# Modeling and Context-Driven Exploitation of Modern Websites

Dimitris E. Simos, Manuel Leithner

COMET Competence Centers for Excellent Technologies
www.ffg.at/comet
SBA Research
MATRIS

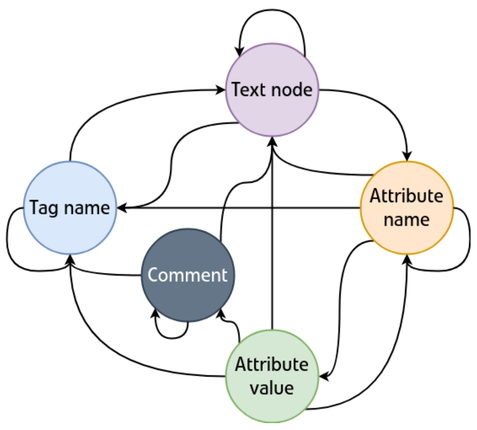
## HYDRA

- Optimizing Context-Driven Black-Box Cross-Site Scripting Exploitation.
- Capable of identifying Reflected and Stored XSS.
- Develops exploit vectors based on target behavior and filters.

### Execution Contexts

HTML Output Context, defined by

- Type
  - Tag name
  - Text
  - Attribute name
  - Attribute value
  - ...
- Tag: `A`, `DIV`, `SCRIPT`, ...
- Attribute: `NONE`, `href`, `onclick`, ...




Some contexts are more desirable:

- Text nodes under `SCRIPT` tags
- `onload`/`onerror` attributes
- Use-case specific, e.g., `IMG src`

### Approach

*Input*

- Target endpoint.
- Execution context weights.
- Transition patterns.

*Process*

1. Perform initial injection
2. Identify execution context
3. Find reachable contexts with higher weight
4. Construct concrete edges from patterns and encodings
5. Extend injection with concrete edges
6. Perform each injection (parallel or serial)
   - Execution context changed → SUCCESS!
   - Otherwise → FAILED
7. Repeat process with successful executions

*Combinatorics*

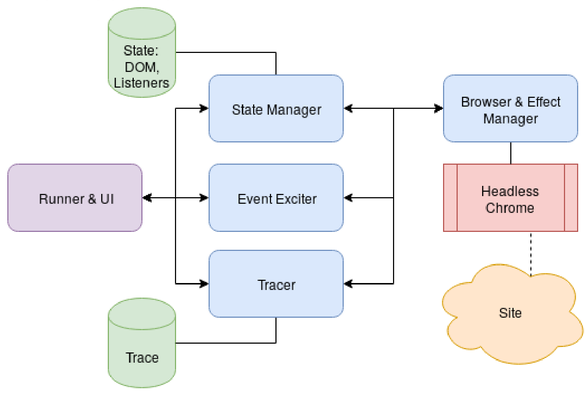
Combinatorial sequences to manage:

- Numerous possible edges
- Various encodings
- Multiple extension steps

## XIEv

- Dynamic Analysis for Model Extraction.
- Compatible with Ajax and other JavaScript-based modifications.
- Navigates site taking all available user actions.
- Records messages, loaded resources, ...
- Additional Use Cases:
  - Crawling
  - Regression Testing
  - Resource Extraction
- Uses `DOMdiff` to classify similar pages.

### Architecture




### Approach

1. Load page
2. Extract DOM and event listeners
3. Select event listener
4. Synthesize user action
5. Record and handle effects
6. Repeat until no further actions possible

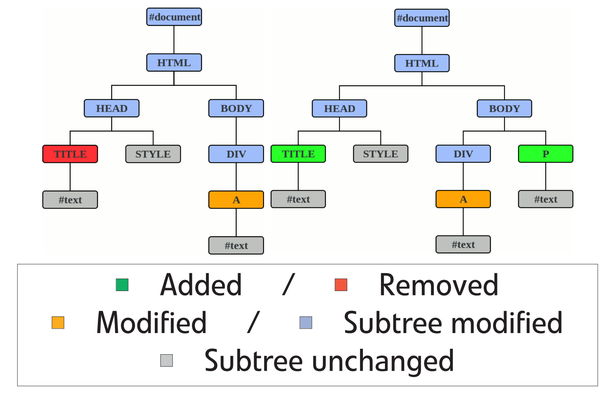
### Effects

- Network requests
- DOM node addition/removal
- Event listener modifications
- CSS animations
- Console messages
- DOM storage
- ...

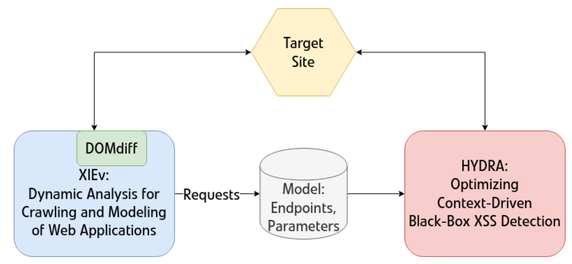
## DOMdiff

- Identifies changes between DOM trees:
  - Additions
  - Deletions
  - Replacements
  - Attribute/text changes
- Classifies pairs of documents:
  - Same: Likely the same page, minor differences
  - Similar: Pages based on the same template
  - Separate: Unrelated pages
- Classifier based on Machine Learning or Decision Trees.
- Supports pages with dynamic content.

### Visualized Output: Old vs New DOM




## Relations

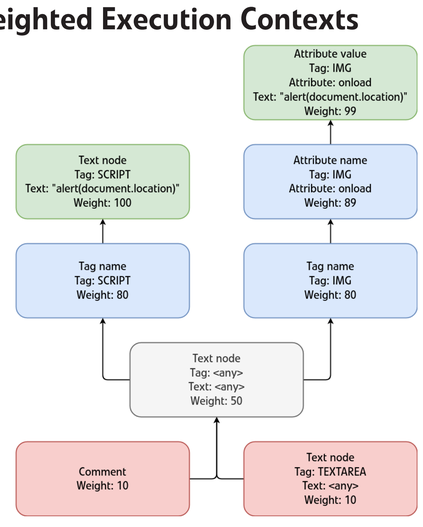



Independent or combined use possible

## HYDRA in Practice

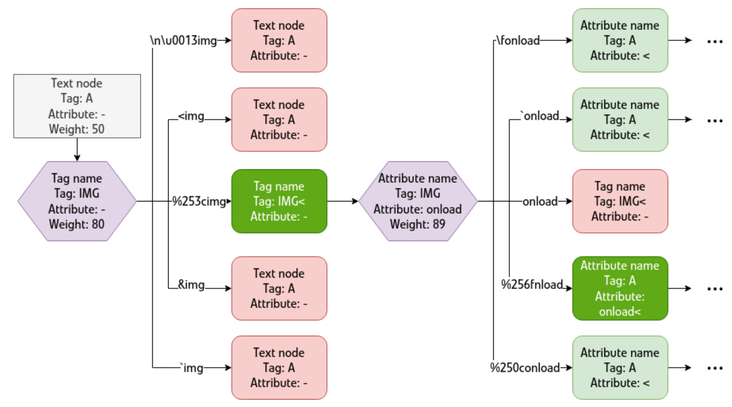
HYDRA navigates the weighted execution context graph towards target states (highest weights) by extending the injection vector with HTML5 parser state graph edges, optionally adding encodings.

### Weighted Execution Contexts




### Example Injection

This example target filters out the < character and suspicious attributes like `onload`, but is susceptible to double-encoding.





- M. Leithner and D. E. Simos, "XIEv: dynamic analysis for crawling and modeling of web applications," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 2020, pp. 2201–2210.
- ——, "Domdiff: Identification and classification of inter-dom modifications," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2018, pp. 262–269.


Bundesministerium Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie
Bundesministerium Digitalisierung und Wirtschaftsstandort
FFG Forschung wirkt.
wirtschaftsagentur wien Ein Fonds der Stadt Wien
Europäische Kommission
FWF Der Wissenschaftsfonds.
netidee OPEN INNOVATIONS

SBA Research (SBA-K1) is a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the federal state of Vienna. The COMET Programme is managed by FFG.