NEW BGP ROUTE LEAKS CLASSIFICATION AND DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUE

By

Salma Abdel Monem Abdel Motaleb Mohamed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Computer Engineering

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2019

NEW BGP ROUTE LEAKS CLASSIFICATION AND DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUE

By
Salma Abdel Monem Abdel Motaleb Mohamed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE in
Computer Engineering

Under the Supervision of

Prof. Dr. Samir I. Shaheen

Professor of Computer
Computer Engineering
Faculty of Engineering, Cairo University

Dr. Ahmed Bashandy

Associate Professor of Computer
Computer Engineering
Faculty of Engineering, Some University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2019

NEW BGP ROUTE LEAKS CLASSIFICATION AND DETECTION USING SUPERVISED MACHINE LEARNING TECHNIQUE

By
Salma Abdel Monem Abdel Motaleb Mohamed

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Computer Engineering

Approved by the Examining Committee:

Prof. Dr. Samir I. Shaheen	Thesis Main Advisor
Dr. Amr G. Wassal	Internal Examiner
Prof. Dr. Mohamed Z. Abd El Mageed	External Examiner

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2019

Engineer's Name:	Salma Abdel Monem Abdel Motaleb Mohamed	
Date of Birth:	08/03/1994	
Nationality	Egyptian	
E-mail:	salmacmpeg@cu.edu.eg	
Phone:	01114302362	
Address:	33 El-Zhoor st, El Motamiz District, 6 October city, Egypt.	
Registration Date:	01/10/2016	
Awarding Date:	29/7/2019	
Degree:	Master of Science	
Department:	Computer Engineering	
Supervisors:	Prof. Samir I. Shaheen Dr. Ahmed Bashandy	
Examiners:	Prof.Dr. Samir I. Shaheen	(Thesis Main Advisor)
	Dr. Amr G. Wassal	(Internal Examiner)
	Dr. Mohamed Zaki Abdel Mageed	(External Examiner)

Title of the thesis:

New BGP Route Leaks Classification and Detection Using Supervised Machine Learning Technique.

Key Words:

BGP; Route Leaks; Leaks Classification; Machine Learning.

Summary:

The route leaks problem is considered one of the unsolved problems in BGP through the previous 15 years. The confidentiality of autonomous systems (ASes) relationships and the lake of advertisement of route leaks incidents are the main two reasons behind this. This thesis solves the route leaks problem relaying on three steps: A new taxonomy for the route leaks types based on their effects to the BGP Traffic not to their ASes relationships is proposed, the first dataset for published real route leaks incidents through the previous years is collected and labeled as route leaks or normal traffic, and the first real-time detection system of route leaks problem using complex features extracted from BGP Update messages only and using Classification algorithms is proposed. The system achieves the best accuracy of 88% and 92% F1 Score, the whole system can detect route leaks upon receiving them in real-time as it runs in less than one second.

Disclaimer

I hereby declare that this thesis is my own original work and that no part of it has been submitted for a degree qualification at any other university or institute.

I further declare that I have appropriately acknowledged all sources used and have cited them in the references section.

Name: Date:

Signature:

Dedication

This thesis is dedicated to my family for their unconditional support, help, and encouragement without whom I would never be the person who am I.

Acknowledgements

I would like to thank Prof. Samir Shaheen and Dr. Ahmed Bashandy for their support, advice, and guidance throughout the thesis. I would like to thank Youssef Ghatas and Eman Hossam for their continuous encouragement, help and their valuable comments and ideas.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

AS	Autonomous System
AS-Path	Autonomous Systems Path
ASN	Autonomous System Number
BGP	Border Gateway Protocol
DAG	Direct Acyclic Graph
DT	Decision Tree classifier
FN	False Negative
FP	False Positive
GBoost	Gradient Boosting classifier
ISP	Internet Service Provider
KNN	K-Nearest Neighbor classifier
NB	Naïve Bayes classifier
RF	Random Forest classifier
SVM	Support Vector Machine classifier
TCP	Transmission Control Protocol
TN	True Negative
TP	True Positive

Abstract

The route leaks problem is considered one of the unsolved BGP problems since more than fifteen years ago. It has a large negative impact on the global internet stability and reliability. A route leak happens when an autonomous system advertises reachability information with a violation of autonomous inter-relationships policies. This problem is hard to prevent due to human errors and misconfigurations, and hard to detect due to the confidentiality of autonomous systems relationships and lake of publicly-advertised datasets.

Traditional solutions to the route leaks problems use one the following methods: the autonomous systems relationships inference, adding information about connection types to the BGP Update messages, or gathering and comparing information from global vantage points. These solutions were not widely applicable because they may be incomplete, have high time and processing cost, depend on third-parties, or requires modification to the existing protocols.

In this study we address the route leaks problem from three aspects, firstly we propose a new taxonomy to the different types of route leaks depending on their effects of the BGP update messages other than existing taxonomy that depends only on the relationships between autonomous systems which could not help much as these types of information are confidential.

Secondly, we gather and label the first route leaks dataset which consists of all the valid route leaks incidents published on the internet. We use this dataset to generate features list which we used as training and testing data to our classifiers. To accelerate the features generation process we use a memory structure called "Patricia Trie" inspired by the structures used in the router's lookup tables. This structure has reduced the average overall searching and updating run time from several hundreds of minutes to very few seconds.

Lastly, we propose a complete real-time widely applicable detection system to the route leaks updates using supervised learning method, and using the classification technique that can be applied to each border router without adding high time or computational overhead. In our study, we apply and compare between six different classifiers namely (Decision Tree, Random Forest Trees, Support Vector Machines, Naïve Bayes, Nearest Neighbor, and Gradient Boosting) classifiers. Our system proves that it can detect and classify route leaks from normal updates on two datasets the "less balanced" and the "more balanced". The system achieves the best accuracy of 88%, 83% and F1 Score of 92%, 98% for the two datasets respectively. These results can be achieved with a system that runs in less than one second.