222	0.81	2.11
Australia	30,870	0.78	2.03
Bahamas	28,762	0.73	1.89

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

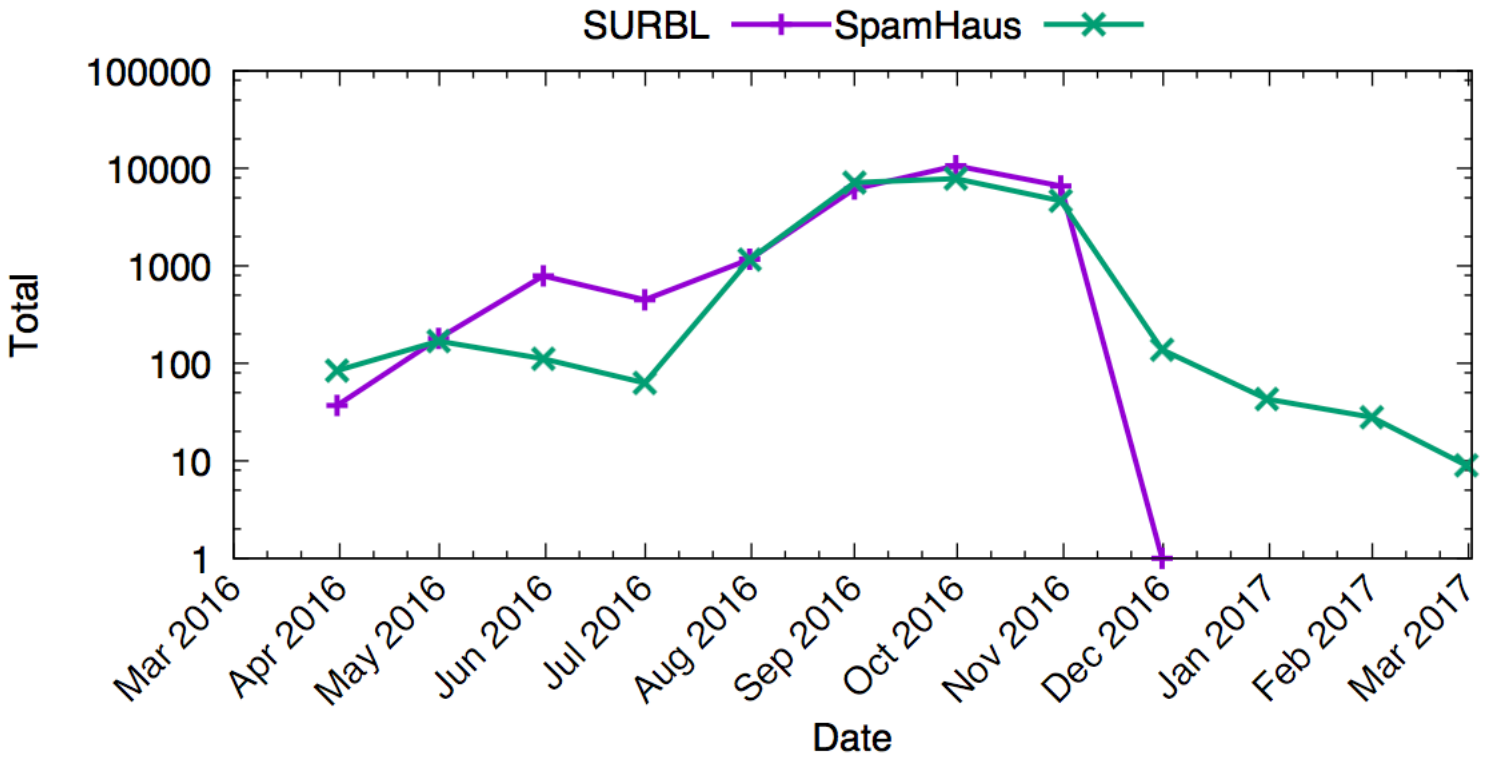
SURBL

new gTLD registrar	#Domains	#Incidents	Percent
Nanjing Imperiosus Technology	26,096	25,991	99.6
Intracom Middle East FZE	20,639	11,254	54.53
Dot Holding Inc.	153	76	49.67
Alpnames Limited	2,623,443	585,839	22.33
Todaynic.com, Inc.	317,534	69,330	21.83
Web Werks India d/b/a ZenRegistry.com	784	146	18.62
Xiamen Nawang Technology Co., Ltd	281,148	42,067	14.96
GMO Internet d/b/a Onamae.com	1,673,447	249,420	14.9
TLD Registrar Solutions Ltd.			
Instra Corporation Pty Ltd.			

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	10,892	30.85
Paknic (Private) Limited	10,512	3,081	29.31
Intracom Middle East FZE	67	16	23.88
AFRIREGISTER S.A.	1,540	266	17.27
Minds and Machines LLC	1,115	171	15.34
OwnRegistrar, Inc.	19,745	2,933	14.85
GMO Internet d/b/a Onamae.com	7,171,201	1,061,902	14.81
GoName.com, Inc	2,662	384	14.43

Registrar Reputation

Nanjing Imperiosus Technology Co. Ltd.



Schedule

- Final report available July 2017

- Incorporate WHOIS data information from Domain Tools
- Inferential analysis of potential relationship with abuse drivers (Regression analysis of abuse in gTLDs)

Questions?