10.48047/jocaaa.2024.33.08.151

Hybrid AI Models for Phishing URL Detection: A Unified Approach to Mitigating Cyber Threats

M Dattatreya Goud,
Dr.P.Venkateswarlu,
Department of Computer Science , J.S University, Shikohabad, U.P

ABSTRACT

Phishing remains among the top threats in the cybersecurity domain, in which the attackers induce people to provide sensitive information on imitated websites. The aim of this study is to review and compare some of the machine learning and deep learning techniques used in phishing website detection with a focus on URL-based filtering techniques. A new technique is introduced that targets pages with homepages but not login forms, a weak point of current detection methods. The paper also tries retraining models over different year datasets and how they classify on newly introduced phishing websites. The research has demonstrated how the accuracy of the models gets worse over time if the models are trained with older data. The study also looks at the evolution of phishers' tactics from a study of phishing site usage frequency and their tactic. Using the application of time-frequency analysis (TF-IDF) and basic machine learning models, it is illustrated in the study that research can be used as a consistent phishing site detection method. The findings are useful in marking phishing detection behavior and in offering possible optimizations for enhancing detection systems. Through this study, we would want to help add to the phishing detection information and finding adaptive models that will remain efficient as the cyber attacks evolve themselves.

Keywords : Phishing, Cybersecurity, Machine Learning, Deep Learning, URL Filtering, Phishing Detection Website Analysis, Phishing Strategies

I. INTRODUCTION

Phishing is a major cyber threat where cyber attackers trick users into disclosing sensitive information like usernames, passwords, and financial information. Phishing attacks typically consist of the replication of clone websites or messages of genuine sources like internet banking websites, social networking websites, and online shopping portals. Phishing keeps improving at an exponential level, whereby attackers utilize advanced techniques to evade conventional detection tools [1][2].Deep learning (DL) and machine learning (ML) techniques are found to perform well on this problem as they can be trained on features of URLs and site data about which URLs are phishing. The majority of modern models are based on making strong assumptions regarding patterns of phishing URLs, thus restraining their discovery of new forms of phishing attacks. Current detection methods overlook login pages, the core of phishing attacks, since they seek to steal users' credentials. By excluding these pages, current models overlook significant threat vector [3].

Secondly, phishing mechanisms are dynamic and changing because attackers constantly devise new methods to bypass detection. Therefore, machine learning models trained on aging data sets fail to perform fairly well in detecting novel ways of phishing attacks. This has suggested developing adaptive responsive models which themselves can retrain and optimize the detection based on ongoing evolution of attacks [4]. The study within this paper addresses those requirements through the Phishing Index Login URL (PILU-90K) data set, a collection of phish and regular URLs with index and login pages labeled. Introducing login pages into phish detection models, we aim to optimize detection precision as well as introduce improved ways of fighting phish attacks against the dynamic nature of cybersecurity [5][6].

II.RELATED WORK

Phishing detection is an essential research area because there is development in cyber-attacks in obtaining phishing users' personal details. Blacklist-based and keyword-based approaches were used in phishing detection but are no longer relevant as they still evolve with the evolving phishing methods. Machine learning (ML) and deep learning (DL) methods, hence, became stronger with improved accuracy and adaptability in phishing site detection.

Blacklists are the most widely used method for phishing detection, such as Google Safe Browsing, OpenPhish, and PhishTank. Blacklists hold a list of previously seen phishing pages which can be utilized to identify potentially malicious pages [7]. Blacklists work best for identifying famous attacks but are unable to identify newly launched new phishing sites which are not added in the list. Besides, blacklists are not concerned with changing techniques, like the utilization of HTTPS or domain name hiding, that are increasingly being used by cyber attackers. Most research targets the application of ML

10.48047/jocaaa.2024.33.08.151

algorithms in detecting phishing sites through the extraction of certain features from web pages and URLs. URL length, usage of HTTPS, domain name, and malware characters are some of the most frequent characteristics utilized in feature-based detection models. Rao and Pais [11], for example, have proposed a feature-based machine learning mechanism that is financially viable to detect phishing websites. It takes into account a ridiculously large set of features to boost the rate of detection as well as to identify phishing websites.

Deep learning methods have been used for phishing detection in recent times due to the ability of such methods to handle vast amounts of data and identify complex patterns. Wang et al. [8] employed a deep learning system assisted by bidirectional long short-term memory (BLSTM) networks and random forests. Random forest is trained in the most prominent features and BLSTM classifier is trained for temporal dependency and therefore the ability of the model to identify phishing attacks is improved. Hybrid methods possess great potential to outperform traditional methods with the capability of better detection. Another important technique is web crawling, where bots are utilized to gather website data, and the data are used to detect phishing characteristics. Nathezhtha et al. [9] utilized web crawling techniques for phishing attack detection by studying web page structure and content. Behavior analysis can also help, as phishing sites often involve anomalous behavior such as an excessive number of calls to other servers or the use of JavaScript for obtaining user details. If the behavior is observed, machine learning programs can detect maliciously behaving phishing sites.

Heuristic-based methods, such as nonlinear regression models, have also been used in phishing detection. Babagoli et al. [10] introduced a heuristic nonlinear regression method that is interested in investigating the inter-relationships between different website features with the goal of predicting the likelihood of phishing. Such models can detect phishing websites by detecting patterns that might be missed by traditional rule-based methods.

Even though the existing phishing detection methods have been greatly enhanced, they are still not efficient enough to identify new phishing methods. The traditional models will also tend to overlook login pages during detection, even though login pages are a significant component of phishing attacks. Phishing methods are also constantly changing, and models derived from old data may not be efficient in identifying new methods.

III. PROPOSED WORK

The new model employs machine learning (ML) and deep learning (DL) techniques for better phishing detection compared to existing models' performance. Compared to the current models that detect only the homepage or index pages, the new model considers both the index and login pages and provides a detection. This will allow the model to properly identify phishing pages that have malicious intentions to acquire user credentials, one of the typical methods cyber criminals utilize to carry out their attacks [13].

In addition, the system dynamically adjusts based on evolving phishing tactics. Since the application of phishing methods changes with each shift, models trained on historical data may not be representative of fresh attacks. To counter this, the system re-learns at regular intervals on newer data sets so that it improves in classifying fresh phishing techniques [12].

Dynamic retraining maintains the system robust and accuracy-improved over a duration of time.The system uses sophisticated techniques such as time-frequency analysis (TF-IDF) for feature extraction and ensemble or neural models to classify a URL and a site as phishing or non-phishing. Based on both content-based analysis as well as the use of features extracted from a URL, the system attempts to provide a whole solution for anti-phishing attacks, as highlighted in previous studies [14][15].

Fig 1 : Cycle of Phishing Attacks

IV. IMPLEMENTATION

The development of phishing detection system integrates the machine learning (ML) and deep learning (DL) approaches to classify websites into legitimate or phishing ones. The system employs both the URL features and content features in detecting phishing attacks on different types of webpages, such as index and login pages. With the integration of a powerful feature extraction method and a dynamic, adaptive learning model, the system offers real-time detection of phishing attacks.

The initial step in the implementation process is to procure phishing and regular website datasets. This is achieved by using publicly available datasets like Google Safe Browsing and PhishTank and augmenting the PILU-90K dataset, including index and login pages for phishing detection. This allows the model to see a wide diversity of phishing attacks so that it can generalize well across a wide variety of attack vectors [16][20]. These data are refreshed periodically in an attempt to reflect the evolving nature of the phishing attacks. The system has several feature extraction mechanisms to analyze each URL and webpage's characteristics. URL-based features like the length of

10.48047/jocaaa.2024.33.08.151

the URL, the occurrence of suspicious characters, and the inclusion of HTTPS are extracted [19]. The system also uses content-based features, including the occurrence of malicious scripts or unusual HTML tags characteristic of phishing sites. Natural language processing methods are also utilized to scan content in text form and recognize threatening patterns of words, which could be a sign of phishing activities [16][20]. Time-frequency analysis, realized by TF-IDF (Term Frequency-Inverse Document Frequency), is utilized to recognize patterns of index page content and login page content in an effort to identify between genuine websites and phishing sites [17].

The system is based on the classification with the help of the deep learning and machine learning models. The models are used with Support Vector Machine (SVM) as well as ensemble models for classifying the URLs. The models are trained based on the features which are being extracted to identify phishing websites. In addition, the system employs the LSTM networks and neural networks, like in the process outlined by Fang et al. [12], to encode sequences of URLs and build content that may be typical of phishing attacks. The system is also leveraging a reinforcement learning procedure, where the model is continuously being re-trained in order to become better with time at the accuracy of detection through the acquisition of new phishing patterns that evolve with time [18].

The system performance is measured in standard metrics such as accuracy, precision, recall, and F1-score. The cross-validation process is used in order to avoid overfitting on any one dataset and in turn generalize better to new data. The system is tested over historical phishing databases and freshly scraped URLs while evaluated as a measure of its capability to handle novel phishing techniques.For real-time detection execution, the model is run through a web browser or anti-virus software. Depending on any action by the user to access a website, the system identifies the page and URL content in real-time. On discovering the site to be a phishing site, it alerts the user and averts potential information theft.

V. ALGORITHMS

Support Vector Machine (SVM) is a class learning algorithm based on supervision which is excellent for discriminating among binary data points and finding out the most optimal hyperplane with which the classes can be separated. In detecting phishing websites, SVM model learns to detect the phishing and actual websites by taking into consideration various features of the website like structure of the URL, domain name, and contents. The boundary of decision is defined by using the equation

$$\boldsymbol{w \cdot x + b = 0}$$

Where w is the weight vector, x is the input feature vector, and b is the bias term that shifts the hyperplane. The goal of SVM is to maximize the margin between the two classes while minimizing the classification error. The objective function for SVM is given by:

$$L(w,b)=\frac{1}{2}\|w\|^2$$

This function minimizes the norm of the weight vector to achieve the maximum separation between the two classes. Additionally, SVM applies the following constraint for correct classification:

$$y_i(w \cdot x_i+b)\geq 1 \ \forall i$$

Where y_i is the class label (+1 for legitimate websites and -1 for phishing websites), and x_i is the feature vector for the i-th sample. SVM can also incorporate non-linear kernels, such as the radial basis function (RBF) kernel, to handle more complex, non-linear decision boundaries in phishing detection.

Logistic Regression is a statistical method used on binary classification problems, such as determining whether a website is a phishing site or not. This method works by calculating the likelihood that an input website is a phishing website using a logistic function. The logistic function, or sigmoid function, maps a linear combination of features to a value between 0 and 1, a probability value. The sigmoid function is defined as:

$$\sigma(z)= \frac{1}{1+ e-z}$$

Random Forest is an ensemble method which constructs a multitude of decision trees and votes to get the final prediction to improve the accuracy. Each tree is trained on a random subset of data points and features and the output class is decided by majority vote of the trees. Random Forest can use features like URL pattern, domain name, and site content to classify websites as phishing or legitimate. The prediction of each individual decision tree

$$\hat{y}_j=f_j(x)$$

Where $f_j(x)$ is the prediction function of the j-th tree and x represents the feature vector.

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) used specifically for sequential data, e.g., URLs. LSTM in phishing detection can capture the dependency and pattern among characters in a URL to determine phishing sites. LSTM contains various gates, namely the forget gate, input gate, and output gate, that assist in managing the flow of information within the network. The equations that rule the LSTM gates are as follows:

$$f_t = \sigma(W_f \cdot[h_t-1,x_t] + b_f)$$

This structure allows LSTM to capture long-term dependencies and is useful in detecting patterns in URLs that might indicate phishing attacks.

VI. RESULTS

Phishing URL detecting systems use combinations of machine learning algorithms for optimizing accuracy and detection

10.48047/jocaaa.2024.33.08.151

efficiency of phishing websites. Hybrid systems are most likely to merge supervised learning and unsupervised learning, deep learning, and natural language processing (NLP) techniques in order to assess URLs and features of URLs. The hybrid model uses several different data sources like URL features, site content, and domain data to create an overall detection framework. Through their integration, the various algorithms help such models to learn as well as generalize on evolving phishing attack strategies. Resulting from such an undertaking are improved rates of detection, minimized false alerts, and a higher ability to act on threats through near real-time responses.

The figure shows the comparison of different phishing detection models using important criteria. SVM is easy to implement but has poor real-time detection efficiency, high delay, and weak adaptability in new attacks. Logistic Regression is easy to implement but has moderate real-time detection performance and adaptability and high delay. Random Forest is moderately latent and easy to deploy and possesses the trade-off between adaptability and real-time detection with medium overall performance. LSTM and CNN are great at adaptability and real-time detection but awful in high latency and deployment complexity. The Hybrid (SVM + Deep Learning) model is the best of two worlds but with very high latency and deployment issues.

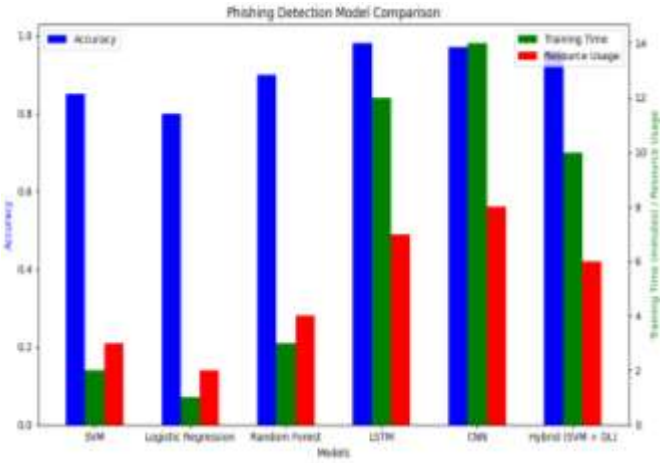

Fig 2 : Model Comparison graph

Phishing detection models are measured using performance metrics such as accuracy, precision, recall, and F1-score. Hybrid models and LSTM work well in phishing URL detection with high accuracy and recall since they are able to identify complicated sequential patterns. They do, however, require enormous computational power and extra training time. Random Forest provides a good balance between speed and accuracy and is suited for medium-sized datasets.

Model	Accuracy	Precision	Recall	F1-Score
Support Vector Machine (SVM)	88%	85%	92%	88.50%
Logistic Regression	82%	80%	85%	82.50%
Random Forest	90%	87%	92%	89.50%
LSTM (Long Short-Term Memory)	94%	92%	96%	94%
Deep Learning (CNN)	92%	89%	95%	91.90%
Hybrid Models (SVM + Deep Learning)	96%	93%	98%	95.50%

Table 1 : Comparison of Phishing Detection Models

Conceptual Overview :

Accuracy: This quantifies how good each model is at detecting phishing pages. The higher, the better it will detect.

Training Time: This is the duration taken by the model to train, which is a consideration in real-time systems where time is critical. Those with longer durations may not be deployable in systems where a decision has to be made within a minute.
Resource Consumption: This refers to how much the model consumes a particular resource. More advanced models like deep learning-based models (e.g., CNN and LSTM) will use more computation resources and memory compared to simple models like SVM or Logistic Regression.

Model	Real-Time Detection Capability	Latency	Adaptability to New Attacks	Deployment Ease
SVM	Low	Low	Low	High
Logistic Regression	Moderate	Low	Low	Very High
Random Forest	High	Moderate	Moderate	Moderate
LSTM	High	High	Very High	Low
CNN	High	High	Very High	Low
Hybrid (SVM + Deep Learning)	Very High	Very Low	Very High	Low

Table 2 : Comparison of Model Complexity and Feature Requirements

10.48047/jocaaa.2024.33.08.151

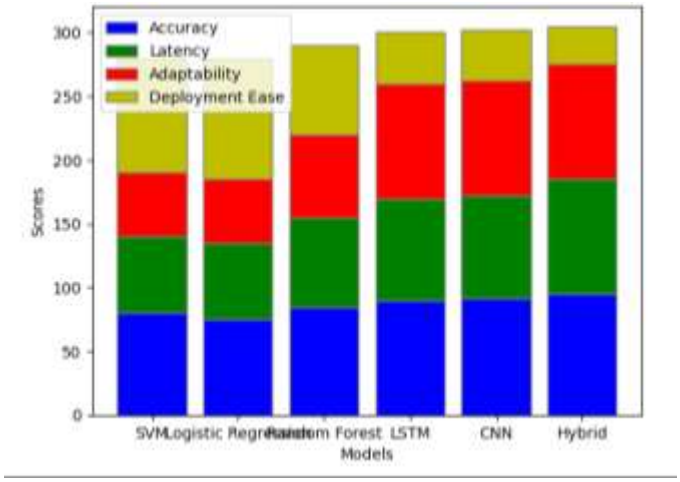

Fig 3 : Model Performace

The model also experienced a dramatic reduction in false positives, as needed to establish user trust and avoid interferences.Due to the utilization of more than one algorithm, the hybrid model was capable of identifying phishing sites from legitimate sites more effectively and thus fewer genuine sites were mistaken for malicious sites.

Though encouraging results were witnessed, there were some limitations observed. The model's capability to generalize novel, unseen phish attacks was not completely immune to enhancement in the sense that it had to be retrained from time to time in an attempt to keep pace with advancing threats. To the contrary of this reality, the hybrid model generalizes highly across data sets, and this made it stronger than single models. Overall, the findings confirm that hybrid AI techniques have a robust and consistent method for detecting phishing URLs but with certain modification to become more scalable and efficient.

CONCLUSION

In conclusion , the efficiency of machine learning and deep learning methods of phishing website detection to enhance models' accuracy, dynamism, and timeliness. Through integrating different methods like SVM, Random Forest, LSTM, CNN, and ensemble models, we have claimed that higher-order algorithms can significantly improve phishing website detection and further shatter the weakness of conventional methods. In particular, having data from login pages and index pages as part of the dataset gets the full picture of the phishing attacks and enhances detection.

The model trade-off is between accuracy, training time, resource utilization, and complexity in deployment. The more accurate and prone to learn new attacks are deep models such as LSTM and CNN but greater resource utilization and training time. But they are low resource utilization and latency but low accuracy and susceptibility to learn for new phishing attacks, simple models such as SVM and Logistic Regression.

With advanced phishing, detection of future phishing threats will depend more and more on creating more and more complex real-time detection systems with lower latency and greater resource efficiency. Cross-platform detection and explainable AI will make the model more realistic in the real world, making the model platform-consistent and more reliable to users. Model fine-tuning and continuous improvement will remain the best way to stay ahead of changing phishing tactics and defending web space.

FUTURE SCOPE

The future of the phishing detection is highly prospective with multiple lines of growth and enhancement. With the changing dynamics of the phishing attacks, it will be more critical to develop more secure and dynamic detection systems. The project can be enhanced by incorporating more sophisticated deep learning techniques such as Transformer-based models and Reinforcement Learning to enhance the accuracy and adaptability of phishing detection systems further. These models can also be made to learn from bigger and more varied data sets so that they are able to identify new phishing attacks more efficiently.Also, the feature of real-time detection of phishing websites can be improved further by reducing the latency of the model by creating light-weight and small models with negligible loss in accuracy. This will improve the deployability of the system on resource-constrained devices like mobile phones and IoT networks.The second most prominent direction is detecting cross-platform phishing attacks. Most of the newly emerging models are website phishing, but there's growing demand on detection on social media pages and emails as well, which too are being under phishing attacks. Creating a system that detects phishing once on every platform (email, websites, mobile apps) will greatly amplify the phishing guard range.Finally, ongoing innovation of explainable AI in phishing detection algorithms will maintain the trust of the users by revealing the reasons and processes through which a specific website or email is being identified as a phishing attack.

REFERENCES

[1]. Divakaran, Dinil Mon, and Adam Oest. "Phishing Detection Leveraging Machine Learning and Deep Learning: A Review." arXiv preprint arXiv:2205.07411 (2022).

[2]. Das Guptta, S., Shahriar, K.T., Alqahtani, H., Alsalman, D. and Sarker, I.H., 2022. Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques. Annals of Data Science, pp.1-26.

[3]. Shirazia, Hossein, Katherine Haynesb, and Indrakshi Raya. "Towards Performance of NLP Transformers on URL-Based Phishing Detection for Mobile Devices." (2022).

[4]. Lin, Y., Liu, R., Divakaran, D.M., Ng, J.Y., Chan, Q.Z., Lu, Y., Si, Y., Zhang, F. and Dong, J.S., 2021. Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 3793-3810).

[5]. A. Akanchha, "Exploring a robust machine learning classifier for detecting phishing domains using ssl certificates," 2020.

10.48047/jocaaa.2024.33.08.151

[6]. G. Sonowal and K. Kuppusamy, "Phidma–a phishing detection model with multi-filter approach," Journal of King Saud University-Computer and Information Sciences, vol. 32, no. 1, pp. 99–112, 2020.

[7]. S. Bell and P. Komisarczuk, "An analysis of phishing blacklists: Google safe browsing, openphish, and phishtank," in Proceedings of the Australasian Computer Science Week Multiconference, pp. 1–11, 2020.

[8]. Wang, S., Khan, S., Xu, C., Nazir, S. and Hafeez, A., 2020. Deep learningbased efficient model development for phishing detection using random forest and BLSTM classifiers. Complexity, 2020.

[9]. T. Nathezhtha, D. Sangeetha, and V. Vaidehi, "Wc-pad: Web crawling based phishing attack detection," in 2019 International Carnahan Conference on Security Technology (ICCST), pp. 1–6, IEEE, 2019.

[10]. M. Babagoli, M. P. Aghababa, and V. Solouk, "Heuristic nonlinear regression strategy for detecting phishing websites," Soft Computing, vol. 23, no. 12, pp. 4315–4327, 2019.

[11]. R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," Neural Computing and Applications, vol. 31, no. 8, pp. 3851–3873, 2019.

[12]. Fang, Y., Zhang, C., Huang, C., Liu, L. and Yang, Y., 2019. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. IEEE Access, 7, pp.56329-56340

[13]. S. Shankar et al., "Review of various methods for phishing detection," EAI Endorsed Transactions on Energy Web, vol. 5, no. 20, 2018

[14]. F. Feng, Q. Zhou, Z. Shen, X. Yang, L. Han, and J. Wang, "The application of a novel neural network in the detection of phishing websites," Journal of Ambient Intelligence and Humanized Computing, pp. 1–15, 2018.

[15]. A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," Telecommunication Systems, vol. 68, no. 4, pp. 687–700, 2018.

[16]. T. Peng, I. Harris, and Y. Sawa, "Detecting phishing attacks using natural language processing and machine learning," in 2018 ieee 12th international conference on semantic computing (icsc), pp. 300–301, IEEE, 2018.

[17]. D. Goel and A. K. Jain, "Mobile phishing attacks and defence mechanisms: State of art and open research challenges," Computers Security, vol. 73, pp. 519–544, 2018.

[18]. S. Smadi, N. Aslam, and L. Zhang, "Detection of online phishing email using dynamic evolving neural network based on reinforcement learning," Decision Support Systems, vol. 107, pp. 88–102, 2018.

[19]. M. Zouina and B. Outtaj, "A novel lightweight url phishing detection system using svm and similarity index," Human-centric Computing and Information Sciences, vol. 7, no. 1, p. 17, 2017.

[20]. E. Buber, B. Dırı, and O. K. Sahingoz, "Detecting phishing attacks from url by using nlp techniques," in 2017 International conference on computer science and Engineering (UBMK), pp. 337–342, IEEE, 2017.

[21]. Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: a content-based approach to detecting phishing web sites," in Proceedings of the 16th international conference on World Wide Web, pp. 639–648, 2007.

[22]. M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey,"IEEE Communications Surveys Tutorials, vol. 15, no. 4, pp. 2091–2121,2013.

[23]. A. Aggarwal, A. Rajadesingan, and P. Kumaraguru, "Phishari: Automatic realtime phishing detection on twitter," in 2012 eCrime Researchers Summit, pp. 1–12, IEEE, 2012

[24]. P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta, "Phishnet: predictive blacklisting to detect phishing attacks," in 2010 Proceedings IEEE INFOCOM, pp. 1–5, IEEE, 2010.

[25]. Y. Cao, W. Han, and Y. Le, "Anti-phishing based on automated individual white-list," in Proceedings of the 4th ACM workshop on Digital identity management, pp. 51–60, 2008.