UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

COS700
Research Proposal

An Active Mode Content Based Stylometric Detection Framework for The Mitigation of Phishing Attacks

Student number: 14163285

Supervisor(s):
Prof Hein S Venter
Dr. Adeyemi R Ikuesan

An Active Mode Content Based Stylometric Detection Framework for The Mitigation of Phishing Attacks

Abstract

Cloud Computing is used to deliver applications through the internet as a service to the consumers. Public cloud resources from Cloud Providers such as Googles services are open to use to the general public. Therefore the Public Cloud is susceptible to cyber crime attacks, particularly Phishing attacks. Spear Phishing attacks in the cloud, are targeted attacks whereby targeted end end-users release sensitive information to attacker, through the attackers use of impersonation of trusted organization. This attack has a higher success rate then the other phishing techniques. This paper will therefore describe a Stylometric Detection Framework, which is derived from an Anti-Spear Phishing Content-based Authorship Identification (ASCAI) framework, to detect and block mismatched emails from a received email body and the studied stylometric style of the original author who the account in the public cloud belongs to. We follow the same methadology as the ASCAI module, but also introduce the latest detection frameworks and block mismatched emails in attempt to increase the accuracy.

Keywords:

Stylometrics, Public Cloud, Phishing, Headers, Body, Detection

1 Introduction

Computing and technology continuously advance in the Information Technology industry to improve processes that assist in meeting business and consumer objectives and improving consumer satisfaction. However, innovated resources do tend to have higher upfront costs. Therefore, not many businesses or consumers can afford to purchase these resources. In addition, with the emergence of the internet in this generation, the expense of storage as well as the consumption of power by computing components has increased [DKCE+16]. Thus, research into the idea of utilizing resources from an innovative and powerful vacant computer remotely immerged. This concept was referred to as Cloud Computing.

Cloud Computing according to the National Institute of Standards and Technology (NIST) is a computing model that exists for the enablement of "ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [MG11]. In hindsight, cloud computing can be viewed as applications that are delivered through the internet as a service to the consumer as well as the reference to hardware and software of data centers that supply these services [ASZ+10]. The services are categorized into the following different models according to NIST [MG11]:

- Software as a service (SAAS), is applied through either a program interface or a web browser to use the supplier's applications that are running on the infrastructure of the cloud.
- Platform as a Service (PaaS), is applied where the consumer deploys a consumer-created application that is supported by the supplier, whereby the user has control of application configuration.
- Infrastructure as a Service (IaaS), is applied when the consumer is equipped with deployment and running of arbitrary software and can, therefore, control the deployed software and storage, in addition, can also have limit control to network components.

For the purpose of this paper, we will look at Software as a Service in closer detail. Cloud computing model is subdivided into 4 deployed infrastructures:

- Public Cloud, exists in the premises of a cloud provider which is open to use of the general public.
- Private Cloud, can exist in either the premises of an exclusive organization or off on a third party organization, which is used exclusively by a single organization or jointly with a third party organization. As long as it's not open to use by the public.
- Community Cloud, is utilized by organizations that share common interests, security or concerns. The environment exists either on the premise of either of the organizations or a third party organization rendering the cloud services to this community of organizations.

- Hybrid Cloud, is the combination of either of the previously stated deployed infrastructures distinctively.

The public cloud is of particular interest to the purpose of this paper, especially when looking into the commercial service platform that exists in the public cloud such as Google, Amazon, and Microsoft. The public cloud service infrastructure has rapidly grown, therefore has also brought forth an increase in unavoidable security issues. In essence, cloud security is a vital vulnerability that is of great concern, and thus even today we still find many organizations, that have not utilized the available cloud services. According to Gartner, a researcher that conducted an investigation around information security issues that are considered in cloud computing listed that privileged access is one of the issues highlighted that public cloud vendors need to consider [SH11]. For the end consumer, sensitive data should only be available to them, however, cloud vendor administrators should have access to a cloud environment for maintainability and security of the environment. Thus sensitive data may also need to be available to administrators.

1.1 Impersonation

To enable the availability to end user, Cloud computing service needs to provide an identity management. This is, however, a great concern in the public cloud environment as the cloud providers identity management are not consistently integrating their platforms with identity services [Bro5]. This is also due to the constant evolution of SaaS. This opens to a renowned cyber crime attack that is probable in exposing the vulnerability. This crime is known as Phishing. Phishing is executed by someone impersonating as a legitimate and reputable organisation or individual targeting any individual or organization to providing sensitive data [ASRS16]. Therefore, attackers performing Phishing attacks usually perform them through email, where they prompt target into giving sensitive information. Attackers also target public cloud infrastructure through PaaS where they inject links that users using Cloud platform as a SaaS would click the link that would redirect the user to a fraudulent web page.

Anti-phishing tools have been developed and utilized by public cloud providers in an attempt to prevent Phishing attacks. Tools like Envelope content splitting (ECS) tool is used to authenticate the attacker sending a link to a fraudulent website, and so protects the recipient from phishing [ASRS16]. There exist tools that also detect Phishing attacks, such as a Phishing tank whereby a database containing Phishing sites are used by other anti-phishing tools to detect phishing. The literature review will focus on some of the anti-phishing tools that were researched for the purpose of detecting Phishing, particularly in the Public Cloud. This will be followed by the identifying the gap in the detection of Phishing attacks that this research paper will address. Authorship detection is the proposed solution based on a Stylometric analysis approach that this paper will investigate.

1.2 Behavioural Biometrics

Behavioural biometrics is the study of users behaviour [SSP18]. The type of behavioural biometry that's of interest in this paper is stylometry as it's the study of users linguistic style in writing. Stylometry has been applied in different frameworks in Cloud platforms with the aid of detecting a phishing attack through attribute authorship, and hence this paper will look to exploiting this feature.

2 Problem Statement

Public cloud environments are susceptible to Spear phishing attacks as it is thought to be the easiest ways for criminals to acquire sensitive information from the end-user as the end-user is seen as the weak point in the system. It takes approximately over 229 days to detect an intrusion with the latest detection models employed in Cloud environments [VEE+15]. This is due to attackers knowledge of impersonating as a legitimate source by mimicking the style of writing format of the organisation.

3 Literature Study

A paper was done by Nazmul Islam, Mohammed Moshiul Hoque and Mohammad Rajib Hossain on the "Automatic Authorship Detection from Bengali Text using Stylometric Approach" [IHH17] investigates writing styles of Bengali writters by collecting writings and blogs from their sample writers. Through analysis, they discovered n-gram features that were useful to detect certain authors. N-grams, also reffered to as shingles, is an adjacent sequence of n items from a given sample of text [BGMZ97]. Thus can be viewed as a probabilistic model in a language base that predicts the next item in a sequence derived from the Markov model [Bri95]. The investigation used unigram, bigram as well as trigram together with parts of speech features such as conjunctions and pronouns on the sample sets of texts provided. Then the use of three machine learning algorithms on the final dataset, namely Naive Bayes, Decision Tree and Random Forest Classifier, were used, with the explanation of how each step in the Random Forest was conducted and the results that got passed into a new document. This is due to the 96% accuracy they received from their research in the Random Forest Classifier. The strength of this research was dataset was reliable was they had selected texts randomly to minimize biasness of the data. It also reached a very high accuracy of 96% from the 3125 literary passages of a sample of 10 prominent writers. However, the approach has a limitation in that was narrowly focussed on blog writings and so proposed method has still need to address writings found in news articles, emails, tweets and other texts found in public cloud environments that are of interests to this research paper.

Rakesh Verma and Nirmala Rai proposed the "Phish-IDetector: Message-Id Based Automatic Phishing Detection" [VR15] which primarily focuses on email headers. The research focussed on observing less than 10 legitimate emails and phishing emails as the research were more drawn to the Message-ID field that is

a universally unique string. The research utilises Machine Learning algorithms together with the properties of Message-IDs onto the n-gram analysis of the Message-IDs. The Random Forest Classifier algorithm performed the best in the research and so results found were utilised with the SMO algorithm (The Sequential Minimal Optimization). The researchers ensured that 100% of the experimental dataset legitimate emails had Message-IDs. This experiment relies on the Message-ID field and so if not existing the experiment will not work, fortunately, it will still raise red flags to the email security. The research reached 99% True Positive Rates. It is important to note that in literature, there is no phishing detection of 100%, therefore the result of 99% was due to the smaller finite sample set of data, and is indeed noted that the exponential increasing of the sample email set together with a higher order of n-grams is difficult to run with different Machine Learning classifiers without the specialized use of big data approaches. This, therefore, shows a limit in the automatic detection system.

Following from this research, Rakesh Verma and Ayman El Aassal state its better to include users in the detection process or send warning to users to warn them of possible attack and thus include user training, when they were introducing "A Correlation-based Analysis and User Participation method for Detecting Phishing Email" [VEE+15]. In this research, a comprehensive method to determine to phish off an email was designed, whereby information extracted from the email header is relevant to the information contained in the email body. This introduced method was executed using two algorithms, namely Header Analysis Algorithm and Matching Algorithm. Within the Matching Algorithm, multiple header fields and sender identity are authenticated using digital signatures in the Domain-Key Identified Mail [Her09]. A sender Policy Framework is imployed to enable verification that the sending mail server is authorized in the domain that appears in the "mail from" address. Other methods like URL Analysis and Semantic Analysis are also used in the research. The research primarily focuses on the header analysis and so if the results pass the rules set for the header, then the email is passed off as legitimate. This deduction, however, is not legitimate as the attacker can still get hold of a legitimate email address by hacking into the organisation's mail server and impersonate as a member of the organisation. Hence future work involves the analysis of email body.

A paper on a Content-Based Authorship Identification Framework that is used for the detection of spear phishing [KIJ11]. The paper introduces a novel framework called the Anti-Spear phishing Content-based Authorship Identification which analyses the message body of the message sender without relying on the sender's ID. Its important to note that this is used specifically for Spear Phishing attacks. Bulk Phishing attacks are generic and target many users, and so many cloud providers contain several software classifiers that have the ability to detect generic nature of bulk phishing attacks. Spear Phishing attacks are targeted and thus difficult to detect to detect due to unique nature. The paper highlights a key point in its motivation into using Content based, that is User ID-based authentication is not helpful in detection as Users read the text in parallel and not sequentially, thus can fall into typo-squatting and cousin-naming. Users also introduce weak authentication in systems, particularly in cloud infrastructures whereby they set very weak passwords as its easier to remember

for them but also easy for a brute force algorithm to hack. In addition, users passwords can be stolen through Keyloggers, and so attackers can impersonate as the user in the system if they have the users credentials. The research aims to provide a software framework, that was otherwise not introduced in the previously mentioned literature. The research also takes a whitelist approach based off of email mining techniques constructed from emails sender stylometric profiles. Thus a unique approach compared to other literature that was a derivation of the black-list approach. 2 proposed modes to calculate similarities between the claimed and the predicted identities are namely the passive and active mode. In the passive mode, both identities are presented to the end-user. In the active mode, in a scenario where a mismatch between identities is found, the content of the message is blocked to the end-user with a warning of a mismatch occurred. The issue with this research is false positives can result if multiple users contribute to an email which the writeprints can be altered from original senders writeprint. Active mode was not explored in this paper as it increases implementation complexity and Software might not have been as accurate as the end-user. The paper followed the Security Content Automation Protocol (SCAP) methodology [MMM+11] due to its high achievement of 100% of classification accuracy on datasets as it makes use of byte-level n-grams and hence useful for natural languages. For this research, a dataset of 289 emails from 12 authors was used. An Accuracy rate was used to measure performance as dataset was evaluated with the use of 10-fold cross-validation. The results found in the research show a maximum accuracy of 83% for a non-greedy n-gram ranking method and 87% maximum accuracy rate of 87% with a focus on n-gram ranking methods. This framework had setbacks and limitations such as the writeprint extraction whereby if no message was previously sent by the original author then the SCAP would attempt to map it to the closest author profile match, which is not desired.

3.1 Aim of this study

Based of the ASCAI framework that the research paper conducted [MMM+11], this papers objective is to implement the ASCAI framework in active mode to develop a white-list of authorized authors stylometric identities. Furthermore, this white-list will be used during the authentication of the daily emails to detect identity impersonation. This is to compare which mode provides the best results.

4 Methodology

Similarly to the ASCAI framework methodology introduced, this paper will follow a similar route whereby:

1. Utilizing known techniques to implement the Identity Extraction module, which will extract the claimed identities from the header field of the message. This may involve NLP (Natural Language Processing) techniques and other metaphone algorithms, which are algorithms for words indexed by their English pronunciation [SSP18] in the aim to detect similarity between words.

2. Using Machine Learning Algorithms to construct stylometric profiles of actual author identities.

3. Construct a mechanism that will construct a stylometric profile of the sending authors identities.

4. A stylometric comparison algorithm mechanism that will compare similarities between claimed identities of the sending author and the predicted identities that was produced from 3.

5. Construct a User Interface that can be used to block the content of the message to the end-user if the comparison algorithm finds a mismatch. This will also facilitate an option for end-user to view content even if the user has been warned.

6. Unit tests will be constructed to test for different algorithms.

The detection algorithm named SEAHound will be explored in addition to the prescribed algorithms found in the ASCAI framework to test and see if we can get a higher accuracy [PHS18].

5 Planning

This is the sprint set up for milestones this research paper will follow. These dates are subject to change. The planning gives a broad clear idea of what will be undertaken and when during the duration of the research.

Project:	A BEHAVIOURAL BIOMETRICS BASED ON STYLOMETRICS IN CLOUD COMPUTING DOMAIN FOR CYBERCRIME DETECTION											
Project Planner												
A legend describing the charting follows:					Done		In Progress		Not Started		Planned Duration	
ID	Current Status	Task description	Duration	Start Date	Estimated Finished Date	Finish Date	Predecessor Tasks	Resource Names	Period 1	Period 2	Period 3	Period 4
									1	2	3	4
1000		**Stylometric Research Paper**		2018/04/01	2018/10/31							
1001		**Research Proposals**		2018/04/01	2018/04/30	2018/05/01						
1002		Gather sources for literature review	20 Days	2018/04/01	2018/04/20	2018/04/21						
1003		Build Research timeline	2 Days	2018/04/13	2018/04/15	2018/04/15						
1004		Identifying research area	1 Day	2018/04/20	2018/04/21	2018/04/25						
1005		Crafting the problem statement	2 Hours	2018/04/20	2018/04/20	2018/04/27						
1006		Literature Review	8 Days	2018/04/20	2018/04/28	2018/04/30	1002					
1007		Conducting first draft of literature review	2 Days	2018/04/20	2018/04/22	2018/04/27						
1008		Review Draft Literature Review	6 Hours	2018/04/22	2018/04/22	2018/04/28						
1009		Enhance Literature Review Cycle	4 Days	2018/04/22	2018/04/26	2018/04/29						
1010		Conduct draft of Literature Review	1 Day	2018/04/22	2018/04/22	2018/04/29						
1011		Review current draft of Literature Review	1 Day	2018/04/22	2018/04/22	2018/04/29						
1012		Conduct Final Literature Review	1 Day	2018/04/27	2018/04/28	2018/04/30						
1013		Conduct Introduction	2 Hours	2018/04/28	2018/04/28	2018/04/29						
1014		Draft Methodolgy	3 Hours	2018/04/28	2018/04/28	2018/04/29						
1015		Review Methodology	1 Hour	2018/04/29	2018/04/29	2018/04/30						
1016		Construct final Methodolgy Draft	6 Hours	2018/04/29	2018/04/29	2018/05/01						
1017		Convert from word to latex	4 Hours	2018/04/30	2018/04/30	2018/05/01						
1018		**Project Report**		2018/05/01	2018/10/31		1001					
1019		Expand literature review		2018/05/02	2018/06/03							
1020		Learn metaphone algorithms		2018/05/04	2018/05/18							
1021		Implement feedback changes from proposal		2018/05/03	2018/05/06							
1022		Learn natrual language processing techniques		2018/05/18	2018/05/25							
1023		Implement Extracrtion module		2018/05/25	2018/06/09							
1024		Set up test email samples that algorithm will learn		2018/06/09	2018/06/10							
1025		Test extraction module		2018/06/10	2018/06/11							
1026		Implement Stylometric profiles from ML algorithms		2018/06/20	2018/07/03							
1027		Construct a run-time stylometric profile module		2018/07/05	2018/07/10							
1028		Construct a stylometric comparison algorithm module		2018/07/12	2018/07/19							
1029		Perform Unit Tests on comparison algorithm module		2018/07/19	2018/07/20							
1030		Construct User Interface to block content of message		2018/07/20	2018/07/25							
1031		Test UI of message		2018/07/25	2018/07/25							
1032		Document findings of results from tests		2018/07/26	2018/07/26							
1033		Conduct first draft of research report		2018/07/30	2018/08/20							
1034		Integrate final draft of proposal		2018/07/30	2018/08/03							
1035		Integrate report findings		2018/08/04	2018/08/14							
1036		Review Research Report		2018/08/15	2018/08/20							
1037		Conduct Final Paper		2018/08/20	2018/08/31							

Figure 1: A Project Plan.

References

[ASRS16] R. Aravindhan, R. Shanmugalakshmi, K. Ramya, and C. Selvan. Certain investigation on web application security: Phishing detection and phishing target discovery. *ICACCS 2016 - 3rd International Conference on Advanced Computing and Communication Systems: Bringing to the Table, Futuristic Technologies from Around the Globe*, 2016.

[ASZ+10] Michael Armbrust, Ion Stoica, Matei Zaharia, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, and Ariel Rabkin. A view of cloud computing. *Communications of the ACM*, 53(4):50, 2010.

[BGMZ97] Andrei Broder, Steven C Glassman, Mark S Manasse, and Geoffrey Zweig. Syntactic clustering of the Web IS (A) n S (B). 29, 1997.

[Bri95] Eric Brill. Transformation-Based Error-Driven Learning and Natural Language Processing : A Case Study in Part-of-Speech Tagging. *Computational Linguistics*, 21(4):543–565, 1995.

[Bro5] Jon Brodkin. problems with saas security. *Network World*, 27(18):1–27, 5.

[DKCE+16] Sevtap Duman, Kubra Kalkan-Cakmakci, Manuel Egele, William Robertson, and Engin Kirda. EmailProfiler: Spearphishing Filtering with Header and Stylometric Features of Emails. *Proceedings - International Computer Software and Applications Conference*, 1:408–416, 2016.

[Her09] Amir Herzberg. DNS-based email sender authentication mechanisms: A critical review. *Computers and Security*, 28(8):731–742, 2009.

[IHH17] Nazmul Islam, Mohammed Moshiul Hoque, and Mohammad Rajib Hossain. Automatic authorship detection from Bengali text using stylometric approach. *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6, 2017.

[KIJ11] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Mitigation of spear phishing attacks: A content-based authorship identification framework. *2011 International Conference for Internet Technology and Secured Transactions, ICITST 2011*, (December):416–421, 2011.

[MG11] Peter Mell and Timothy Grance. The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology. *National Institute of Standards and Technology, Information Technology Laboratory*, 145:7, 2011.

[MMM+11] Isla S. Mackenzie, Brian J. Mantay, Patrick G. Mcdonnell, Li Wei, and Thomas M. Macdonald. Managing security and privacy concerns over data storage in healthcare research. *Pharmacoepidemiology and Drug Safety*, 20(8):885–893, 2011.

[PHS18] Tianrui Peng, Ian G Harris, and Yuki Sawa. Detecting Phishing Attacks Using Natural Language Processing and Machine Learning. 2018.

[SH11] Farhan Bashir Fb Shaikh and Sajjad Haider. Security threats in cloud computing. *2011 International Conference for Internet Technology and Secured Transactions*, (December):214–219, 2011.

[SSP18] Howard Shrobe, David L Shrier, and Alex Pentland. *New Solutions for Cybersecurity*. MIT Press, 1 edition, 2018.

[VEE+15] Rakesh Verma, Ayman El, Aassal Ensias, Avenue Mohamed, Ben Abdellah, and Regragui Rabat. Comprehensive Method for Detecting Phishing Emails Using Correlation-based Analysis and User Participation. pages 155–157, 2015.

[VR15] R Verma and N Rai. Phish-IDetector: Message-ID based automatic phishing detection. *2015 12th International Joint Conference on e-Business and Telecommunications (ICETE)*, 04:427–434, 2015.