International Conference on Recent Trends in Science & Technology-2021 (ICRTST - 2021)
Organised by: ATME College of Engineering, Mysuru, INDIA

Detection of Malicious Content or Web Links Related to Cyber Frauds

Sushma V
Assistant Professor
Department of CSE
ATMECE, Mysuru

Rachana G S
UG Student
Department of CSE
ATMECE, Mysuru

Nandakishor B M
UG Student
Department of CSE
ATMECE, Mysuru

Nishchal R
UG Student
Department of CSE
ATMECE, Mysuru

Tejas K
UG Student
Department of CSE
ATMECE, Mysuru

Abstract—Malicious URLs are being extensively used to mount numerous cyber attacks together with spamming, phishing andmalware. Detection of malicious URLs and identification of threat sorts lead to critical stage in these attacks.
Knowing the kind of a threat permits estimation of severity of the attack and helps adopt an effective countermeasure. Existing strategies typically detect malicious URLs of one attack kind. During this cyber fraud, wetend to propose method using cyber frauds to detect malicious URLs of all the popular attack sorts and identify the nature of attack a malicious URL attempts to launch.

Keywords- Malicious URLs, Domain NameSystem, Cyber Frauds.

I. INTRODUCTION

While the World Wide Web has become a killer application on the web, it has also additionally brought in an immense risk of cyber-attacks. Challenges have used this web(net) as a vehicle to deliver malicious attacks such as phishing, spamming, and malware infection. As an example, phishing generally involves sending an email seemingly from a trustworthy supply to trick people to click a URL (Uniform Resource Locator) contained within the email that links to a counterfeit webpage.

To address Web-based attacks, an amazing effort has been directed towards detection of malicious URLs. A standard step is to use a blacklist of malicious URLs, which might be constructed from various sources, significantly human feedbacks that are extremely correct however long. Blacklisting incurs no false positives; however it is effective just for noted malicious URLs. It cannot detect unknown malicious URLs. The very nature of actual match in blacklisting renders it simple and easy to be evaded.

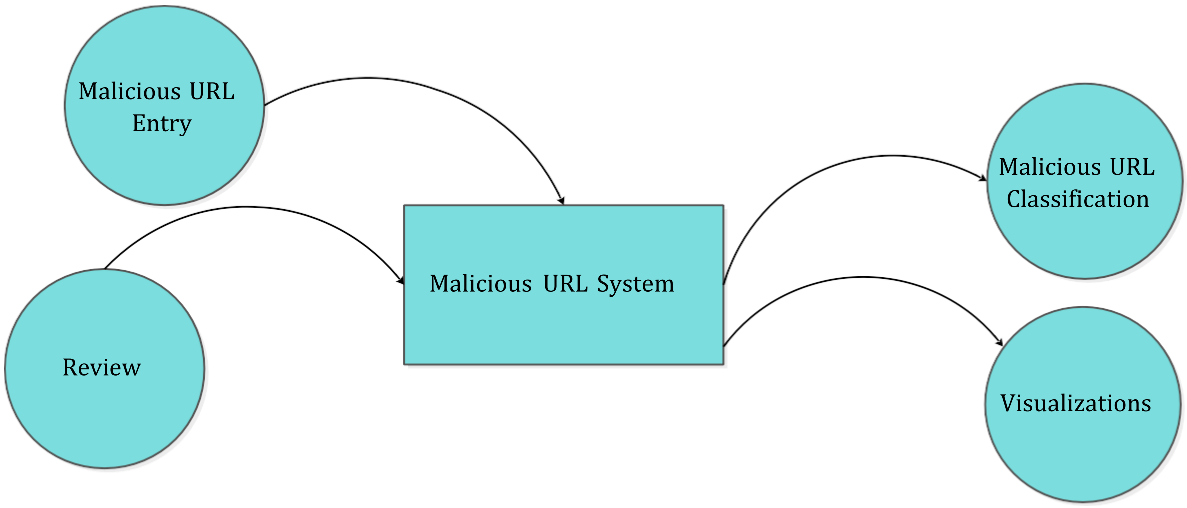
II. BACKGROUND

Detailed design starts after the system design phase is completed and the system design has been certified through the review. The purpose of this segment is to broaden the internal logic of each of the modules identified during system design.

In the system design, the focus is on identifying the modules, whereas during detailed design the focus is on designing the logic for the modules. In different phrases in device layout interest is on what additives are needed, whilst in specified layout how the additives may be carried out withinside the software program is the issue.

Data Flow Diagram graphically representing the functions, or processes, which capture, manipulate, store, and distribute data between a system and its environment and between components of a system. The visible illustration makes it an awesome verbal exchange device among User and System designer. Structure of DFD permits beginning from a vast assessment and extend it to a hierarchy of specified diagrams. DFD has often been used due to the following reasons:

- Logical information flow of the system
- Determination of physical system construction requirements
- Simplicity of notation
- Establishment of manual and automated systems requirements



III. LITERATURE SURVEY

Malicious URL, a.k.a. malicious website, is a common and serious threat to cybersecurity. Malicious URLs hostunsolicited content (spam, phishing, drive-by exploits, etc.) and lure unsuspecting users to become victims of scams (monetary loss, theft of private information, and malware installation), and cause losses of billions of dollars every year. It is imperative to detect and act on such threats in a timely manner. Traditionally, this detection is done mostly through the usage of blacklists. There are many popular websites which host a list of blacklisted websites, e.g. Phis Tank. The blacklisting technique lack in two aspects, blacklists might not be exhaustive and do not detect a newly generated phishing website. In, this cyber frauds we have compared different learning techniques for the phishing URL classification task and achieved the highest accuracy of 98% for Naïve Bayes Classifier with a precision=1, recall = .95 and F1 - Score= .9

Phishing has been easy and effective way for trickery and deception on the Internet. While solutions such as URL blacklisting have been effective to some degree, their reliance on exact match with the blacklisted entries makes iteasy for attackers to evade. We start with the observation that attackers often employ simple modifications Our system, Phish net, exploits this observation using two components. In the first component, we propose five heuristics to enumerate simple combinations of known phishing sites to discover new phishing URLs.

For years security machine learning research has promised to obviate the need for signature based detection by automatically learning to detect indicators of attack. Unfortunately, this vision hasn't come to fruition: in fact,developing and maintaining today's security machine learning systems can require engineering resources that arecomparable to that of signature-based detection systems, due in part to the need to develop and continuously tune the "features" these machine learning systems look at as attacks evolve. Deep learning, a subfield of machine learning, promises to change this by operating on raw input signals and automating the process of feature design and extraction. In this paper we propose the eXpose neural network, which uses a deep learning approach we have developed to take generic, raw short character strings as input .

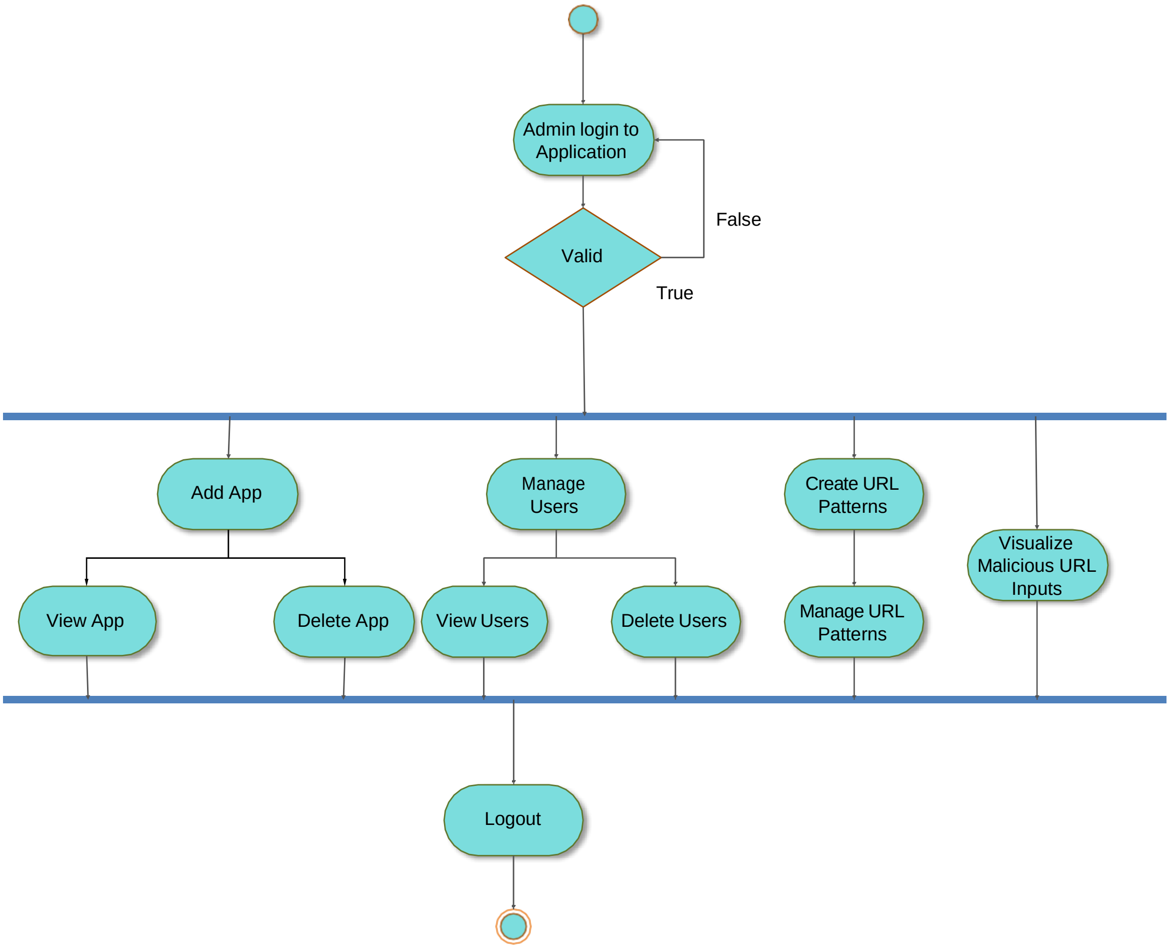
Embedding malicious URLs in e-mails is one of the most common web threats facing the internet community today. Malicious URLs have been widely used to mount various cyber- attacks like spear phishing, pharming, phishing and malware. By falsely claiming to be a trustworthy entity, users are lured into clicking on these compromised links to divulge vital information such as usernames, passwords, or credit card details andunknowingly succumb to identity theft. Hence, the detection of malicious URLs in e-mails is very essential so asto help internet users implement safe practices and as well prevent them from becoming victims of fraud. This cyber frauds explores how malicious links in e-mails can be detected from the lexical and host-based features of their URLs to protect users from identity theft attacks. This research uses Naïve Bayesian classifier as a probabilistic model to detect if a URL is malicious or legitimate.

IV. PROPOSED MODEL

Machine learning uses multiple comparison layers within which every layer will do non- linear projection to betold representations of multiple levels of abstraction within the URLs knowledge and they area applied to several cyber security applications. The most contributions of the projected work are:

1. Elaborated and Detailed investigation and analysis of assorted benchmark machine learning architecture areaunit performed for malicious URL detection.

2. Numerous kind of data sets like UCI dataset URL 2016 datasets are employed in the experimental analysis tofind out how generalizable the models area unit. The difference between the time-split and random-split of knowledge splitting methods will be shown within the experimental analysis.



CONCLUSION & FUTURE ENHANCEMENTS

Malicious URL detection plays a very important role for several cyber security applications, and clearly machine learning approaches are a promising direction. Here, we conducted a comprehensive and systematic survey on Malicious URL Detection using machine learning techniques. We tend to offer a scientific formulation of Malicious URL detection from a machine learning perspective, and then detailed the discussions of existing studies for malicious URL detection particularly in the forms of developing new feature representations, and designing new learning algorithms for resolving the malicious URL detection tasks. In this survey, we categorized most, if not all, the existing contributions for malicious URL detection in literature, and also identified the requirements and challenges for developing Malicious URL Detection as a service for real-world cybersecurity applications. Finally, we have a tendency to highlight some sensible problems for the appliance domain and indicated some necessary open problems for further additional research investigation. Especially, despite the in depth studies and therefore the tremendous progress achieved within the past few years, automateddetection of malicious URLs using machine learning remains a very difficult and challenging open problem.

V. REFERENCES

[1] P. Prakash, M. Kumar, R. R. Kompella and M. Gupta, "Phishnet: Predictive blacklisting to detectphishing attacks", *Proc. IEEE INFOCOM*, pp. 1-5, Mar. 2010.

[2] D. Sahoo, C. Liu and S. C. H. Hoi, Malicious URL detection using machine learning: A survey, 2017, [online] Available: https:// arxiv.org/abs/1701.07179.

[3] J. Saxe and K. Berlin, eXpose: A character-level convolutional neural network with embeddings for detecting malicious URLs file paths and registry keys, 2017,.

[4] B. Cui, S. He, X. Yao and P. Shi, "Malicious URL detection with feature extraction based on machinelearning", *Int. J. High Perform. Comput. Netw.*, vol. 12, pp. 166-178, 2018.

[5] B. Sun, M. Akiyama, T. Yagi, M. Hatada and T. Mori, "Automating URL blacklist generation with similarity search approach", *IEICE Trans. Inf. Syst.*, vol. E99-D, no. 4, pp. 873-882, 2016.

[6] CHUNG, Y.-J., TOYODA, M., AND KITSUREGAWA, M. Identifying spam link generators formonitoring emerging web spam. In WICOW: Proceedings of the 4th workshop on Information credibility (2010).