black hat EUROPE 2022

DECEMBER 7-8, 2022

BRIEFINGS

Bridging the Gap Between Research and Practice in Intelligently Bypassing WAF

Cheng Chi
Sangfor Technologies Inc.

Cheng Chi

- Technical Expert at Sangfor Technologies
- Research interest: AI Security, Data-driven Security
- Published at: USENIX Security, TPAMI, NeurIPS, CVPR and AAAI etc.
- Homepage: https://chicheng123.github.io/

Agenda

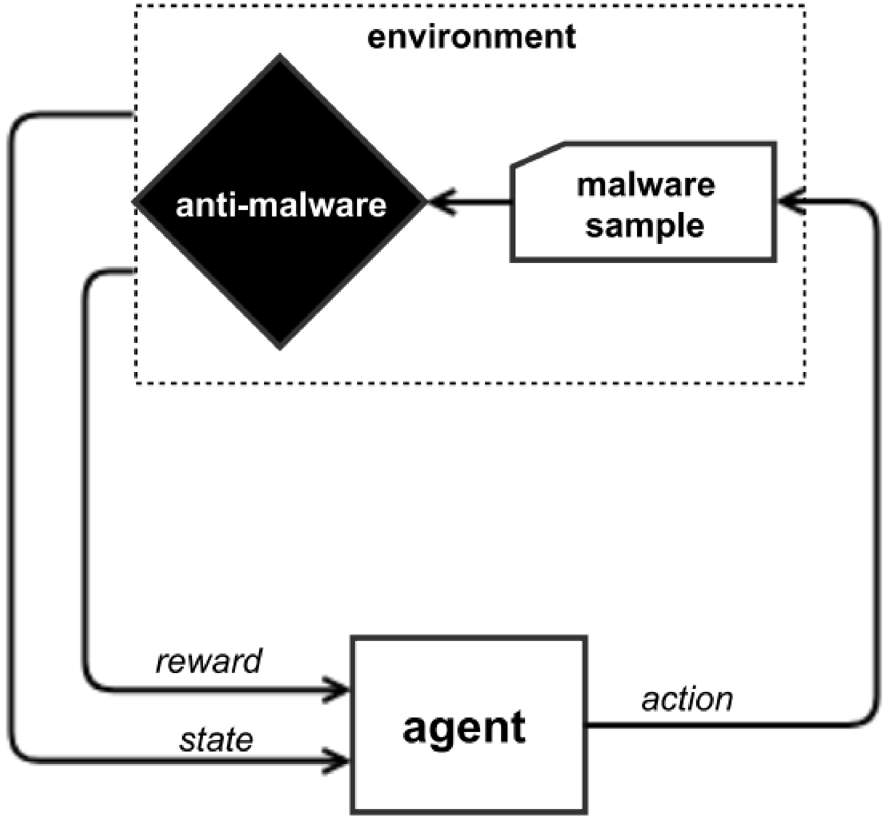
- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

Agenda

- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

AI-enabled cyber attack

Bot vs. Bot: Evading Machine Learning Malware Detection.
Hyrum Anderson. Blackhat USA 2017.



- **Environment**
  - A malware sample (*Windows PE*)
  - Buffet of malware mutations
    - *preserve format & functionality*
  - Reward from static malware classifier
- **Agent**
  - Input: **environment state** (*malware bytes*)
  - Output: **action** (*stochastic*)
  - Feedback: **reward** (*AV reports benign*)

Functionality-preserving mutations:

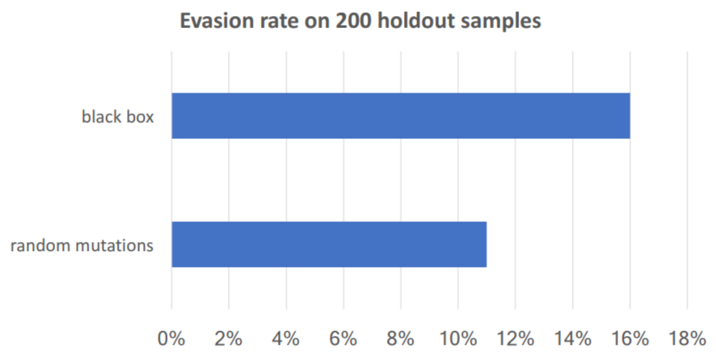
- **Create**
  - New Entry Point (w/ trampoline)
  - New Sections
- **Add**
  - Random Imports
  - Random bytes to PE overlay
  - Bytes to end of section
- **Modify**
  - Random sections to common name
  - (break) signature
  - Debug info
  - UPX pack / unpack
  - Header checksum
  - Signature

Features

- Static Windows PE file features compressed to 2350 dimensions
  - General file information (size)
  - Header info
  - Section characteristics
  - Imported/exported functions
  - Strings
  - File byte and entropy histograms
- Feed a neural network to choose the best action for the given "state"

Evasion Results

- Agent training: 15 hours for 100K trials (~10K games x 10 turns ea.)
- Using malware samples from VirusShare

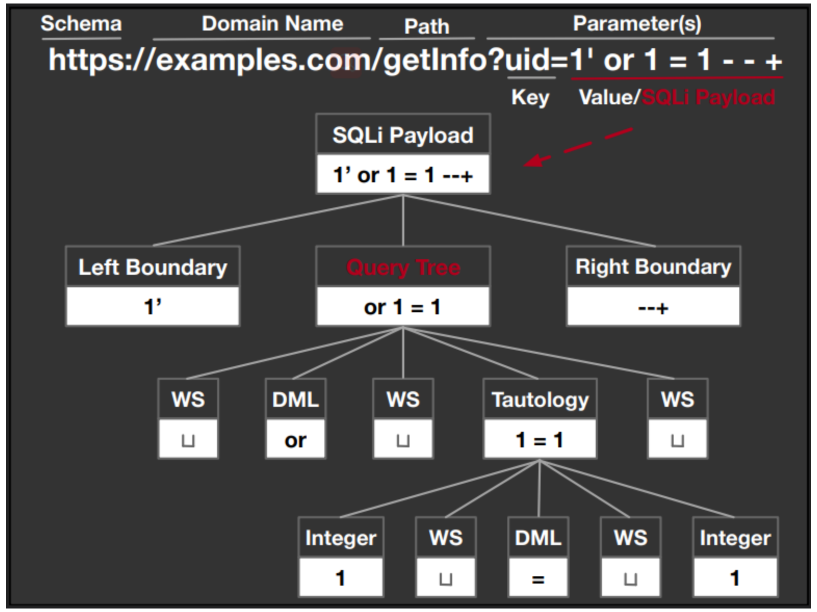


Cross-evasion: detection rate on VirusTotal (average)
- from 35/62 (original)
- to 25/62 (evade)

AI-enabled cyber attack

AutoSpear: Towards Automatically Bypassing and Inspecting Web Application Firewalls.
Zhenqing Qu, et al. Blackhat Asia 2022.

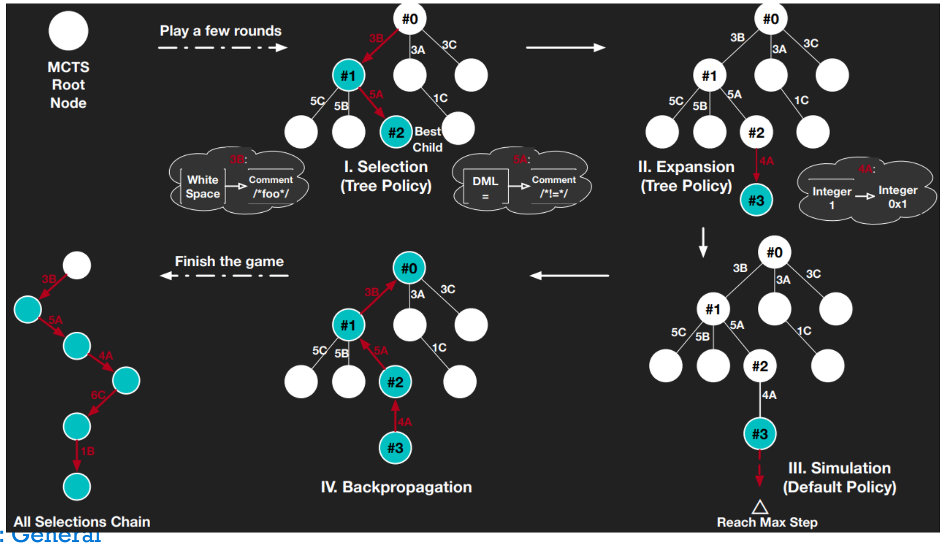
(1) Hierarchical Tree Representation



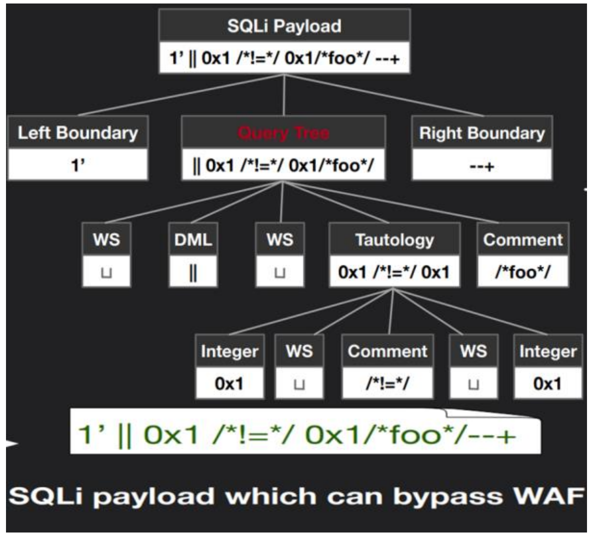
(2) Mutation with Context-free Grammar

Operator	Example
Case Swapping	or 1 = 1 → oR 1 = 1
Whitespace Substitution*	or 1 = 1 → \tor1\n=1
Comment Injection*	or 1 = 1 → /*foo*/or 1 =/*bar*/1
Comment Rewriting	/*foo*/or 1 = 1 → /*1.png*/or 1 = 1
Integer Encoding	or 1 = 1 → or 0x1 = 1
Operator Swapping	or 1 = 1 → or 1 like 1
Logical Invariant	or 1 = 1 → or 1 = 1 and 'a' = 'a'
Inline Comment	or 1 = 1 → /*!or/ 1 = 1 union select → /*!union*/ /*!50000select*/
Where Rewriting	where xxx → where xxx and True where xxx → where (select 0) or xxx
DML Substitution*	or 1 = 1 → || 1 = 1 and name = 'foo' → && name = 'foo'
Tautology Substitution	1 = 1 → 'foo' = 'foo' '1' = '1' → 2 <> 3 1 = 1 → (select ord('r') regexp 114) = 0x1

(3) Monte-carlo Tree Search Guided Searching

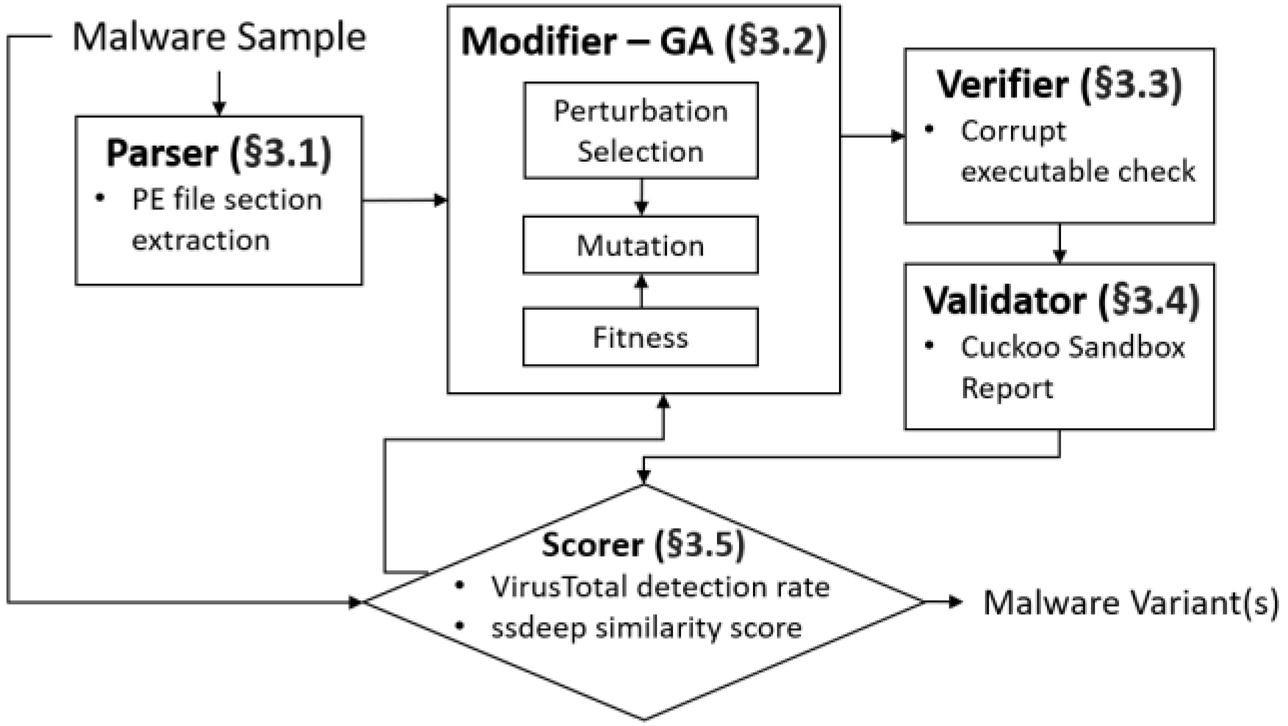


(4) Payload Reconstruction



AI-enabled cyber attack

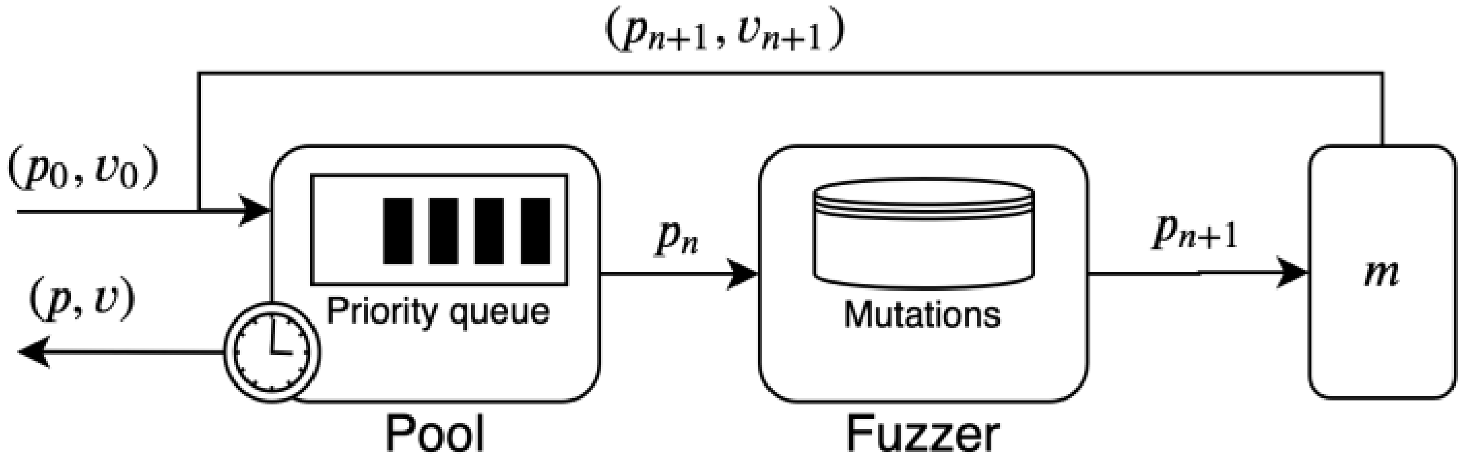
FUMVar: A Practical Framework for Generating Fully-working and Unseen Malware Variants. Beomjin Jin, et al. ACM SAC 2021.



Types	Name	Abbr	Description
NBC	Overlay append	OA	Adds the random length of zeroes to the end of the binary.
	DOS header	DH	Change the field values in DOS header.
	DOS stub	DS	Change DOS stub to random byte sequence.
	COFF header	CH	Change the field values in COFF header.
	Optional header	OH	Change the field values in Optional header.
	Data directory	DD	Change the field values in Data directory.
	Rich header	RH	Insert a new content info Rich header.
BC	Section rename	SRN	Change the name of a randomly selected section.
	Section add	SAD	Add a new section with random sequence of bytes.
	Section append	SAP	Append random bytes sequence to section content.
	Code cave inject	CI	Inject some random byte sequence to code cave.
	Pack	PA	Pack the file with UPX tool.
	Unpack	UP	Unpack the file with UPX tool.

AI-enabled cyber attack

WAF-A-MoLE: Evading Web Application Firewalls through Adversarial Machine Learning.
Luca Demetrio, et al.




```
input:  Model m, Payload p0, Threshold t
output: head(Q)

Q := create_priority_queue()
v := classify(m, p0)
enqueue(Q, p0, v)
while v > t
  p := mutate(head(Q))
  v := classify(m, p)
  enqueue(Q, p, v)
```

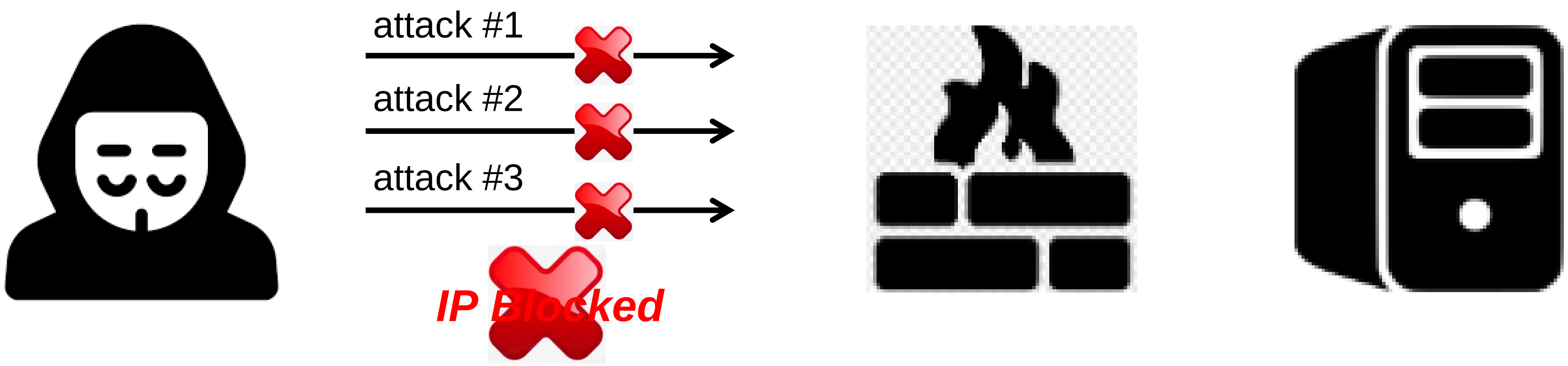
Agenda

- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

Two Key Issue Between Research and Practice

1. Research works focus more on sample generation, relying on huge number of attempts

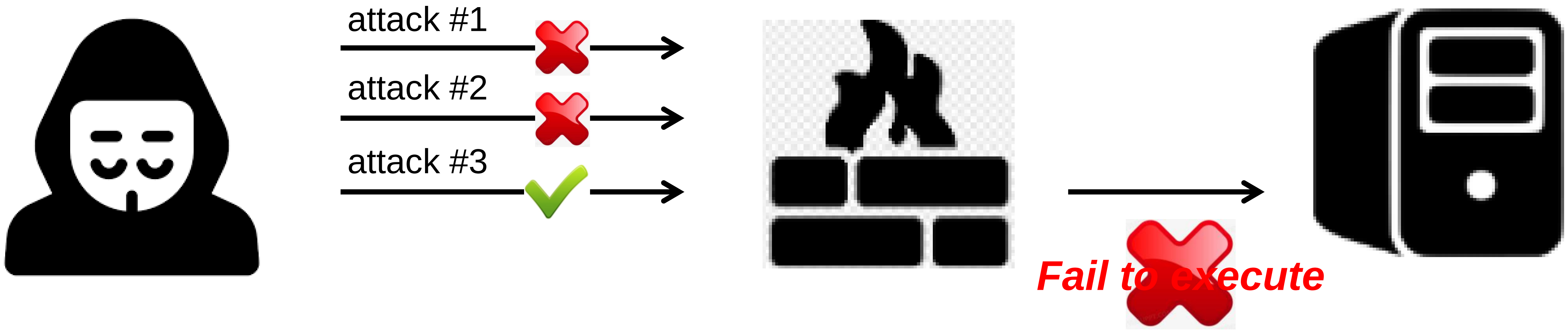
- Reinforcement learning or other AI-enabled methods need huge number of iteration to learn
- Current commercial WAFs have the blocking IP function. If too many attack attempts are made, the IP will be blocked.



Two Key Issue Between Research and Practice

2. Traditional web fuzz methods change the semantics of the payload

- Payloads may bypass the WAF after mutation, but cannot execute the attack correctly on the servers.



Agenda

- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

Takeaways

Our research is helpful to security researchers, WAF vendors and penetration test engineers

- Security researchers can use some basic mutation operations to obtain more complex bypass patterns, which is very helpful to the research and utilization of vulnerabilities.
- WAF vendors can use our algorithm framework to fight against their own WAF, search for bypass points and fix them. It can achieve the promotion effect of attack and defense.
- Penetration test engineers can directly use our model for penetration testing, which can automatically bypass the specified WAFs.

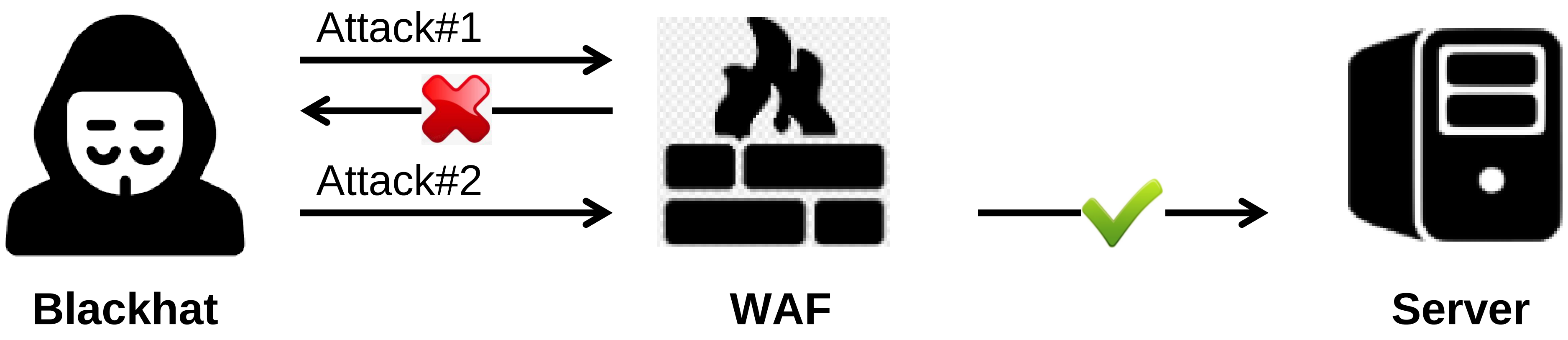
Agenda

- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

Our Solution

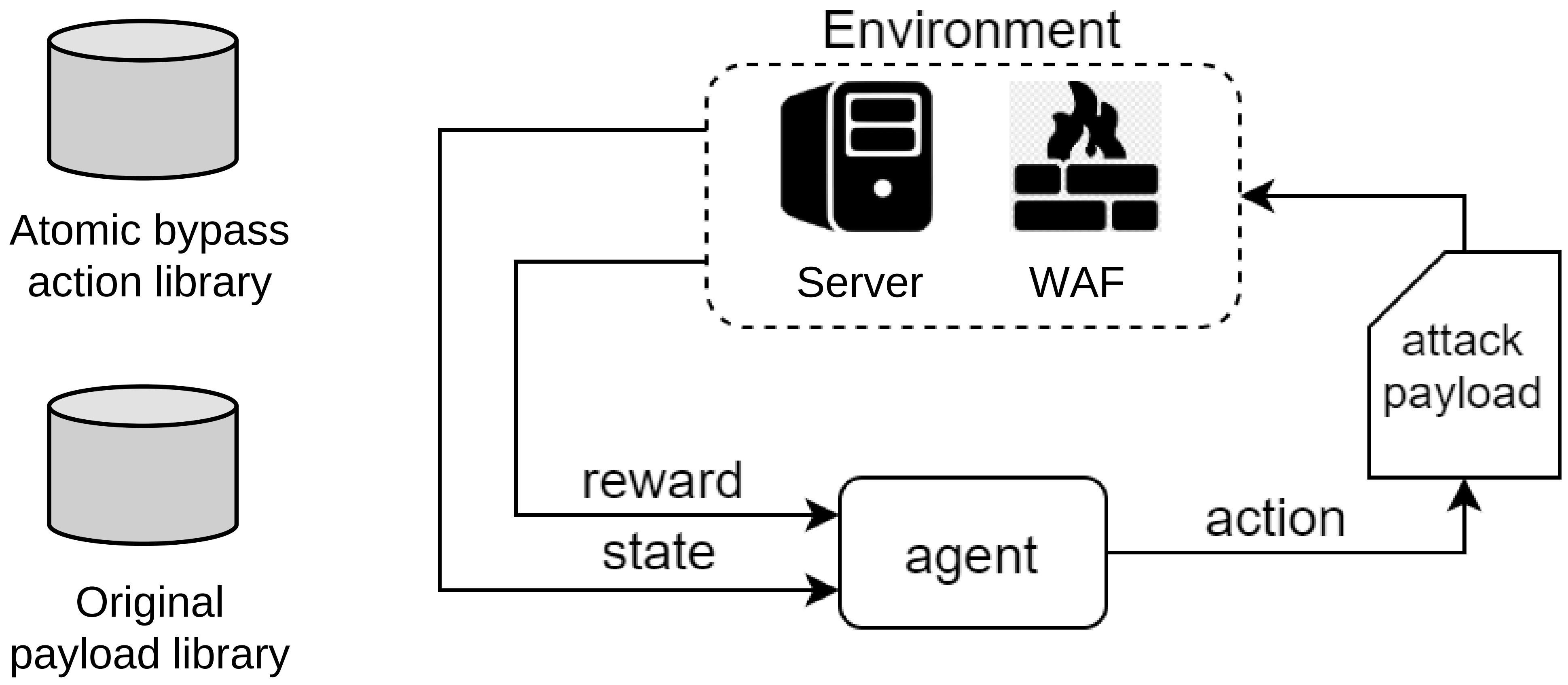
We first review the workflow of how human bypass WAF

Found that the whole process is very similar to reinforcement learning(RL)

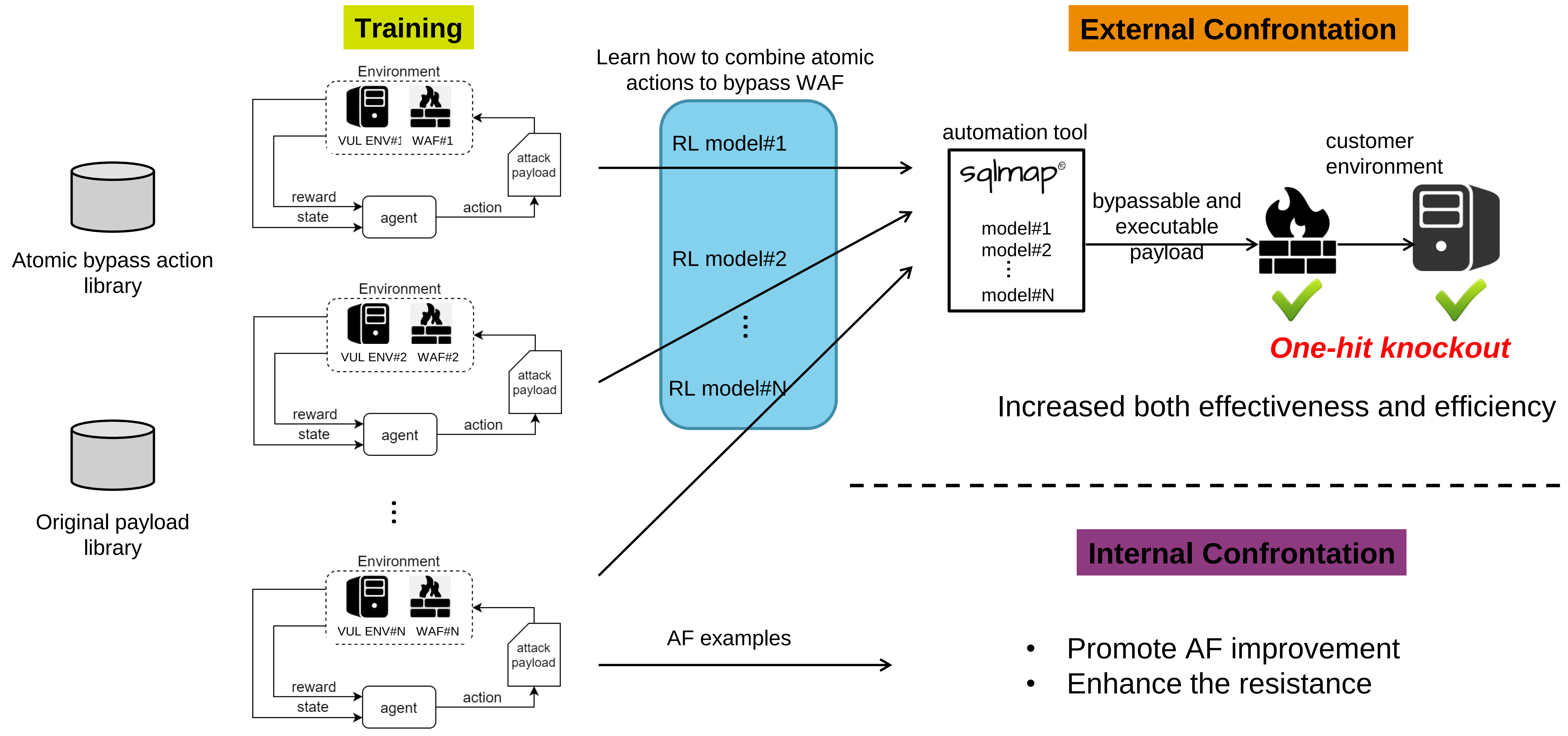


Our Solution

So we design the whole scheme based on RL



Our Solution



Internal Confrontation

- Promote AF improvement
- Enhance the resistance

Our Solution

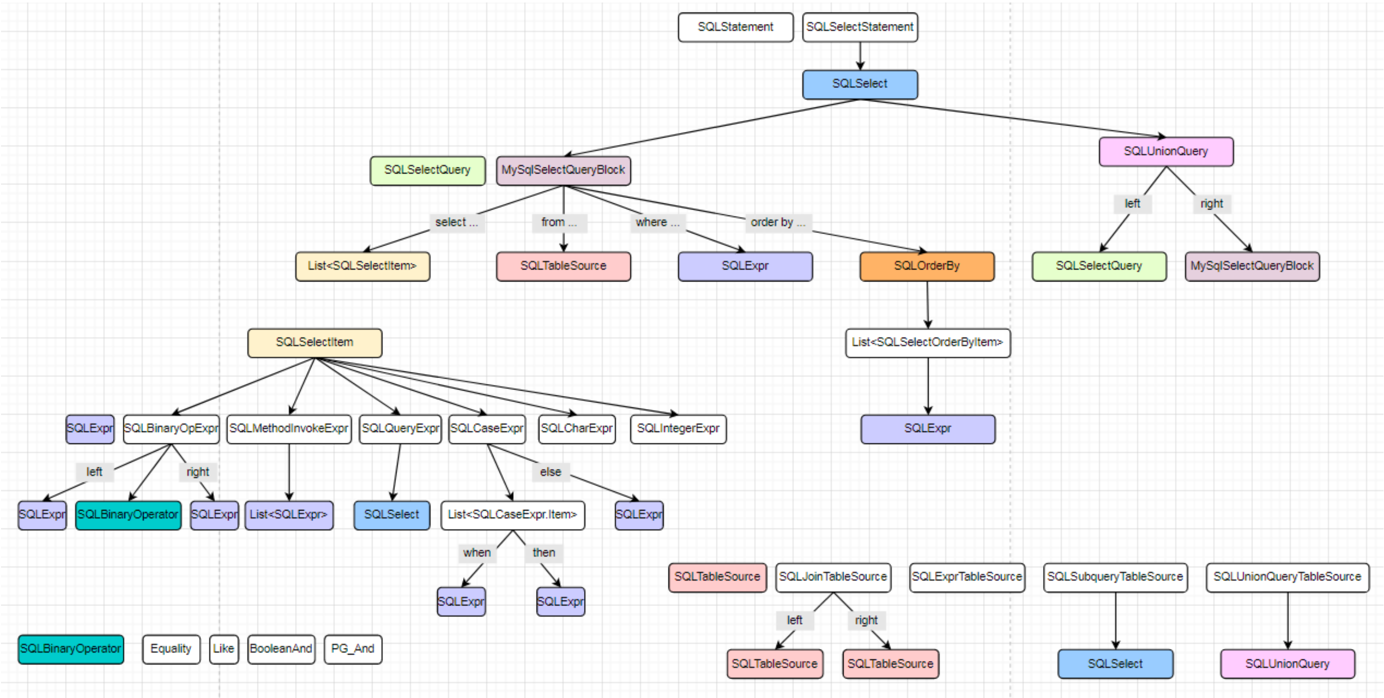
To maintain the semantic of payloads and improve the execution success of the mutated payload:

- **We propose a mutation method based on grammar and lexical analysis:**

1. Payload is parsed into a lexical token sequence and a syntax tree
2. mutations are performed on the token sequence and syntax tree to maintain the semantic invariance of the payload

- **In the reward design, we consider:**

1. Whether WAF can be bypassed or not
2. Whether the payload can execute correctly



Agenda

- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

Algorithm Design

Proximal Policy Optimization Algorithms

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov
OpenAI
{joschu, filip, prafulla, alec, oleg}@openai.com

Importance Sampling

$$E_{x\sim p}[f(x)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x^i)$$

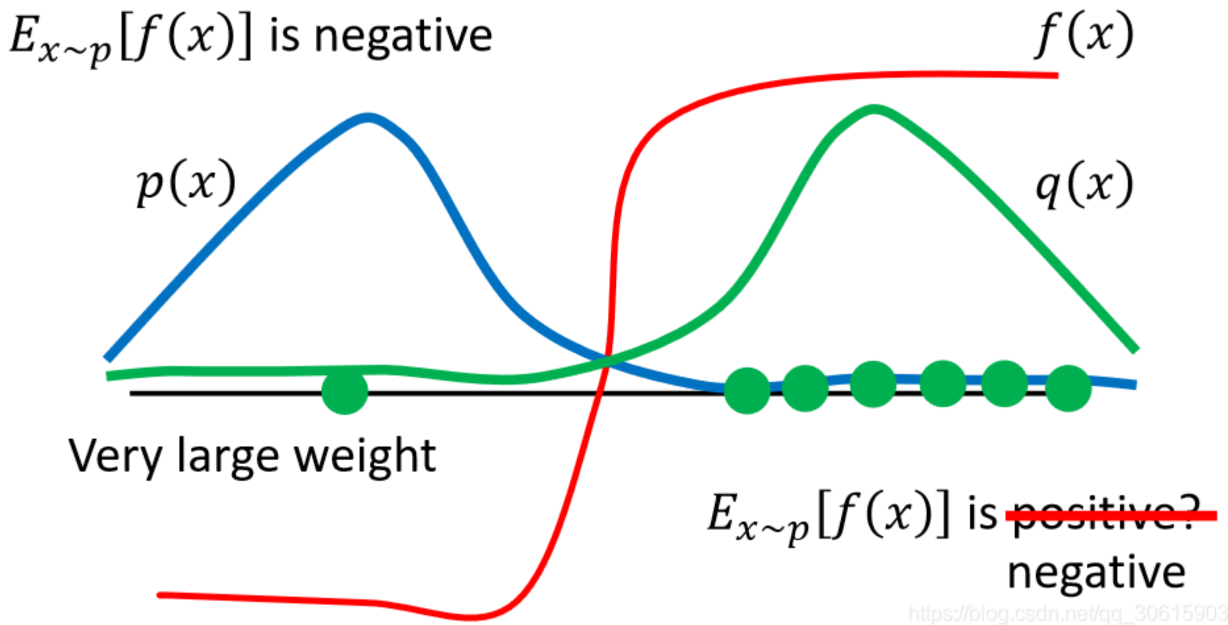
x^i is sampled from $p(x)$

We only have x^i sampled from $q(x)$

$$= \int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx = E_{x\sim q}[f(x)\frac{p(x)}{q(x)}]$$

Importance weight

$$E_{x\sim p}[f(x)] = E_{x\sim q}[f(x)\frac{p(x)}{q(x)}]$$



Agenda

- AI-enabled Cyber Attack
- Two Key Issue Between Research and Practice
- Takeaways
- Our Solution
- Algorithm Design
- Experimental Results

Experimental Results

Internal Confrontation

- **Black box:**

Generate 2W+ bypassable and executable confrontational samples

Extract 20+ bypass modes

Further promote the improvement of our WAF

- **White box:**

Use more feedback information from WAF to drive the RL agent to discover more bypass modes

More than 50 bypass modes have been found

Experimental Results

External Confrontation

- Against several mainstream WAF, we have almost 100% bypass success rate
- Function is integrated into the SQLmap tool
- The effect of one-click injection has been verified in dozens of customers

Thanks

Contact: