Co-funded by the
Erasmus+ Programme
of the European Union

Pathway in Enterprise Systems Engineering (PENS)

"The Name of the Rose"

What Statistical Analysis can reveal about the threats hiding behind domain names

Davide Ariu - University of Cagliari

February 20th, 2019

…or even better

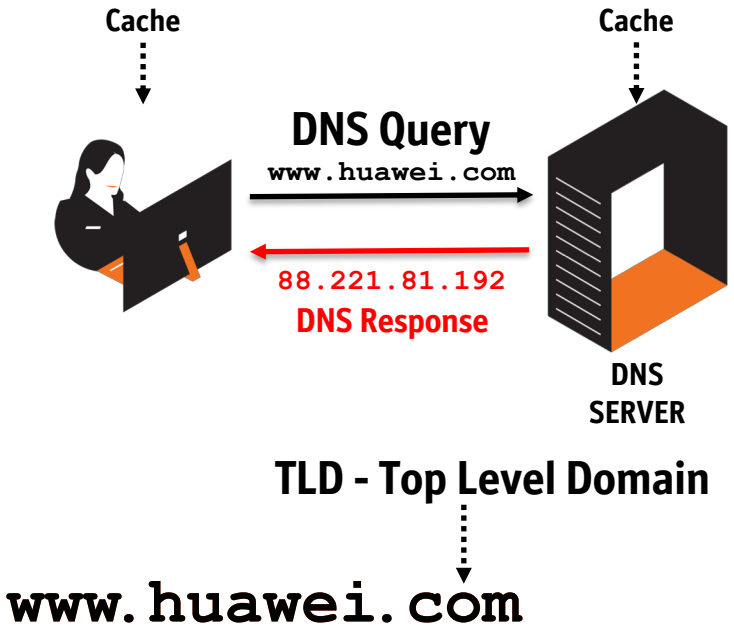
- **A (very quick) review of the DNS infrastructure**
  - Main components
  - Working Principles

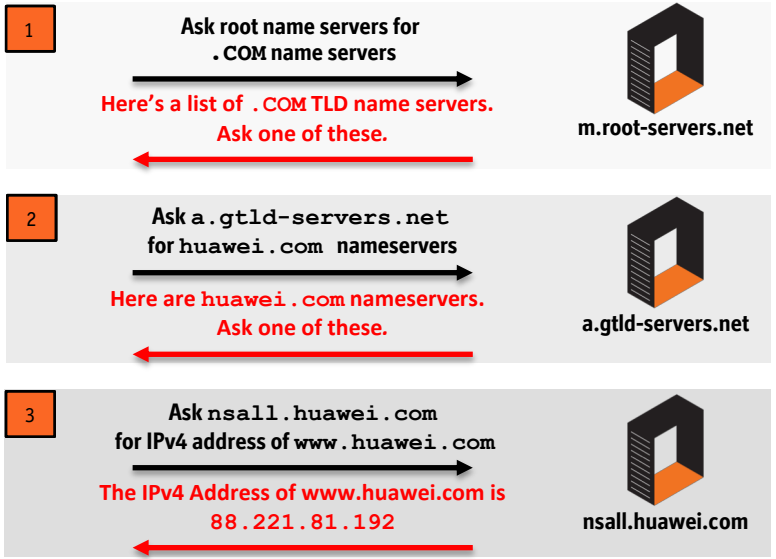
- **An overview of different kinds of abuses against such infrastructure**, which enable a wide variety of cyber attacks

Co-funded by the Erasmus+ Programme of the European Union

Domain Name System (DNS) – Principles

- With DNS we generically refer to the *infrastructure* (servers, protocol, resolvers, caches, etc.) which allows to translate (resolve) domain names into IPv(4/6) addresses

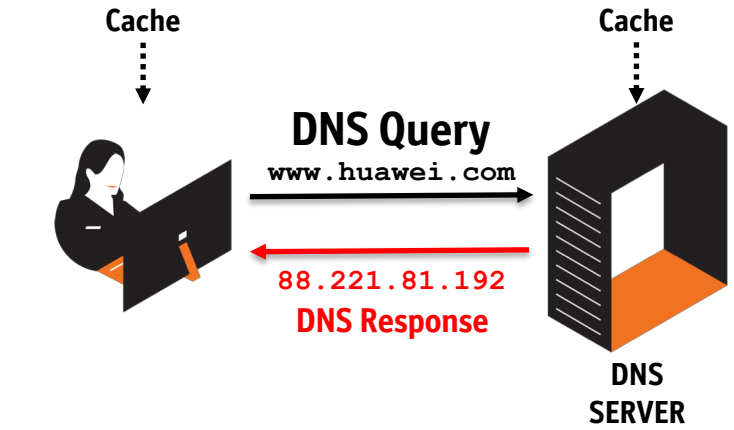


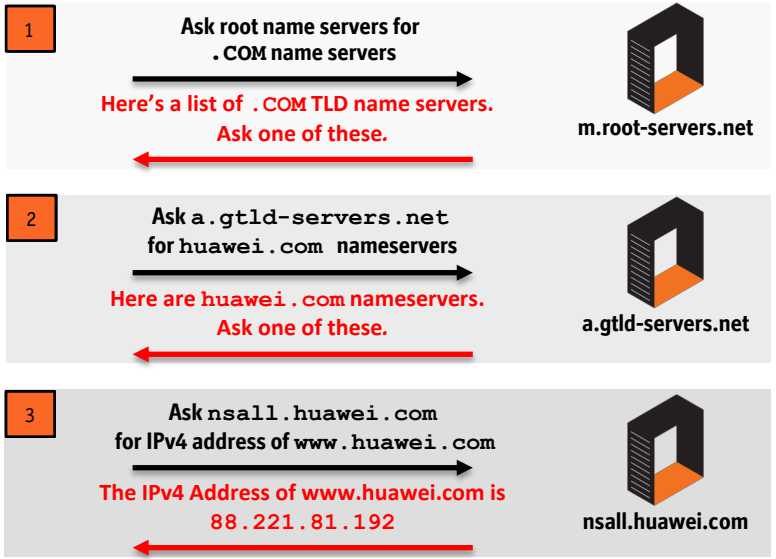


- **Distributed**
- **Hierarchical**
- **Recursive**

Domain Name System (DNS) – Principles

- With DNS we generically refer to the *infrastructure* (servers, protocol, resolvers, caches, etc.) which allows to translate (resolve) domain names into IPv(4/6) addresses



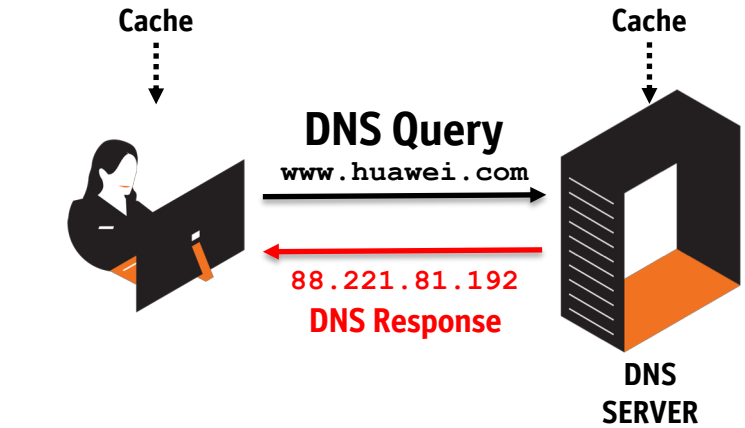


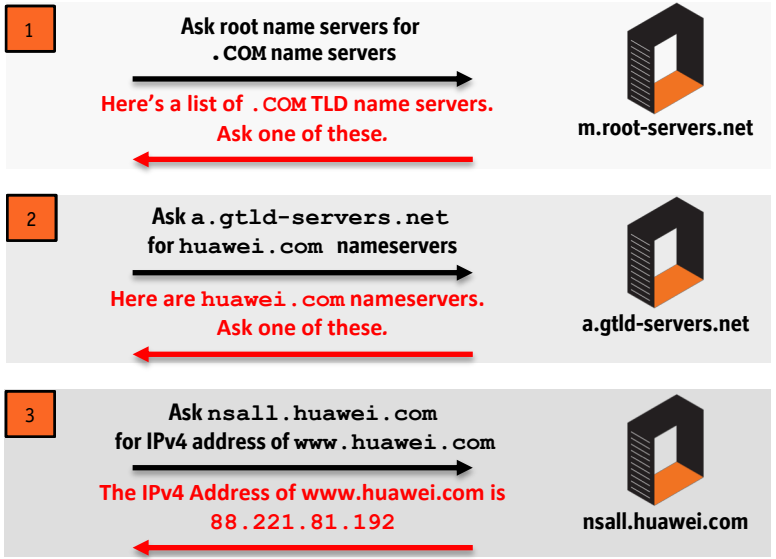
- **Distributed**
- **Hierarchical**
- **Recursive**

Co-funded by the Erasmus+ Programme of the European Union

Domain Name System (DNS) – Principles

- With DNS we generically refer to the *infrastructure* (servers, protocol, resolvers, caches, etc.) which allows to translate (resolve) domain names into IPv(4/6) addresses





- **Distributed**
- **Hierarchical**
- **Recursive**

Co-funded by the Erasmus+ Programme of the European Union

A closer look at Domain Names

A.k.a. FQDN – Fully Qualified Domain Name

TLD - Top Level Domains

- **gTLD (generic)** - .com, .net, .org, .info, etc.
- **ccTLD (country-code)** - .it, .de, .us, .ru, etc.
- **New gTLD** - .movie, .radio, .top, .online , etc.

- >1200 gTLD introduced after 2013
- Operated by
  - **International** (gTLD, New gTLD) **registries** (ICANN, Verisign, Neustar, Afilias, etc.)
  - **National** (ccTLD) **registries** (e.g. GARR in Italy)
- Each TLD defines a specific «zone» of the DNS
  - E.g. the file containing the list of .com domain names is called «.com zone file»

- Identifies a specific host on the Internet
- Consisting of ASCII characters only
  - Letters
  - Digit
  - Hypens (not at the begin/end of the string)
- **Resolved from right to left**
  1. **Top Level Domain**

www.huawei.com

Co-funded by the Erasmus+ Programme of the European Union

A closer look at Domain Names

A.k.a. FQDN – Fully Qualified Domain Name

Second Level Domain

- It is the *domain name* actually registered by a private/public entity (the **Registrant**)
- **Registries** maintain a list of the 2LD registered per TLD (zone files)
- **Registrars** are the subjects who act as intermediaries between the **Registrants** and the **Registries**
  - **Billing**
  - **Whois data**

- Identifies a specific host on the Internet
- Consisting of ASCII characters only
  - Letters
  - Digit
  - Hypens (not at the begin/end of the string)
- **Resolved from right to left**
  1. Top Level Domain
  2. **Second Level Domain**

www.huawei.com

Co-funded by the Erasmus+ Programme of the European Union

A closer look at Domain Names

A.k.a. FQDN – Fully Qualified Domain Name

Hostname

- Identifies a specific host under a **2LD**
- Also called **Third Level Domain**
- Completely under control of the the **2LD Registrant** (the entity which has registered the 2LD)
- It appears only on the DNS server responsible (authoritative) for the corresponding **2LD**
  - **No public registry of 3LD exist**

- **Issue: points are among the characters allowed in the 3LD**

- Identifies a specific host on the Internet
- Consisting of ASCII characters only
  - Letters
  - Digit
  - Hypens (not at the begin/end of the string)
- **Resolved from right to left**
  1. Top Level Domain
  2. Second Level Domain
  3. **Hostname**

www.huawei.com

Co-funded by the Erasmus+ Programme of the European Union

A closer look at Domain Names

A first ssue

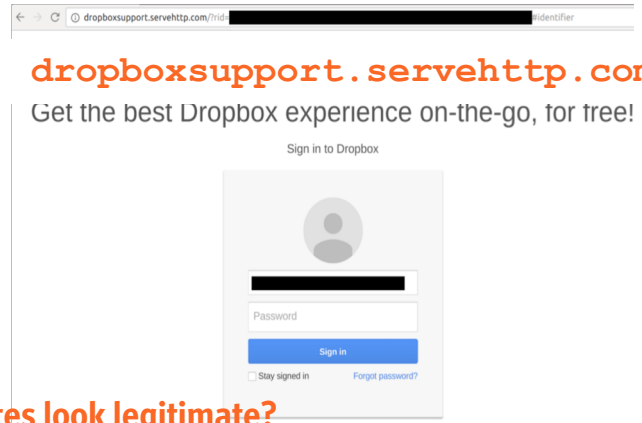
- **Issue: points are among the characters allowed in the 3LD**

www.huawei.com.maldomain.net

- **Resolved from right to left**
  1. Top Level Domain
  2. Second Level Domain
  3. **Hostname**

Co-funded by the Erasmus+ Programme of the European Union

Role of the DNS in the modern Cyber-attacks



- The DNS has a pivotal role to support a large variety of cyber-attacks.

1. **Phishing.**
   - Well crafted domain names help to make phishing[*1] campaigns more effective
2. **Scams**
   - Similarly to phishing, other kind of scams are vehiculated by using well-crafted domain names
3. **Botnets**
   - to make botnet more resilient, malware writers avoid to hardcode the IP addresses of the C&C servers and dropzones
     - **They use domain names** instead
     - Often include into malware algorithms able to generate domain names dynamically → **Domain Fluxing**
     - Often use mechanisms to change frequently the association between domain names and IP addresses → **IP fluxing**

Do these websites look legitimate?

www.ryanaĭr.com
www.heineṇken.com
www.alitalıa.com

*1 Source: **John Scott-Railton et Alii,** *Nile Phish: Large-Scale Phishing Campaign Targeting Egyptian Civil Society, CitizenLab.org, 2017*

Co-funded by the Erasmus+ Programme of the European Union

About this presentation

1. **What is this presentation about:**
   - Attacks which can be spotted through (statistical) analysis of the domain names and of information gathered during the resolution process
     - Domain Fluxing
     - IP Fluxing
     - Domain *Squatting
   - **Examples taken from real traffic**
     - >3 years of real traffic monitoring
       - ≃ 300.000.000 DNS requests/day
       - ≃ 2.5M domain names/day
       - ≃ 430k unique IPs/day
2. **Attacks <u>not</u> covered by the presentation**
   - Abuses of the DNS Protocol (Spoofing)
   - DDoS Attacks against DNS
   - Covert channels over DNS

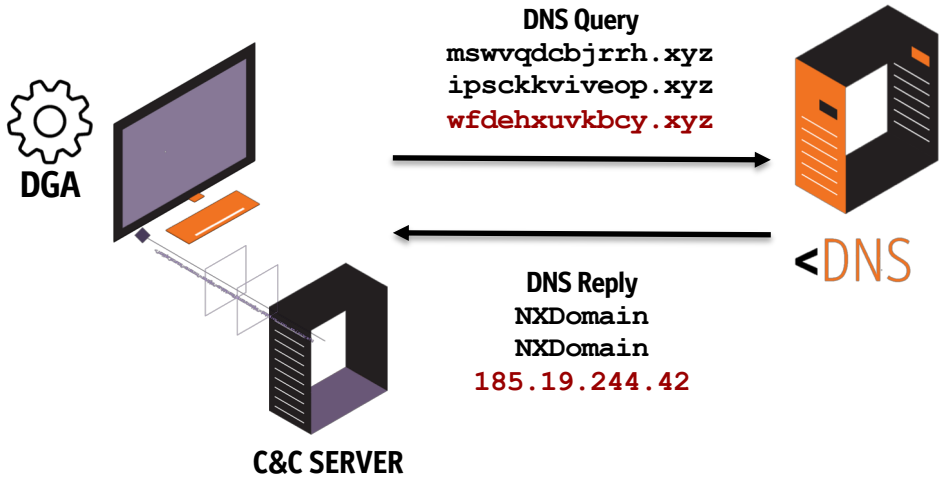
Co-funded by the Erasmus+ Programme of the European Union

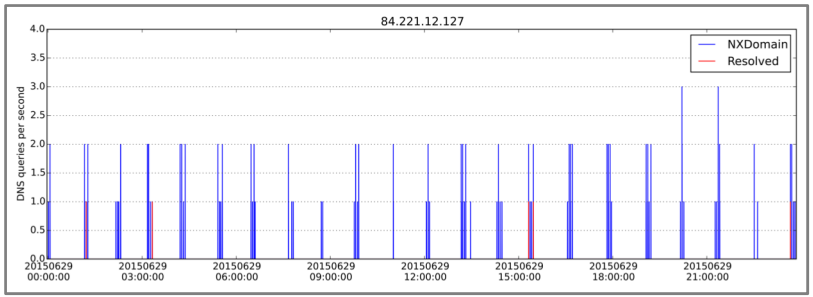
Domain Fluxing

Co-funded by the
Erasmus+ Programme
of the European Union

Domain Fluxing

- **Domain Fluxing** refers to attacks where domain names are algorithmically generated (eventually in a high number → Flux)
  - A **DGA** (**Domain Generation Algorithm**) is a component modern malware use to generate a large number of domain names and then use a small subset for actual C&C communication.
    - **Allows botmasters to harden the infrastructure of their botnets**





Co-funded by the Erasmus+ Programme of the European Union

Algorithmically Generated Domains (AGD)

- DGA (Domain Generation Algorithm) is a component modern malware use to generate a large number of domain names and then use a small subset for actual C&C communication.
  - **Allows botmasters to harden the infrastructure of their botnets**
- The generated domains are computed based on a given *seed* (shared secret between botmasters and the bots)
- **Advantages**:
  - **Blacklisting becomes harder**
    - **Static** domain **blacklisting** becomes **ineffective**
    - No-domains hardcoded in the malware binary
    - The generated **domains** are **time dependent**
    - Short-lived domain allow to evade domain reputation services
  - **The botnet is more resilient to takedown efforts**
    - The botmaster needs to control just 1 domain
    - Defenders need to control all the domains
      - The increase in the number of TLD (>1000) makes this more challenging
      - E.g. Conficker Version C spreaded its domains over 113 TLDs (takedown required coordination of 30 different organisations)

Co-funded by the Erasmus+ Programme of the European Union

About Domain Generation Algorithms - 1

- Uses a **seed** to coordinate with the criminal organisation
  - Timestamp (HTTP request to widely used services)
  - Trending topics on Twitter (**Torpig Botnet**)
  - Foreign exchange reference rates (**Bedep malware family**)
  - Victim's HW information (**Jiripbot**)
  - Strings obfuscated in the malware code

- Generation Scheme
  - **Arithmetic-based** → DGAs calculate a sequence of values that either have:
    - a direct ASCII representation usable for a domain name
    - designate an offset in one or more hard-coded arrays
  - **Hash-based DGAs** use the hexdigest representation of a hash (MD5, SHA256) to produce an AGD.
  - **Wordlist-based DGAs** concatenate a sequence of words from one or more wordlists
    - either embedded in the malware binary or obtained from a publicly accessible source
  - **Permutation-based** DGAs derive all possible AGDs through permutation of an initial domain name.

Source: D. Plohmann et. Al., *A Comprehensive Measurement Study of Domain Generating Malware*, 2016.

Co-funded by the Erasmus+ Programme of the European Union

About Domain Generation Algorithms - 2

- Domain generation layout
  - Dictionary based domain names (E.g. jeannettegeorgeson.net)
  - Alphabetic layout (E.g. isrkvfkoxnwsatkdb.ki)
  - Numeric layout (E.g. 9af4a478.net)
  - Alphanumeric layout (E.g. txjcbx3oejncdvg4.com)

- DGA may use different sets of TLD under which to register domain names
  - Conficker uses 123 TLDs; Necurs uses 43 TLDs

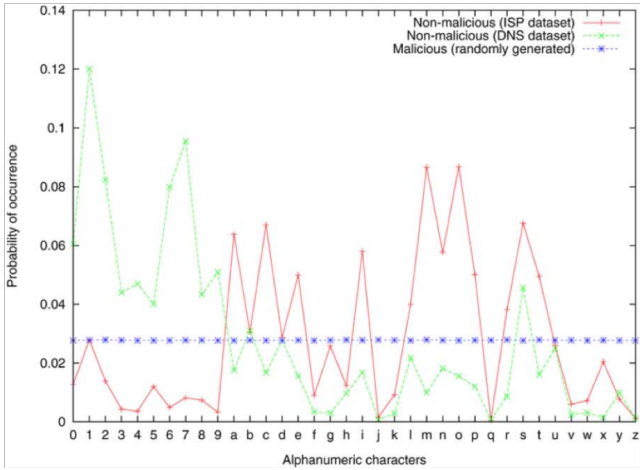
- The domain generation rate ranges from douzens to thousands of domains per day!
  - **Conficker.C creates 50k domains per day**
  - **Pykspa 1 generates a list of 5000 domains every two days**
  - **Flashback DGA generates 5 domains per day**

- Domain validity
  - Typically disjunct (only a single set of domains is valid at each point in time)
  - Ranges from hours to months

Source: D. Plohmann et. Al., *A Comprehensive Measurement Study of Domain Generating Malware*, 2016.

Detecting DGA Malware with Passive DNS Analysis

- **Distribution of alphanumeric characters**[*1]
  - Motivation: algorithmically generated domains differ significantly from legitimate (human) generated ones
    - Kullback-Leibler Divergence
      - Calculated on Unigrams, Bigrams
    - Jaccard Index
    - Edit Distance



- **Streams of unsuccessful domain name resolutions (or NXDomains)** [*2]
  - Motivation: domain names generated by to the same algorithm exhibit similar characteristics and are queried by overlapping sets of hosts.
    - N-grams
    - Entropy in the character distribution
    - Domain name length, number of domain levels, number of TLDs

Source*1: S. Yadav et. Al., *Detecting Algorithmically Generated Domain-Flux Attacks With DNS Traffic Analysis*, 2012.
Source*2: M. Antonakakis, *From Throw-Away Traffic to Bots:Detecting the Rise of DGA-Based Malware, Usenix S.S., 2012.*

Co-funded by the Erasmus+ Programme of the European Union

Challenge - Hostname (3LD) Domain Fluxing

abilities.lolklubü.org

accredited.lolklubü.org

ac.lolklubü.org

ac4-qaamdb-clu.lolklubü.org

adobe.lolklubü.org

akamai.lolklubü.org

amazon-dc.lolklubü.org

marriage.lolklubü.org

moviecontrol.lolklubü.org

.
.
.

young.lolklubü.org

wu.lolklubü.org

zongapiserv.lolklubü.org

- >100 Domains Generated in ≃ 6 minutes (July 19th, 2018)
- Challenge:
  - The variable part is on the Hostname (3LD)
    - **Only visible if one has the chance to observe the traffic**

Co-funded by the Erasmus+ Programme of the European Union

IP Fluxing

Co-funded by the
Erasmus+ Programme
of the European Union

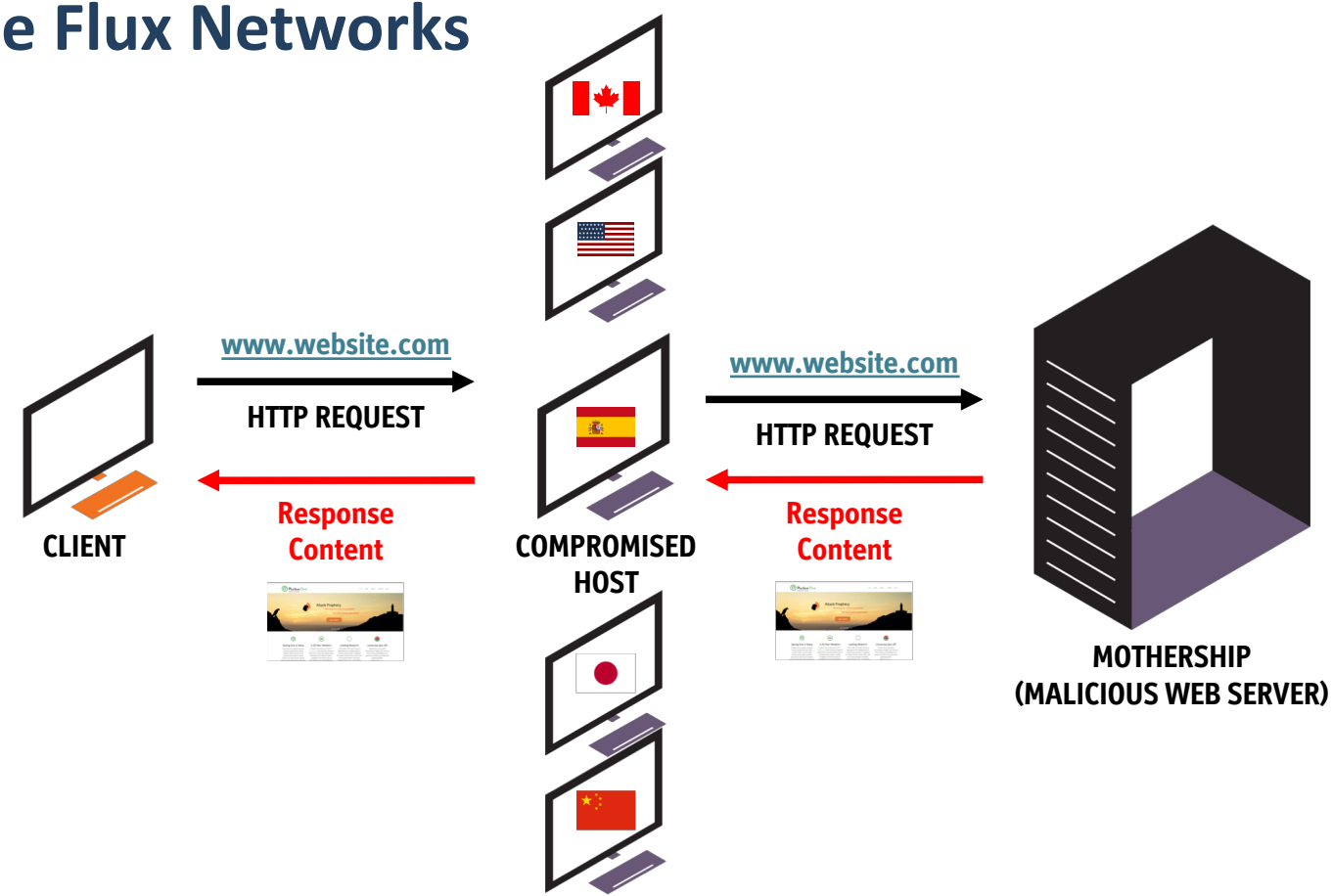
Design of Fast Flux Service Network

- **Fast-Flux Service Networks** exploit an architecture which is conceptually similar to that of **Content Delivery Networks**

- The key idea is to construct **a distributed proxy network on top of compromised machines** that redirects traffic through these proxies to a central site,which hosts the actual content.
  - Taking down any of the proxies does not effect the availability of the central site (**Mothership**)
  - The **attacker always returns a different set of IP addresses for a DNS query** and thus distributes the traffic over the whole proxy network.
  - This leads to an **increased resilience** since taking down such schemes usually needs cooperation with a domain name registrar.
  - The **Mothership** itself becomes really **hard to track**

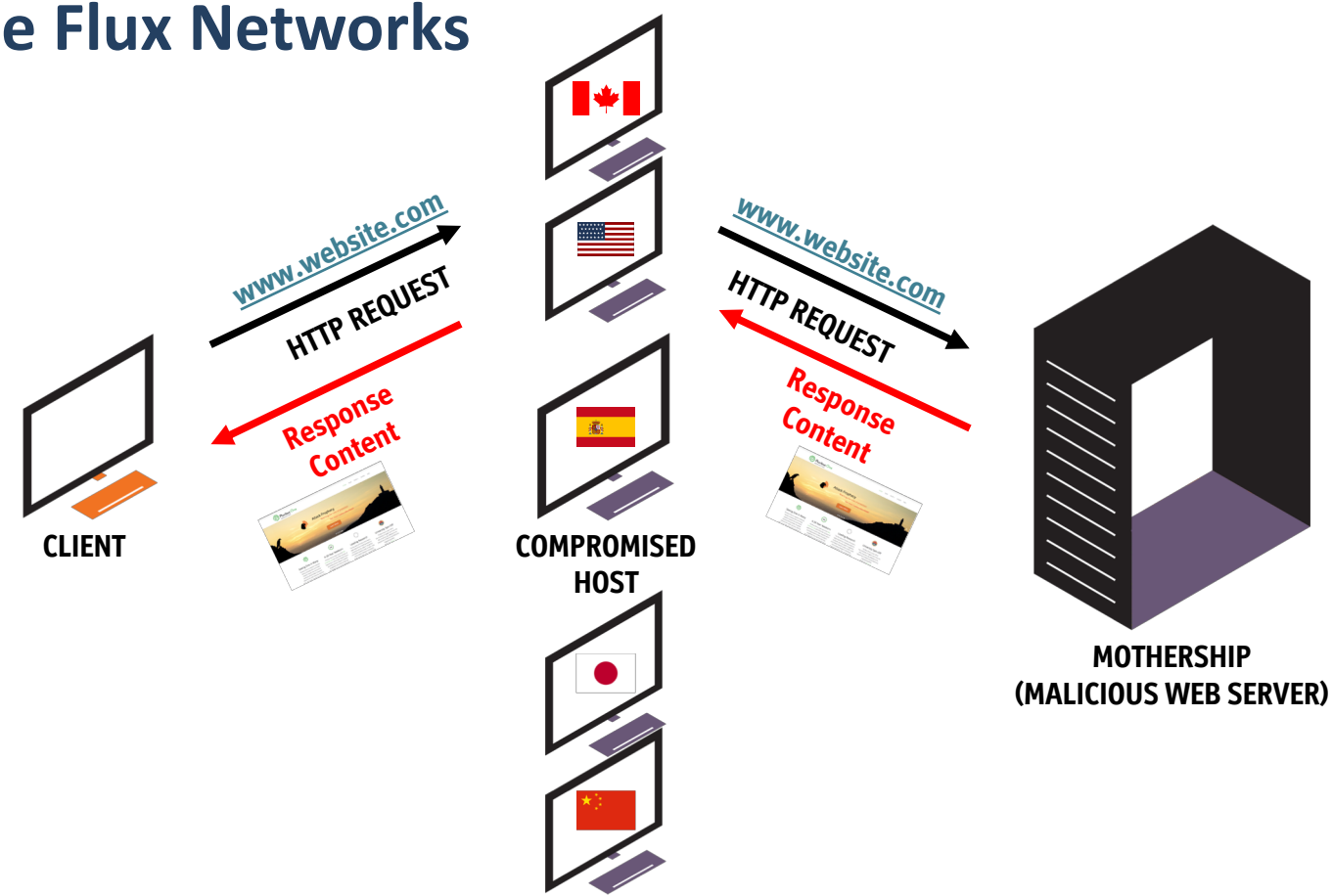
Co-funded by the Erasmus+ Programme of the European Union

Single Flux Networks



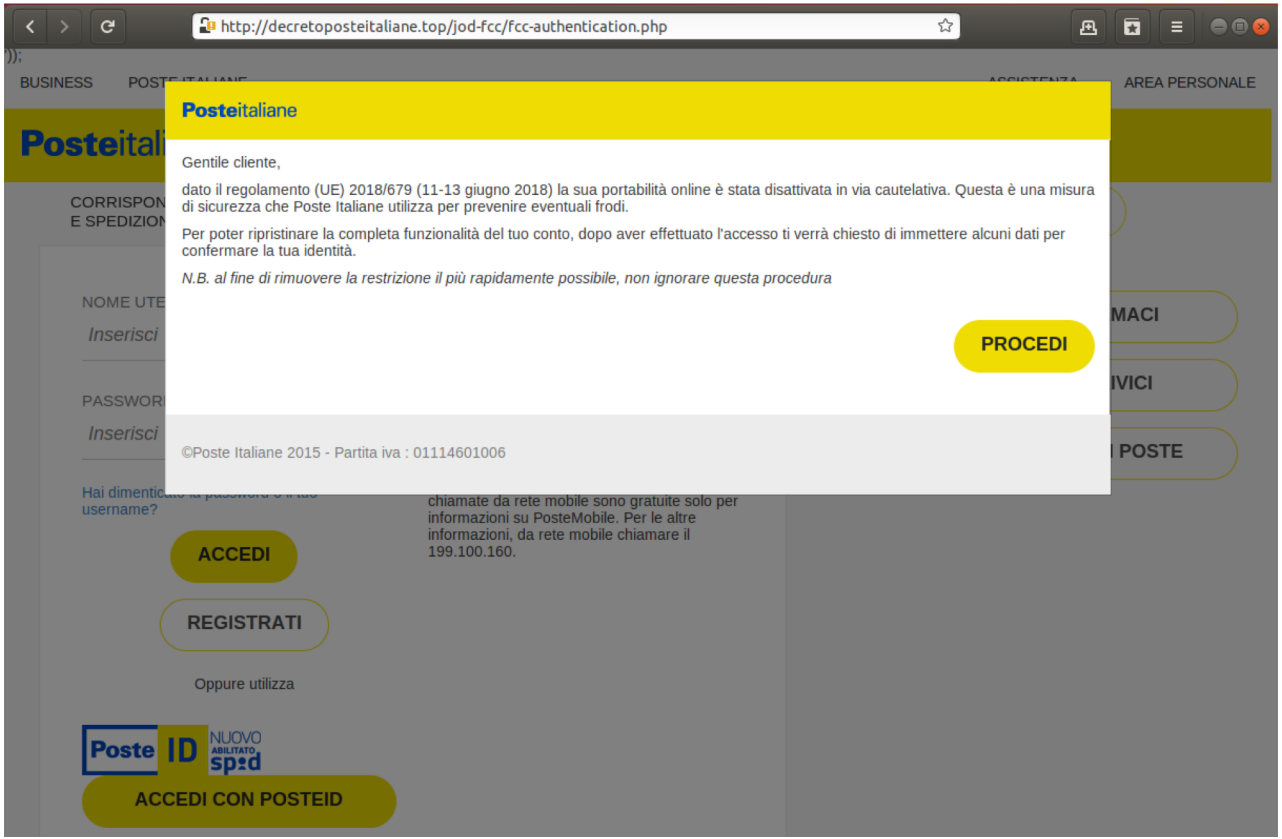
Co-funded by the Erasmus+ Programme of the European Union

Single Flux Networks



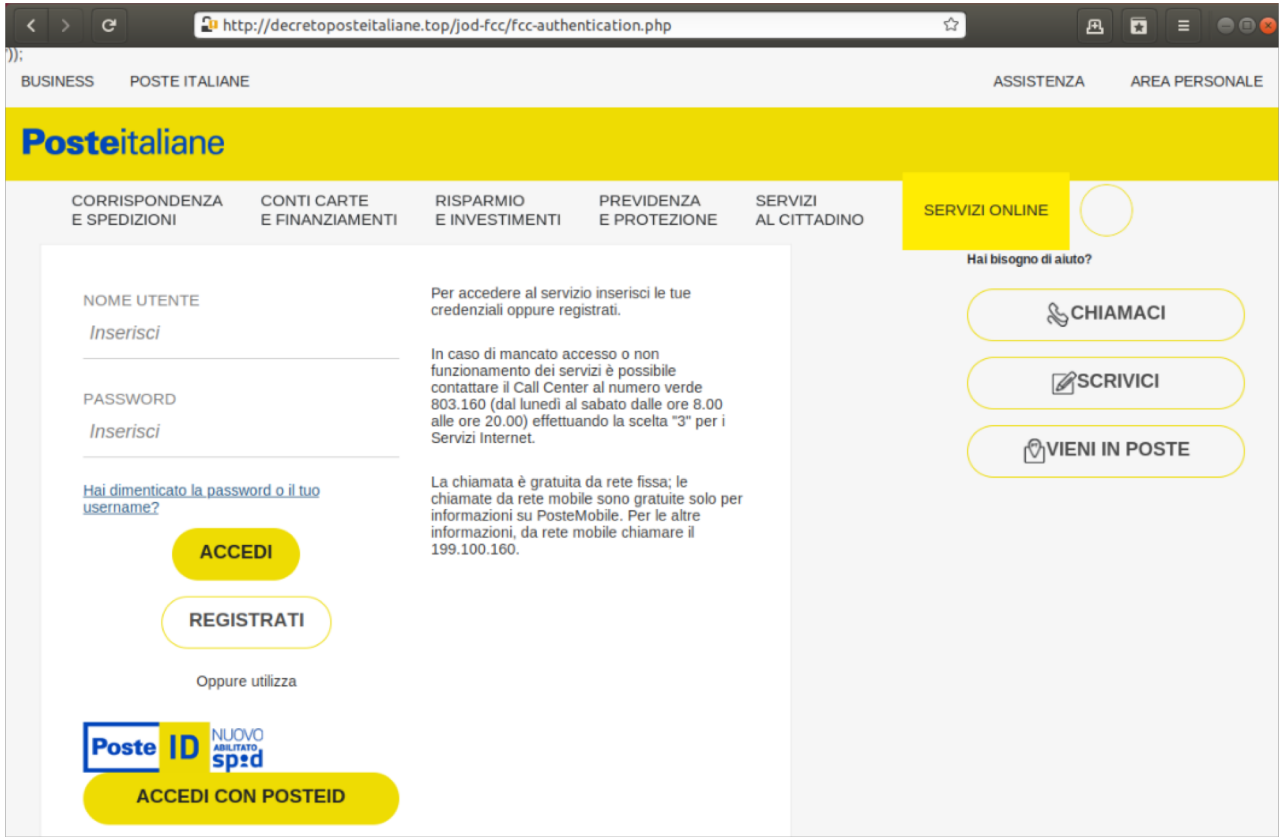
Co-funded by the Erasmus+ Programme of the European Union

A Fast-Flux Domain – decretoposteitaliane.top



Co-funded by the Erasmus+ Programme of the European Union

A Fast-Flux Domain – decretoposteitaliane.top



Co-funded by the Erasmus+ Programme of the European Union

A Fast-Flux Domain – decretoposteitaliane.top

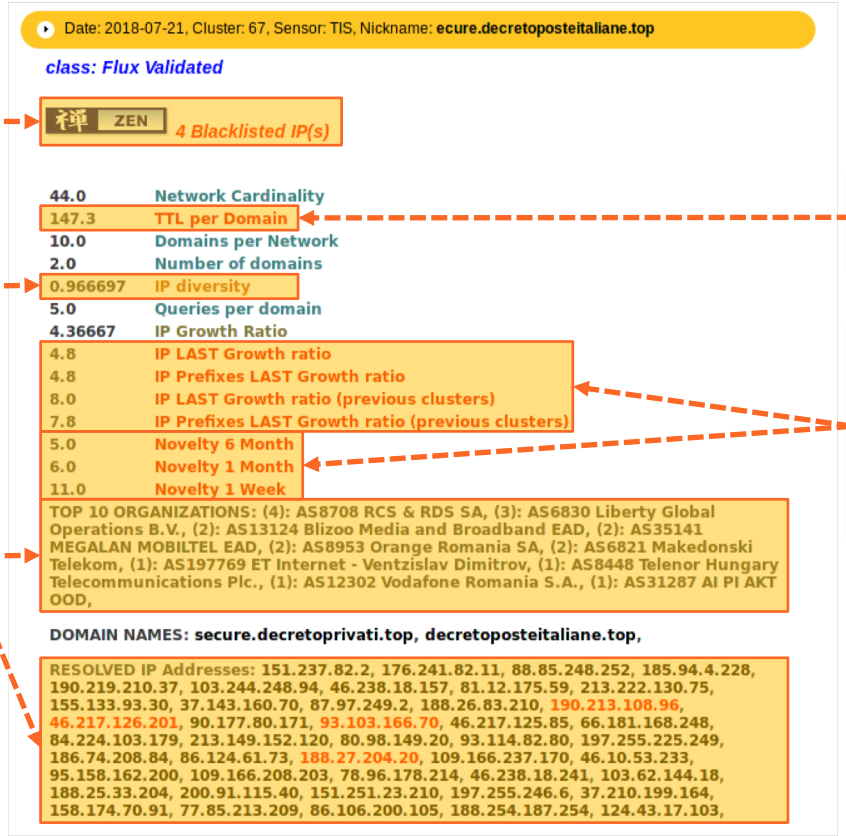
Blacklisted IPs

High IP diversity (entropy of IP /24 prefixes)

Resolved IP addresses spread over many different Autonomous Systems / Organizations.

Time To Live (TTL) << than in legitimate networks

High ratio of previously unseen IP addresses (novelty).
The number of IPs as well as the number of networks they do belong tends to grow.



Co-funded by the Erasmus+ Programme of the European Union

Domain Squatting

Co-funded by the
Erasmus+ Programme
of the European Union

Cybersquatting

Anticybersquatting Consumer Protection Act – U.S. 1999

- *A **person** shall be **liable** in a civil action **by the owner of a mark if** […]*
  - *‘(ii) **registers**, traffics in, or uses **a domain name that**—*
    - *‘(I) **in the case of a mark** that is distinctive at the time of registration of the domain name, **is identical or confusingly similar to that mark**; […]*
    - *‘(III) **is a trademark, word**, or name protected by reason of section 706 of title 18, United States Code, or section 220506 of title 36, United States Code.*

- **Cybersquatting is used to make cyber-attacks more effective (less noticeable), and in particular:**
  - **Phishing**
    - Well crafted domain names help to make phishing campaigns more effective
  - **Scams**
    - Are vehiculated by using well-crafted domain names
    - Often using social networks and messaging platforms (e.g. Whatsapp)

Co-funded by the Erasmus+ Programme of the European Union

Squatting Types

Domain Name	Squatting Type
youtube.com	Original Domain
yewtube.com	Homophone-Based Squatting
youtubg.com	Bitsquatting
YOUTUBE.COM	Homograph-Based Squatting
youtubee.com	Typosquatting

- **Typosquatting** generally refers to the practice of registering domains which are minor typographical variations of popular domain names.
- **Possible replacement strategies:**
  - **Missing Dot** → wwwyoutube.com
  - **Character Omission** → yutube.com
  - **Character Duplication** → youutube.com
  - **Character Permutation** → yuotube.com
  - **Character Substitution** → yoyube.com

Source: P. Kintis et. Al., *Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse, ACM CCS, 2017*
Source: Farsight Security, Global Internationalized Domain Name Homograph Report, Q2-2018.

Co-funded by the Erasmus+ Programme of the European Union

Google & Facebook Monitoring - Results

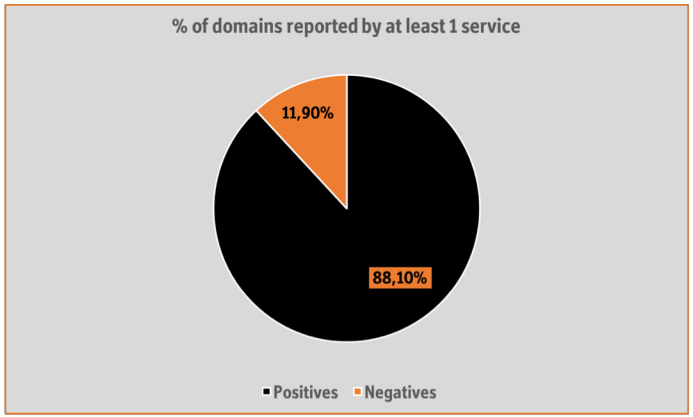
- We developed a detection algorithm based on **string similarity measures**
  - No network features have been used

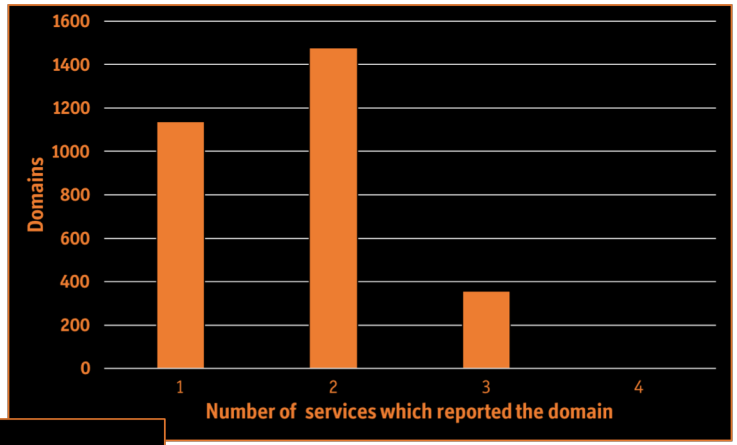
- **We monitored for a period of 4 months** (August 1st – November 30, 2016) ***google.com*** & ***facebook.com*** and their subdomains
  - About 1.4 unique domain names requested every day (approximately 30GB/day of rDNS data…) by the customers of a major Italian ISP

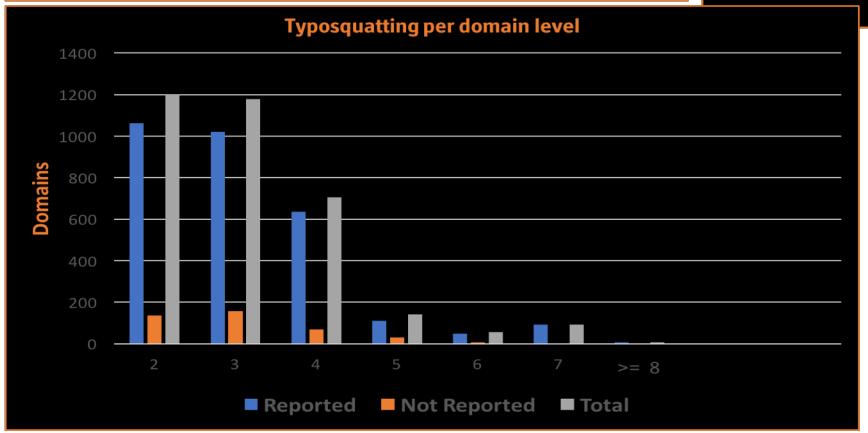
- We validated our detection results:
  - **Comparing them against 4 reputable analysis services:**
    - VirusTotal - www.virustotal.com
    - IBM X-Force Exchange - **exchange.xforce.ibmcloud.com**
    - DNSBL – **www.dnsbl.info**
    - Google Safebrowsing - **developers.google.com/safe-browsing/**
  - **Considering a domain malicious either if:**
    - It is blacklisted
    - One of the IPs on which it resolves is blacklisted

Co-funded by the Erasmus+ Programme of the European Union

Results Related to Facebook.com





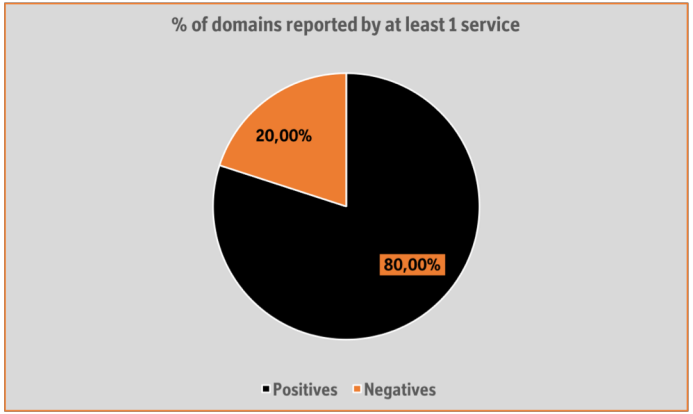


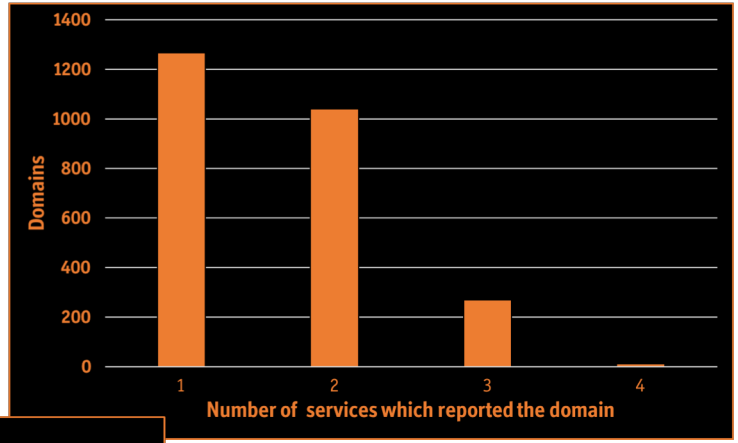
3.380 facebook.com typo-domains

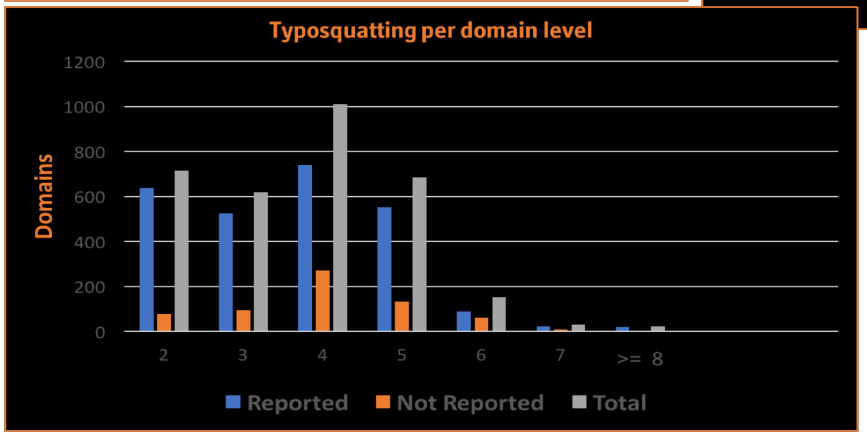
2.978 considered as positives by at least 1 validation service

Co-funded by the Erasmus+ Programme of the European Union

Results Related to Google.com







3.237 google.com typo-domains

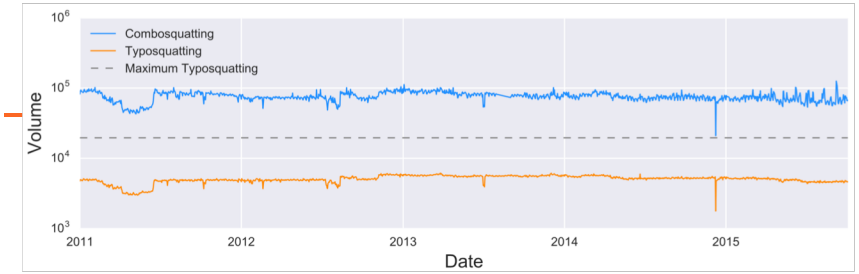
2.590 considered as positives by at least 1 validation service

Co-funded by the Erasmus+ Programme of the European Union

Squatting Types - Combosquatting

Domain Name	Squatting Type
youtube.com	Original Domain
yewtube.com	Homophone-Based Squatting
youtubg.com	Bitsquatting
YOUTUBE.COM	Homograph-Based Squatting
youtubee.com	Typosquatting
youtube-login.com	Combosquatting

- **Combosquatting**
  - Attackers register domains that combine a popular trademark with one or more phrases:
    - better**facebook.com**



Source: P. Kintis et. Al., *Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse, ACM CCS, 2017*
Source: Farsight Security, Global Internationalized Domain Name Homograph Report, Q2-2018.

Co-funded by the Erasmus+ Programme of the European Union

Squatting Types - Combosquatting

Domain Name	Squatting Type
youtube.com	Original Domain
yewtube.com	Homophone-Based Squatting
youtubg.com	Bitsquatting
YOUTUBE.COM	Homograph-Based Squatting
youtubee.com	Typosquatting
youtube-login.com	Combosquatting

- **Combosquatting**
  - Attackers register domains that combine a popular trademark with one or more phrases:
    - better**facebook.com**
    - **youtube**-live.**com**
  - **Higher degree of freedom**
  - **More expensive for the attacker, as:**
    - Lacks of generative models
    - Requires miscreants to coerce users to visit combosquatting domains

Source: P. Kintis et. Al., *Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse, ACM CCS, 2017*
Source: Farsight Security, Global Internationalized Domain Name Homograph Report, Q2-2018.

Co-funded by the Erasmus+ Programme of the European Union

Squatting Types - Combosquatting

Domain Name	Squatting Type
youtube.com	Original Domain
yewtube.com	Homophone-Based Squatting
youtubg.com	Bitsquatting
YOUTUBE.COM	Homograph-Based Squatting
youtubee.com	Typosquatting
youtube-login.com	Combosquatting

- **Combosquatting**
  - Attackers register domains that combine a popular trademark with one or more phrases:

Source: P. Kintis et. Al., *Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse, ACM CCS, 2017*
Source: Farsight Security, Global Internationalized Domain Name Homograph Report, Q2-2018.

Co-funded by the Erasmus+ Programme of the European Union

Squatting Types – Homogliph Based Squatting

Domain Name	Squatting Type
youtube.com	Original Domain
yewtube.com	Homophone-Based Squatting
youtubg.com	Bitsquatting
YOUTUBE.COM	Homograph-Based Squatting
youtubee.com	Typosquatting
youtube-login.com	Combosquatting
youṭube.com	Homogliph-Based Squatting

- **Homogliph based squatting**
  - Attackers abuse Internationalized Domain Names
  - Introduced by ICANN in 2003 to enable people to use domain names in local languages and scripts (e.g. Arabic, Chinese, Cyrillic or Devanagari).

Source: P. Kintis et. Al., *Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse, ACM CCS, 2017*
Source: Farsight Security, Global Internationalized Domain Name Homograph Report, Q2-2018.

Co-funded by the Erasmus+ Programme of the European Union

Squatting Types – Homogliph Based Squatting

Domain Name	Squatting Type
youtube.com	Original Domain
yewtube.com	Homophone-Based Squatting
youtubg.com	Bitsquatting
YOUTUBE.COM	Homograph-Based Squatting
youtubee.com	Typosquatting
youtube-login.com	Combosquatting
youṭube.com	Homogliph-Based Squatting

- **Homogliph based squatting**
  - Attackers abuse Internationalized Domain Names
  - **At DNS level, domains are encoded using ASCII characters only**
    - Punycode encoding is used
    - E.g. **youṭube.com** is represented as **xn--youube-k17b.com**
  - **Homogliph Examples**
    - Cyrillic « **а** »
    - Lating jota « **ι** »

Source: P. Kintis et. Al., *Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse, ACM CCS, 2017*
Source: Farsight Security, Global Internationalized Domain Name Homograph Report, Q2-2018.

Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.alitalıa.com

Valido solo per oggi-

Stiamo dando 2 biglietti gratuiti per celebrare il nostro decimo anniversario!

Biglietti rimanenti : Caricamento

Si prega di prendere parte al sondaggio prima:

Domanda 1: Hai mai viaggiato con noi ?

Sì

Non

Non ricordo

Mi Piace · Commenta · Condividi

e altri 65.059

Rachel Singleton Grazie Alitalia, ora posso andare per la mia luna di miele

Mi Piace · Risposta · 50 Â· Proprio ora

Julia Schröder Ho appena fatto una prova e mi chiedo, ho ottenuto il mio 2 biglietti gratuiti

Mi Piace · Risposta · 16 Â· Proprio ora

Michael Drechsler Mi piace volare con loro, sono i migliori

Mi Piace · Risposta · 27 Â· Proprio ora

Johanna Grunewald Grazie Alitalia.

Mi Piace · Risposta · 18 Â· Proprio ora

Scrivi un commento...

Maggiori informazioni

* Questa offerta è valida solo per tempo limitato.

Alitalia-2018

Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.ryanaịr.com

Valido solo per oggi-

Ryanair dà 2 biglietti gratuiti per celebrare il 34 ° anniversario.

Biglietti rimanenti : Caricamento

Si prega di prendere parte al sondaggio prima:

Domanda 1: Hai mai viaggiato con noi?

Sì

Non

Mi Piace Commenta Condividi

e altri 65.059

Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.ryanaīr.com

Valid for today only-

Ryanair donne 2 billets gratuits pour célébrer le 34e anniversaire.

Billets restants : Chargement

S'il vous plaît prendre part à l'enquête d'abord:

Question 1: Avez-vous déjà voyagé avec nous ?

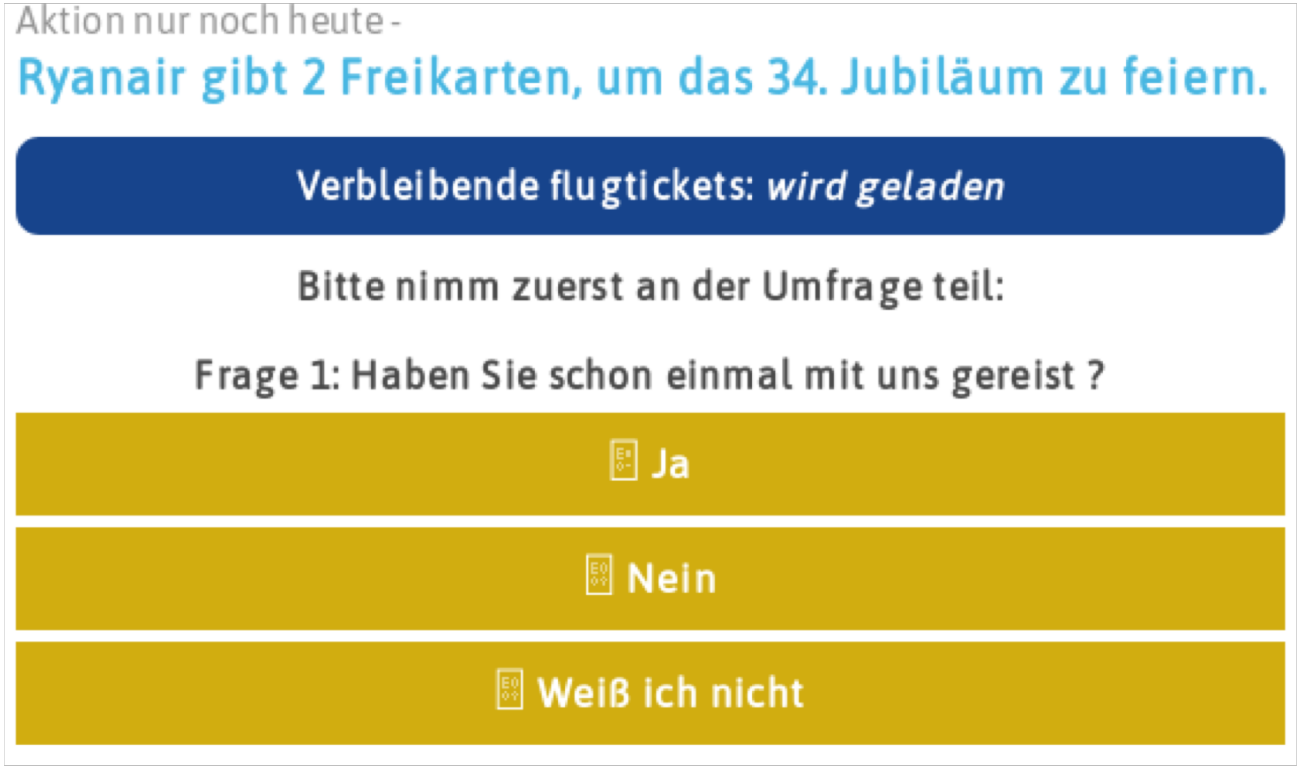
Oui

Non

Ne me souviens pas

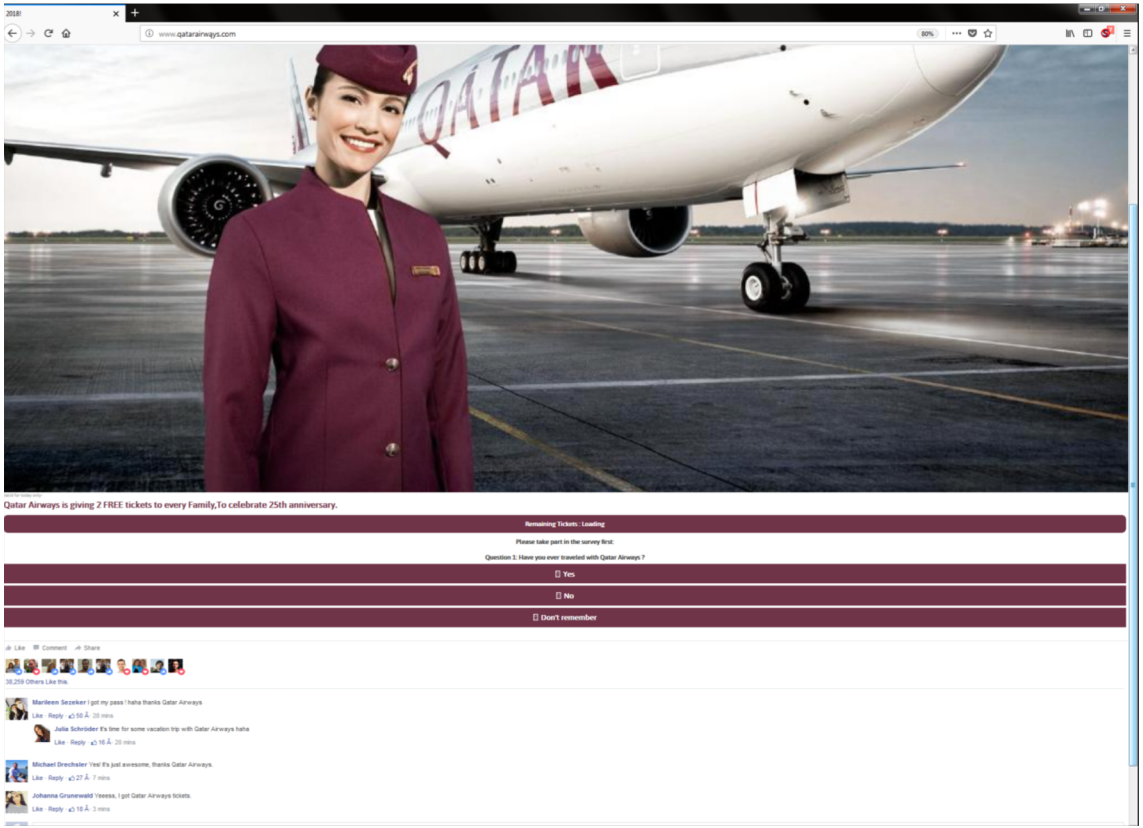
Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.ryanair.com



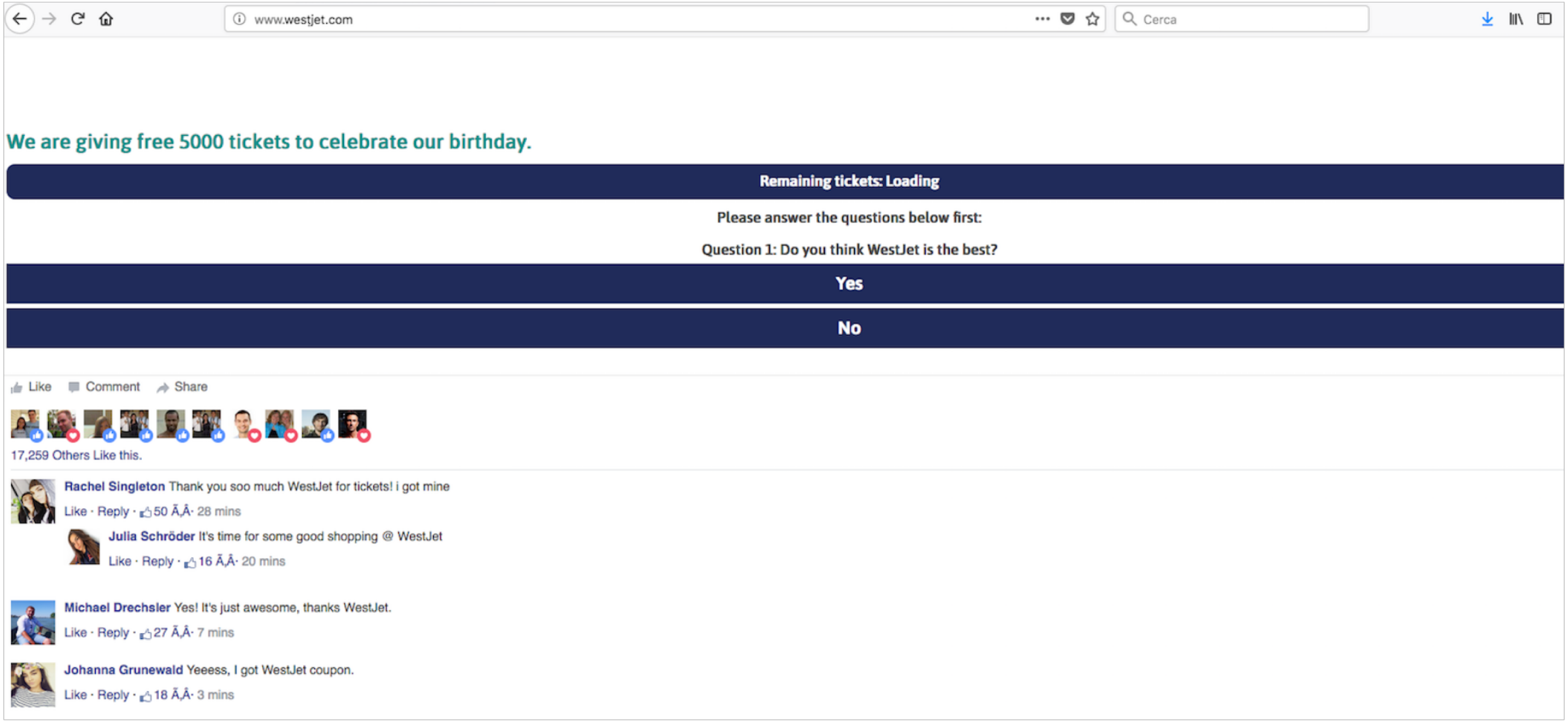
Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.qatarairways.com



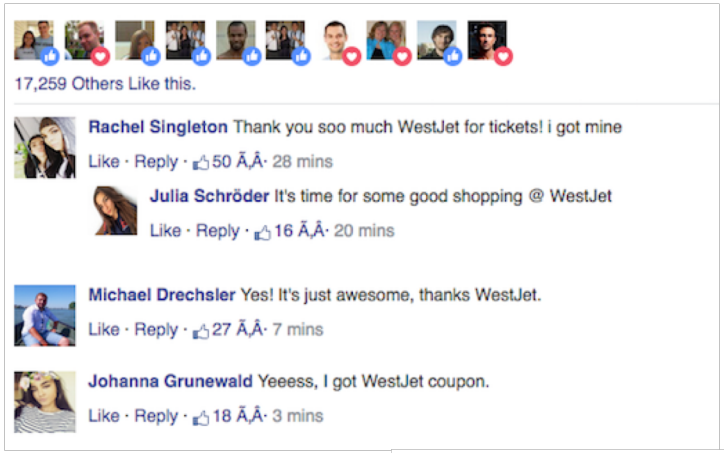
Co-funded by the Erasmus+ Programme of the European Union

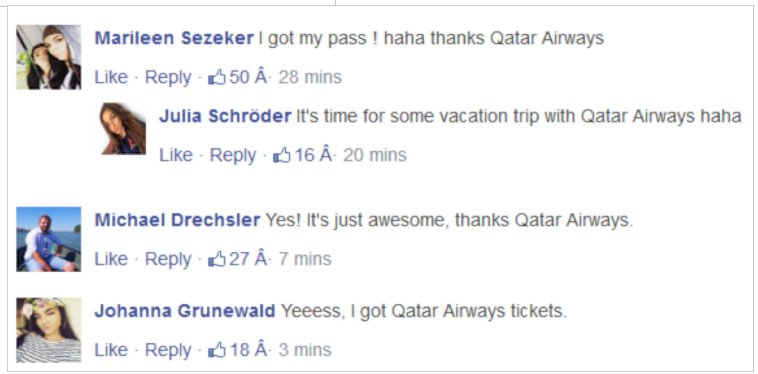
Homogliph Based Squatting – www.westjet.com



Co-funded by the Erasmus+ Programme of the European Union

Look at the winners…



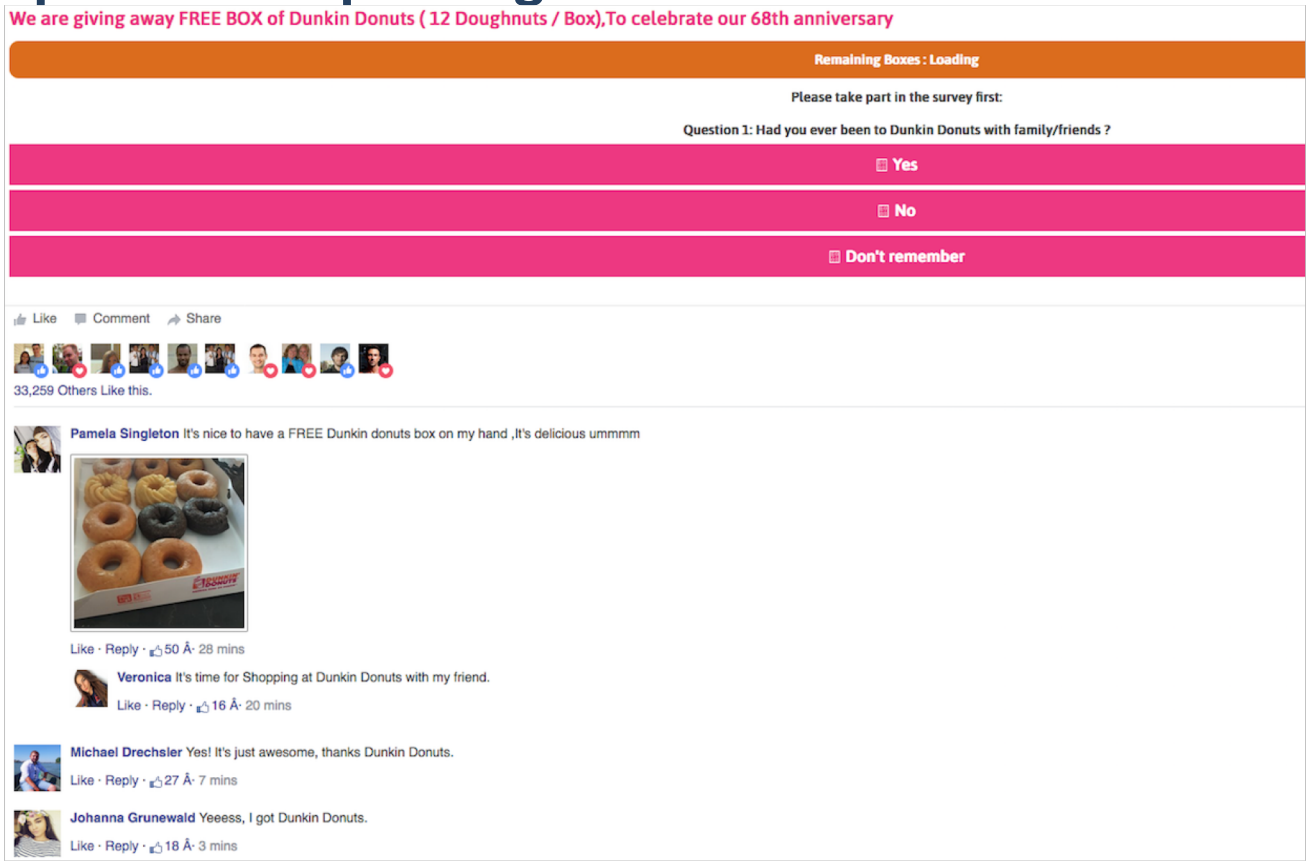


Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.nutęlla.com

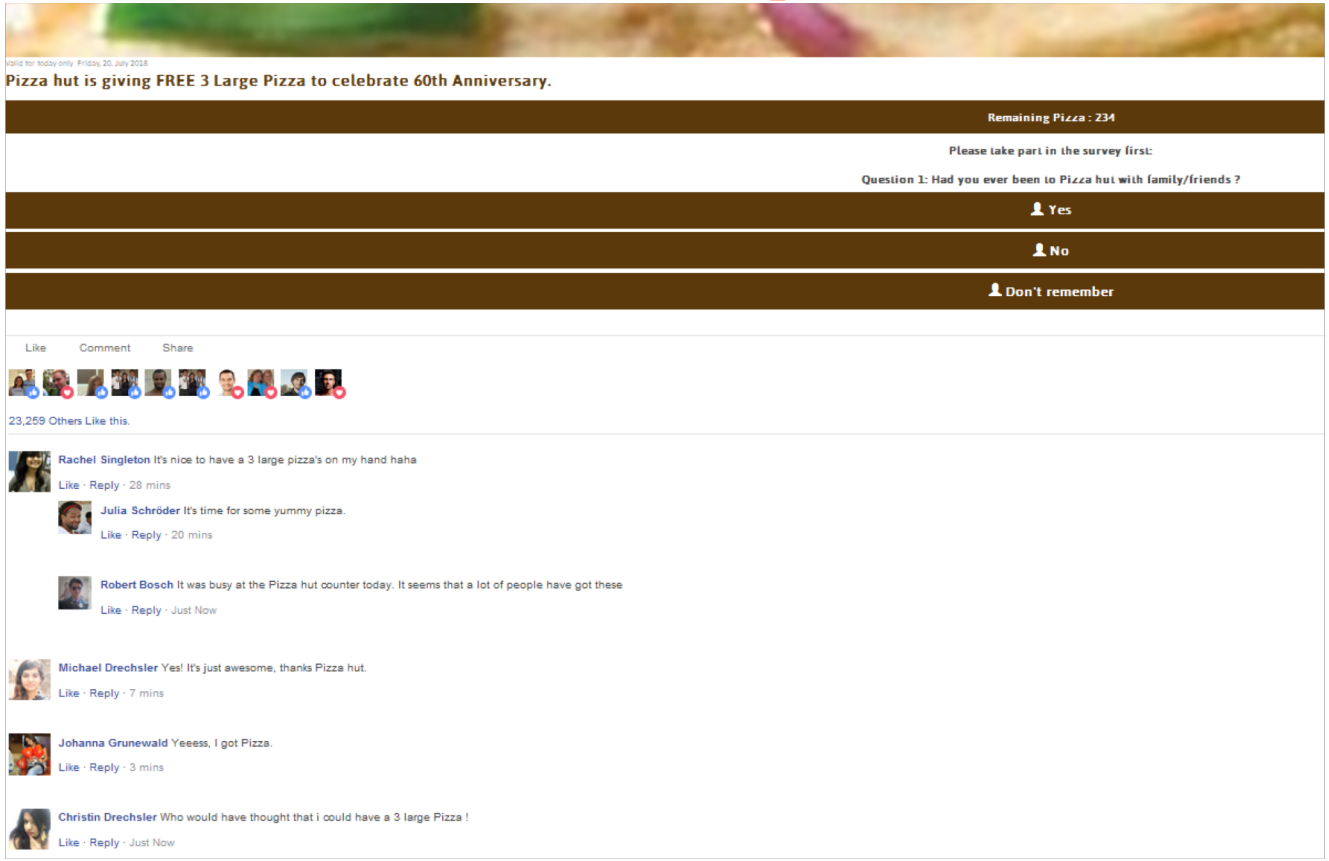
Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.dunkindonuts.com



Co-funded by the Erasmus+ Programme of the European Union

Homogliph Based Squatting – www.pizzahut.online



Co-funded by the Erasmus+ Programme of the European Union

Lessons Learned

- The DNS traffic is a primary source of information regarding what is happening on the network
  - **Always worth to look at it!**
- The analysis should be done on site:
  - External blacklisting services are really effective only against large scale campaigns:
    - Short-lived domains can easily evade them
    - **You should monitor and analyse your own traffic!**
- Challenges
  - Ad-hoc detection algorithms for every specific threat
    - Many multi-faceted threats to face
  - Validation of the results

Co-funded by the Erasmus+ Programme of the European Union

Thank you for your attention!

Co-funded by the
Erasmus+ Programme
of the European Union