UNIVERSITÀ DI TRENTO

Cyber Security for Europe

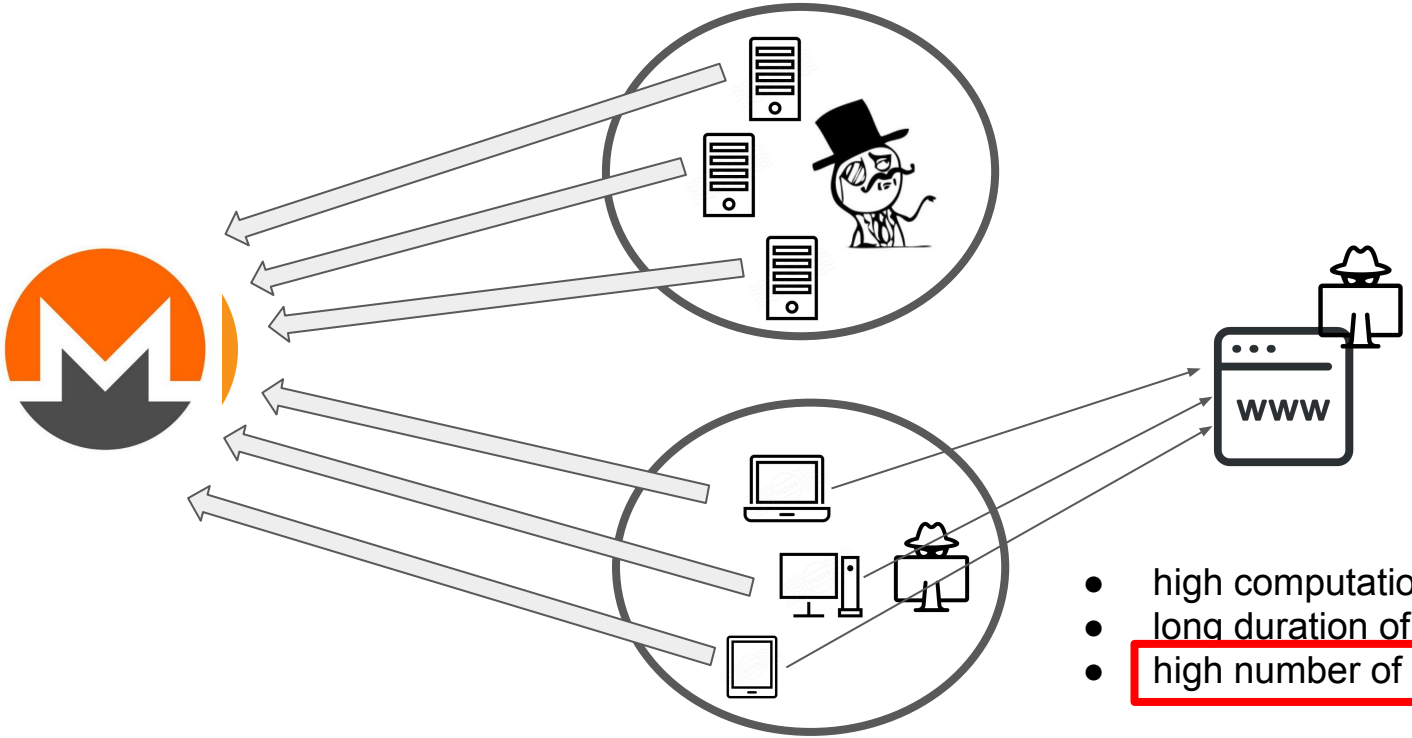
Are You a Favorite Target For Cryptojacking? A Case-Control Study On The Cryptojacking Ecosystem

Giorgio Di Tizio
joint work with Chan Nam Ngo

IEEE EuroS&P WACCO'20

This work is part of CyberSec4Europe (cybersec4europe.eu), a project that has received funding from the European Union's Horizon 2020 research and innovation programme H2020-SU-ICT-03-2018 under grant agreement No. 830929.

Cryptocurrency and Cryptojacking in a Nutshell



- high computational power
- long duration of the visit
- high number of visitors

The Attacker’s Strategy - high number of visitors

- This is the easiest controllable variable, thus attacker must compromise either:
  - a well-known, and highly likely secure, website **OR**
  - a high number of less popular, but at the same time, potentially less secure, websites
- Attackers want to *maximize profit and minimize the effort*
- Require to identify some common characteristics that can be exploited in mass

The Research Problem and Hypothesis

- **RQ:** *Are there certain technical characteristics of a website that may increase (decrease) the likelihood of being compromised for cryptojacking campaigns? (but not WHY)*

- **H1:** E.g. are websites based on NGINX more likely to be compromised than websites based on APACHE HTTP SERVER ?

- **H2:** E.g. are websites based on more likely to be compromised than websites without CMS?

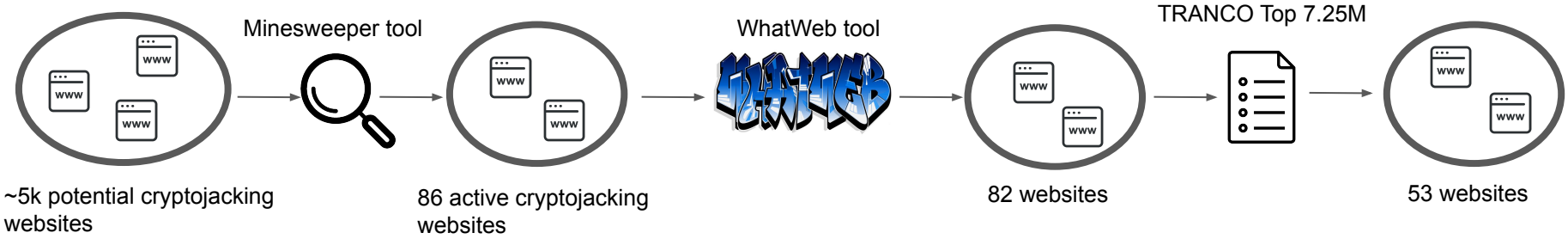
- **H3:** E.g. are websites that hide software information less likely to be compromised?

Case-control Study vs Experiment

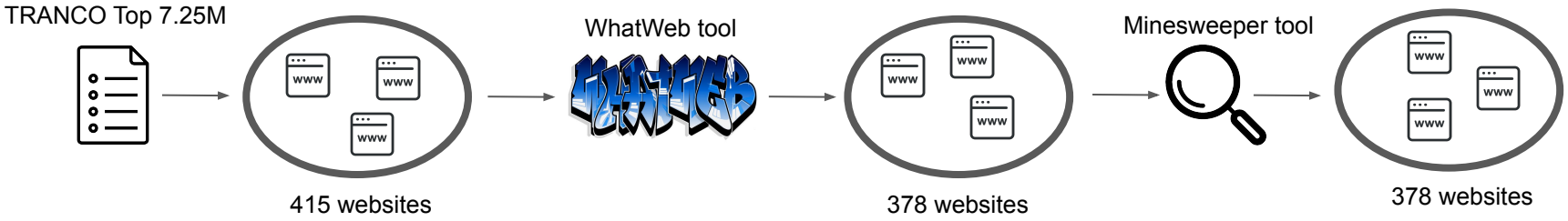
- How can we answer to these questions?
  - Experiment -> are not always possible: ask subjects to smoke to see if they die from cancer
  - Case-control -> retrospective analysis
- In case-control studies the **case** group is compared to the **control** group:
  - **case group:** subjects that present the observed effect (e.g. cancer, cryptojacking activity)
  - **control group:** subjects chosen randomly from a population w/ similar characteristics of the **case** that do not present the observed effect
  - **risk factor:** the explanation of the presence of the observed effect (e.g. smoking, CMS Drupal)
- Good to measure correlation between an observation and a presumed risk factor
- Not good for causation -> non-observable factors that can influence the process

Data Collection - Case and Control group

CASE:



CONTROL:



Preliminary Results - Odds ratio

- **H1** relative to Apache
  - Odds ratio (Microsoft IIS) ~ 1 CI:(0.27,3.88)-> neither a positive nor a negative risk factor
  - Odds ratio (NGINX) ~ 1.6 CI:(0.76,3,37) -> possibly **positive** risk factor

- **H2** compared to no CMS
  - Odds ratio ( ) ~ 1.32 CI:(0.71,2.43) -> possibly **positive** risk factor
  - Odds ratio (Drupal) ~ 2 CI:(0.39,9.69) -> possibly **positive** risk factor

- **H3** compared to not hiding CMS, Server, and application framework type
  - Odds ratio ~ 0.27 CI:(0.03,2.11) -> possibly **negative** risk factor

- ***Github:*** *https://github.com/giorgioditizio/risk_cryptojacking*

Limitations and Future Work

- Currently the results are **not statistically** significant -> increase the size of the case and control
  - We are currently crawling to collect more data
- Extend the analysis on visible characteristics associated with hardening (e.g. security headers like *CSP, X-XXS-Protection,* etc.)
- Study if attacker's technology preferences change depending on the malicious activity (e.g. phishing *vs* cryptojacking *vs* drive-by download, etc.)