PHISHWEB – a Progressive, Multi-Layered System for Phishing Websites Detection
Lucas Torrealba A. (1 & 2), P. Casas (1), Javier Bustos-Jiménez(2)
(1) AIT Austrian Institute of Technology, Vienna, Austria
(2) NIC Chile Research Labs, University of Chile

Motivation & Challenge
- Why phishing – malicious domain detection is important?
  - Compromises the security of users.
  - By the Internet Crime Complaint Center (IC3) Report “Almost 2 billion lost to BEC scams in 2020”
  - By the phishing landscape the number of phishing attacks and domain names reported for phishing trended up over the yearly period.

Types of Malicious Domains
- The domain names used for phishing could be (1) Hacked Domain Names or (2) Malicious Domain Names. The second can be classified into:
  - Typo: Domains that are registered with a name that is an intentionally misspelled version of other sites. e.g. facebook-policyform442.tk
  - Homophony: Domains in which international characters or symbols are included that look like a character in the original domain. e.g. amazon.com
  - AGD: Domains that are generated by DGA, which can dynamically generate several random domain names with a random seed. These domains are mainly used by botnets and malware to avoid detection and blocking. e.g. wmmncfyswc.com

Dataset
- PhishStorm: Dataset used in [1] for phishing detection. 96.018 URLs. 50% benign URLs and 50% phishing URLs.
- Phishing-Benign: Dataset used in [2] for phishing detection. 38.800 URLs. 50% benign - 50% phishing.
- DGA: Dataset used in [3] for algorithmically generated malicious domain detection. 675.000 domain names. 50% non-AGD and 50% AGD.

The data in 2 first datasets corresponds to URLs. This is why a data processing is performed where only both the protocol and the path are removed, leaving only the domain name.

This process causes loss of information as we can see:

Results & Future Work
- Remove distractions for SOD analysis
- Add redirection checker in SOD
- Replace by similar character (J, Q, Z)
- Add new features (ASN, DNS, content-html, etc.)