



復旦大學  
FUDAN UNIVERSITY



JOHNS HOPKINS  
CAREY BUSINESS SCHOOL



清華大學


# TextExerciser: Feedback-driven Text Input Exercising for Android Applications

Yuyu He<sup>\*,1</sup>, **Lei Zhang**<sup>\*,1</sup>, Zhemin Yang<sup>1</sup>, Yinzhi Cao<sup>2</sup>, Keke Lian<sup>1</sup>, Shuai Li<sup>1</sup>,  
Wei Yang<sup>3</sup>, Zhibo Zhang<sup>1</sup>, Min Yang<sup>1</sup>, Yuan Zhang<sup>1</sup>, Haixin Duan<sup>4</sup>

1. Fudan University
2. Johns Hopkins University,
3. University of Texas at Dallas,
4. Tsinghua University

\*. The first two authors have contributed equally to this work.

P: 0/1 dx: 0.0 dy: 0.0 Xv: 0.0 Yv: 0.0 Prs: 0.40 Size: 0.01



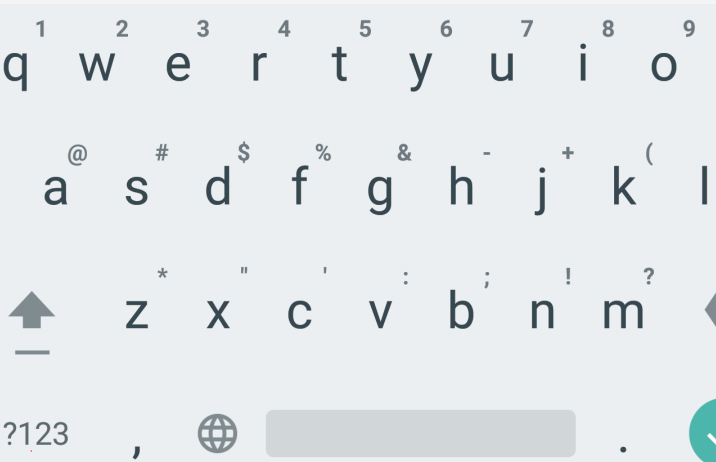
# You'll need a password

Make sure it's 6 characters or more.

.....

Your password must be at least 6 characters.

Next



1 2 3 4 5 6 7 8 9  
q w e r t y u i o  
@ # \$ % & - + ( )  
↑ z x c v b n m  
?123 , .

P: 0/1 dx: 0.0 dy: 0.0 Xv: 0.0 Yv: 0.0 Prs: 0.75 Size: 0.02

P: 0/1 dx: 0.0 dy: 0.0 Xv: 0.0 Yv: 0.0 Prs: 0.32 Size: 0.01

## Sign up

SIGN UP

Your Name

Email

Password (min. 6 characters)


Your Birthday\*

\*Telling us your age will allow us to connect you with other parents

By signing up, you agree to our [Terms](#) and [Privacy Policy](#).

P: 0/1 dx: 25.0 dy: 26.0 Xv: 0.0 Yv: 0.0 Prs: 0.26 Size: 0.0

10:28



### Clone WhatsWeb

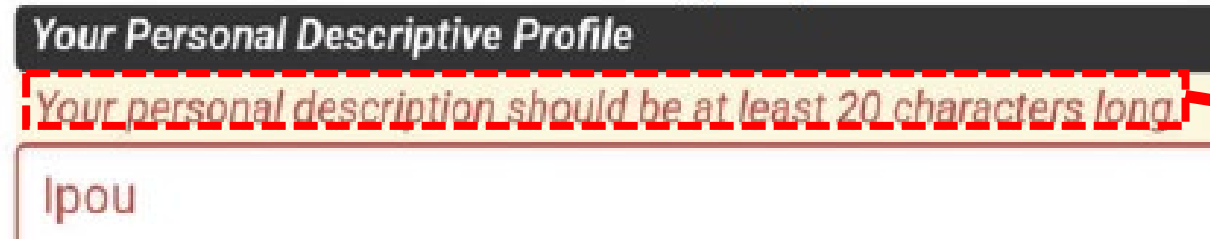
Enter 5 Digit password

SAVE CANCEL

Page 5

# Problem

- Text-based inputs of UI exerciser
  - UI exerciser
    - Automatically drives android apps to reach different code branches so that dynamic analysis can be improved
  - Text-based inputs



Constraint of the input field

- Problem

How to generate valid text inputs?

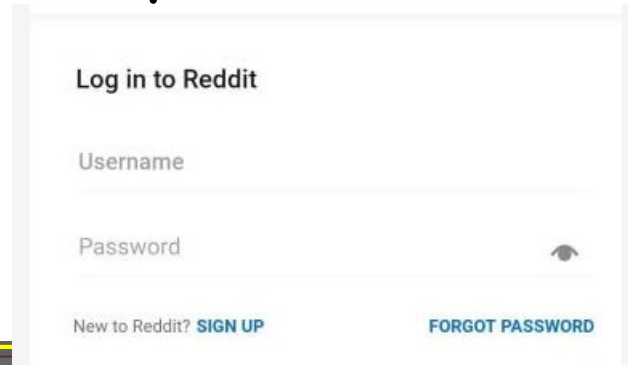
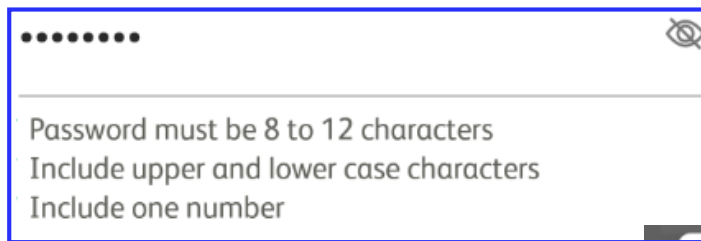
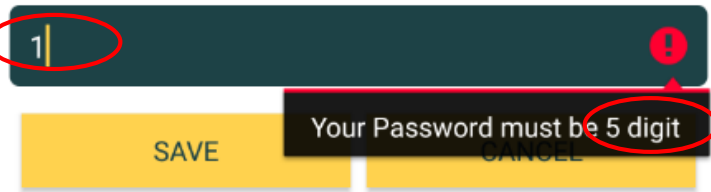
# State-of-the-art

- Summarize input patterns to pre-define inputs
  - AppsPlayground [CODASPY' 13]
  - Arnatovich et al. [APSEC' 16]
- Machine learning to automatically learn inputs
  - Liu et al. [ICSE' 17]
- Symbolic execution to extract input constraints from app code and utilizes a solver to generate inputs
  - Mobolic [SPE' 18]
- Third-party login to bypass inputs
  - Authscope [CCS' 17]

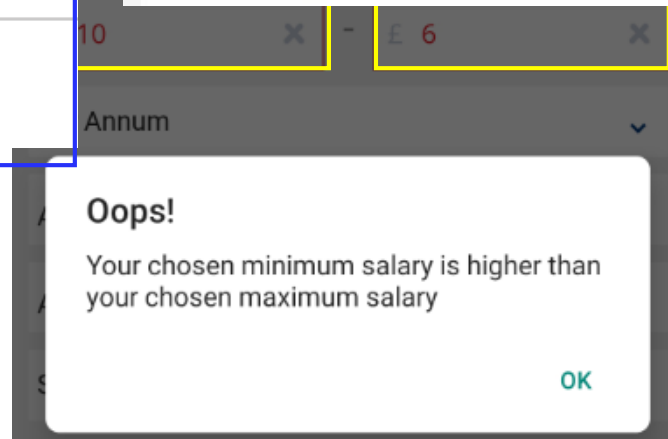
# But wait...

- Different apps have various ways to constrain text inputs

Restrictions of same type inputs are different



Many apps don't support third-party login

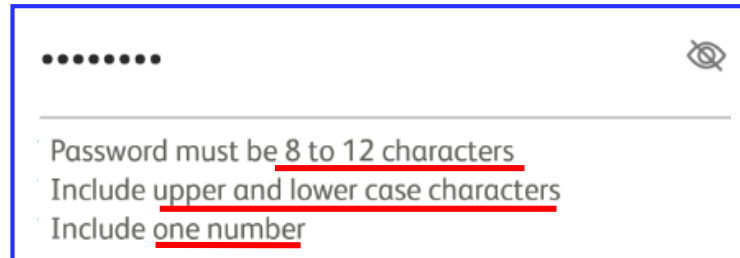


Two input fields are correlated

If generated input doesn't satisfy input constraints, the existing exercisers will stuck !

# Our Work

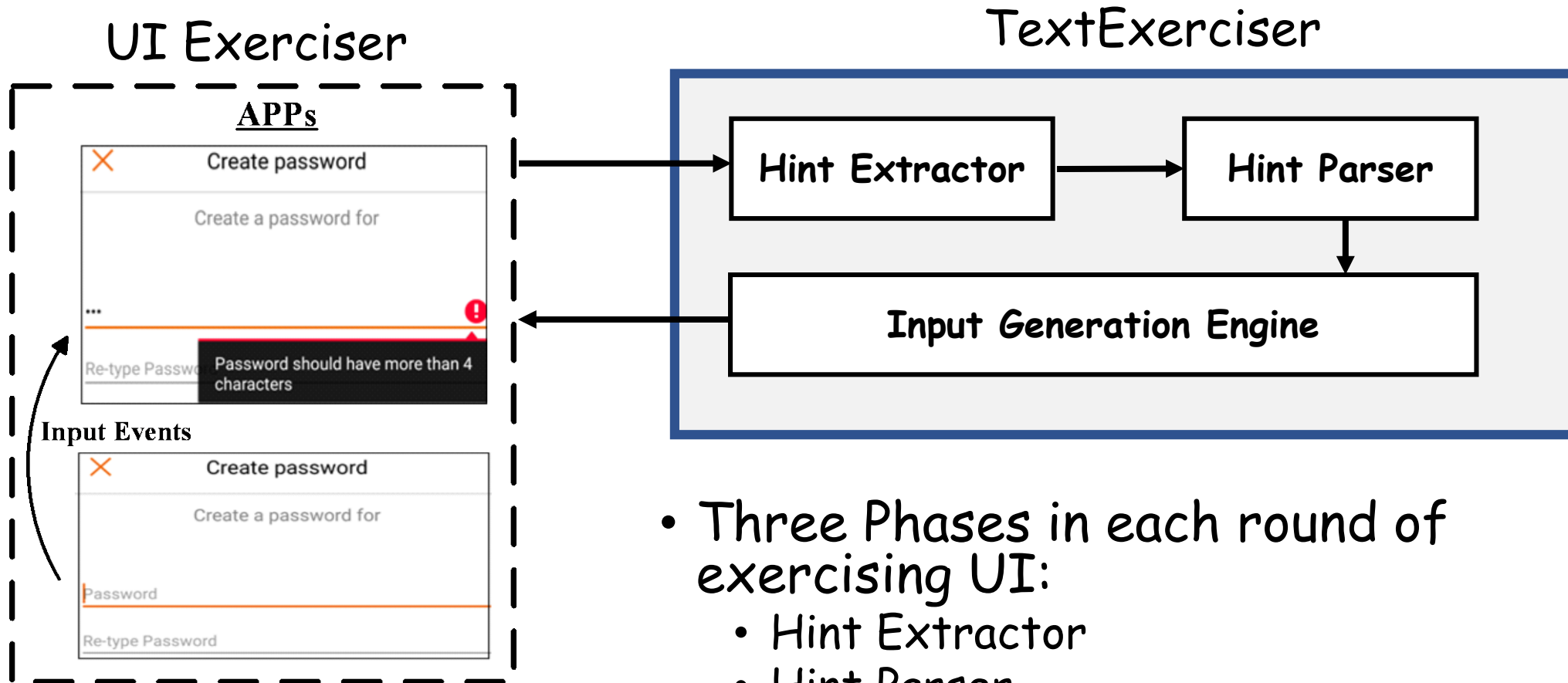
- TextExerciser
  - Feedback-driven text input generator
  - Iteratively generates inputs based on hints shown on UI
- Key Insight
  - The feedback information (i.e., hints shown on UI) coming from app servers can guide the input generation of UI exercisers.



# Contributions

- We propose the first feedback-driven input exerciser that iteratively generates text inputs using a constraint solver based on hints from the target app.
- We implement a prototype of our text input exerciser and the source code of TextExerciser is available at GitHub (<https://github.com/yyyyHe/TextExerciser>).
- We evaluate the performance of TextExerciser on popular Google Play apps. The evaluation result shows that TextExerciser achieves higher code coverage than state-of-the-art tools and also finds more privacy leaks and vulnerabilities when combined with existing dynamic analysis tools.

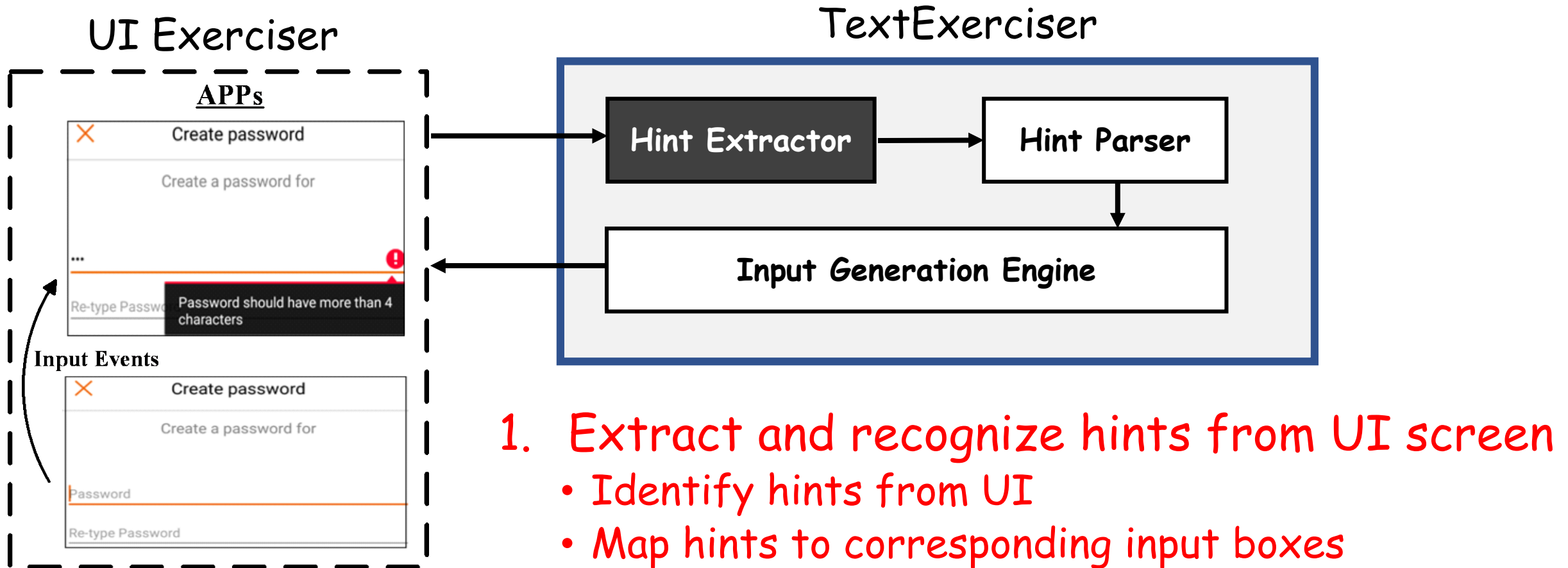
# TextExerciser Overview



- Three Phases in each round of exercising UI:
  - Hint Extractor
  - Hint Parser
  - Input generation Engine



# TextExerciser Overview



# Hints Extractor

- Identify hints

- Hints on UIs

- Dynamic hints

- Appear when users touch the input box or inject a wrong input

- Static hints

- Appear together with input fields

- Hint Classifier

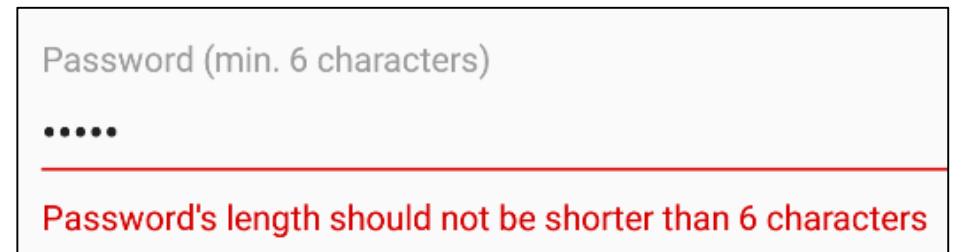
- Positive samples

- Dynamic hints

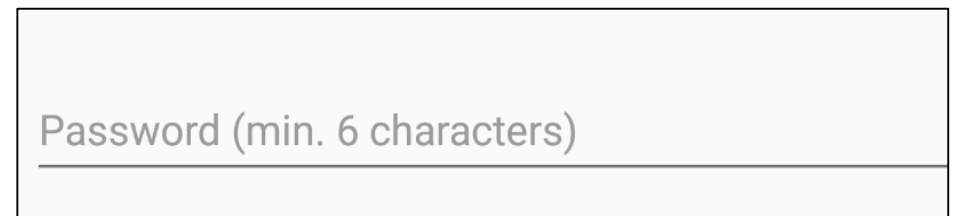
- Negative samples

- Static texts

- texts on UIs which have no input fields.



Dynamic hint example: A password input field with the label "Password (min. 6 characters)". The input contains five dots. A red error message is displayed below the field: "Password's length should not be shorter than 6 characters". A blue arrow points from the "Dynamic hints" section of the text to this screenshot.



Static hint example: A password input field with the label "Password (min. 6 characters)". The input is empty. A blue arrow points from the "Static hints" section of the text to this screenshot.

# Hints Extractor

- Map hints to corresponding input fields

- Keywords mapping

- Text related to the input field
    - Keywords in a hint

The same keywords

A screenshot of a form field. The label 'Gamer ID' is in green. The input contains 'Gravatio&++++tter2'. Below the input, a red dashed box highlights a hint: 'Gamer ID must be 3 to 20 letters and numbers'. To the right of the hint, the text '20 / 20' is visible. A red arrow points from the text 'The same keywords' to the red dashed box.

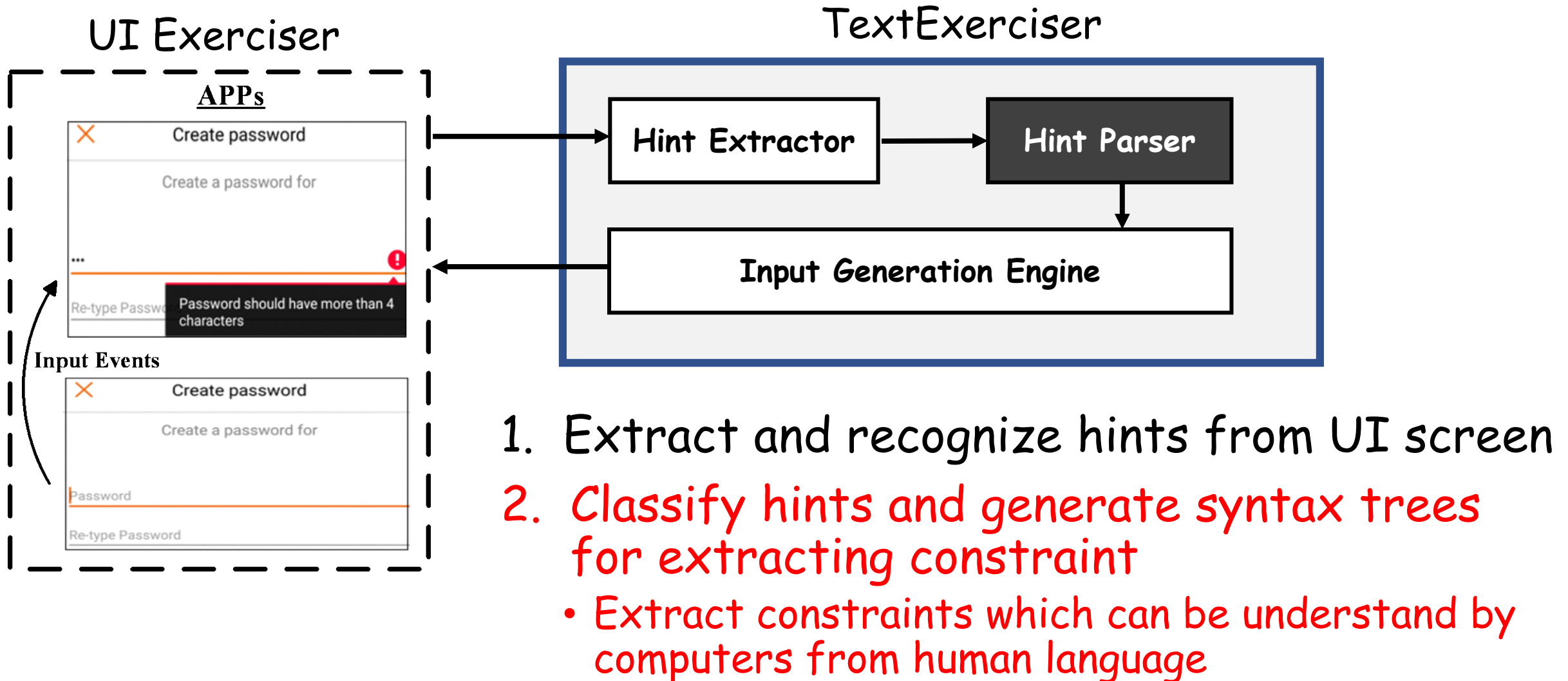
- Shortest-distance

- Map each hint to its nearest input fields
    - Multiple hints can be mapped to a single input field

Shortest distance between them

A screenshot of two form fields. The top field contains 'qq@1.1'. The bottom field contains '.....'. Below the bottom field, a red dashed box highlights a shared hint: 'Minimum of 8 characters ✓ 1 uppercase letter ✓ 1 lowercase letter ✓ 1 number ✓'. A red arrow points from the text 'Shortest distance between them' to the red dashed box.

# TextExerciser Overview



# Hint Parser

- Key observation:

Similar semantics



Similar syntax structure

- Classify hints

- Classifier

- Multi-Text CNN and RNN classifier[1]

- Training data

- 1,200 popular apps
- Manual labelled 1,548 hints
- 4 major, 10 minor, 18 sub-minor

- [1]: <https://github.com/jiegzhan/multi-class-text-classification-cnn-rnn>.

MajorCategory	MinorCategory	SubMinorCategory
Precise Single-field	Length Constraints	The lower bound of input length
		The upper bound of input length
		A range of input length
		A fixed input length
	Existence Constraints	Input should contain certain characters
		Input should not contain certain characters
	Value Constraints	The lower bound of value
		The upper bound of value
		A range of value
Fuzzy Single-field	Length Constraints	Require longer input
		Require shorter input
	Value Constraints	Require larger value
		Require smaller value
	Non-directional Constraints	Non-directional Constraints on invalid input
Precise Joint-fields	Equivalence Constraints	The equivalence of two input fields
	Non-repetitive Constraints	Value of two input fields can't be the same
	Value Restriction	The comparison of values in two input fields
Fuzzy Joint-fields	Non-directional Constraints	The relationship of the two field need domain knowledge

# Hint Parser

- Generate Syntax Trees via Stanford Parser

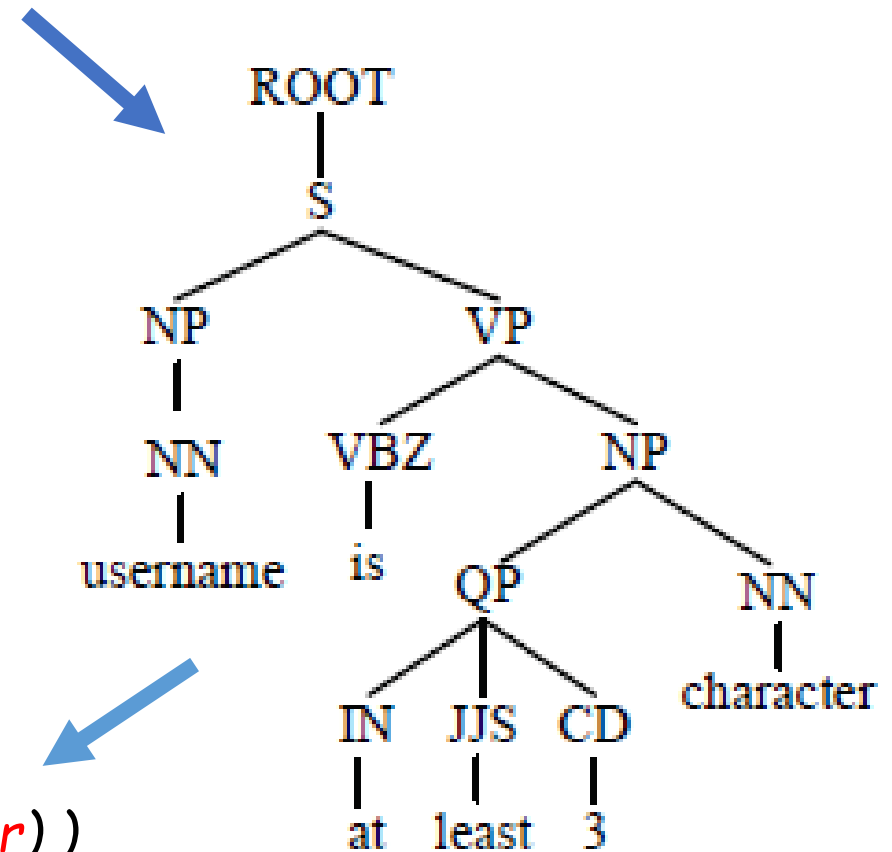
- Constraint representation

- Traverse trees and extract constraint

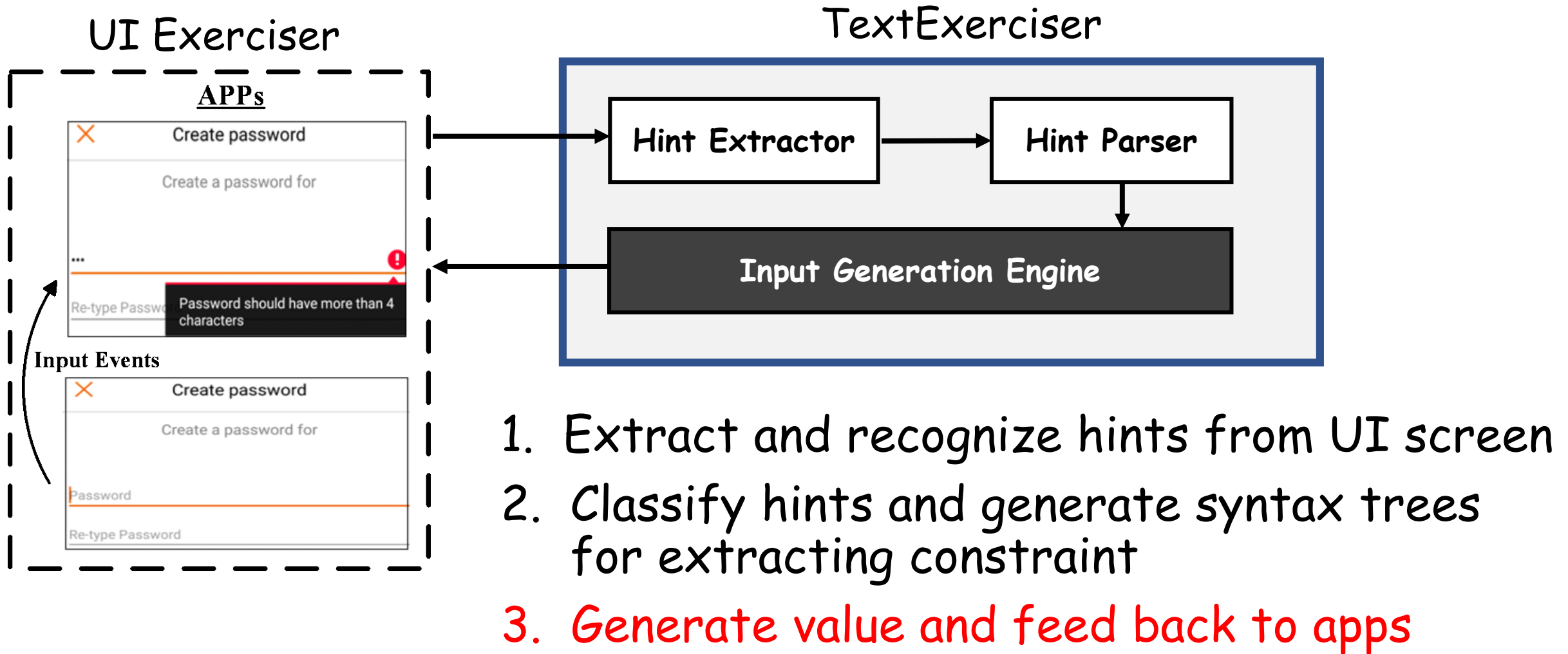
- *Length*: range of length
  - *Content*: restricts certain format
    - E.g. digital, character
  - *Value*: range of value
    - E.g. your weight must between 10 and 999

- Constraint representation

- `LengthConstraint ( username; Range(6; Infity) )`
- `ContentConstraint ( username; Format(character) )`



# TextExerciser Overview



# Input Generation Engine

- **Generate inputs**
  - **Solve constraints**
    - Obtain concrete values from constrain representation
    - Use Z3StrSolver to generate input
    - Joint-field input
      - Generate one, apply the constraint to the other
  - **External sources**
    - Email, phone
      - Pre-register
    - Pin codes
      - Fetch from receive message via email /phone



# Evaluation

- RQ1: is TextExerciser more effective than existing tools in exercising Android apps?
- RQ2: can TextExerciser improve existing dynamic analysis of Android apps?
- RQ3: is TextExerciser efficient for generating text input for popular Android apps?

# Datasets

- Top 500 apps from all the categories except for games from Google Play
- Training set
  - 1200 apps for manually label hints
- Testing sets
  - Small dataset (40 apps)
    - Instrumented by method measurement tool, i.e. Ella
  - Large dataset (6000 apps)

# State-of-the-art Tools

- Combine TextExerciser with Open-sourced UI Exercisers

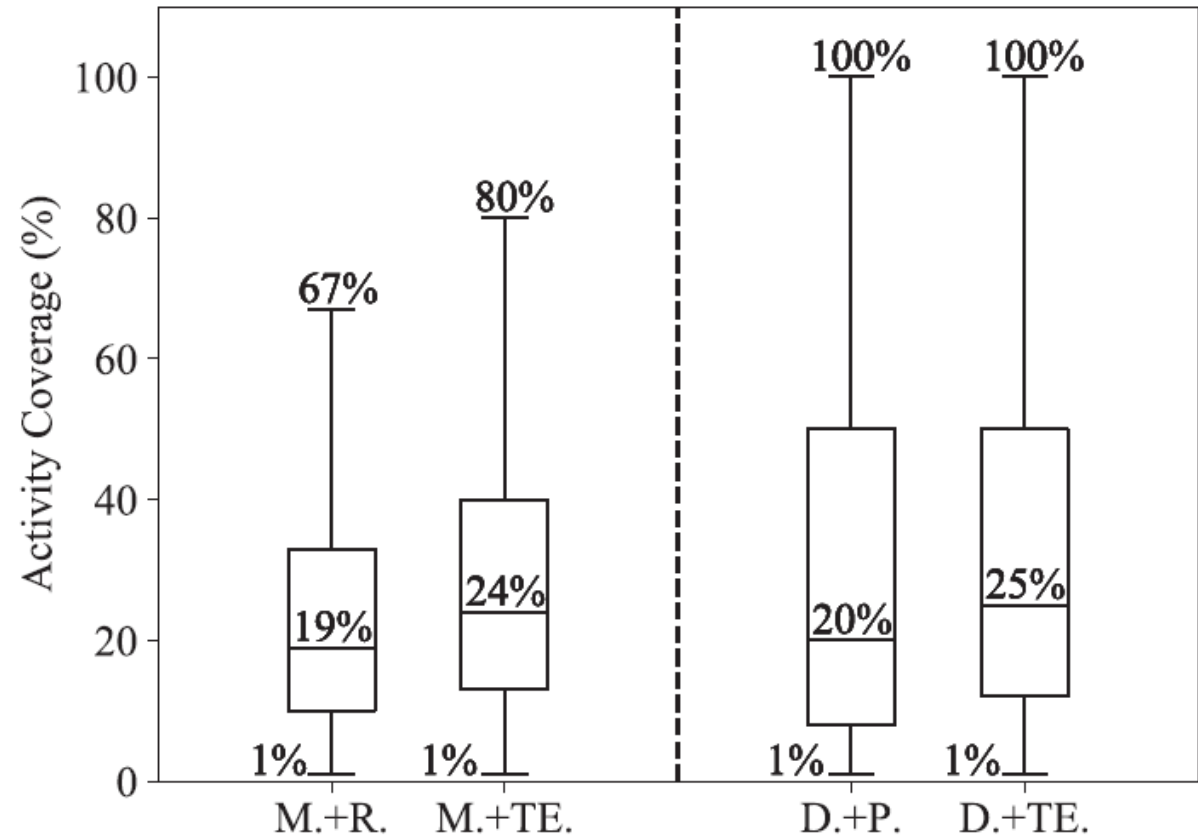
Tool	Text Input Strategy	Abbreviation	Combination
Monkey	Random clicking	M. + R.	M. + TE.
DroidBot	Predefined	D. + P.	D. + TE.
Stoat	Random	St. + R.	St. + TE.

# Code Coverage

- Comparison with State-of-the-art Testing Tools
- Small dataset (40 apps)
  - Method and Activity coverage
  - Mitigate randomness
    - Fixed seed for Monkey
    - Run 3 times

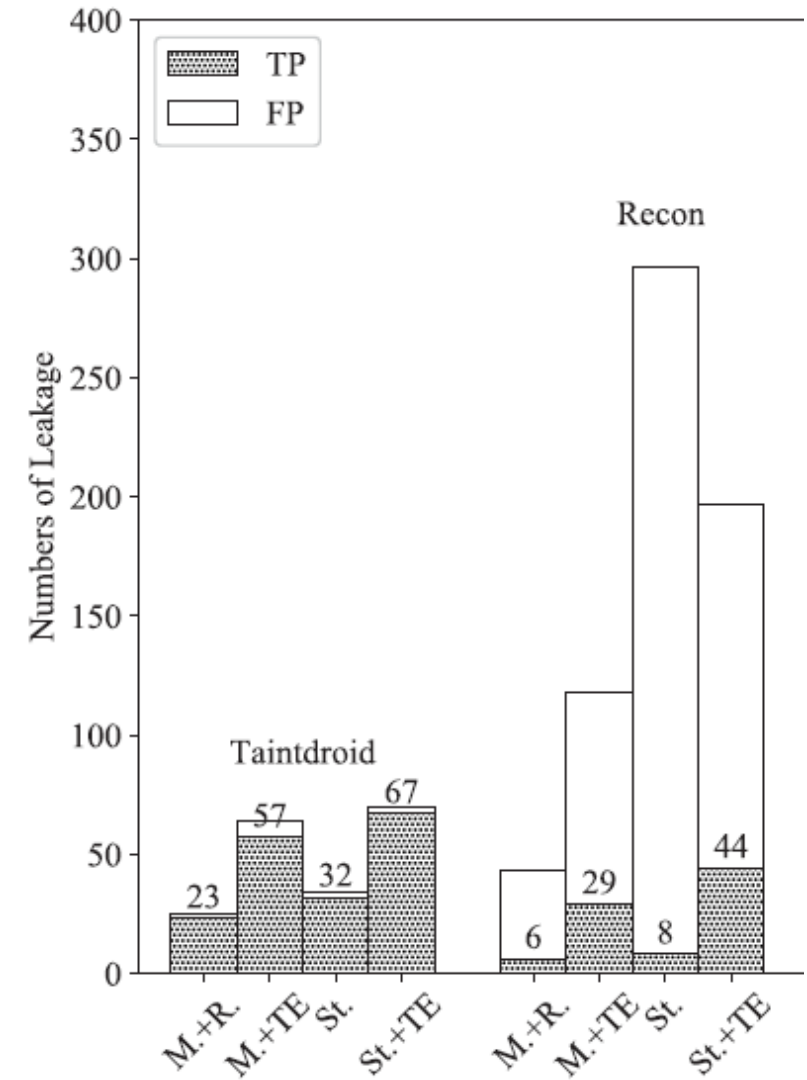
Combination	Activity Improve	Method Improve
M. + TE.	48.5%	29.0%
D. + TE.	45.3%	26.4%
St. + TE.	37.0%	20.2%

- Large dataset(6000apps)



# Behavior Coverage

- Work with dynamic analysis tool for privacy leak detection
  - Taintdroid: taint analysis
  - Recon: traffic analysis
- In small dataset
  - More privacy leaks with the help of TextExerciser



# Cases

- Insecure configuration of SSL communication
- Transfer of User Credential or Private Information in HTTP

App Name	#Downloads	Description
<i>Previously-unknown Vulnerabilities:</i>		
BlackWhiteMeet	100,000+	Doesn't verify signature in https
Coco	10,000,000+	Leak user credential in http
10times	100,000+	Leak user location and device info in http
Yippi	100,000+	Change user password in http
Saviry	100,000+	Modify user profile in http
Eskimi	1,000,000+	Leak user credential and profile in http

# Efficient of TextExerciser

- Performance of TextExerciser
  - Hint identification classifier
    - Accuracy (90.2%), Precision (89.4%) , Recall (90.2%)
  - Hint parser
    - Coverage (87.3%)
- Number of trials in generation
  - First-round success rate : 95.1%
  - Most are finished in three rounds
  - 1.2% exceed 30 trials (limitation)
    - Most of them require certain external knowledge to solve
      - e.g. invite code

# Conclusions

- TextExerciser: iterative, feedback-driven text input exerciser, generates text inputs for Android apps.
- Implement its prototype and the source code is available at <https://github.com/yyyyHe/TextExerciser>.
- Can improve code/behavior coverage of existing exercisers and dynamic analysis tools



# Thanks !

## Q&A



复旦白泽战队

• [zx1@fudan.edu.cn](mailto:zx1@fudan.edu.cn)