Fraunhofer FKIE

botconf2016

Analysis of Free Movies and Series Websites Guided by User Search Terms

Botconf 2016 - Lyon, France

Luis Alberto Benthin Sanguino, Martin Clauß
luis.alberto.benthin.sanguino@fkie.fraunhofer.de
martin.clauss@fkie.fraunhofer.de

CA&D | Cyber Analysis and Defense

Agenda

Fraunhofer FKIE

botconf2016

1. Motivation
2. Evaluation Method
3. Results
4. Conclusion

Motivation

Fraunhofer FKIE

botconf2016

1. **Why do we analyse websites?**
2. **Why a reduced subset of the web?**
3. **Why Free Movies and Series (FMS) websites?**

Motivation

Fraunhofer FKIE

botconf2016

1. Why do we analyse websites?

- Websites used by cybercriminals to infect users with malware

⇒ **Why do cybercriminals use websites as infection vector?**

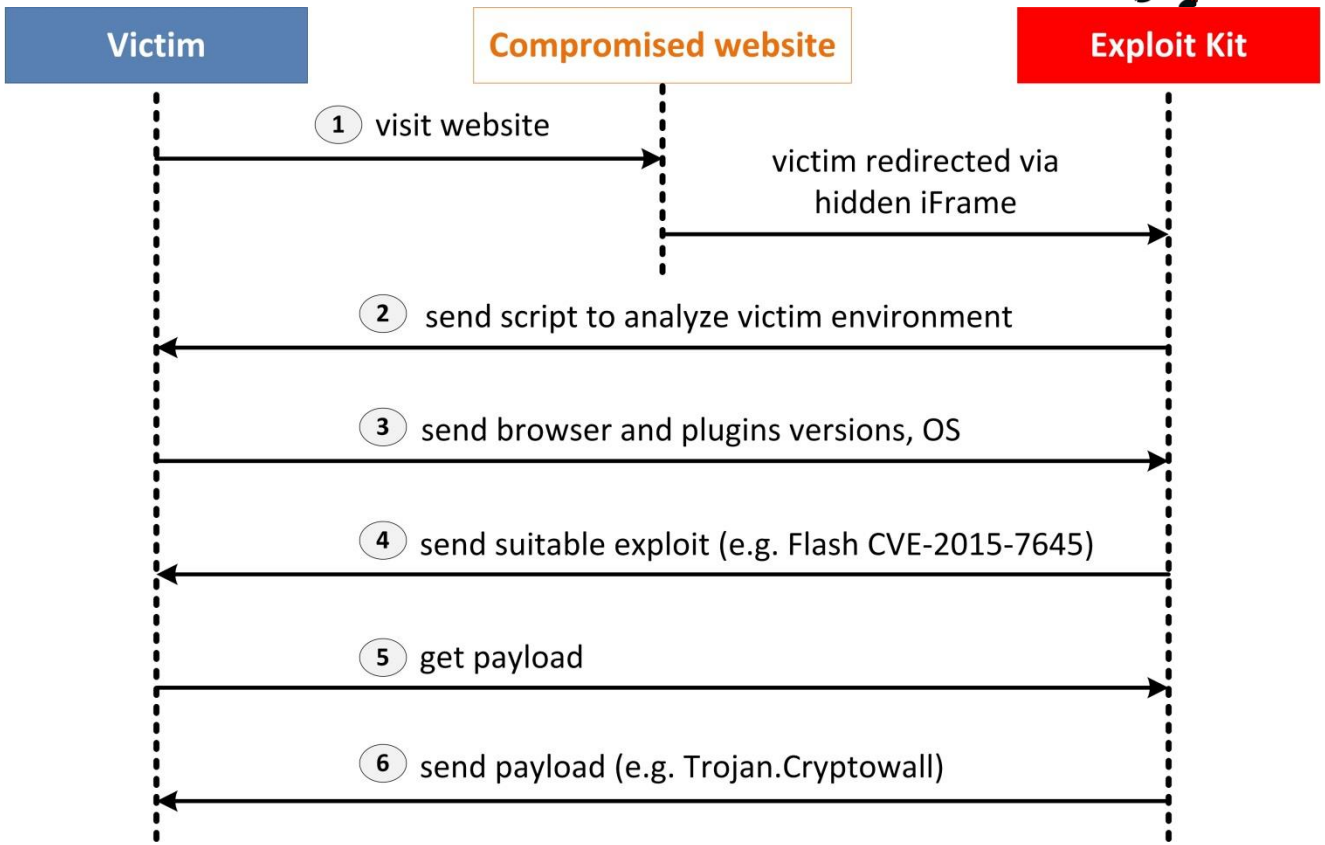
  - **Efficient**
    - Massive number of possible victims
  - **Effective**
    - **Server side** Legitimate websites can be compromised
      - Via e.g. vulnerable CMS
    - **Client side** Web users can be infected
      - Via vulnerable versions of e.g. IE, Flash, PDF

Motivation

Fraunhofer FKIE

botconf2016

- **How do web users get infected?**



Motivation

Fraunhofer FKIE

botconf2016

2. Why do we focus on a subset of the web?

- 5 billion pages and 1.1 billion websites [1, 2]

  ⇒ Detection and analysis resources must be spent efficiently

3. Why Free Movies and Series (FMS) websites?

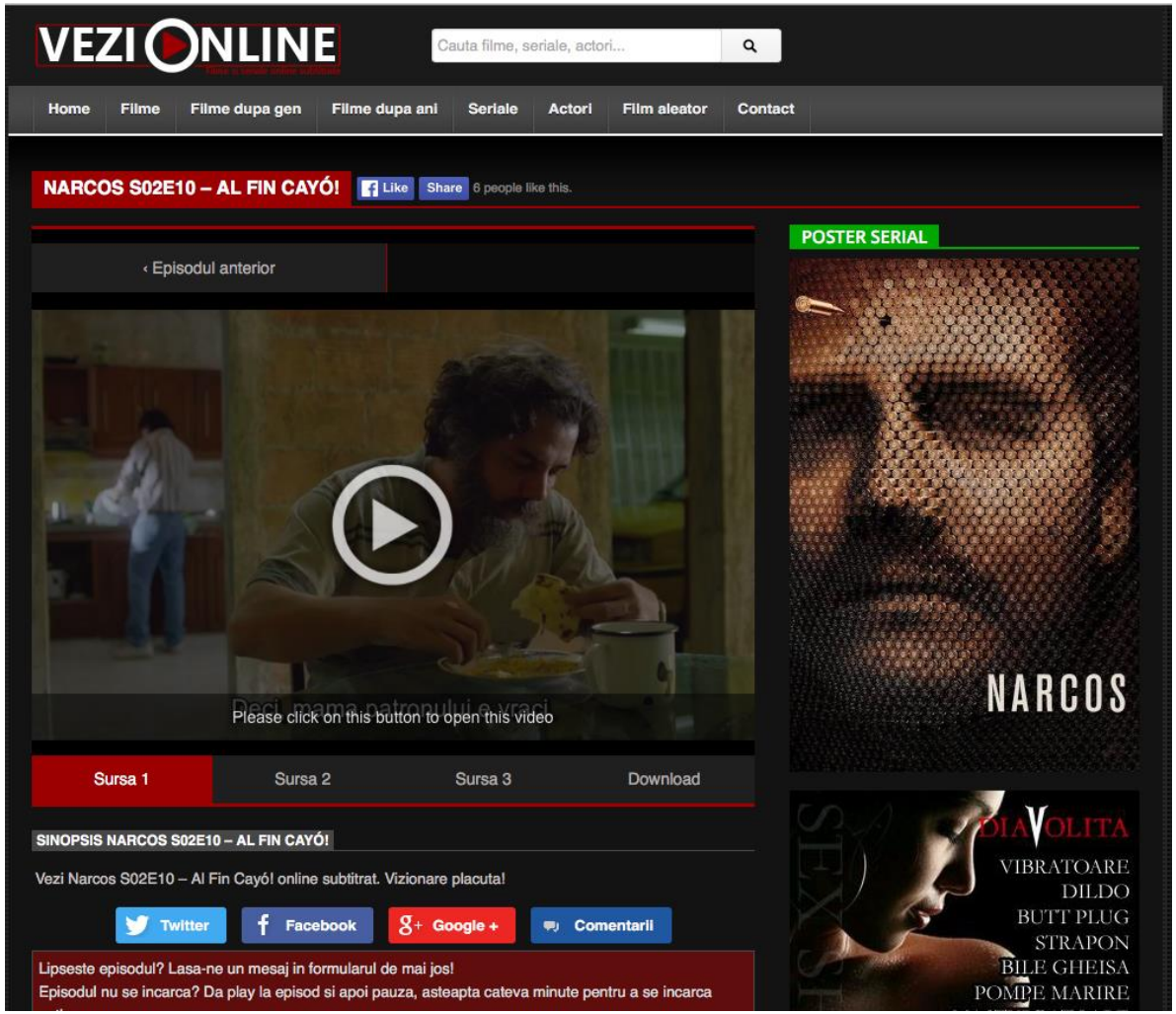
- Free entertainment content attract millions of web users
  - Why to pay, if it can be obtained for free?
- Cybercriminals use free services websites to expose users to malware
- Manual analysis showed that FMS websites try to trick users
  - Malicious Flash installer
  - Malicious video converter
  - Possible scam

[1] http://www.worldwidewebsize.com
[2] http://www.internetlivestats.com/total-number-of-websites

Motivation

Fraunhofer FKIE

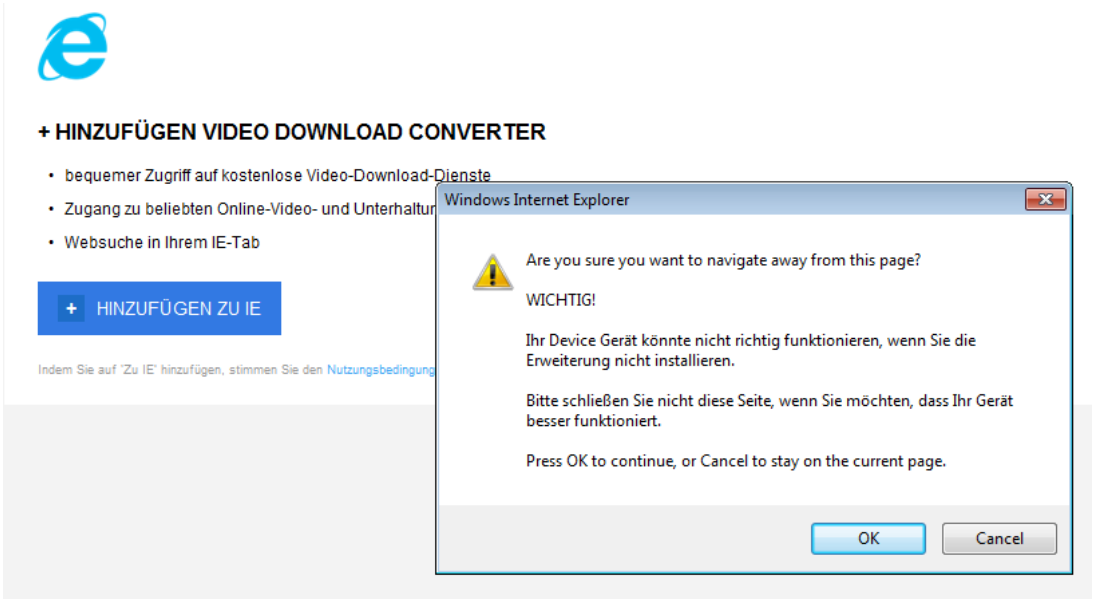
botconf2016



Motivation

Fraunhofer FKIE

botconf2016



virustotal

SHA256: 7edad9ba600a2b20a86574829a0246a8502297c337122f966aba3cbd2e73f3ca

Dateiname: VideoDownloadConvert.b3d5b696659745cb9cede9cb5b0f0083.exe

Erkennungsrate: 26 / 57

Analyse-Datum: 2016-10-27 22:23:05 UTC (vor 4 Tage, 12 Stunden)

0 0

Analyse | File detail | Zusätzliche Informationen | Kommentare 0 | Bewertungen | Verhaltens-Informationen

Antivirus	Ergebnis	Aktualisierung
AVware	MyWebSearch.J (v)	20161027
Antiy-AVL	GrayWare[:not-a-virus]/Win32.StartPage.gen	20161027
Arcabit	PUP.WebToolbar.MyWebSearch	20161027
Avast	Win32:PUP-gen [PUP]	20161027

Motivation

Fraunhofer FKIE

botconf2016

virustotal

SHA256: a785f4a2d12c8df69307ecea082074fa6cf0c068a2e732e73234c0301e35e0fe

Dateiname: adobe_flash_setup.exe

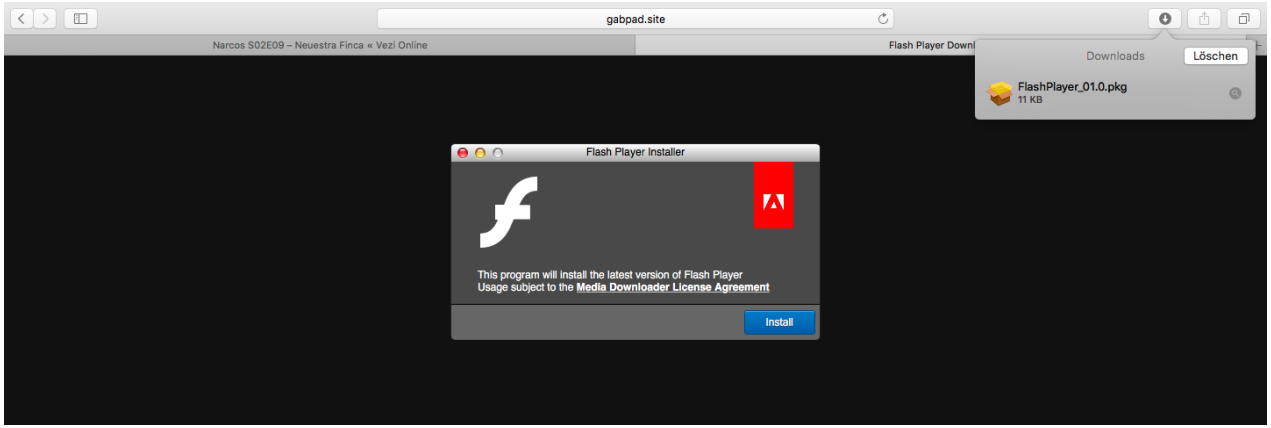
Erkennungsrate: 13 / 56

Analyse-Datum: 2016-11-01 10:41:14 UTC (vor 0 Minuten)

0 0

Analyse | File detail | Zusätzliche Informationen | Kommentare | Bewertungen

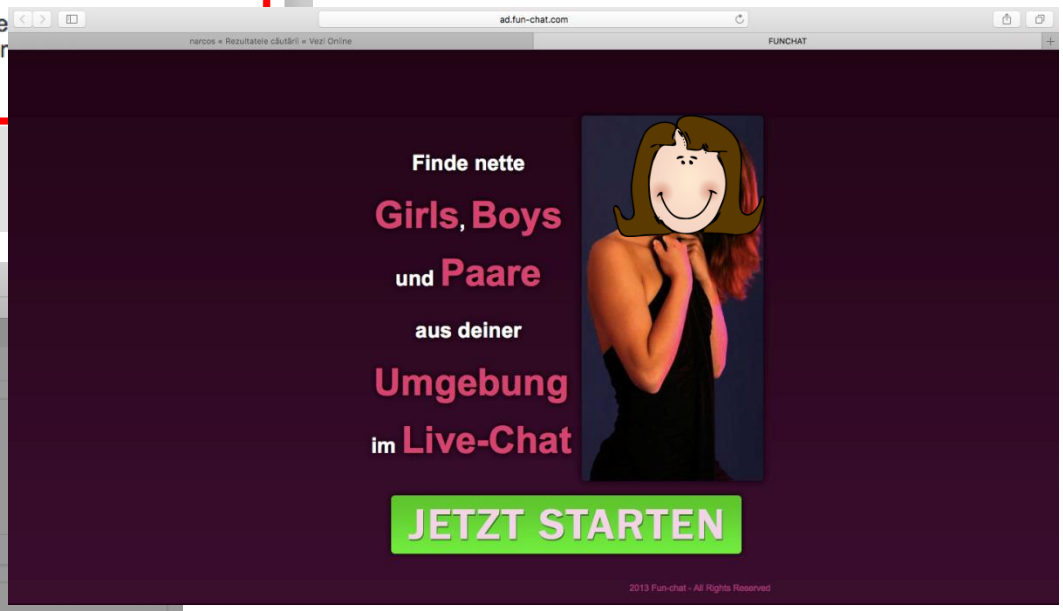
Antivirus	Ergebnis	Aktualisierung
AVG	Generic.F15	20161101
Antiy-AVL	GrayWare[Adware]/Win32.InstallCore.gena	20161101
Bkav	W32.HfsAdware.8CF1	20161031
CAT-QuickHeal	PUA.Ookod.Gen	20161101

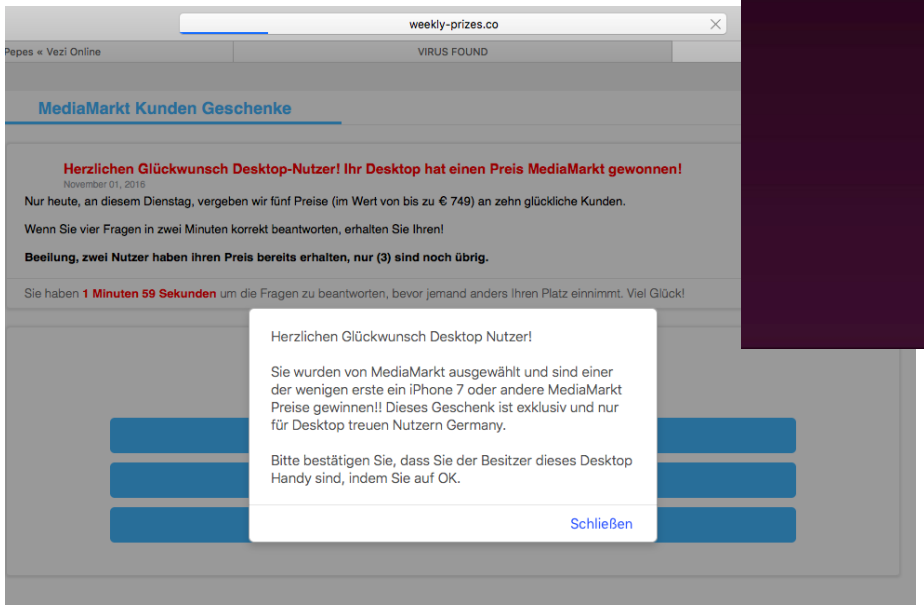


Motivation

Fraunhofer FKIE

botconf2016





Goals

Fraunhofer FKIE

botconf2016

Goals of our research

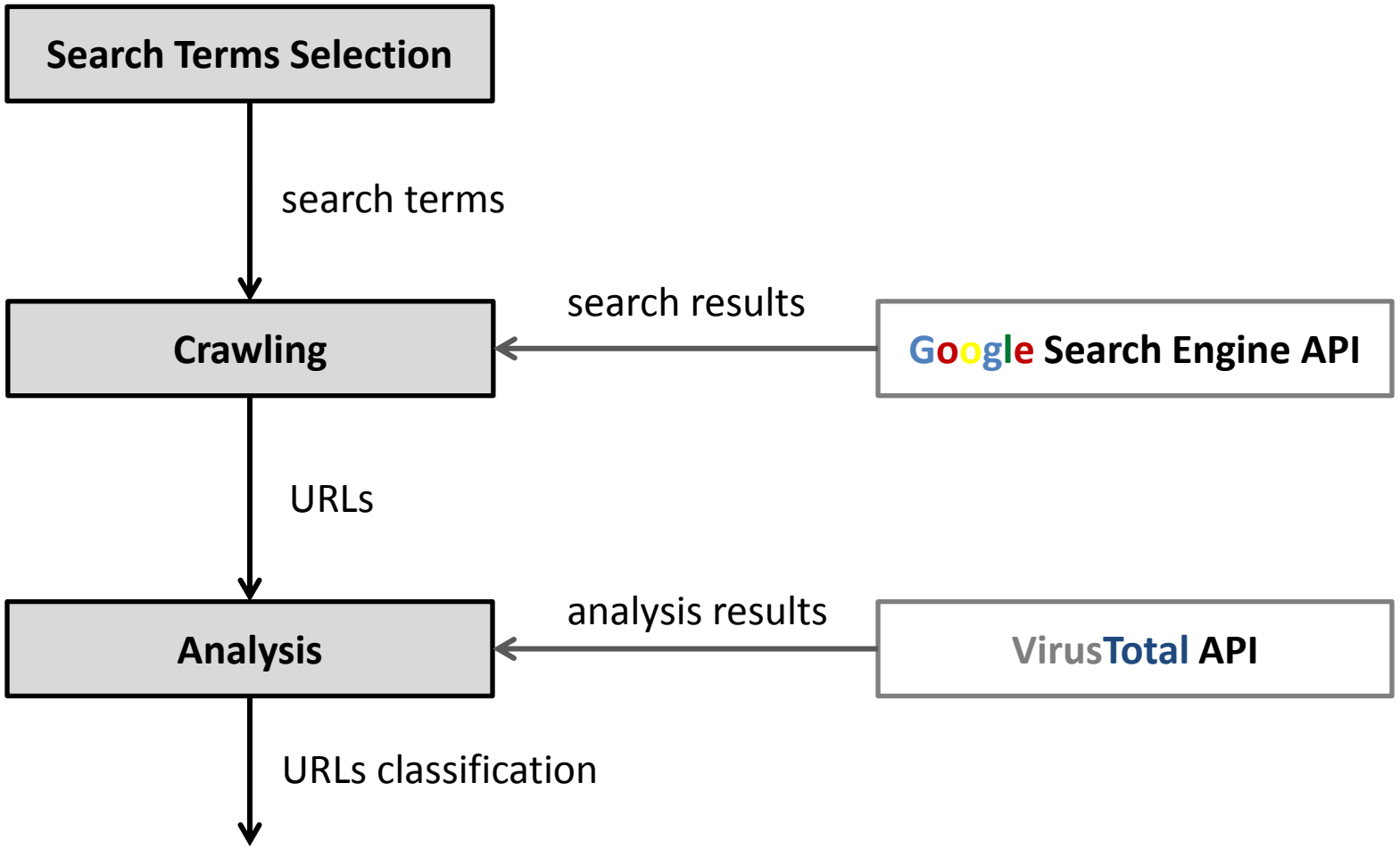
1. **Hypothesis:** FMS websites are more dangerous than other web categories
2. Approach to analyze potentially malicious websites categories guided by **user search terms**
3. Easy-to-implement infrastructure to analyze website categories

Evaluation Method

Fraunhofer FKIE

botconf2016

3-Phase method



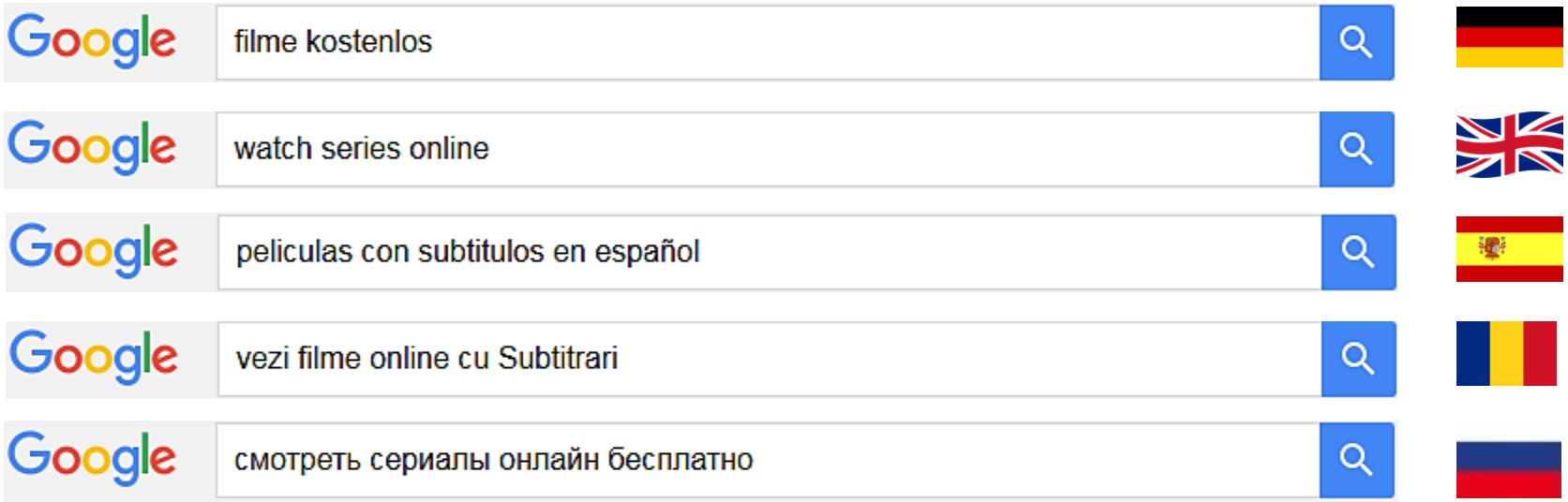
Evaluation Method

Fraunhofer FKIE

botconf2016

Search terms selection – FMS category

1. Native speaker search terms



- **Output** 25 search terms

Evaluation Method

Fraunhofer FKIE

botconf2016

Search terms selection – FMS category

2. Google Trends

- **Output** 50 search terms

Evaluation Method

Fraunhofer FKIE

botconf2016

Search terms selection – Other categories

- **Google Trends (GT15) 2015 report**

Category	Search term e.g.	Category	Search term e.g.
1. Music Artists	Adele	**4. Loss**	Bobbi Kristina
2. Global News	Charlie Hebdo	**5. Consumer Tech**	iPhone 6S
3. Sporting Events	Copa America	**6. People**	Lamar Odom

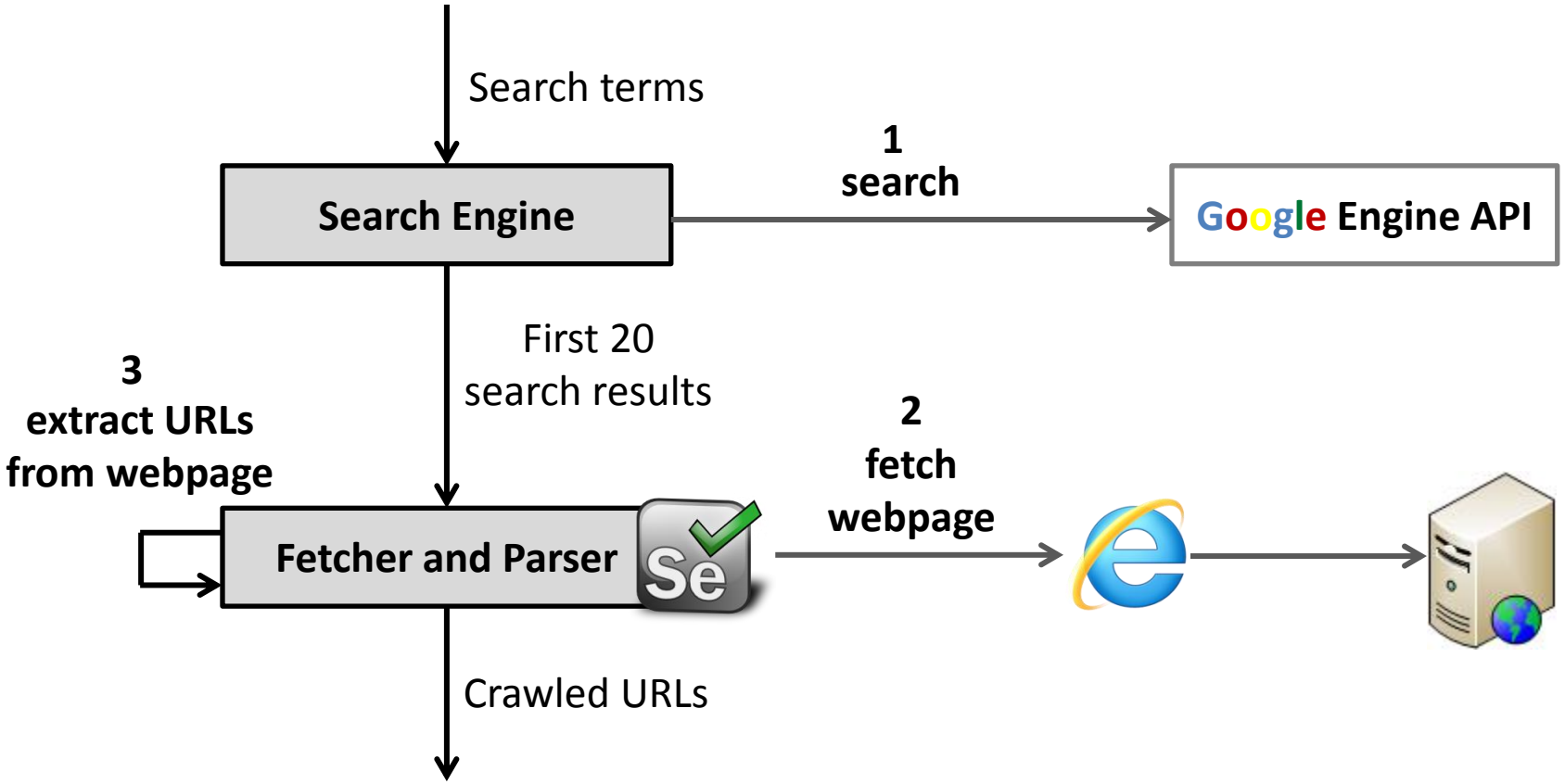
  - **Output** 60 search terms

Evaluation Method

Fraunhofer FKIE

botconf2016

Phase 2: Crawling



Selenium: web browser automation tool, http://docs.seleniumhq.org/.

Evaluation Method

Fraunhofer FKIE

botconf2016

Phase 3: Analysis

Fraunhofer FKIE

botconf2016

Interpretation of VirusTotal Results

- **Positives**
  - Number of scanners that classified a URL as:
    - malicious site
    - malware site
    - phishing site
- **Positive URL**
  - positives > 0
- **Positive Rate**
  - Percent of positive URLs in relation with the number of analyzed URLs

Results

Fraunhofer FKIE

botconf2016

FMS Category

	Scanned	# Positives	Positive Rate (%)
URLs	18110	1261	**6.96**
Domains	1157	117	**10.11**

URLs

Lang.	Positive Rate (%)
ES	**11.35**
RO	9.25
EN	8.16
RU	3.66
DE	**0.34**

Domains

Lang.	Positive Rate (%)
ES	**13.27**
RO	13.13
RU	11.64
EN	11.11
DE	**3.36**

Results

Fraunhofer FKIE

botconf2016

GT15 Categories

	Scanned	# Positives	Positive Rate (%)
URLs	34793	178	**0.51**
Domains	3103	33	**1.06**

URLs

Category	Positive Rate (%)
Music artists	**1.89**
Consumer tech	0.79
Sporting events	0.21
People	0.19
Loss	0.14
Global news	**0.11**

Domains

Category	Positive Rate (%)
Consumer tech	**3.10**
Music artists	1.55
Global news	1.18
Sporting events	1.09
Loss	0.84
People	**0.60**

Results

Fraunhofer FKIE

botconf2016

FMS category vs. GT15 categories

URLs

Category	Positive Rate (%)
FMS	6.96
GT15	0.51

Domains

Category	Positive Rate (%)
FMS	10.11
GT15	1.06

Results

Fraunhofer FKIE

botconf2016

Positive Domain Geolocation - FMS Category

- **38 % hidden** behind **CloudFlare** [3] DNS service
  - We uncovered 57 % using **CrimeFlare** [4] database
- EU 36 %
- USA 32 %
- Russia 9 %

[3] https://www.cloudflare.com
[4] http://www.crimeflare.com

Fraunhofer FKIE

botconf2016

Conclusions

- ✓ FMS category more malicious than GT15 categories
  - Hypothesis proven
- ✓ Analysis based on real user search terms
  - Analyzed websites are likely to be visited by web users
- ✓ Spanish search terms with the highest positive rate
- ✓ Proposed method requires low-computational costs to be implemented

Future work

- Determine the causes why FMS websites were classified as malicious
- Why ES search terms have the highest positive rate?

Thank you!

Luis Alberto Benthin Sanguino, Martin Clauß
luis.alberto.benthin.sanguino@fkie.fraunhofer.de
martin.clauss@fkie.fraunhofer.de

CA&D | Cyber Analysis and Defense