STORIFY : Protecting Privacy of Social Media Post with Hash-based Anonymity

JACKIE CHIN YONG AN

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITY MALAYSIA SABAH
2022

UMS
UNIVERSITI MALAYSIA SABAH

STORIFY : Protecting Privacy of Social Media Post with Hash-based Anonymity

JACKIE CHIN YONG AN

THESIS SUBMITTED IN PARTIAL FULFILLMENT
FOR THE DEGREE OF BACHELOR OF COMPUTER
SCIENCE WITH HONOURS
(NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITY MALAYSIA SABAH 2022

NAME	**:** JACKIE CHIN YONG AN
MATRIC NUMBER	**:** BI18110258
TITLE	**:** STORIFY: Protecting Privacy of Social Media Post with Hash-Bashed Anonymity
DEGREE	**:** BACHELOR OF COMPUTER WITH HONOURS (NETWORK ENGINEERING)
VIVA'S DATE	**:** 25th January 2022

CERTIFIED BY;

1. **SUPERVISOR** Signature

Dr. Tan Soo Fun

DR TAN SOO FUN
SENIOR LECTURER
Faculty of Computing and Informatics
Universiti Malaysia Sabah

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries and references, which have been duly acknowledged.

25th JANUARY 2022

JACKIE CHIN YONG AN
BI18110258

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor Dr Tan Soo Fun for the continuous support of my final year project entitled Storify to protect privacy of social media post with hash-based anonymity and for her guidance along the way. Her guidance has helped me in all the time of research and writing of this thesis. My completion of this project would not be accomplished without the support of my supervisors, friends and family. Last but not least, I would also like to thank my parents for their caring and support throughout this final year project.

JACKIE CHIN YONG AN
25th January 2022

ABSTRACT

In year 2020, the number of users using social networking has exceeded 3.6 billions out of 7.6 billions people, which is 47% of the populations. There are some famous social media applications that are widely used worldwide, which are Facebook, Twitter, Snapchat and many more. However, social media has some negative effect which has been labelled as a "likely culprit" that leads youngsters to depression, social isolation and have suicidal thoughts. Lack of anonymous features of recent social media apps that allow users to express their feelings without getting hurt by cyber bullying has urged recently to protect user's privacy. This project proposes a social media application, called as "Storify", that aimed to protect privacy of social media post with hash-based Anonymity. The objectives of this project are to investigate a lightweight hash-based anonymity algorithm in protecting the privacy of social media post, in terms of computation speed and ciphertext size, to design and develop the proposed Storify mobile social media application by using prototype development approach and to evaluate the performance of the developed Storify mobile social media application by using System Usability Scale (SUS) approach. Questionnaire and quantitative survey is used to gather user requirements. The hash-based algorithm is conducted through testing. The outcome of this proposed project is to confirm the lightweight hash-based anonymity algorithm with the fastest computation speed and greater ciphertext size.

ABSTRAK

STORIFY: MELINDUNGI PRIVASI SIARAN MEDIA SOSIAL DENGAN ANONIMITI BERASASKAN HASH

Pada tahun 2020, bilangan pengguna yang menggunakan rangkaian sosial telah melebihi 3.6 bilion daripada 7.6 bilion orang, iaitu 47% daripada populasi. Terdapat beberapa aplikasi media sosial terkenal yang digunakan secara meluas di seluruh dunia iaitu Facebook, Twitter, Snapchat dan banyak lagi. Walau bagaimanapun, media sosial mempunyai beberapa kesan negatif yang telah dilabelkan sebagai "kemungkinan penyebab" yang membawa anak muda kepada kemurungan, pengasingan sosial dan mempunyai pemikiran untuk membunuh diri. Kekurangan ciri tanpa nama aplikasi media sosial terkini yang membolehkan pengguna meluahkan perasaan mereka tanpa terluka oleh buli siber telah menggesa baru-baru ini untuk melindungi privasi pengguna. Projek ini mencadangkan aplikasi media sosial, dipanggil "Storify", yang bertujuan untuk melindungi privasi siaran media sosial dengan Anonymity berasaskan cincang. Objektif projek ini adalah untuk menyiasat algoritma anonimiti berasaskan hash yang ringan dalam melindungi privasi siaran media sosial, dari segi kelajuan pengiraan dan saiz teks sifir, untuk mereka bentuk dan membangunkan aplikasi media sosial mudah alih Storify yang dicadangkan dengan menggunakan pendekatan pembangunan prototaip dan untuk menilai prestasi aplikasi media sosial mudah alih Storify yang dibangunkan dengan menggunakan pendekatan Skala Kebolehgunaan Sistem (SUS). Soal selidik dan tinjauan kuantitatif digunakan untuk mengumpul keperluan pengguna. Algoritma berasaskan hash dijalankan melalui ujian. Hasil daripada projek yang dicadangkan ini adalah untuk mengesahkan algoritma anonimiti berasaskan hash ringan dengan kelajuan pengiraan terpantas dan saiz teks sifir yang lebih besar.

TABLE OF CONTENTS

TITLE	Page
DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	iv
ABSTRAK	v
TABLE OF CONTENTS	vi
LIST OF TABLES	xiii
LIST OF FIGURES	ix
LIST OF APPENDICES	xiv
CHAPTER 1: INTRODUCTION	1
1.1 Introduction	1
1.2 Problem Background	1
1.3 Problem Statement	2
1.4 Project Objective	3
1.5 Project Scope	3
1.6 Organization of the Report	4
1.7 Conclusion	5
CHAPTER 2: LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Review on Anonymous Communication Protocols	6
2.2.1 Onion Routing	7
2.2.2 Freenet	8
2.2.3 Hash-based Anonymity	9
2.3 Hash-based Anonymity Algorithms and Related Work	11
2.3.1 QUARK	11
2.3.2 PHOTON	13
2.3.3 SPONGENT	15
2.3.4 GLUON	16
2.3.5 SipHash	17
2.3.6 Comparison of Lightweight Hash-based Anonymity Algorithms	17

2.4 Review on Existing Systems 18

CHAPTER 3: METHODOLOGY 34

3.1 Introduction 34

3.2 Methodology 34

3.3 Software and Hardware Requirement 38

3.4 Conclusion 38

CHAPTER 4: SYSTEM DESIGN AND ANALYSIS 39

4.1 Introduction 39

4.2 Data Flow Diagram 39

4.2.1 DFD Level 0 39

4.2.2 DFD Level 1 40

4.3 ERD Diagram 46

4.4 User Interface Design 49

4.5 User Requirement Analysis 51

CHAPTER 5: EXPERIMENT IMPLEMENTATION 61

5.1 Introduction 61

5.2 Experimental Setup and Testing 61

5.3 Implementation 63

5.3.1 Implementation of SipHash 63

5.3.2 Implementation of Photon 64

5.3.3 Implementation of Spongent 64

5.3.4 System Interface for User Module 65

5.3.5 System Interface for Admin Module 71

5.3.6 Implementation of Database 73

CHAPTER 6: EXPERIMENT RESULT AND DISCUSSION 75

6.1 Introduction 75

6.2 Investigation on Computation Speed of SipHash, Photon and Spongent 75

6.3 Ciphertext Size 76

CHAPTER 7: TESTING AND EVALUATION 77

7.1 Introduction 77

7.2 Unit Testing for Storify 77

7.3 Integration Testing for Storify 83

7.4 System Testing for Storify 85
7.5 System Usability Scale 87
7.6 Conclusion 91
CHAPTER 8: CONCLUSION 92
8.1 Introduction 92
8.2 Unit Testing 92
8.3 Limitation of the Project 93
8.4 Future Work 93
REFERENCES 94
APPENDICES 97

LIST OF FIGURES

CONTENT

	Page
Figure 2.1: Onion Routing Circuit	7
Figure 2.2: The Sponge Construction	13
Figure 2.3: The Extended Sponge Framework	14
Figure 2.4: f function of the GLUON family	16
Figure 2.5: The calculation of f function	16
Figure 2.6 & 2.7: Screenshot of Yik Yak	19
Figure 2.8: Login and Signup of Snapchat	21
Figure 2.9: Storage Access Permission	21
Figure 2.10: Location Request	22
Figure 2.11: Stories	22
Figure 2.12: User Profile	23
Figure 2.13: Stories Management	23
Figure 2.14: Discover Session	24
Figure 2.15: Chat Session	24
Figure 2.16: Chatting Interface	25
Figure 2.17: MD5 utilized in TikTok	26
Figure 2.18: Terms and Policy of Tik Tok	27
Figure 2.19: Field of Interests	27
Figure 2.20: Language	28
Figure 2.21: Username	28
Figure 2.22: User Profile	29

Figure 2.23: Settings 29

Figure 2.24: Home 30

Figure 2.25: Inbox 30

Figure 2.26: Discover 31

Figure 2.27: Camera 31

Figure 3.1 & 3.2: Flowchart of Methodology 37

Figure 4.1: Storify DFD Level 0 39

Figure 4.2: Registration of Storify 40

Figure 4.3: Login of Storify 41

Figure 4.4: User Profile Modification in Storify 42

Figure 4.5: User Identity Hash Using Hashing Algorithm 43

Figure 4.6: Social Media Posts 44

Figure 4.7: Chatting Session 45

Figure 4.8: ERD Diagram of Storify 46

Figure 4.9: Login Page 49

Figure 4.10: Profile Page 49

Figure 4.11: Join Session Page 50

Figure 4.12: Homepage 50

Figure 4.13 & 4.14: Usage of Social Media Application 51

Figure 4.15: Information Provided During Registration of Existing Application 52

Figure 4.16: Social Media With Anonymous Posting 53

Figure 4.17: Importance of Securing User Profile 53

Figure 4.18: Full Access To Respondents' Devices 54

Figure 4.19: Identity Theft in Social Media Applications 55

Figure 4.20: Vigilance Across All Users' Private Information 55

Figure 4.21: Anonymous Social Media Application in Helping Students in Campus 56

Figure 4.22: Privacy of Post Shared 57

Figure 4.23: Language Preferred 57

Figure 4.24: Frequency of Login 58

Figure 4.25: Remove Inappropriate Posts 59

Figure 4.26: Posting History 59

Figure 4.27: Information to Be Kept As Anonymous 60

Figure 5.1: Hash Implementation Flowchart 61

Figure 5.2: Implementation of SipHash 63

Figure 5.3: Implementation of Photon 64

Figure 5.4: Implementation of Spongent 64

Figure 5.5: User Login 65

Figure 5.6: User Registration 65

Figure 5.7: User Profile 66

Figure 5.8: User Profile Modification 66

Figure 5.9: OTP Verification 67

Figure 5.10: Password Modification 67

Figure 5.11: Main Page 68

Figure 5.12: Forum Detail 68

Figure 5.13: Add Forum 69

Figure 5.14: Chatting Room 69

Figure 5.15: Notification 70

Figure 5.16: Admin Login 71

Figure 5.17: Admin Main Page 71

Figure 5.18: Forum Validation 72

Figure 5.19: Authentication 73

Figure 5.20: Firestore Database 73

Figure 5.21: Realtime Database 74

Figure 5.22: Storage 74

Figure 6.1: Computation Speed of Three Selected Hash Algorithm 75

Figure 7.1: System Usability Scale for Functionality 87

Figure 7.2: System Usability Scale for GUI 88

Figure 7.3: System Usability Scale for User-friendly Interface 88

Figure 7.4: System Usability Scale for Operational Performance 89

Figure 7.5: System Usability Scale for The Concept of Storify 89

Figure 7.6: System Usability Scale for Anonymity Requirement Satisfaction 90

Figure 7.7: System Usability Scale for Bug Issue in Storify 90

Figure 7.8: Feedback for Storify 91

LIST OF TABLES

CONTENT

Page

Table 1.1: Modules of Storify 4

Table 2.1: Anonymity Properties of Freenet 9

Table 2.2: Comparison of Anonymous Communication Protocols 10

Table 2.3: Comparison of Types of QUARK 12

Table 2.4: Software Performance in cycles per bytes of the PHOTON variants for long messages 14

Table 2.5: Variants of SPONGENT 15

Table 2.6: Feature of SipHash 17

Table 2.7: Comparison of Lightweight Hash-Based Anonymity Algorithm 17

Table 2.8: Comparison of Existing Applications and The Proposed Application 32

Table 4.1: Data Dictionary of ERD Diagram 47

Table 5.1: Dataset for Hashing 62

Table 7.1: Unit Testing for User Registration and Login Module 77

Table 7.2: Unit Testing of User Profile Modification Module 78

Table 7.3: Unit Testing of Forum and Chatting Module 80

Table 7.4: Unit Testing of Notification Module 81

Table 7.5: Unit Testing of Admin Module 82

Table 7.6: Integration Testing for Login and Registration Module for User Side 83

Table 7.7: Integration Testing between Forum Posting and Chatting Module for User Side 84

Table 7.8: Integration Testing for Login and Registration Module for User Side 85

LIST OF APPENDICES

CONTENT

	Page
APPENDIX A: FYP Proposal Feedback	97
APPENDIX B: Questionnaire	99
APPENDIX C: FYP Progress Feedback	100
APPENDIX D: Experiment Result	101

CHAPTER 1

INTRODUCTION

1.1 Introduction

Social media applications have been growing at a tremendous rate, and the adoption rate has been skyrocketing, which leads to a huge increase in users in less than ten years (Al-Deen & Hendricks, 2011). However, there are limitations on privacy protection in existing social media applications. For instance, existing social media platforms such as Facebook do not allow users to hash their information and post something anonymously. The system that to be proposed in this proposal will be a social media platform named Storify that will be able to protect users' private information and identities using lightweight hash-based anonymity. Other than that, users are allowed to share posts anonymously without the fear of privacy leakage. The reason for choosing this title for the proposal is that cybercrime and cyberbullying are getting out of control on social media platforms. Because of the essence of social media, data sharing and trust is an unavoidable aspect of the applications. These two factors come together to form a perfect cyberstorm (McGuire, 2019).

1.2 Problem Background

Most of the social media platforms that can be found online nowadays are quite famous, and some are even a daily basis for people to keep up with updates about the world. Almost every application supports 24-hour-only stories, which will disappear after 24 hours after posting them; as for Snapchat, not only stories that will get disappeared automatically, so as the messages, they will be gone after 24 hours (Kahn & Martinez, 2020). However, most of the content shared is traceable and recordable in some ways; therefore, privacy is not protected well (Burns, 2019). There is also another example of anonymous posting, which is UMS confession. This platform is based on a google form that anyone can submit, and their identities will not be revealed. After that, administrators will get the contents that have been submitted and post them online. This platform can protect students' privacy; nonetheless, bad content such as bully, name-calling posting are not filtered out and will be just posted online by administrators. Also, this platform requires an administrator to keep posting; otherwise, the content that students submitted will not be posted online.

As it is too complicated and time-taking to share a post anonymously on an existing application, a lightweight hash-based anonymity algorithm will be used in this project that allows users to hash their identities first before proceeding to post something online. A hashing function is chosen other than other Anonymous Communication Protocols (ACP) because it is more appropriate and applicable in this project. Other than that, it is an irreversible action in which attackers are unable to decrypt it easily. Likewise, a lightweight hash-based anonymity algorithm will not consume too much memory of the device and is compatible with all devices.

Therefore, there is a lack of privacy by anonymizing identities and social media posts among all the applications today. Lightweight hash-based anonymity is lacking in any existing social media application; hence, users' identities are unable to be hashed and protected, which will lead to exposure of their identities when they post on social media and might cause unwanted events such as cyberbullying and identity theft.

1.3 Problem Statements

I) Lack of Anonymity Properties in Protecting Social Media Posts

There is a lack of anonymity properties in social media that has been widely used today. For example, for Facebook, there is no feature or function that has anonymity properties, and all of the postings are not being posted without revealing identities. However, inside Facebook, a page can be created, and admin can post forums that are submitted by others through google form without revealing their identities. For instance, the UMS confession Facebook page allows students to practice on anonymous posting, but this required the admin's integrity in keeping the poster's identity privately and admin to be active to update posting frequently. Therefore, in the proposed application, a lightweight hash-based anonymity algorithm will be used to encrypt users' identities. By hashing the identities of the users, cybercrime such as identity theft, fraud, and hacking will not occur in the platform, and they could join the anonymous session after that to continue browsing the newsfeed with hashed identity.

II) Lack of Lightweight Hashing Algorithm

There is a lack of a lightweight hashing algorithm in most of the existing social media applications. A lightweight hashing algorithm, which is designed to cope with the rapid expansion of advanced technologies, is employed to provide Authentication Encryption (AE), which guarantees the privacy and integrity of data. For example, in this project, AE is likely to be used to protect users' identity when they want to go anonymous online. Also, the lightweight hashing algorithm also uses smaller internal states, short blocks (64-bit), and key sizes, leading to smaller RAM consumption. Thus, it can be used to provide a security solution for resource-limited devices. This could greatly help to protect students in UMS when they are using this project without leaking their personal information to unwanted forces.

1.4 Project Objectives

I. To investigate a lightweight hash-based anonymity algorithm in protecting the privacy of social media posts in terms of computation speed and ciphertext size.

II. To design and develop the proposed Storify mobile social media application with a lightweight hash-based anonymity algorithm by using a prototype development approach.

III. To evaluate the usability performance of the developed Storify mobile social media application by using the System Usability Scale (SUS) approach.

1.5 Project Scope

The target users of this proposed application - Storify, are students from every faculty in Universiti Malaysia Sabah (UMS). The research embedded elements discussed are to protect user's identities to prevent unwanted events. The users's anonymity is scoped at MAC address layer and it will not concern on the network layer.

Table 1.1: Modules of Storify

Module	Description	Target Users/Roles
User Registration and Login	Users will be asked to register before login to the application	Students and System Admin
Personal Information Modification	Users are able to edit their profile and verify their information	Students
Undergo Identity Hash to Join Anonymous Session	Users are able to hash their identity to join the anonymous session	Students
Chatting Room	Users are able to chat with others without revealing their identities	Students
Forums	Users are able to share posts and comments on each other posts	Students
Validate Forums	Admin can remove or keep the reported forum	Admin

1.6 Organization of The Report

Chapter 1 describes the introduction of the project, problem background, problem statements, problem objectives, project scope with modules, and conclusion. For chapter 2, a literature review will be described In the literature review, Anonymous Communication Protocols will be reviewed and some of the existing social media applications. As for chapter 3, a methodology will be shown in this chapter. For chapter 4, the system analysis and design will be shown and also the questionnaire results. For chapter 5, experiment on the algorithms will be discussed in details in this chapter. For chapter 6, the result of the experiment will be shown and discussed. Last but not least, for chapter 7, a conclusion will be deducted in this chapter.

1.7 Conclusion

In a nutshell, the expected outcome of this proposed project is Storify social media application with a lightweight hash-based anonymity algorithm with the fastest computation speed and greater ciphertext size that allows campus students to share their problems and challenges anonymously.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this chapter, some anonymous communication protocols will be reviewed, and the most applicable protocol will be chosen to continue this project. After choosing the best protocol, five algorithms of the protocol will be studied, and three of them will be researched in more details afterward. In addition, three existing social media applications are reviewed to find out their advantages and limitations in order to create Storify in a better way.

2.2 Review On Anonymous Communication Protocols

Anonymous communication protocols are used to provide anonymity to online users on the internet to protect their identities. Anonymous communication protocols have been employed over the last decades due to the success of the Tor network – one of the anonymous communication protocols (Backes *et al.*, 2013). Solid privacy is one of the three important properties for anonymous communication protocols other than scalability and reliability. Powerful anonymity refers to a system's ability to defend users' identities from an attacker capable of limitless passive attacks while only risking a small percentage of active attacks (Goel *et al.*, 2003).

Some anonymous communication protocols are widely used, such as Onion Routing, Freenet, and hash-based anonymity.

2.2.1 Onion Routing

Onion Routing protocol, which is based upon the idea of mixes, is more robust than singleproxy methods for anonymous communication. A collection of proxies communicating over encrypted channels cooperate in Onion Routing to forward data to a receiver. In other words, messages are going through layers of encryption just like an onion with many layers to uncover the data’s destination. Data is wrapped in a series of encrypted layers that are peeled one by one at proxies along the way to the recipient. It is used to prevent others from seeking the network and servers that have been contacting (Onion Routing, 2018). Each onion router is presumed to be aware of the identities and public keys of the other onion routers. The sender begins by choosing a route through the other onion routers to the receiver. Each layer is encrypted with the public key of the corresponding router to avoid sending another onion, and each onion router pair uses a locally unique anonymous communication identifier (aci). The layers are stripped off one by one as the packet travels through the onion routers’ path until it reaches the last router in the path, where the data is forwarded directly to the receiver (Shields & Levine, 2000). However, Onion Routing is not employed at every host, and its anonymity can be broken down using timing analysis – trace and log connections between computers.

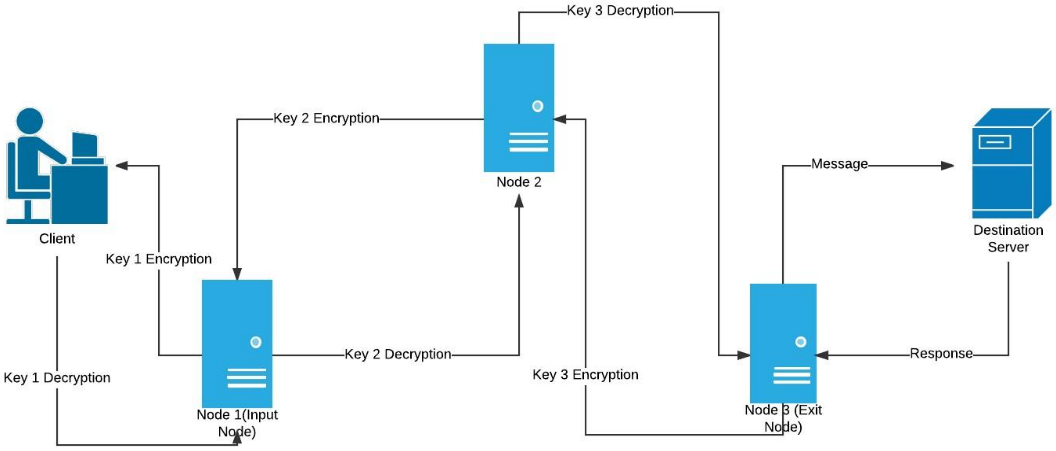

Figure 2.1: Onion Routing Circuit

Source: https://www.geeksforgeeks.org/onion-routing/

Figure 2.1 demonstrated the client with the access of Key 1, Key 2, and Key 3 encrypts the message thrice under three layers like an onion to be peeled one by one at a time. Then, the encrypted message is sent to Node 1, which is the only one that has an address of Node 2 and Key 1. Therefore, it decrypts the message using the given key and passing it to Node 2 since it still has two more layers. Node 2 has Key 2 and the address of Node 1 and Node 3; thus, it decrypts the message then sends it to Node 3 because it still has one more layer to be peeled.

After that, Node 3 decrypts the last layer, sends it to the destination server, and asks for a response. The response from the server then passes through the same nodes in a reverse direction to encrypt it with the same layer using their specific keys. Last but not least, the triple encrypted response is then sent to the client to get decrypted since the client has access to all three keys (Nigam, 2018). However, onion routing will not be used in this project as it requires another browser to make it anonymous, which is the Tor browser, and this project is an application that is compatible with all devices.

2.2.2 Freenet

Freenet is a platform for censorship-resistant communication, with a packet-oriented protocol, and it only implements self-contained messages. Each message includes a 64-bit transaction ID that is generated randomly, a hops-to-live limit, and a depth counter. The ID of the messages is not be determined as unique, but the chances of collision are also extremely low. The depth will be incremented at each hop and used by the replying node to set-hops-to-live high enough to reach the recipient. Hops-to-live is set on the originated message and will be decremented at each hop to prevent messages from being forwarded indefinitely. The aim of this design is to allow for flexibility in the choice of message transport mechanisms, whether TCP, UDP, or packet radio. A transport method is combined with a transport-specific identifier, such as an IP address and port number, to form a node address.

A Freenet transaction begins with a Request Handshake, in which a message is sent from one node to another, containing the desired return address of the receiving node. If the remote node is active and responding to requests, it will answer with a Reply Handshake, indicating which protocol version it supports. Handshakes are remembered for few hours. However, if the request is unsuccessful, the remote node with reply with a Reply.nNotFound and will be passed along upstream, unless