ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

CRF Based Secured Framework for Filtering Malicious Traffic

PRATHAP.C[1], SIRAJUN MUNIRA@SHAMIMA.S[2], KALAIVIZHI.P[3], STELLA.S[4]

Assistant Professor, Dept of UG studies in Engineering, Christ college of Engineering and technology, Puducherry, India[1]

UG Scholar, Dept of UG studies in Engineering, Christ college of Engineering and technology, Puducherry, India[2]

UG Scholar, Dept of UG studies in Engineering, Christ college of Engineering and technology, Puducherry, India[3]

UG Scholar, Dept of UG studies in Engineering, Christ college of Engineering and technology, Puducherry, India[4]

***Abstract*— Network security contains the provisions and policies adopted by a network administrator to prevent and monitor unauthorized access. It involves the authorization of access to data in a network. Filtering capabilities are available in access control lists (ACLs). It is typically stored in Ternary Content Addressable Memory (TCAM), whereas the size and cost of TCAM puts a limit on the number of filters, and this parallel access and reduces the number of lookups per forwarded is not expected to change in the near future. In this paper, we present a secured framework for filtering Malicious Traffic. This filtering framework is designed using CRF, where Conditional models are discriminative probabilistic systems that are used to model the conditional distribution over a set of random variables. Such models have been extensively used in the natural language processing tasks. Conditional models offer a better framework as they do not make any unwarranted assumptions on the observations and can be used to model rich overlapping features among the visible observations. The framework is designed such a way that the CRF are extensively trained by the models and then involved in the purpose of filtering malicious traffic in network. Our proposed method overcomes the problems in existing systems. We prove that our system is better than the existing system in both the terms of accuracy and efficiency**

Keywords—Filtering Malicious Traffic, DShield.org, Prefix filtering, Network security, CRF based filtering.

I.INTRODUCTION

Task of protecting networks from malicious traffics such as spammers, bots, internet worms, worst-case threats, flash crowds , malicious codes and denial of service attacks (DOS).These are the platform for launching malicious traffic. These activities cause problems on normal sources, ranging from simple traffic to operational, financial and political damage to official places, organizations, and critical infrastructure. In most of these years, they have rising in sophistication, and automation, largely enabled by botnets, volume, which are used as the platform for launching these attacks. Protecting a host or network from malicious traffic is a hard problem that requires the coordination of several complementary components, including business, legal and technical solutions at the application and levels of business. Filtering support from the network is a fundamental building block in this effort. With this traffic on the rise, the attackers have a keen interest in identifying networks and hosts that are responsible for a significant portion of malicious activities or expose vulnerabilities [1]. Software defined networking is an approach to building computer networks that separates and abstracts elements of these systems. This system that makes decisions about where traffic is sent from the underlying system that forwards traffic to the selected destination. The inventors and vendors of these systems claim that this technology simplifies networking this architecture allows network administrators to have programmable central control of network traffic without requiring physical access to the network's hardware devices. Intrusion detection provide as the important detection of identifying malicious activities and determing the activities, worst-case problem, seriousness etc. Detection system is a set of program that analyses what happen during the indication of particular IP address. Detection system uses the prefix based filtering to find the attack traces. The detection system can be classified according to two types they are anomaly detection and misuse detection. Anomaly detection can be based on statistical model to find the novel attacks in the system. Misuse detection is used to identify the intrusion scenario [2].now-a-days filtering the attacks are the main problem in the network system. One of the more sophisticated filtering

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

tools means which attacks come with advanced penetration method to identify the attackers has the advanced penetration method to defeat the installed security system. An Intrusion detection system (**IDS**) is software and/or hardware designed to detect unwanted attempts at accessing, manipulating, and/or disabling of computer mainly through a network, such as the Internet. These attempts may take the form of attacks, as examples, by crackers, malware and/or disgruntled employees. IDS cannot directly detect attacks within properly encrypted traffic. An intrusion detection system is used to detect several types of malicious behaviors that can compromise the security and trust of a computer system. This includes network attacks against vulnerable services, data driven attacks on applications, host based attacks such as privilege escalation, unauthorized logins and access to sensitive files, and viruses Detection system has the inherent requirements to find the many attacks with minimum number of false to defeat the network system. An accurate system that can handle the large number of malicious traffic and slow in the filtering of particular attacks in the system. To overcome the weakness of single prefix-based system the number of framework have been proposed, which describes the collaborative use of the network based systems [3].The main purpose of this framework is to identify the source prefix to filter a malicious attacks. The problem of blocking malicious traffic on the internet via source-based filtering. In particular filtering via access control list they are available at a routers today, but are a scarce resource because they are stored in the expensive Ternary Content Addressable Memory (TCAM).By filtering source prefixes instead of individual IP addresses helps to reduce the no of filters, but comes also at the cost of blocking legitimate traffic originating from the filter prefixes. This project shows how to optimally choose which source prefix to filter for a variety of realistic attack scenario and operator policies.

Problem Statements – A critical problem in filtering a malicious traffic on the Internet via source-based filtering. In particular, we consider filtering via ACLs. Aggregation helps to reduce the number of filters, but comes also at the cost of blocking legitimate traffic originating from the filtered prefixes. Filtering capabilities are already available at routers today via access control lists (ACLs). ACLs enable a router to match a packet header against predefined rules. ACLs are typically stored in ternary content addressable memory (TCAM).packet. However, TCAM is more expensive and consumes more space and power than conventional memory. The size and cost of TCAM puts a limit on the number of filters, and this - parallel access and reduces the number of lookups per forwarded is not expected to change in the near future.

II.RELATED WORK

In the malicious traffic, it is hard problem that which requires the cooperation of several components, including detector and migration techniques, as well as the architectural aspects'. We optimize the use of filtering a mechanism that already exists on the internet today and a necessary building block of any bigger solution. More especially, we focus on the optimal selection of which prefixes to block. The filtering rules can be then propagating by filtering the protocols which will control the high bandwidth of aggregation in the network. And it is ideally installed on routers as close to the attack sources as possible. However, the protocol typically assumes the ability to filter traffic at arbitrarily fine granularity and focus on where to place the filters. Therefore, they are complementary but orthogonal to this work. The Detection of malicious traffic is an important problem, but out of the scope. The source of legitimate traffic are also assume known and which is used for assessing the collateral damage-e.g., Web server or ISPs typically keep historical data and to know their important customers. We also consider addresses in the blacklist to not be spoofs. Where TCAM puts hard limit on the number of ACLs, there is no hard limit on the number of firewall rules in software. However there is still an incentive to minimize a number and thus any associated performance penalty for high performance packet classification. There is a body of work on firewall rule management and configuration of the firewall policy advisor, which aims at detecting anomalies, while we focus on resources on resource allocation.

Several rules have demonstrated that malicious sources exhibit spatial and temporal clustering such as by using the uncleanliness to predict future botnet addresses ,the spatial-temporal characteristics of internet malicious sources, analyzing large DDoS attack using multiple data use, understanding the network level behavior of spammers, and exploiting network structure for proactive spam mitigation and also the highly predicting blacklisting.

In order to deal with dynamic malicious traffic of IP addresses that how dynamic are IP addresses? Whereas IP prefix rather than the individual IP address are typically considered

The clustering, in combination with the fact that the distribution of addresses as well as the other statistical characteristics differ for good and bad traffic, have been exploited in the past for detection and the migration of malicious traffic, such as the e.g., spam by understanding the network level behavior of spammers, exploiting network structure for proactive spam mitigation and also the highly predicting blacklisting or DDOS by on filtering of DDoS attack based on source addresses prefixes. The work in on filtering of DDoS attack based on source addresses prefixes which are studied in source prefixes

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

filtering for classification and blocking of DDOS traffic, which is closely related to our problem of FLOODING. The selection of prefixes of this filtering was done heuristic, thus the large collateral is leading to damage. We tackle analytically the optimal source prefix selection so as to minimize collateral damage in contrast. Furthermore, we provide a more general framework for formulate and optimally solving a family of related problems, including but it is not limited to FLOODING.

The optimal filtering of source address prefixes as models and algorithm. In this new contribution includes the formulation and optional solution of the time varying version of the filtering problem; an extended evaluation section that simulating all filtering problems over *Dshield.org* logs, which includes FLOODING and DIST-FLOODING, which were not evaluated in the optimal filtering solution and additional proofs, complexity analysis, and simulation results. We also studied the optimal range-based filtering in the filtering source of unwanted traffic, where malicious source addresses were aggregated into continuous ranges of the IP address, instead of prefixes, and we developed greedy solutions.

III.EXISTING METHOD

The existing system uses the problem of blocking malicious traffic on the Internet via source-based filtering. In particular, we consider filtering via access control list (ACLs). The purpose of blocking malicious traffic, a filter is a simple ACL rule that denies access to a source IP address. Aggregation will helps us to reduce the number of filters, but comes also at the cost of blocking legitimate traffic originating from the filtered prefixes. Filtering capabilities are already available at routers today via access control lists (ACLs) [4]. ACLs enable a router to match a packet header against predefined rules. ACLs are typically stored in ternary content addressable memory (TCAM) packet [5]. However, TCAM is more expensive and consumes more space and power than conventional memory. The size and cost of TCAM puts a limit on the number of filters, and this - parallel access and reduces the number of lookups per forwarded is not expected to change in the near future. Then, there was five specific problems that correspond to different attack scenarios and operator policies:

- BLOCK-ALL:

Blocking all the addresses in the black list to minimize the collateral damage.

- BLOCK-SOME:

Blocking some addresses in blacklist to minimize the total cost of the attack.

- TIME-VARYING BLOCK-ALL/SOME:

Blocking all/some addresses in a time-varying blacklist. And it contains operations like addition, deletion, adjustment, multiple addresses.

- FLOODING:

The flooding contains, blocking flows during a DDoS flooding attack to meet bandwidth constraints

- DIST-FLOODING:

The distributed filtering across several routers during flooding.

IV. PROPOSED WORK

We propose an efficient filtering mechanism in a network using CRF (Conditional Random Field) [6].Generally these CRF are probabilistic systems that are used to model the conditional distribution over a set of random variables. In our framework we use CRF in labeling the malicious nodes. Conditional Random fields (CRFs) is used to model the conditional distribution over a set of random variables. Conditional random fields exploit the sequence structure in the observations without making unwarranted assumptions, which results in better classification [7]. Hence, in Conditional random fields for building intrusion detection systems. The figure 1 shows CRF models with different attributes. It shows inter-dependence between the features which are used to find the conditional dependence.

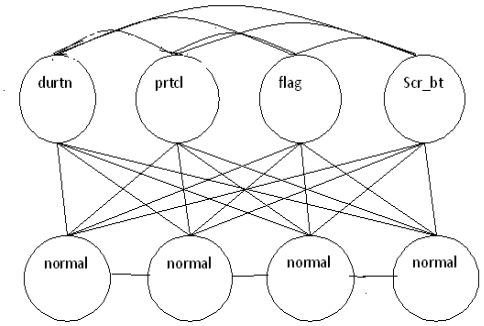

Figure. 1 CRF model

Method description

Conditional random fields have been effectively used for a variety of tasks including gene prediction, determining secondary structures of protein sequences, part of speech tagging, text segmentation, shallow parsing, named entity recognition, object recognition, intrusion detection and many others. Conditional random fields exploit the sequence structure in the observations without making unwarranted assumptions, which results in better classification. Hence, in Conditional random fields for

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

building intrusion detection systems. Conditional models are probabilistic systems that are used to model the conditional distribution over a set of random variables. Such models have been extensively used in the natural language processing tasks [8]. Conditional models have been extensively used in the natural language processing tasks. Conditional models offer a better framework as they do not make any unwarranted assumptions on the observations and can be used to model rich overlapping features among the visible observations [9].

V. MODULE DESCRIPTION

A. Network Creation Module

In this module we construct a network using socket programming, as shown in our Architecture. Where the users can send data to other nodes/network by using the options given. The user node will be listing all the nodes which are connected to the network. The sender can able to select the node name and then send the data.

B. CRF Construction module

In this module, we train the system by using the logs from Dshield.org. Given past logs of malicious activity collected at various locations. Predict sources likely to send malicious traffic to each victim network in the future. So in this module we train the system using CRF as shown in the figure2 by using the logs of older data collected from Dshiled.org [10]. In this module we construct the algorithm

Step 1: Select the 'n' layers needed for the whole filter

Step 2: Code the system for detecting various types of attacks and alerts for respective attacks.

Step 3: Integrate the developed CRF with the proposed system

Step 4: Specify each type of alert on which category it falls, so that user can easily recognize the attack type.

Step 5: Test the system using Attack Simulation module, by sending different attacks to the proposed system.

Step 6: Generate the log system where the attacks and other filter type's details are maintained continuously.

C. Optimal Source based filtering module

In this module we design Framework for optimal filter selection

- defined various filtering problems
- designed efficient algorithms to solve them
- Lead to significant improvements on real datasets
- Compared to non-optimized filter selection, to generic Clustering, or to uncoordinated routers
- Because of clustering of malicious sources

D. Evaluation module

In evaluation module, the evaluation nodes list the details of the malicious node and the good nodes. This node is designed as such it will be refreshed for a few seconds of period to update the information on each and every second. This node acts as a evaluation node as since it evaluates the nodes from malicious ones.

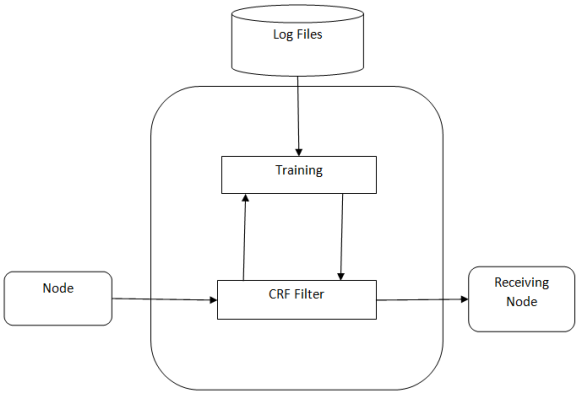

Figure. 2 Training diagram for CRF

VI. CONCLUSION

In this paper, we introduce a framework for source-Based filtering. The framework is provided at the theory of the CRF-based Filtering problem and provides a novel extension to it. Within it, we formulate practical problems, presented in increasing order of complexity. For each problem, we designed optimal CRF-based algorithms that are also small-complexity. We rooted our algorithms over Dshield.org logs and demonstrate that they bring significant benefit compared to non-optimized filter selection or to generic clustering algorithms. A work insight behind that benefit is that our algorithms exploit the spatial and temporal clustering formed by the source of malicious traffic.

VII. FUTURE ENHANCEMENT

As future enhancement, we propose an efficient System Using CRF technique. We propose an efficient filtering mechanism in a network using CRF (Conditional Random Filed). In our framework we use CRF in labeling the malicious nodes using the features of malicious nodes and so it protects the network from any

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 2, February 2013

types of attacks. Also in future we to manipulate the IP addresses with the same short prefix as the actual source to avoid the outbound traffic filtering.

REFERENCES

1. Z. Chen, C. Ji, and P. Barford, "Spatial-temporal characteristics of internet malicious sources," in *Proc. IEEE.*, Phoenix, May 2008.
2. P.K.Girwalkar, "A comparison review on Intrusion Detection using Bayesian Networks and Conditional Random Fields" in ICRAET 12,May 2012.
3. B.Bhanu chander, K.Radhika, D.Jamuna, "An approach on Layered framework on IDS" in AJCSIT 2012.
4. "Understanding ACL on Catalyst 6500 series switches," Cisco Systems, San Jose, CA, 2003.
5. Fabio Soldo, Athina Markopoulou, Katerina Argyraki "Filtering Malicious IP Source" Attack Infocom 2009.
6. Sargur Srihari., "Conditional Random Fields" as Introduction in 2012.
7. Hanna M. Wallach., "Conditional Random Fields: An Introduction" in February 24, 2004
8. C. Sutton and A. McCallum., "An Introduction to Conditional Random Fields for Relational Learning," in 2006.
9. Roman Klinger, Katrin Tomanek.," Classical Probabilistic Models and Conditional Random Fields" Algorithm Engineering Report TR07-2-013 in December 2007 ISSN 1864-4503.
10. Fabio Soldo, Katerina Argyraki, and Athina Markopoulou.,"An optimal source based filtering of malicious traffic" in April 2012