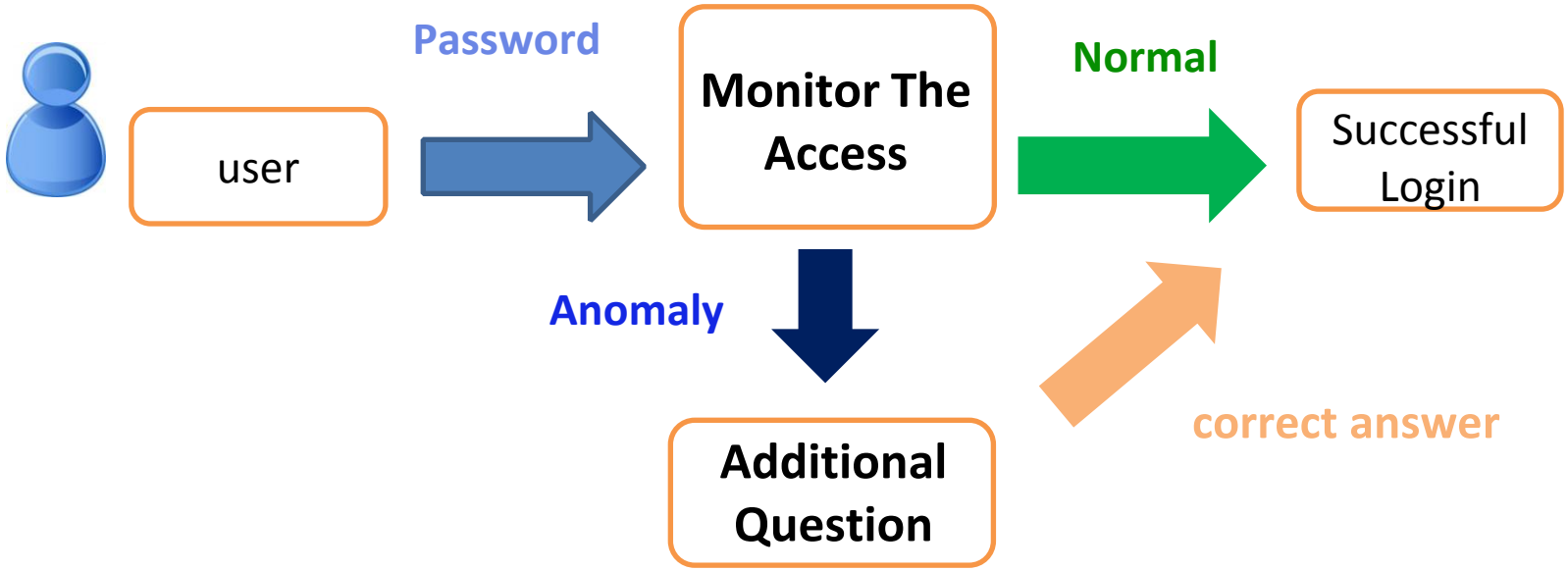
Automatic Generation of Authentication Questions from Private Messages

Ming Li, Keishi Tajima
Kyoto University

Background

- Many SNS or email services monitor the access patterns of users.
- Require the accessing user to answer an additional question when they detect some anomaly.



Existing Approaches to KBA (Knowledge-Based Authentication)

1. static KBA

e.g.:question registered by users

- Some users do not register questions

- Many users use the same question-answer pairs in many services

2. dynamic KBA

e.g.:name of a person in a photo posted in the closed group [Yardi et al.2008]

- Fake users can answer by searching for similar images on the web

Our Approach

- Our method automatically generates such authentication questions for an account or group by using the messages in that account or group.

- Our method shows one of massages with substituting one noun with a blank.

e.g.: **It's our kid's birthday! I baked a _____!**

We need to choose a word that is difficult to guess for fake users.

Two factors in the selection of the word to hide

1. For each candidate noun, we compute its co-occurrence degrees on the Web with other words in the same message.

   e.g.: **It's our kid's birthday! I baked a cookie!**

   C(cookie, kid) = 0.2
   C(cookie, birthday) = 0.3

   It approximates the proability that the fake users infer the word "cookie" from other words in the message.

Two factors in the selection of the word to hide

2. Our system collects coordinate terms (co-hyponyms) of each candidate noun, and calculate the same co-occurrence degree of them.

e.g.: **It's our kid's birthday! I baked a cookie!**

coordinate terms of cookie = {cake, pie}

C(cookie, kid) = 0.2
C(cookie, birthday) = 0.3

$\leq$

C(cake, kid) = 0.3
C(cake, birthday) = 0.5

We choose the candidate that has many coordinate terms with higher co-occurrence degrees.

Two ways to approximate the probability that fake users infer the correct answer

$$C(t_1|t_2) = \frac{|D(t_1) \cap D(t_2)|}{|D(t_2)|}$$

co-occurrence degree

$$P_{all}(t_q|m) = 1 - \prod_{t \in words(m), t \neq t_q} (1 - C(t_q|t))$$

It approximates the probability that fake users infer the correct answer from at least one word in the message.

$$P_{max}(t_q|m) = \max_{t \in words(m), t \neq t_q} C(t_q|t)$$

It approximates the probability that fake users infer the correct answer from the word that seems to be the most useful clue.

Method that uses coordinate terms

We compute $P(t|m)$ for each coordinate noun *t,* and compute the score of *tq*, denoted by *S(tq)*, by the formula below

$$S(t_q) = \sum_{t \in CO(t_q), P(t|m) > P(t_q|m)} P(t|m)$$

- That is, we sum up $P(t|m)$ of coordinated terms that seem to fake users more likely to be the answer than t_q
- We select a noun t_q with the highest score $S(t_q)$
- To compute $P(t|m)$, we use either $P_{all}(t|m)$ or $P_{max}(t|m)$

Experiment

4 methods we tested in our experiments

	Do not use coordinate terms	Use coordinate terms
P_all	method 1	method 3
P_max	method 2	method 4

Experiment

- Dataset

  — 80 messages from a closed group of Facebook

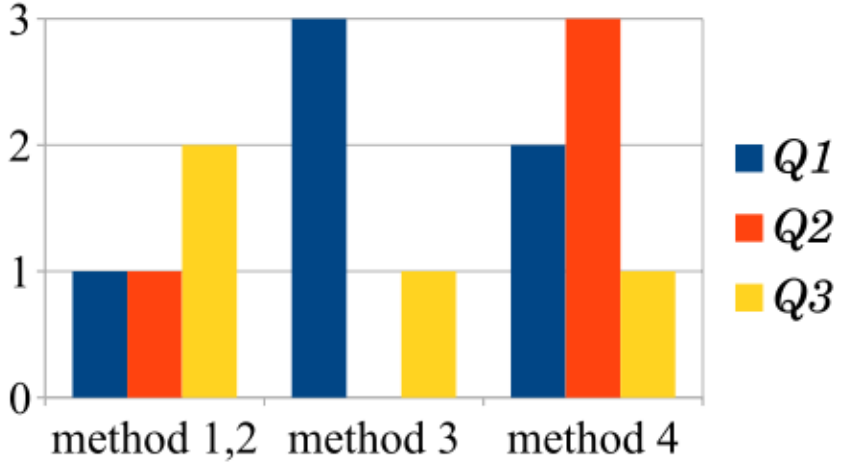
- Two groups of users

  1. Those who have read the messages in the group.
  2. Those who have not read the messages.

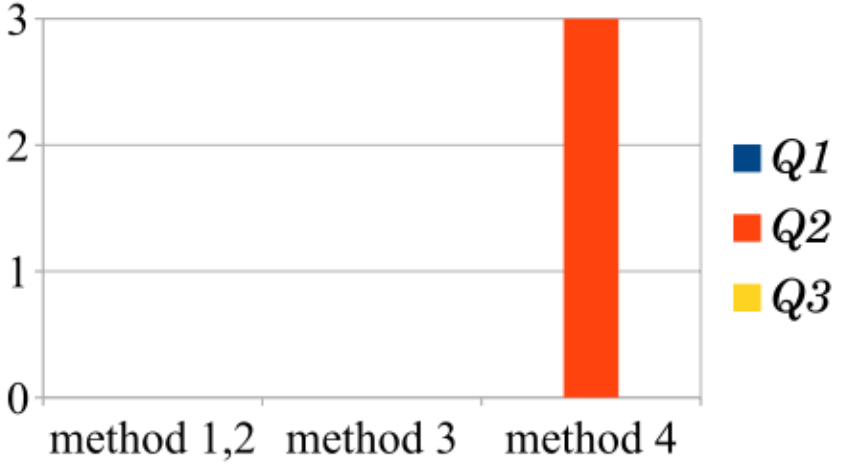
- Evaluation Method

  — We compare the ratio of correct answers in the group 1 and the group 2.

Experiment Result



Users who read messages



Users who did not read messages

The Number of users who gave the correct answer

Conclusion

- We proposed four methods that generate authentication questions by removing one noun in a private message, and compared them by a preliminary experiment.
- We are planning experiments of these methods with a bigger data set.