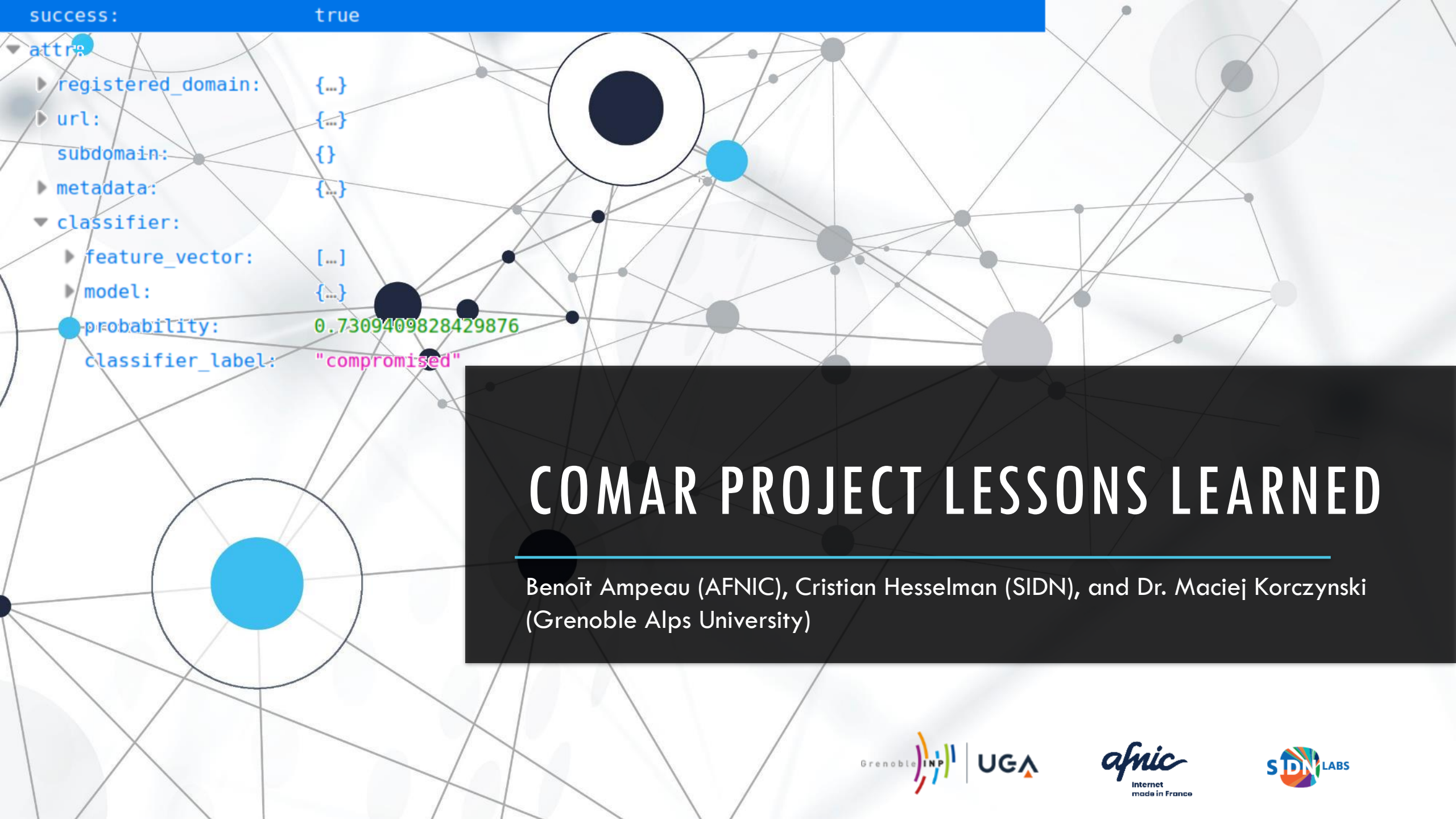


COMAR PROJECT LESSONS LEARNED

Benoît Ampeau (AFNIC), Cristian Hesselman (SIDN), and Dr. Maciej Korczynski (Grenoble Alps University)

COMAR[1]: CLASSIFICATION OF COMPROMISED VERSUS MALICIOUSLY REGISTERED DOMAINS

COMAR is a **machine-learning system** that distinguishes maliciously registered domains from compromised domain names

- Domain names **registered by cybercriminals** solely for fraudulent purposes
- **Benign but hacked domain names** exploited at the hosting level, mainly by taking advantage of vulnerabilities in web applications.

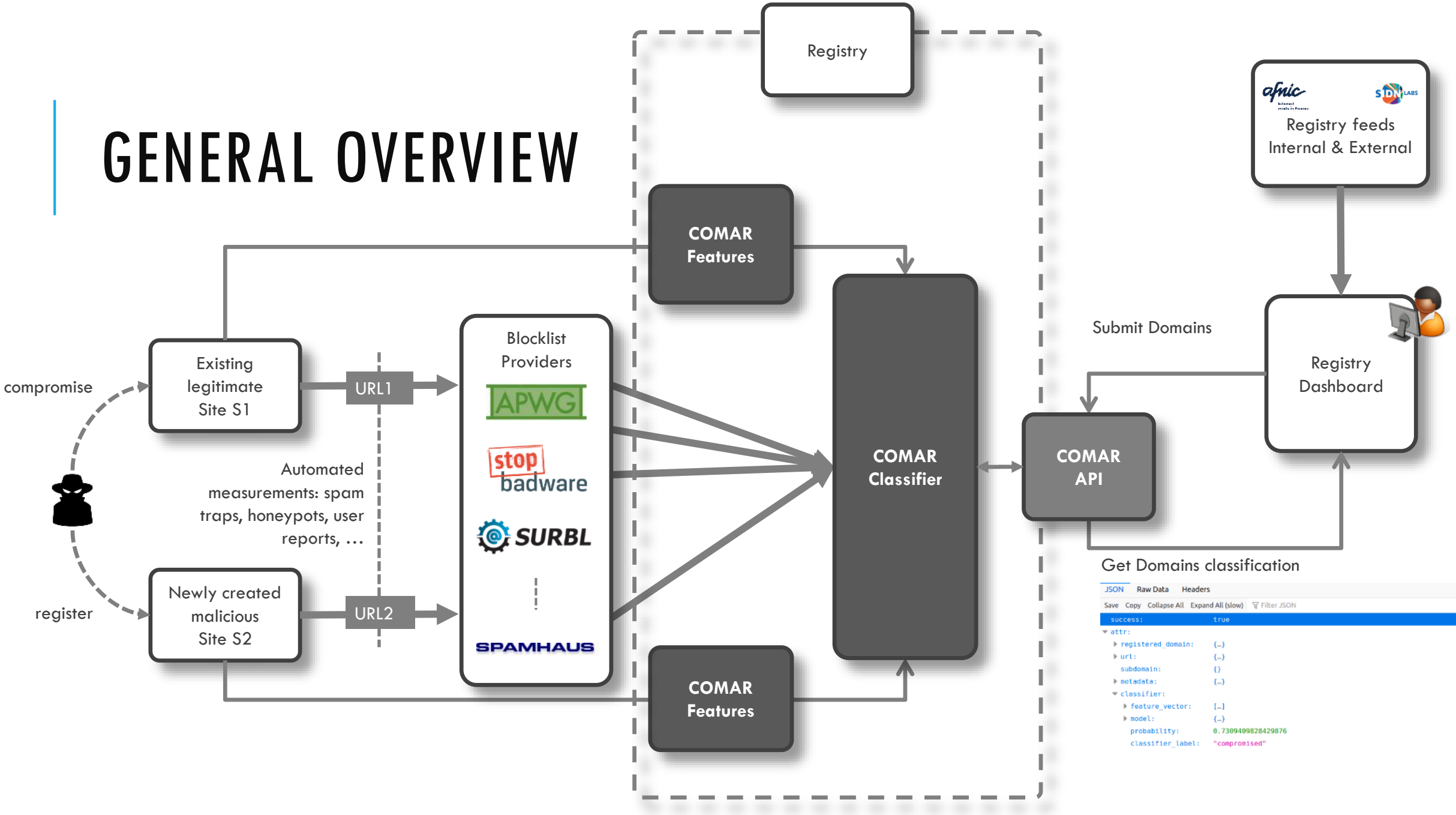
Relevant for **anti-abuse teams** because they need to be handled differently

- Simply take maliciously registered domain names and their content offline
- Can't take offline, need to work with hosting providers to get the vulnerabilities fixed

Dataset: **blocklists** from security companies such as APWG

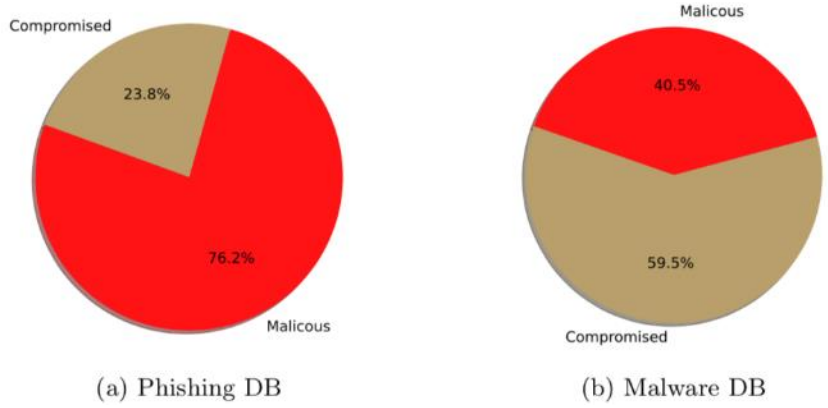
(1) COMAR: Classification of Compromised versus Maliciously Registered Domains" Sourena Maroofi and Maciej Korczyński and Cristian Hesselman and Benoit Ampeau and Andrzej Duda, 2020 IEEE European Symposium on Security and Privacy (EuroS&P), 2020.

GENERAL OVERVIEW



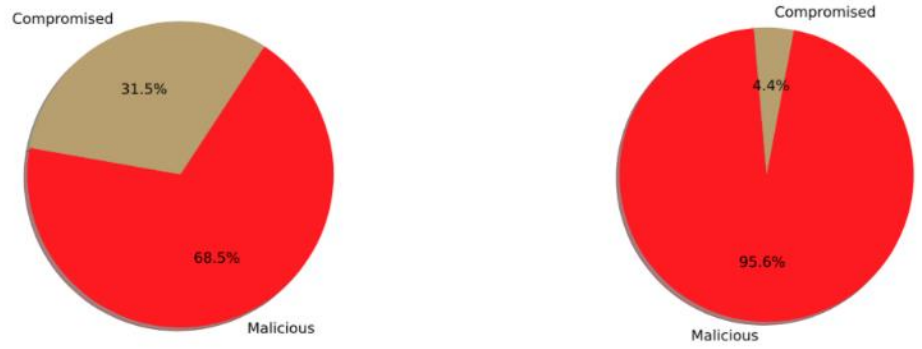
TLD LEVEL ANALYSIS

PHISHING AND MALWARE DELIVERY URLS COLLECTED THROUGH THE FIRST SIX MONTHS OF 2021

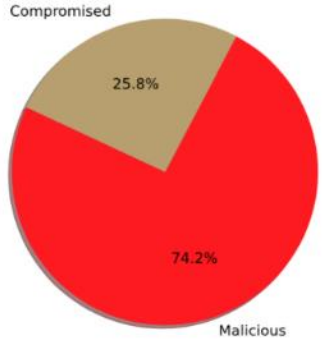


(a) Phishing DB (b) Malware DB

Overall classification of Malware and Phishing URLs



(a) Legacy gTLDs (b) New gTLDs



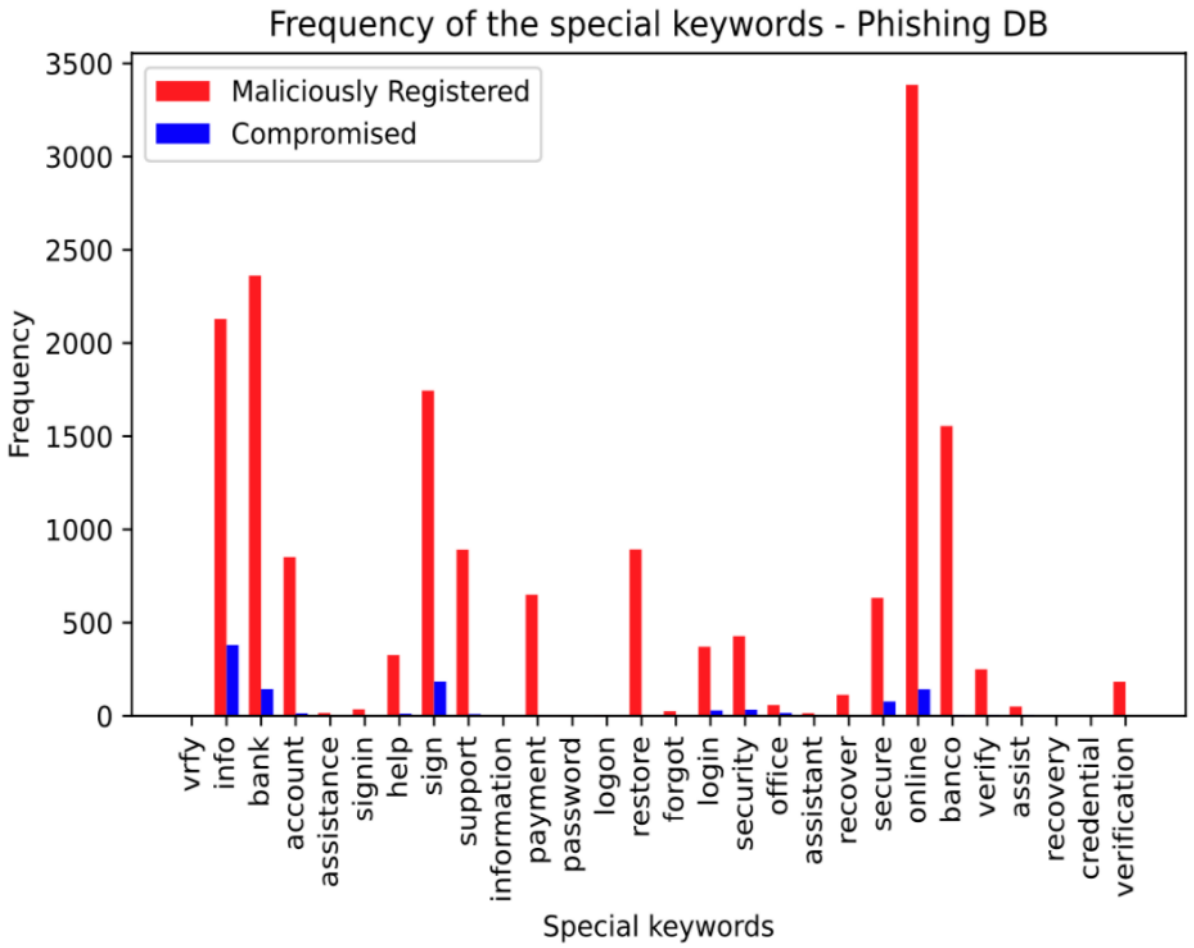
(c) ccTLDs

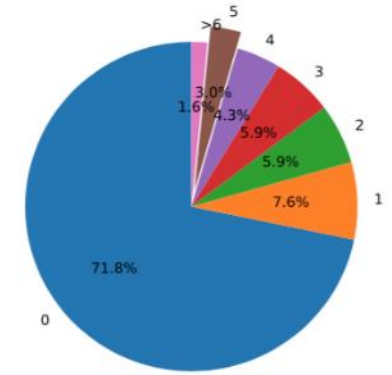
Classification of phishing domains for different types of TLDs

Datasets:

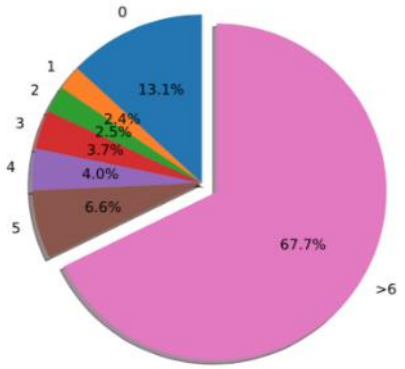
36,260 URLs (with unique underlying domain names)
35,519 unique phishing URLs
741 unique malware delivery URLs
Phishing URLs from APWG and PhishTank
Malware delivery from URLhaus.

FEATURE-LEVEL ANALYSIS





(a) Maliciously registered



(b) Compromised

Number of technologies for maliciously registered and compromised domains

IMPACT ACHIEVED

AFNIC, SIDN, and wider TLD ecosystem: **new in-depth insights** into attackers' behavior, allowing security practices to become more proactive

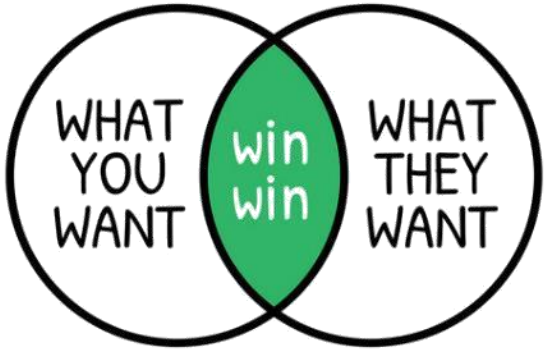
Example: specific strings/keywords in maliciously registered domain names and wider range of technologies are an indicator for compromised domains.

AFNIC and SIDN: more efficient and more effective incident response handling through **automatic classification** of maliciously registered vs. compromised domain names

AFNIC and SIDN: increased NIS2-readiness because of reinforced anti-abuse capabilities

Grenoble Alps University: further **reinforced research results** because of industry-relevant problem and exploitation of COMAR classifier at AFNIC and SIDN

LESSONS LEARNED

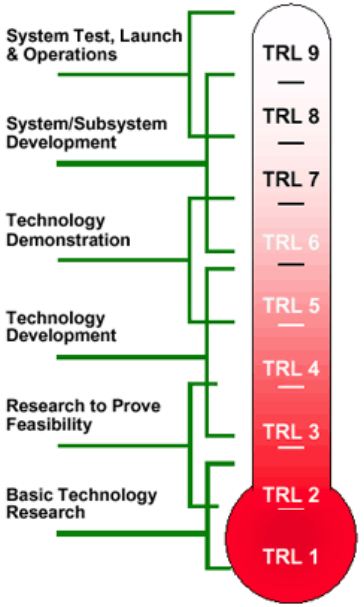


Projects collaboration between CENTR members and academic institution **just works !**

- **Academics gain from our vision, expertise and experience**
- **TLDs operator gain from research mindset and workforce**

However, you need to:

- **Be patient.** It takes time and flexibility to explore new territory, give researchers that time and space. Research is totally different from product development.
- **Be aware of the time required for administrative and legal issues**: initial agreement, selection/hiring PhD student, data sharing agreement. (No, we did not mention Covid-19… ;))
- From the beginning, **define an ambition level** for both research (e.g., the COMAR classifier at around TRL6) and the path to production after the project (TRL8-9)
- Like for any other project, **follow up** with meetings, projects outcomes, presentations, workshops



NEXT STEPS

Integrating COMAR engine into operational services and processes at SIDN and AFNIC

Currently testing COMAR into existing internal abuse toolset and office desk process such as technical DNS active/passive monitoring, registrant whois information scoring, registrar/registrant registration scoring (RegistryAI at Afnic), ...

Planning to share more data on both .nl and .fr TLDs soon…

QUESTIONS AND DISCUSSION

https://comar-project.univ-grenoble-alpes.fr/