

# Web Application Security

Fang Yu

Software Security Lab.  
Department of Management Information Systems  
College of Commerce, National Chengchi University  
<http://soslab.nccu.edu.tw>

FloLac Talk, August 14, 2023



## About Me

Yu, Fang

- 2014-present: Associate Professor, Department of Management Information Systems, National Chengchi University
- 2010-2014: Assistant Professor, Department of Management Information Systems, National Chengchi University
- 2005-2010: Ph.D. and M.S., Department of Computer Science, University of California at Santa Barbara
- 2001-2005: Institute of Information Science, Academia Sinica
- 1994-2000: M.B.A. and B.B.A., Department of Information Management, National Taiwan University



## Book Reference

- *String Analysis for Software Verification and Security*  
Tevfik Bultan, Fang Yu, Muath Alkhalaf, Abdulbaki Aydin. [Springer. 2018]
- <https://www.springer.com/gp/book/9783319686684>



## More Recent Work

- *Parameterized Model Counting for String and Numeric Constraints*  
Abdulbaki Aydin, William Eiers, Lucas Bang, Tegan Brennan, Miroslav Gavrilov, Tevfik Bultan and Fang Yu. [ACM ESEC/FSE '18]
- *A Symbolic Model Checking Approach to the Analysis of String and Length Constraints*  
Hung-En Wang, Shih-Yu Chen, Fang Yu, Jie-Hong R. Jiang. [ACM ASE'18]
- *Static API Call Vulnerability Detection in iOS Applications*  
Chun-Han Lin, Fang Yu, Jie-Hong Jiang, and Tevfik Bultan. [ACM/IEEE ICSE'18]
- *Optimal Sanitization Synthesis for Web Application Vulnerability Repair*  
Fang Yu, ChinYuan Shueh, ChunHan Lin, YuFang Chen, BowYaw Wang, Tevfik Bultan. [ACM ISSTA'16]
- *String Analysis via Automata Manipulation with Logic Circuit Representation*  
HungEn Wang, ThungLin Tsai, ChunHan Lin, Fang Yu, JieHong R Jiang. [CAV'16]



## Automatic Verification of **String Manipulating Programs**

Web Applications = String Manipulating Programs



# Web Applications

Web applications are used extensively in many areas

- Commerce: online banking, online shopping, etc.
- Entertainment: online game, music and videos, etc.
- Interaction: social networks



# Web Applications

We may rely on web applications more in the future

- Health Records: Google Health, Microsoft HealthVault
- Controlling and monitoring national infrastructures: Google Powermeter



# Web Applications

Web software is also rapidly replacing desktop applications.





# One Major Road Block

Web applications are **not trustworthy!**

Web applications are notorious for security vulnerabilities

- Their global accessibility makes them a target for many malicious users

Web applications are becoming increasingly dominant and their use in safety critical areas is increasing

- Their trustworthiness is becoming a critical issue



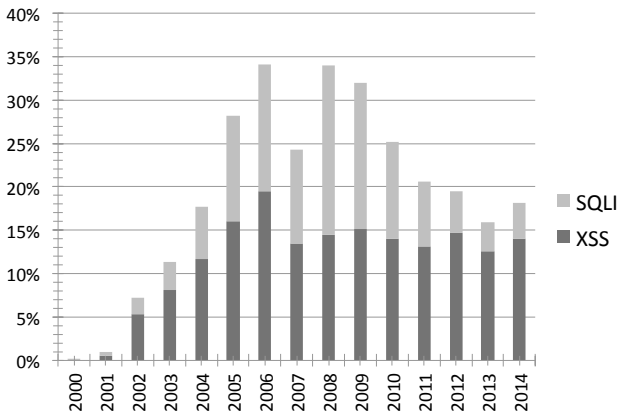
# Web Application Vulnerabilities

- The top two vulnerabilities of the Open Web Application Security Project (OWASP)'s top ten list in 2007, 2010, 2013, and 2017
  - 1 Cross Site Scripting (XSS)
  - 2 Injection Flaws (such as SQL Injection)



## Web Application Vulnerabilities

Percentage of the Cross-site Scripting (XSS) and SQL Injection (SQLI) vulnerabilities among all the computer security vulnerabilities reported in the CVE repository.



## Why are web applications error prone?

Extensive string manipulation:

- Web applications use extensive string manipulation
  - To construct html pages, to construct database queries in SQL, to construct system commands
- The user input comes in string form and must be validated and sanitized before it can be used
  - This requires the use of complex string manipulation functions such as string-replace
- String manipulation is error prone



# SQL Injection

## Exploits of a Mom.



Source: XKCD.com



# SQL Injection

Access students' data by \$name (from a **user input**).

| 1:<?php

| 2: \$name =\$\_GET["name"];

| 3: \$user\_data = \$db->query('SELECT \* FROM students  
WHERE name = "\$name" ');

| 4:??>



# SQL Injection

```
| 1:<?php  
| 2: $name = $_GET["name"];  
| 3: $user_data = $db->query('SELECT * FROM students  
|   WHERE name = "Robert '); DROP TABLE students; - -');  
| 4:?>
```



## Cross Site Scripting (XSS) Attack

A PHP Example:

```
| 1:<?php  
| 2: $www = $_GET["www"];  
| 3: $_otherinfo = "URL";  
| 4: echo "<td>" . $_otherinfo . ": " . $www . "</td>";  
| 5:?>
```

- The *echo* statement in line 4 can contain a Cross Site Scripting (XSS) vulnerability





# XSS Attack

An attacker may provide an input that contains `<script` and execute the malicious script.

- | 1: <?php
- | 2: \$www = <script ... >;
- | 3: \$\_otherinfo = "URL";
- | 4: echo "<td>" . \$\_otherinfo . ": " . <script ... > .  
" </td>";
- | 5: ?>



## Is it Vulnerable?

A simple **taint analysis**, e.g., [Huang et al. WWW04], would report this segment as vulnerable using *taint propagation*.

```
| 1:<?php  
| 2: $www = $_GET["www"];  
| 3: $l_otherinfo = "URL";  
| 4: echo "<td>" . $l_otherinfo . ": " . $www. "</td>";  
| 5:?>
```



## Is it Vulnerable?

Add a sanitization routine at line `s`.

```
| 1:<?php  
| 2: $www = $_GET["www"];  
| 3: $_l_otherinfo = "URL";  
| s: $www = ereg_replace("[^A-Za-z0-9 .-@://]", "", $www);  
| 4: echo "<td>" . $_l_otherinfo . ": " . $www . "</td>";  
| 5:?>
```

- Taint analysis will assume that `$www` is **untainted** after the routine, and conclude that the segment is **not** vulnerable.



## Sanitization Routines are Erroneous

However, `ereg_replace("[^A-Za-z0-9 .-@:/]", "", $www)`; does not sanitize the input properly.

- Removes all characters that are not in { A-Za-z0-9 .-@:/ }.
- `.-@` denotes all characters between "." and "@" (including "<" and ">")
- `.-@` should be `.\-@`



## A buggy sanitization routine

```
| 1:<?php  
| 2: $www = <script ... >;  
| 3: $l_otherinfo = "URL";  
| s: $www = ereg_replace("[^A-Za-z0-9 .-@://]", "", $www);  
| 4: echo "<td>" . $l_otherinfo . ": " . <script ... > .  
|   "</td>";  
| 5:?>
```

- A buggy sanitization routine used in MyEasyMarket-4.1 that causes a vulnerable point at line 218 in trans.php [Balzarotti et al., S&P'08]
- Our string analysis identifies that the segment is vulnerable with respect to the attack pattern:  $\Sigma^* \langle \text{script} \Sigma^* \rangle$ .



## Eliminate Vulnerabilities

Input `<!sc+rip!t ...>` does not match the attack pattern  $\Sigma^* \langle \text{script} \Sigma^* \rangle$ , but still can cause an attack

- | 1: `<?php`
- | 2: `$www = <!sc+rip!t ...>;`
- | 3: `$_otherinfo = "URL";`
- | s: `$www = ereg_replace("[^A-Za-z0-9 .-@://]", "", <!sc+rip!t ...>);`
- | 4: `echo " <td> " . $_otherinfo . " : " . <script ...> . " </td>";`
- | 5: `?>`



# Eliminate Vulnerabilities

- We generate **vulnerability signature** that characterizes **all** malicious inputs that may generate attacks (with respect to the attack pattern)
- The vulnerability signature for `$_GET["www"]` is  $\Sigma^* < \alpha^* s \alpha^* c \alpha^* r \alpha^* i \alpha^* p \alpha^* t \Sigma^*$ , where  $\alpha \notin \{ A-Za-z0-9 \cdot - @ : / \}$  and  $\Sigma$  is any ASCII character
- Any string accepted by this signature can cause an attack
- Any string that dose not match this signature will **not** cause an attack. I.e., **one can filter out all malicious inputs using our signature**



## Prove the Absence of Vulnerabilities

Fix the buggy routine by inserting the escape character `\`.

```
| 1:<?php
| 2: $www = $_GET["www"];
| 3: $l_otherinfo = "URL";
| s': $www = ereg_replace("[^A-Za-z0-9 .\-\@://]", "", $www);
| 4: echo "<td>" . $l_otherinfo . ": " . $www . "</td>";
| 5:?>
```

Using our approach, this segment is **proven** not to be vulnerable against the XSS attack pattern:  $\Sigma^* \langle \text{script} \Sigma^*$ .





## Multiple Inputs?

Things can be more complicated while there are **multiple inputs**.

| 1: <?php

| 2: \$www = \$\_GET["www"];

| 3: \$l\_otherinfo = \$\_GET["other"];

| 4: echo "<td>" . \$l\_otherinfo . ": " . \$www . "</td>";

| 5: ?>

- An attack string can be contributed from one input, another input, or their combination
- We can generate **relational vulnerability signatures** and automatically synthesize effective patches.



# String Analysis

- String analysis determines all possible values that a string expression can take during any program execution
- Using string analysis we can identify all possible input values of the sensitive functions. Then we can check if inputs of sensitive functions can contain attack strings
- If string analysis determines that the intersection of the attack pattern and possible inputs of the sensitive function is empty. Then we can conclude that the program is secure
- If the intersection is not empty, then we can again use string analysis to generate a vulnerability signature that characterizes all malicious inputs

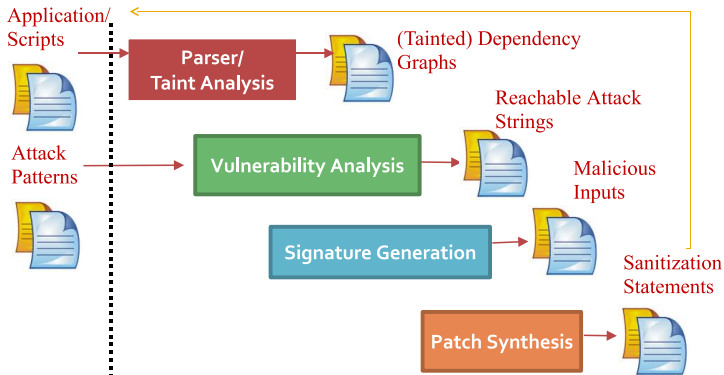


# Automata-based String Analysis

- Finite State Automata can be used to characterize sets of string values
- We use automata based string analysis
  - Associate each string expression in the program with an automaton
  - The automaton accepts an over approximation of all possible values that the string expression can take during program execution
- Using this automata representation we symbolically execute the program, only paying attention to string manipulation operations
- Attack patterns are specified as regular expressions



# String Analysis Stages



## Front End

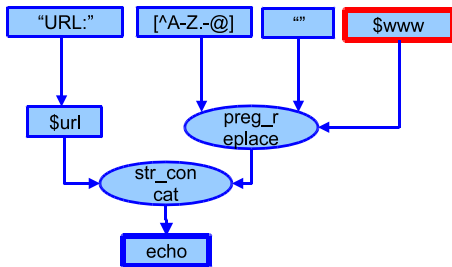
Consider the following segment.

```
| <?php  
| 1: $www = $_GET["www"];  
| 2: $url = "URL:";  
| 3: $www = preg_replace("[^A-Z.-@]", "", $www);  
| 4: echo $url. $www;  
| ?>
```



## Front End

A dependency graph specifies how the values of input nodes flow to a sink node (i.e., a sensitive function)



NEXT: Compute all possible values of a sink node



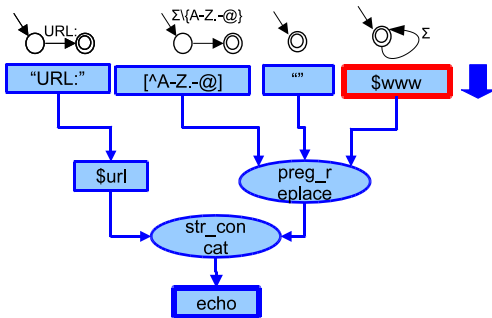
## Detecting Vulnerabilities

- Associates each node with an **automaton** that accepts an over approximation of its possible values
- Uses automata-based **forward** symbolic analysis to identify the possible values of each node
- Uses *post-image* computations of string operations:
  - $\text{postConcat}(M_1, M_2)$  returns  $M$ , where  $M = M_1.M_2$
  - $\text{postReplace}(M_1, M_2, M_3)$  returns  $M$ , where  $M = \text{REPLACE}(M_1, M_2, M_3)$



## Forward Analysis

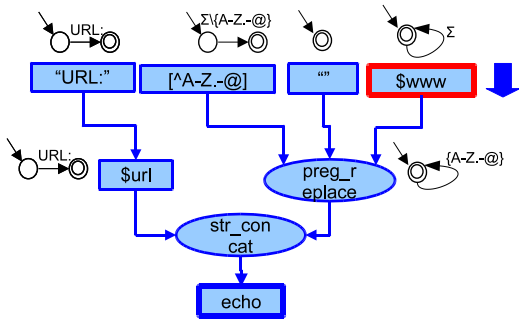
- Allows *arbitrary* values, i.e.,  $\Sigma^*$ , from user inputs
- Propagates post-images to next nodes iteratively until a fixed point is reached





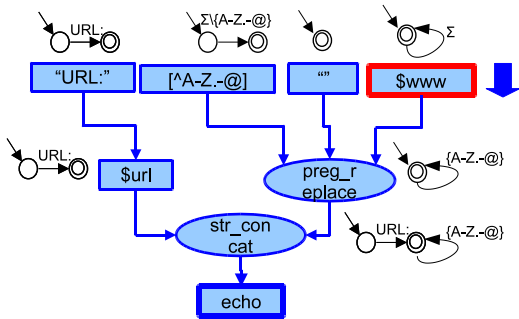
## Forward Analysis

- At the first iteration, for the replace node, we call `postReplace( $\Sigma^*$ ,  $\Sigma \setminus \{A-Z.-@\}$ , "")`



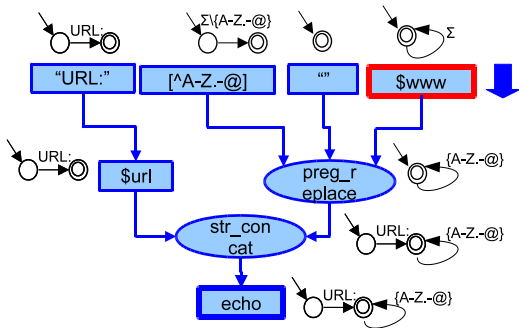
## Forward Analysis

- At the second iteration, we call `postConcat("URL:", {A-Z.-@}*)`



# Forward Analysis

- The third iteration is a simple assignment
- After the third iteration, we reach a fixed point

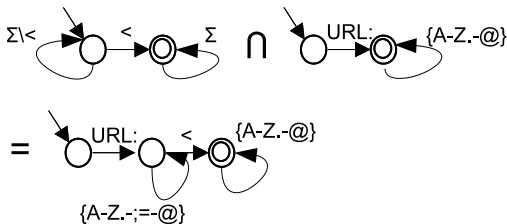


NEXT: Is it vulnerable?



## Detecting Vulnerabilities

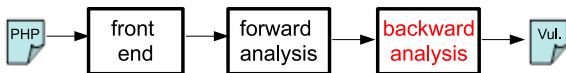
- We know all possible values of the **sink node (echo)**
- Given an attack pattern, e.g.,  $(\Sigma \setminus \langle \rangle)^* \langle \Sigma^*$ , if the intersection is not an empty set, the program is vulnerable. Otherwise, it is not vulnerable with respect to the attack pattern



NEXT: What are the malicious inputs?

## Generating Vulnerability Signatures

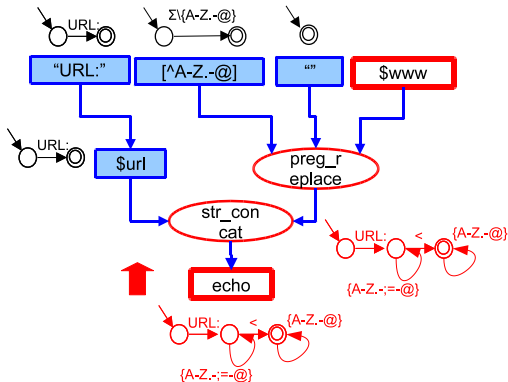
- A vulnerability signature is a characterization that includes **all malicious inputs** that can be used to generate attack strings
- Uses **backward** analysis starting from the sink node
- Uses *pre-image* computations on string operations:
  - $\text{preConcatPrefix}(M, M_2)$  returns  $M_1$  and  $\text{preConcatSuffix}(M, M_1)$  returns  $M_2$ , where  $M = M_1.M_2$ .
  - $\text{preReplace}(M, M_2, M_3)$  returns  $M_1$ , where  $M = \text{REPLACE}(M_1, M_2, M_3)$ .





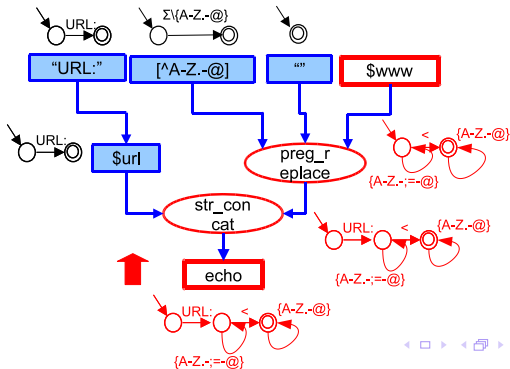
# Backward Analysis

- The first iteration is a simple assignment.



# Backward Analysis

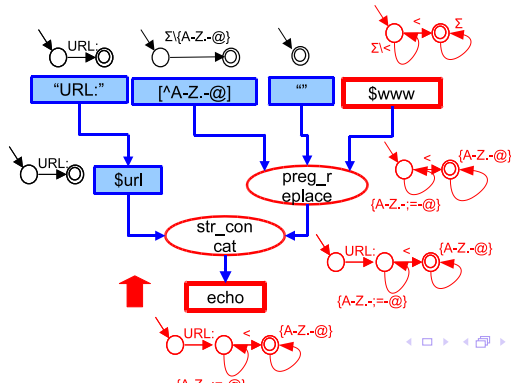
- At the second iteration, we call  $\text{preConcatSuffix}(\text{URL} : \{A-Z.-;=-@ \}^* < \{A-Z.-@ \}^*, \text{"URL:"})$ .
- $M = M_1.M_2$





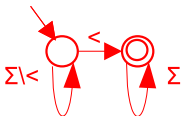
## Backward Analysis

- We call  $\text{preReplace}(\{A-Z.-;=-@\}^* < \{A-Z.-@\}^*, \Sigma \setminus \{A-Z.-@\}, "")$  at the third iteration.
- $M = \text{replace}(M_1, M_2, M_3)$
- After the third iteration, we reach a fixed point.



# Vulnerability Signatures

- The vulnerability signature is the result of the input node, which includes all possible malicious inputs
- An input that does not match this signature cannot exploit the vulnerability



NEXT: How to detect and prevent malicious inputs



## Patch Vulnerable Applications

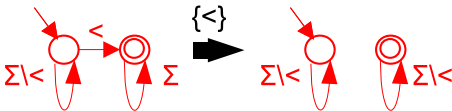
- Match-and-block: A patch that checks if the input string matches the vulnerability signature and halts the execution if it does
- Match-and-sanitize: A patch that checks if the input string matches the vulnerability signature and modifies the input if it does



# Sanitize

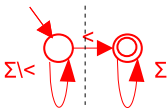
The idea is to modify the input by deleting certain characters (as little as possible) so that it does not match the vulnerability signature

- Given a DFA, an **alphabet cut** is a set of characters that after "removing" the edges that are associated with the characters in the set, the modified DFA does not accept any non-empty string



## Find An Alphabet Cut

- Finding a minimum alphabet cut of a DFA is an NP-hard problem (one can reduce the vertex cover problem to this problem)
- We apply a min-cut algorithm to find a cut that separates the initial state and the final states of the DFA
- We give higher weight to edges that are associated with alpha-numeric characters
- The set of characters that are associated with the edges of the min cut is an alphabet cut



{<} is an alphabet cut



## Patch Vulnerable Applications

A match-and-sanitize patch: If the input matches the vulnerability signature, delete all characters in the alphabet cut

```
| <?php  
| if (preg_match('/[<]*<.*\/',$_GET["www"]))  
| $_GET["www"] = preg_replace(Ò<Ó,"",$_GET["www"]);  
| 1: $www = $_GET["www"];  
| 2: $url = "URL:";  
| 3: $www = preg_replace("[^A-Z.-@]","", $www);  
| 4: echo $url. $www;  
| ?>
```



# Automatic Verification of String Manipulating Programs

- Symbolic String Vulnerability Analysis
- Relational String Analysis
- Composite String Analysis



## Relational String Analysis

Instead of multiple *single-track* DFAs, we use *one multi-track DFA*, where each track represents the values of one string variable.

Using multi-track DFAs we are able to:

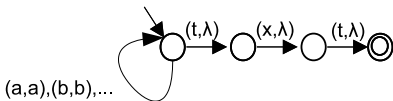
- Identify the *relations* among string variables
- Generate relational vulnerability signatures for multiple user inputs of a vulnerable application
- Prove properties that depend on relations among string variables, e.g.,  $\$file = \$usr.txt$  (while the user is *Fang*, the open file is *Fang.txt*)
- Summarize procedures
- Improve the precision of the path-sensitive analysis





## Multi-track Automata

- Let  $X$  (the first track),  $Y$  (the second track), be two string variables
- $\lambda$  is a padding symbol
- A multi-track automaton that encodes  $X = Y.txt$



## Relational Vulnerability Signature

- Performs forward analysis using multi-track automata to generate relational vulnerability signatures
- Each track represents one user input
- An auxiliary track represents the values of the current node
- Each constant node is a single track automaton (the auxiliary track) accepting the constant string
- Each user input node is a two track automaton (an input track + the auxiliary track) accepting strings that two tracks have the same value



## Relational Vulnerability Signature

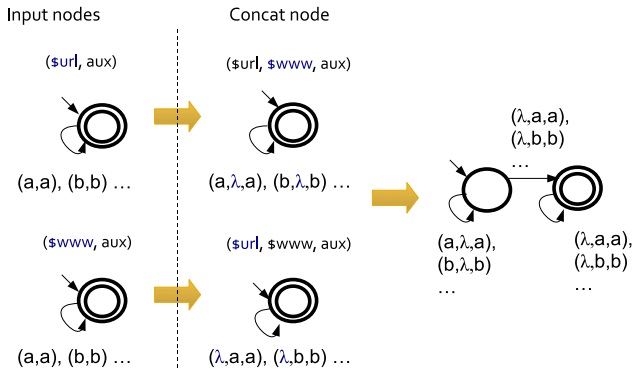
Consider a simple example having multiple user inputs

```
| <?php  
| 1: $www = $_GET["www"];  
| 2: $url = $_GET["url"];  
| 3: echo $url. $www;  
| ?>
```

Let the attack pattern be  $(\Sigma \setminus \langle \rangle)^* \langle \rangle \Sigma^*$



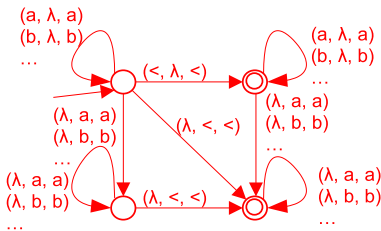
# Signature Generation



## Relational Vulnerability Signature

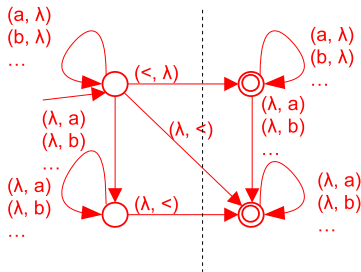
Upon termination, intersects the auxiliary track with the attack pattern

- A multi-track automaton: ( $\$url$ ,  $\$www$ ,  $aux$ )
- Identifies the fact that the concatenation of two inputs contains  $<$



## Relational Vulnerability Signature

- Projects away the auxiliary track
- Finds a min-cut
- This min-cut identifies the alphabet cuts:
  - $\{<\}$  for the first track ( $\$url$ )
  - $\{<\}$  for the second track ( $\$www$ )



## Patch Vulnerable Applications with Multi Inputs

Patch: If the inputs match the signature, delete its alphabet cut

```
| <?php  
| if (preg_match('/[<^ <]*<.*\/', $_GET["url"].$_GET["www"]))  
| {  
|   $_GET["url"] = preg_replace("<","",$_GET["url"]);  
|   $_GET["www"] = preg_replace("<","",$_GET["www"]);  
| }  
| 1: $www = $_GET["www"];  
| 2: $url = $_GET["url"];  
| 3: echo $url. $www;  
| ?>
```



## Other Technical Issues

To conduct relational string analysis, we need a meaningful "intersection" of multi-track automata

- **Intersection** are closed under **aligned** multi-track automata
- $\lambda$ s are **right justified** in all tracks, e.g.,  $ab\lambda\lambda$  instead of  $a\lambda b\lambda$
- However, there exist unaligned multi-track automata that are **not describable** by aligned ones
- We propose an alignment algorithm that constructs aligned automata which **under/over approximate** unaligned ones





## Other Technical Issues

Modeling Word Equations:

- **Intractability of  $X = cZ$** : The number of states of the corresponding aligned multi-track DFA is **exponential** to the length of  $c$ .
- **Irregularity of  $X = YZ$** :  $X = YZ$  is not describable by an aligned multi-track automata

We have proven the above results and proposed a **conservative** analysis.



# Automatic Verification of String Manipulating Programs

- Symbolic String Vulnerability Analysis
- Relational String Verification
- Composite String Analysis



# Composite Verification

We aim to extend our string analysis techniques to analyze systems that have **unbounded string and integer variables**.

We propose a composite static analysis approach that combines **string analysis** and **size analysis**.



# String Analysis

*Static String Analysis*: At each program point, statically compute the possible values of **each string variable**.

The values of each string variable are over approximated as a regular language accepted by a **string automaton** [Yu et al. SPIN08].

String analysis can be used to detect **web vulnerabilities** like SQL Command Injection [Wassermann et al, PLDI07] and Cross Site Scripting (XSS) attacks [Wassermann et al., ICSE08].



## Size Analysis

*Integer Analysis:* At each program point, statically compute the possible states of the values of **all integer variables**.

These infinite states are symbolically over-approximated as linear arithmetic constraints that can be represented as **an arithmetic automaton**

Integer analysis can be used to perform **Size Analysis** by representing lengths of string variables as integer variables.



## What is Missing?

Consider the following segment.

- 1: <?php
- 2: \$www = \$\_GET["www"];
- 3: \$l\_otherinfo = "URL";
- 4: \$www = ereg\_replace("[^A-Za-z0-9 ./-@://]", "", \$www);
- 5: if(strlen(\$www) < \$limit)
- 6: echo "<td>" . \$l\_otherinfo . ": " . \$www . "</td>";
- 7: ?>



## What is Missing?

If we perform **size analysis solely**, after line 4, we do not know the length of \$www.

- 1:<?php
- 2: \$www = \$\_GET["www"];
- 3: \$\_otherinfo = "URL";
- 4: **\$www = ereg\_replace("[^A-Za-z0-9 ./-@://]", "", \$www);**
- 5: if(strlen(\$www) < \$limit)
- 6: echo "<td>" . \$\_otherinfo . ": " . \$www . "</td>";
- 7: ?>



## What is Missing?

If we perform **string analysis solely**, at line 5, we cannot check/enforce the branch condition.

- 1:<?php
- 2: \$www = \$\_GET["www"];
- 3: \$\_otherinfo = "URL";
- 4: \$www = ereg\_replace("[^A-Za-z0-9 ./-@://]", "", \$www);
- 5: **if(strlen(\$www) < \$limit)**
- 6: echo "<td>" . \$\_otherinfo . ": " . \$www . "</td>";
- 7:?>





# What is Missing?

We need a **composite analysis** that combines string analysis with size analysis.

Challenge: How to transfer information between string automata and arithmetic automata?



## Some Facts about String Automata

- A **string automaton** is a single-track DFA that accepts a regular language, whose length forms a **semi-linear set**, .e.g.,  $\{4, 6\} \cup \{2 + 3k \mid k \geq 0\}$
- The unary encoding of a semi-linear set is uniquely identified by a **unary automaton**
- The unary automaton can be constructed by replacing the alphabet of a string automaton with a unary alphabet



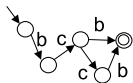
## Some Facts about Arithmetic Automata

- An **arithmetic automaton** is a multi-track DFA, where each track represents the value of one variable over a binary alphabet
- If the language of an arithmetic automaton satisfies a **Presburger formula**, the value of each variable forms a semi-linear set
- The semi-linear set is accepted by the **binary automaton** that projects away all other tracks from the arithmetic automaton

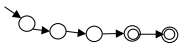


# An Overview

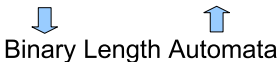
To connect the dots, we propose a novel algorithm to convert unary automata to binary automata and vice versa.



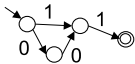
String Automata



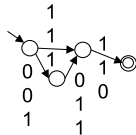
Unary Length Automata



Binary Length Automata



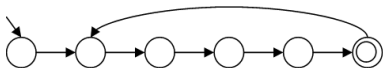
Arithmetic Automata



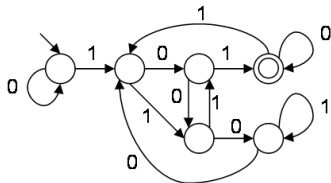
## An Example of Length Automata

Consider a string automaton that accepts  $(great)^+$ .  
The length set is  $\{5 + 5k | k \geq 0\}$ .

- 5: in unary 11111, in binary 101, from lsb **101**.
- 1000: in binary 1111101000, from lsb **0001011111**.



(a) Unary



(b) Binary

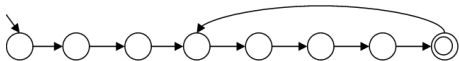


## Another Example of Length Automata

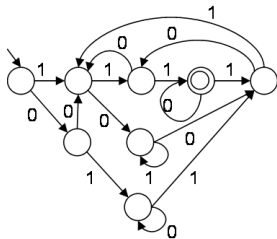
Consider a string automaton that accepts  $(great)^+cs$ .

The length set is  $\{7 + 5k | k \geq 0\}$ .

- 7: in unary 1111111, in binary 1100, from lsb **0011**.
- 107: in binary 1101011, from lsb **1101011**.
- 1077: in binary 10000110101, from lsb **10101100001**.



(c) Unary



(d) Binary



# STRANGER Tool

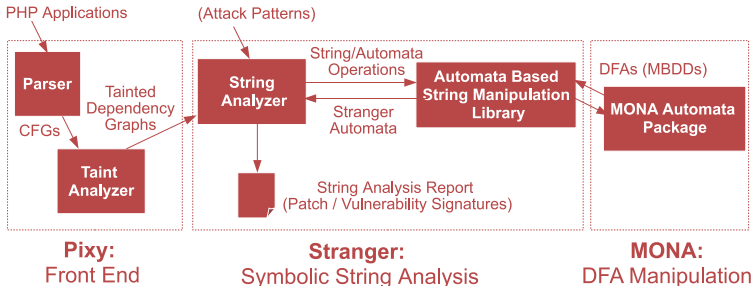
We have developed **STRANGER** (**STR**ing **A**utomato**N**  
**G**enerato**R**)

- A public automata-based string analysis tool for PHP
- Takes a PHP application (and attack patterns) as input, and automatically analyzes all its scripts and outputs the possible XSS, SQL Injection, or MFE vulnerabilities in the application



# STRANGER Tool

- Uses Pixy [Jovanovic et al., 2006] as a front end
- Uses MONA [Klarlund and Møller, 2001] automata package for automata manipulation



The tool, detailed documents, and several benchmarks are available: <http://www.cs.ucsb.edu/~vlab/stranger>.





# STRANGER Tool

A case study on Schoolmate 1.5.4

- 63 php files containing 8000+ lines of code
- Intel Core 2 Due 2.5 GHz with 4GB of memory running Linux Ubuntu 8.04
- STRANGER took **22 minutes / 281MB** to reveal **153 XSS** from 898 sinks
- After manual inspection, we found **105 actual vulnerabilities (false positive rate: 31.3%)**
- We inserted patches for all actual vulnerabilities
- Stranger proved that our patches are correct with respect to the attack pattern we are using



# STRANGER Tool

Another case study on SimpGB-1.49.0, a PHP guestbook web application

- 153 php files containing 44000+ lines of code
- Intel Core 2 Due 2.5 GHz with 4GB of memory running Linux Ubuntu 8.04
- For all executable entries, STRANGER took
  - 231 minutes to reveal 304 XSS from 15115 sinks,
  - 175 minutes to reveal 172 SQLI from 1082 sinks, and
  - 151 minutes to reveal 26 MFE from 236 sinks



## Related Work on String Analysis

- String analysis based on context free grammars: [Christensen et al., SAS'03] [Minamide, WWW'05]
- String analysis based on symbolic execution: [Bjorner et al., TACAS'09]
- Bounded string analysis: [Kiezun et al., ISSTA'09]
- Automata based string analysis: [Xiang et al., COMPSAC'07] [Shannon et al., MUTATION'07] [Barlzarotti et al. S&P'08][Veneas et al., POPL'15][Wang et al. CAV'16]
- String constraint solving: [CVC4] [Z3, Z3-Str, Z3-Str2,2016] [SSS, S3P] [Norn] [Slog, Slender (Wang et al. CAV'16, ASE'18)]
- Application of string analysis to web applications: [Wassermann and Su, PLDI'07, ICSE'08] [Halfond and Orso, ASE'05, ICSE'06]



## Related Work on Size Analysis and Composite Analysis

- Size analysis : [Dor et al., SIGPLAN Notice'03] [Hughes et al., POPL'96]  
[Chin et al., ICSE'05] [Yu et al., FSE'07] [Yang et al., CAV'08]
- Composite analysis:
  - Composite Framework: [Bultan et al., TOSEM'00]
  - Symbolic Execution: [Xu et al., ISSTA'08] [Saxena et al., UCB-TR'10]
  - Abstract Interpretation: [Gulwani et al., POPL'08] [Halbwachs et al., PLDI'08]
- Model Counting Analysis: [William et al., ESEC/FSE'18]



## Related Work on Vulnerability Signature Generation

- Test input/Attack generation: [Wassermann et al., ISSTA'08] [Kiezun et al., ICSE'09]
- Vulnerability signature generation: [Brumley et al., S&P'06] [Brumley et al., CSF'07] [Costa et al., SOSP'07][Yu et al. ISSTA'16]



## Take Away: Future Direction

To have impact.

- What will be the most dominant software platform?
- What will be the major roadblock?
- What will be the key techniques?



## Take Away: Software Dependability

- Software will become parts of our live.
  - Web applications and services
  - Mobile applications
  - IoT applications
  - Smart contract applications
  - Machine learning applications
- Security and dependability will be the major roadblock
- An **automatic and scalable** verification framework to achieve *dependability of web applications*



Thank you for your attention.

Questions?

