SQL Injection Detection
Using Fuzzy Logic and Naïve Bayes

FIT5190 IT Research Method

- **Group 6**
- **HU Ying – 2819******
- **Date: 25/05/2017**

Southeast University-Monash University Joint Graduate School

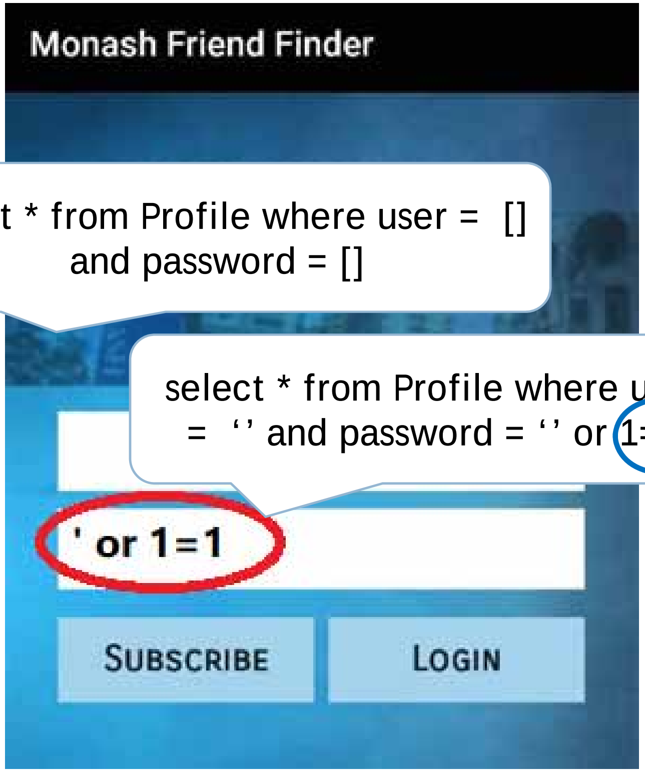
Objectives

- This research proposes a novel detection method for SQL injection using fuzzy logic and the Naïve-Bayes Algorithm.

Content

- Background
- Existing Detection Methods
- Proposed Method
- Conclusion

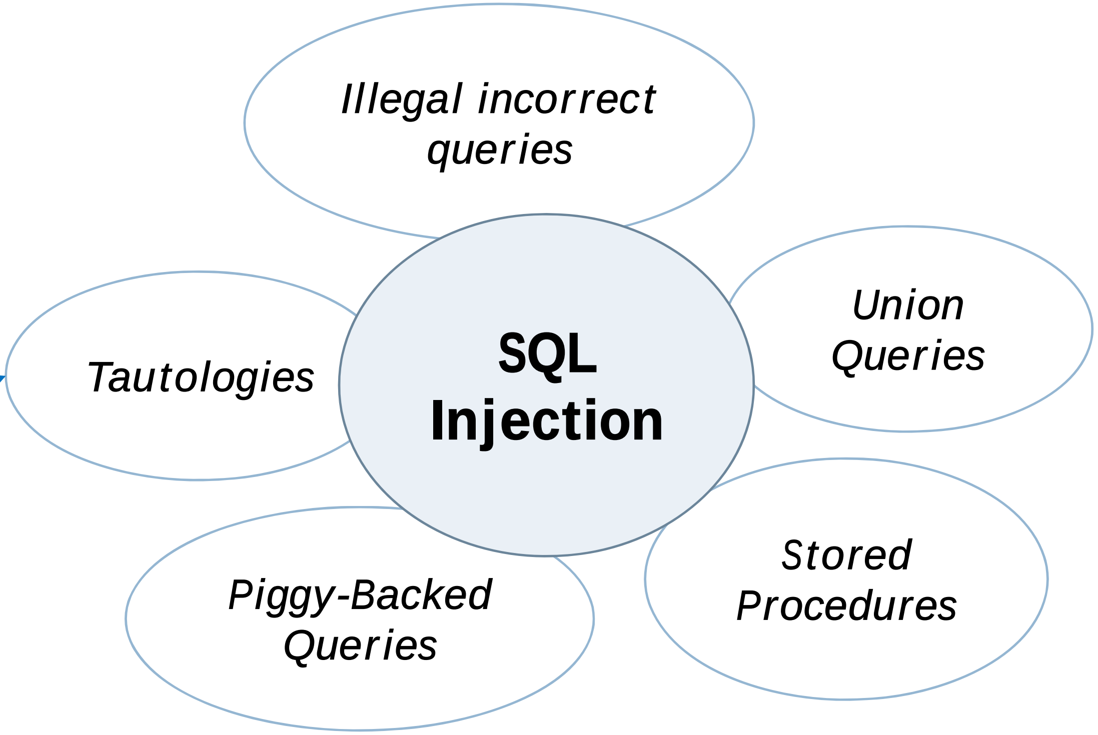
Background



select * from Profile where user = [] and password = []

select * from Profile where user = ‘’ and password = ‘’ or 1=1

Identity equation



Existing Detection Methods

Pattern matching methods based on
Aho-Corasick Algorithm v
phase and dynamatic ph

Removing the attribute values in SQL
operate exclusive OR on
ed and pre-defined

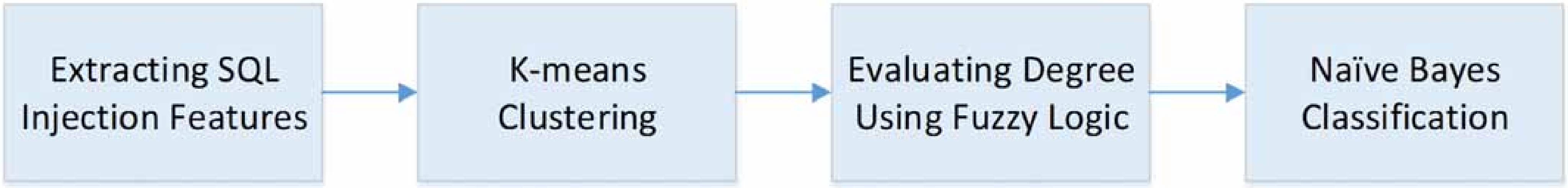
But …

- *Lack Flexibility and Scalability*
- *May lose efficiency with variety of malicious codes explosively increasing*

Estimating the query re
generated queries and co
that of normal query [3]

SQL queries during the
and debugging phases to
injection vulnerabilities,
such as Sania [4]

Proposed Method



Extracting SQL Injection Features

K-means Clustering

Evaluating Degree Using Fuzzy Logic

Naïve Bayes Classification

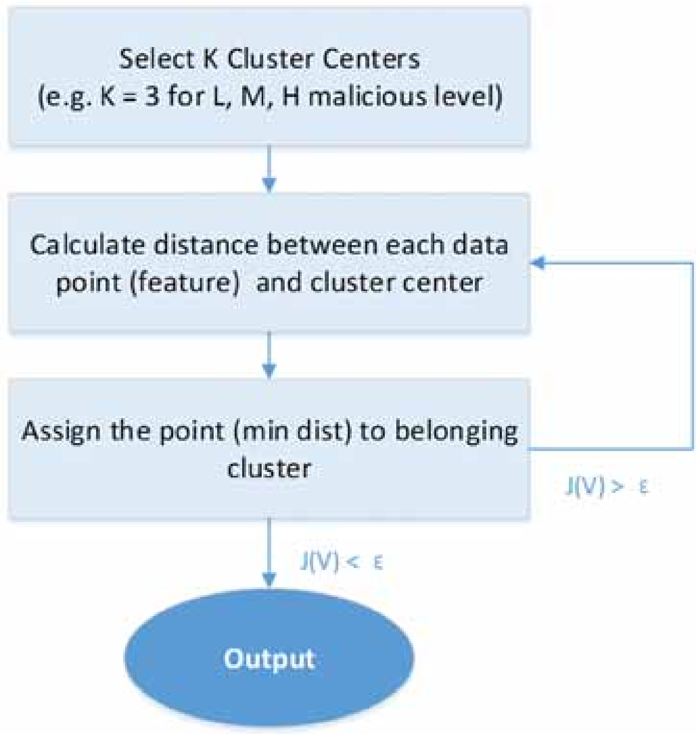
- Process of Proposed Method

Proposed Method (Continued)

1 *Extracting SQL Injection Features*

- f1: Frequencies of dangerous characters
  - *e.g. --, #, /*, ‘, “, | |, =*
- f2: Frequencies of dangerous tokens
  - *e.g. rename, drop, delete, insert, exec*
- f3: Length of SQL statement
- f4: Existence of statements always true
  - *e.g. 1=1, @=@, 123=123*
- … *(More features. Here 7 features are used)*

2 *K-means Clustering*



Select K Cluster Centers
(e.g. K = 3 for L, M, H malicious level)

Calculate distance between each data point (feature) and cluster center

Assign the point (min dist) to belonging cluster

Output

Proposed Method (Continued)

3 *Evaluating Degree Using Fuzzy Logic*

fuzzy Logic is an efficient and flexible method for managing degrees of uncertainty in attack detection.

7 features as Input, (L, H, M)

- *e.g.* $\mu_L(f1) = \frac{1}{1} + \frac{0.5}{2} + \frac{0}{3}$

Triangular function:

$$f(x;a,c) = \max(\min(\frac{x-a}{b-a}, \frac{c-x}{c-b}),0)$$

- *a, b, c – lower, center , upper limits of a cluster*

Fuzzy Rule

- IF f1 is H AND f2 is M THEN the Degree of SQL query is H
- … (More)

4 *Naïve Bayes Classification*

$$P(h|D) = \frac{P(D|h)\overset{\Sigma\mu_R(h)}{\boxed{P(h)}}}{P(D)}$$

Conclusion

- Adaption capability to detect new types of attacks
- The utility of fuzziness lessens the influence of the quality of training dataset

References

1. Prabakar, M., Karthikeyan, M., & Marimuthu, K. (2013). An efficient technique for preventing SQL injection attack using pattern matching algorithm. 2013 International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 503-506.
2. Lee, I, Jeong, S, Yeo, S, & Moon, J. (2012). A novel method for SQL injection attack detection based on removing SQL query attribute values. Mathematical and Computer Modelling, 55(1 2), 58-68.
3. Jang, Y. S., & Choi, J. Y. (2014). Detecting SQL injection attacks using query result size. Computers & Security, 44, 104-118.
4. Sania: Syntactic and Semantic Analysis for Automated Testing against SQL Injection. Twenty-Third Annual Computer Security Applications Conference, 2007, ACSAC 2007. 107-117.