SaTC: CORE: Small: Understanding, Measuring, and Defending against Malicious Web Crawlers

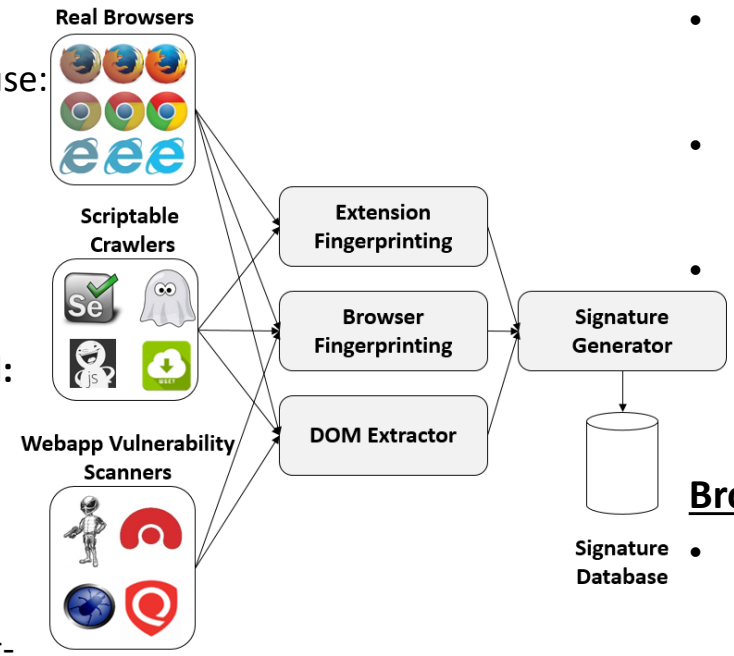
Stony Brook University

Problem:

- Malicious web crawlers scour the web searching for websites to abuse:
  - Content scraping
  - Vulnerability exploitation
  - Data exfiltration
  - Fake account creation

Our approach:

- **Discover web crawlers in the wild:** Flexible network of honeysites designed to attract malicious crawlers with no content for real users
- **Detecting and defending against malicious crawlers**: Fingerprinting-based and network-based detection. Reverse-proxy design to block malicious traffic
- **Protocols:** New protocols for crawler access control to supplement or replace robots.txt. Reduce the currently unconstrained capabilities of malicious crawlers



Challenges:

- No ground-truth for malicious crawlers
- Management of an ever-expanding network of honeysites
- Evasions by sophisticated crawlers

Scientific Impact:

- Advance the community's understanding of malicious web crawlers
- Ground-truth data of malicious crawler activity that does not require manual filtering
- Research into accurate and performant detection of malicious bots using traps and fingerprinting techniques

Broader Impact:

- Increase the security of all websites and user trust of the web platform
- Defend user data by blocking malicious bots before they exploit existing vulnerabilities
- Educational opportunities for underrepresented groups through SBU's WISE program

Award Number: 1813974 (2018 – 2021)
Institution: Stony Brook University
Nick Nikiforakis (nick@cs.stonybrook.edu)