Global Effectiveness of Eco Monitoring Networks

Sumin Li[1,a], Xiuqin Pan[2,b]

[1]School of Information Engineering, Minzu University of China, Beijing, 100081, China

[2]School of Information Engineering, Minzu University of China, Beijing, 100081, China

[a]email:Lsm.cun@163.com, [b]email:sandycode@163.com

Keywords: Eco monitoring networks; Conditional Renyi Entropy; DShield

Abstract. This document studies the eco monitoring networks. Based on the prior work: the local effectiveness is proved, this paper focus on the global effectiveness of eco monitoring network. Information theoretical measures, i.e., conditional non-uniformity factor is applied. Although our work eventually quantify the ineffectiveness of global view, it discuss the possible reasons. Large-scale data is provided by DShield used to access the effectiveness of a real eco monitoring network.

Introduction

The eco monitoring networks is a kind of community-based network. In the world there are several such networks for example DShield (Internet Storm Center)[1], PREDICT (Protected Repository for the Defense Infrastructure Against Cyber Threats) [2], and DIMES (Distributed Internet Monitoring infrastructure) [3].

All data used in this paper comes from DShield. Such information: “source IP address A, source port number A1, destination IP address B, destination port number B1, monitor C, time T” is included in a scan packet from the direct DShield data.

Some prior work is studied on malicious/infected sources [4][5][6]. Through the saliency of a local view, the effectiveness of local inference is proved in[7]. The following will discuss the global effectiveness of eco monitoring networks, ie., the effectiveness of network-wide monitoring.

Global Effectiveness

Would the effectiveness of local views result in an effective eco monitoring network? In other words, would voluntary participations, i.e., good wills of the community, result in effective network-wide monitoring? In this section, we characterize the effectiveness of an eco monitoring network through conditional Renyi entropy, and measure the effectiveness of the real eco network using DShield data.

Conditional Renyi Entropy

We adopt Renyi information entropy [8] to measure the effectiveness of an eco monitoring network. In the prior work [9], Renyi entropy has been used to quantify a worse case scenario of random attacks that exploit the non-uniformity of malicious sources. In this work, Renyi entropy is extended to monitors. In particular, Renyi Entropy of a local view given monitor j is

$$H_v(P_j)=\frac{1}{1-v}\log_2\sum_{i=1}^{n}\left[P(X=i|Y=j)\right]^v, \qquad (1)$$

for v = 0,1, 2, and 1 ≤ j ≤ k . v is the order of the Renyi entropy, and v = 0, 2 are of interest here. $H_0(P_j)=\log_2 n$ is the Renyi entropy of order zero. When $n=2^l$, $H_0(P_j)=l$ is the prefix length of /l subnets. Renyi entropy of order 2 (Renyi entropy in short), $H_2(P_j)=\log_2 n-\log_2\beta_j(n)$, characterizes the randomness remaining in a malicious source distribution given a local view. Hence, the Renyi entropy is related to the conditional non-uniformity factor in a non-linear fashion [9].

The larger βj (n) is, the smaller the Renyi entropy, the more non-uniform the malicious distribution in view of a monitor, and less randomness remaining in a local view. The (expected) conditional Renyi entropy is defined by considering all monitors as follows [8].

Definition: Conditional Renyi entropy [8]. Conditional Renyi entropy is

$$E[H_v(P_j)] = \frac{1}{1-v}\sum_{j=1}^{k} P(Y=j)\log_2 \sum_{i=1}^{n}[P(X=i|Y=j)]^v , \tag{2}$$

where the expectation is over monitors. For v =2 , we obtain the following relation.
Lemma 1: $E[H_2(P_j)] = \log_2 n - E_Y(\log_2 \beta_j(n))$.

The proof of the lemma follows the definition of conditional non-uniformity factor in Equation 3[7] and that of conditional Renyi entropy in Equation 2 as well as simple algebraic manipulations; is thus omitted.

This expression represents the conditional Renyi entropy into two computing terms. The first is the conditional Renyi entropy of an uniform distribution over the network address space. A uniform distribution assumes that each subnet is equally likely to be a malicious source. For$/l$ subnets, there are $n = 2^l / l$ subnets in the network address space; and l , i.e., $\log_2 n$, is the dimension of the network space for malicious source locations. When l increase, the dimension of the network space increases, the amount of uncertainty increases as for where malicious sources may be. The growth rate l is the fastest as it corresponds to the most uncertainty in a uniform malicious source distribution. Hence, the first term $\log_2 n$ is referred to as the dimension uncertainty of the network address space.

The second term $E_Y(\log_2 \beta_j(n))$ can be regarded as the amount of information in bits that is captured by a typical local view in an eco monitoring network. This term takes two factors into consideration. One is the saliency β j (n) of the local view of an individual monitor. The other is the expectation over all monitors in an eco monitoring network. The expectation takes into consideration that monitors can receive malicious scan packets from different sources, and the number of scan packets received can vary from monitors to monitors.

Consider an extreme case of a most non-uniform distribution where each monitor sees one unique location of malicious sources, i.e., P(X=i | Y=j)=1 if i =i0 ,i0∈[1, n] ; and P(X=i | Y=j)=0 , otherwise. Then $\log_2 \beta_j(n)$ takes the maximum value log2n . This implies that there is no uncertainty left in the locations of malicious sources in view of a monitor since the monitor sees a unique peak as a source distribution. Hence the expected saliency compensates completely the dimensional uncertainty, and the conditional Renyi entropy is zero. Another extreme case is when all monitors see a uniform source distribution, i.e., $P(X=i|Y=j) = \frac{1}{n}, \forall i, j$. Then $\log_2 \beta_j(n) = 0$, and the expected Renyi entropy is the largest, i.e., log2 n . This implies that the saliency of the local views can not compensate any dimension uncertainty. In general, since $E_Y(\log_2 \beta_j(2^l)) \le \log_2 2^l$ and $\beta_j(2^l)$ is a non-decreasing function of l as shown in Property 2, we can obtain a lemma below.

Lemma 2: $0 \le E_Y[H_2(P_j)] \le l$, where 1⩽l .
Hence, the difference $E[H_2(P_j)] = \log_2 n - E_Y(\log_2 \beta_j(n))$ characterizes a trade-off between the dimension uncertainty and the saliency of local views: As the dimension l increase, the uncertainly of the network space increases. Meanwhile, the information bits $E_Y(\log_2 \beta_j(n))$captured by local views increase also, since monitors can see a more detailed malicious source distribution. Whether or not an increase in the saliency of local views can compensate more uncertainty in a larger network space determines the global effectiveness of an eco monitoring network. We use the growth rate of $E_Y(\log_2 \beta_j 2^l)$, i.e., the rate of $E[H_2(P_j)] = l - E_Y[\log_2 \beta_j(2^l)]$, with respect to l to quantify the effectiveness of an eco monitoring network.

Definition: An eco monitoring network is considered as effective if $\frac{E_Y[\log_2 \beta_j(2^l)]}{l} = o(l)$ This definition means that an eco monitoring network is effective if the information bits $E_Y[\log_2 \beta_j(2^l)]$ of local views can nearly compensate the dimension uncertainty.

B.Effectiveness of Eco Monitoring Network

Do the information bits captured by local views indeed compensate the dimension uncertainty in the real eco monitoring network? We evaluate empirical conditional Renyi entropy using DShield data. The empirical conditional Renyi entropy is for /l subnets and 1≤l≤k ,

$$\hat{E}\left[H_2\left(\hat{P}_j\right)\right] = l - \hat{E}_Y\left(\log_2 \hat{\beta}_j\left(2^l\right)\right). \tag{3}$$

Figure 1 shows the estimated conditional Renyi entropy. The conditional Renyi entropy increases piece-wise linearly with respect to l for /l subnets. This implies that more and more uncertainties arise when a monitor looks “deeper into the network”, i.e., at longer prefixes of malicious source locations. Furthermore, the uncertainty grows at the linear rates. Hence the eco monitoring network is ineffective by definition.

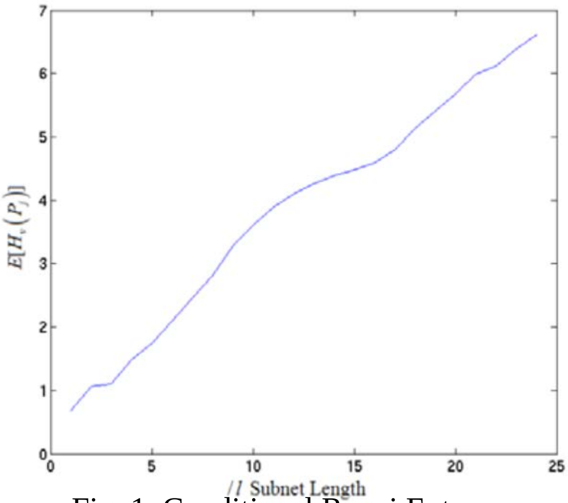

Fig. 1. Conditional Renyi Entropy

So far we have obtained the finding: the ineffectiveness of the eco monitoring network although the effectiveness of the local views is improved in [7]. Are these two findings contradictory? Observing Figure 1 carefully, we see that the slope of empirical conditional Renyi entropy is less than 1. This implies that the linear growth of information bits $\hat{E}_Y\left(\log_2 \hat{\beta}_j\left(2^l\right)\right)$ with respect to l is slower than l; thus, the dimension uncertainty dominates the conditional Renyi entropy. Therefore, as the network address space becomes larger, and the saliency of local views is insufficient for compensating the dimension uncertainty. This results in a twofold characteristic of the eco monitoring network: local views are salient individually but insufficient to result in an effective eco monitoring network system-wide.

Discussion

Why is the eco monitoring ineffective for the inference task? As this results in a worthy topic for further research, we discuss here a few possibilities.

The first is a relative small size of the eco monitoring network compared with the network address space being monitored. The number of monitors we found from DShield in the three months is less than three thousands. However, the monitored network address space for IPv4 is in the order of billions. The relative small size results from the number of voluntary participants. Community-based eco-systems are often moderate in size, and in fact, DShield is one of the largest. Hence, more study is needed to understand whether the relative small size poses limitations to eco monitoring networks.

The second issue is the topology of the eco monitoring network. The privacy of participants does not allow reconstruction of the topology of the eco monitoring network from DShield data. However, monitors from voluntary organizations are often clustered at a certain locations and lacking at the others (see DIMES [3] for a similar example). Hence, it would be helpful to understand whether or not such clustered participants would be able to monitor an entire network space.

Summary

This work has studied the issue in community-based eco monitoring networks: The effectiveness of an eco monitoring network system-wide. These issues have been studied in a setting of inference of malicious source locations. The conditional Renyi information entropy has been applied to quantify the effectiveness. Large-scale malicious scan measurements from DShield have been used to understand the effectiveness of a real eco monitoring network.

Our study has resulted in the finding: Although the local inference by individually monitors is effective[7], the eco monitoring network is ineffective system-wide. The observation is that the increase of information gain in local views is not fast enough to compensate the dimension uncertainty of malicious sources as the network address (prefix) space becomes larger.

The contribution here is an understanding based on a quantification using information theoretical measures. One direction is to understand whether a moderate number of participants and their clustered locations result in the ineffectiveness of eco monitoring networks. Another direction is to study other types of inference results shared by organizations in eco monitoring networks.

Acknowledgement

This work was supported by " Top university, Top discipline".

References

[1] Internet Storm Center (DShield), http://www.dshield.org/.

[2] http://www.predict.org

[3] Y. Shavitt and E. Shir, “DIMES: Let the Internet Measures Itself,” ACM(Sigcomm) Computer Communication Review, Vol. 35, Issue 5, pp. 71-74, Oct. 2005.

[4] M. Allman, P. Barford, B. Krishnamuthy, J. Wang, “Tracking the Role of Adversaries in Measuring Unwanted Traffic,” Proc. Of The Unix Second Workshop on Steps to Reducing Unwanted Traffic on the Internet, July 2006.

[5] Z. Chen, C. Ji, and P. Barford, “Spatial-temporal characteristics of malicious sources,” Proc. of INFOCOM’08 Mini-Conference, Phoenix, AZ, April 2008.

[6] F. Soldo, A. Le, and A. Markopoulou, “Predicative Blacklisting as an Implicit Recommendation System,” Proc. Infocom 2010.

[7] S. Li. “Understanding Effectiveness of Eco Monitoring Networks: Information Theoretical Perspective”, Advanced Materials Research, Vol. 765 – 767, pp. 2213-2219,Sep. 2013.

[8] J. N. Tsitsiklis, “Decentralized Detection,” in Advances in Signal Processing, Vol. 2, H. V. Poor and J. B. Thomas, editors, JAI Press, 1993, pp. 297-344.

[9] Z. Chen and C. Ji, “Information Theoretical View of Network-Aware Attacks,” IEEE Trans. Information Security and Forensics, Vol. 4, Issue 3, 530-541, Sept. 2009.