Threshold Estimation in Self-Destructing Scheme Using Regression Analysis

19th Asia-Pacific Network Operations and Management Sympsium

APNOMS 2017

Networking Lab, Department of Computer Science and Engineering, Kyung Hee University

Presented by : Young Ki Kim
Date : September 28, 2017

KYUNG HEE UNIVERSITY

NETWORKING LAB

Outline

- Introduction
- Related Work
- Proposed Scheme
- Scenario
- Experiment Result
- Conclusion

Introduction

- As users increasingly use and store personal information in cloud storage, research on privacy protection models is becoming more important.
- A Self-Destructing Scheme has been proposed to prevent the decryption of encrypted user data after a certain period of time using a DHT network.
- However, the existing privacy protection model does not mention the method of setting the threshold value considering the availability and security of the data.
- Therefore, in this paper, we propose an optimal threshold finding method considering both data availability and security of privacy protection model by applying regression analysis.

Related Works

- **Vanish System**

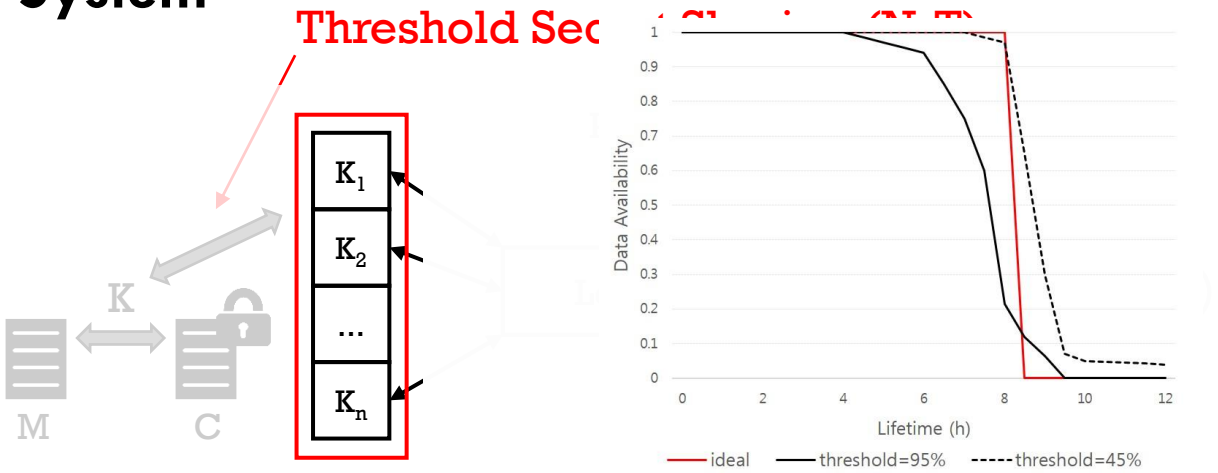


Fig. 1. Vanish: Self-Destructing Scheme

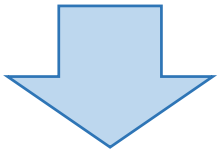
$$\text{Threshold Ratio} = \frac{Num.of\ subkeys\ to\ create\ the\ key\ K}{Total\ num.of\ subkeys}$$

How can we determine the Optimal Threshold Ratio?

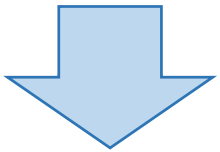
Roxana Geambasu, Tadayoshi Kohno, Amit A. Levy and Henry M.Levy, "Vanish: Increasing data privacy with self-destructing data," USENIX Security Symposium, June 2009, pp.299-316.

Proposed Scheme

Optimal Threshold Ratio prediction considering User-specified Time, Key Size, Availability and Security



Inference of results based on the relationship between Independent Variables



Suggest a solution using Regression Analysis

Proposed Scheme

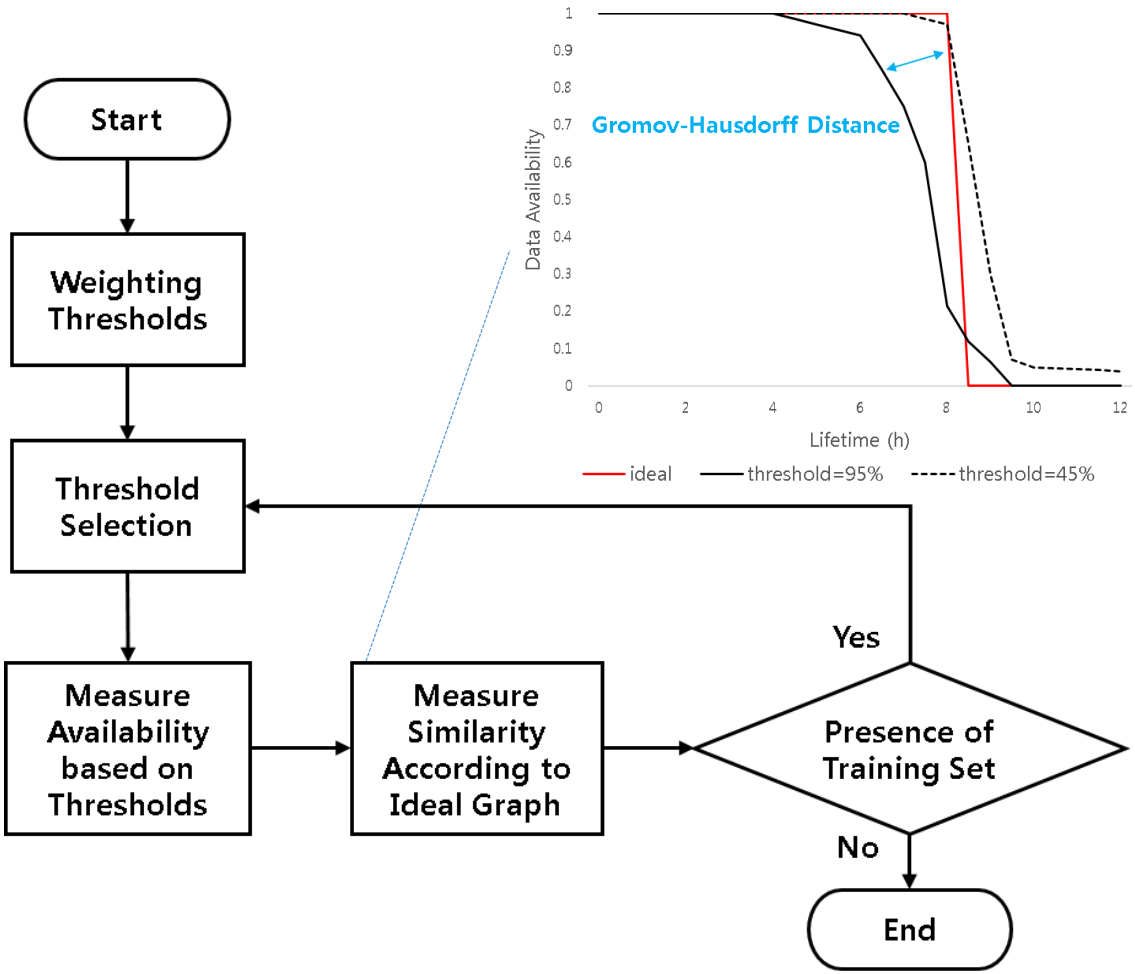


Fig. 3. Flowchart of Process for Regression Analysis

Proposed Scheme

- We classify prediction process into training and testing to improve the accuracy. (Training Set : Testing Set = 7 : 3)

Algorithm 1 Training Phase

```
1: if there is training set then
2:   extract data information from training set
3:   N = total number of key shares
4:   T = threshold ratio
5:   set current time
6:   measure data availability graph
7:   if the user-specified time is expires then
8:     F = similarity of availability graph with ideal
9:     save the result of training set
10:  else
11:    wait until the user-specified time is expires
12:  end if
13: else
14:  calculate nonlinear regression equation
15: end if
```

<Training Phase>

Algorithm 2 Testing Phase

```
1: if there is training set then
2:   extract data information from testing set
3:   N = total number of key shares
4:   T = threshold ratio
5:   NE = closet N from training set
6:   TE = estimated threshold ratio
7:   set current time
8:   measure data availability graph
9:   if the user-specified time is expires then
10:    F = similarity of availability graph with ideal
11:    FE = similarity of estimated availability graph
12:    save the result of testing set
13:  else
14:    wait until the user-specified time is expires
15:  end if
16: else
17:  update nonlinear regression equation
18: end if
```

<Testing Phase>

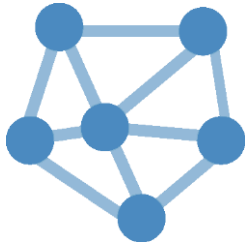
Scenario

Total Num. of Key Shares
&
Threshold Ratio

Graph Similarity Analysis

Regression Analysis Modeling

Distribute Key Shares



User-specified Time

Experiment Result

- Simple command line tool using Levenberg-Marquardt algorithm for regression modeling (https://github.com/claudejrogers/curvefit)

Table 1: Experiment Environment

DHT Network Type	**Vuze(Azureus)**
Num. of Training Set	**1000**
Num. of Key Shares for each Training Set	**100**

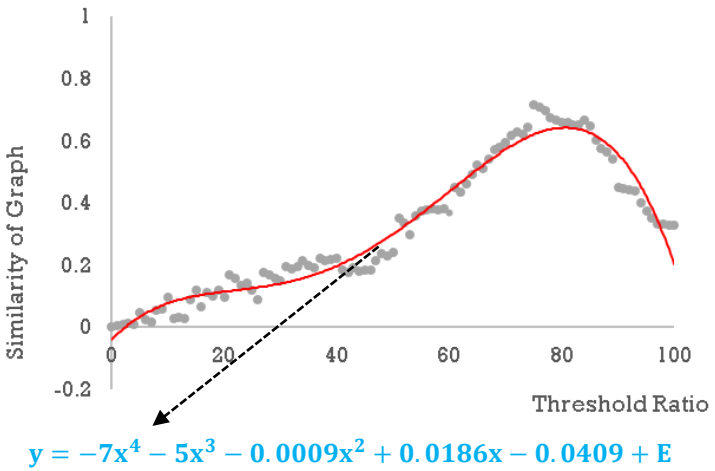


Fig. 4. Regression Analysis According to Similarity Graph

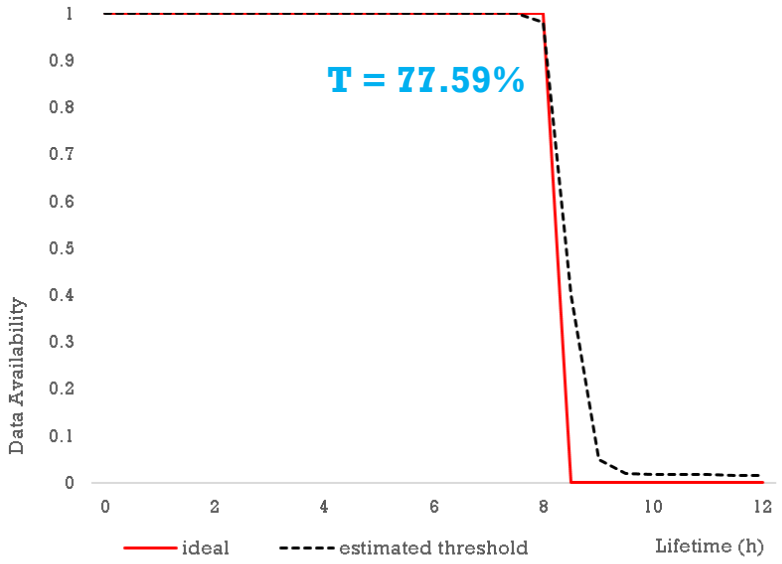


Fig. 5. Availability Graph According to Estimated Threshold

Conclusion

- In this paper, we propose a method to find optimal threshold in Self-Destructing Scheme considering both availability and security of data by applying regression analysis.
- In addition, the availability of the data with the estimated threshold value by the regression analysis and the result are sufficiently reliable.
- However, there is a limitation that the threshold estimation method proposed in this paper takes much time in the predicting process.
- Therefore, in the future, we will study methods to minimize the time required for the predicting process by using approaches such as clustering or parallel computation.

References

- [1] Mark D. Ryan, "Cloud computing privacy concerns on out doorstep," Communications of the ACM, vol. 54, no. 1, January 2011, pp. 36-38.
- [2] A. Shamir, "How to share a secret," Communications Magazine, ACM, vol. 22, no. 11, November 1979, pp. 612-613.
- [3] Roxana Geambasu, Tadayoshi Kohno, Amit A. Levy and Henry M.Levy, "Vanish: Increasing data privacy with self-destructing data," USENIX Security Symposium, June 2009, pp.299-316.
- [4] Guojun Wang, Fengshun Yue, Qin Liu, "A secure self-destructing scheme for electronic data," Journal of Computer and System Sciences, vol. 79, no. 2, March 2013, pp.279-290.
- [5] Gene Golub, Victor Pereyra, "Seperable nonlinear least squares: the variable projection method and its applications," Inverse Problems, vol. 19, no. 2, February 2003, R1.
- [6] S. Rhea, D. Geels, T. Roscoe and J. Kubiatowicz, "Handling churn in a DHT," USENIX Annual Technical Conference, vol. 6, December 2004, pp.127-140.
- [7] Alexander M. Bronstein, Michael M. Bronstein, Ron Kimmel, Mona Mahmoudi and Guillermo Sapiro, "A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching," International Journal of Computer Vision, vol. 89, no. 2, September 2010, pp.266-286.

Thank You !!!

Q & A