ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Solutions for Finding Network Traffic Measurements and Analysis

T.Durga Prasad
M.Tech Student,
Department of CSE,
Sree Rama institute of Technology and Science,
Kuppenakuntla, Penuballi, Khammam,TS India.

P.Spoorthi
Assistant Professor,
Department of CSE,
Sree Rama institute of Technology and Science,
Kuppenakuntla, Penuballi, Khammam,TS India.

ABSTRACT:

Online network traffic measurements and analysis is critical for detecting and preventing any real-time anomalies in the network. We propose, implement, and evaluate an online, adaptive measurement platform, which utilizes real-time traffic analysis results to refine subsequent traffic measurements.Central to our solution is the concept of Multi-Resolution Tiling (MRT), a heuristic approach that performs sequential analysis of traffic data to zoom into traffic subregions of interest. However, MRT is sensitive to transient traffic spikes. In this paper, we propose three novel traffic streaming algorithms that overcome the limitations of MRT and can cater to varying degrees of computational and storage budgets, detection latency, and accuracy of query response. We evaluate our streaming algorithms on a highly parallel and programmable hardware as well as a traditional software-based platforms. The algorithms demonstrate significant accuracy improvement over MRT in detecting anomalies consisting of synthetic hard-to-track elephant flows and global icebergs. Our proposed algorithms maintain the worst-case complexities of the MRT while incurring only a moderate increase in average resource utilization.

Index Terms:

Classification algorithms, computer network management, intrusion detection.

INTRODUCTION:

ACCURATE traffic measurement and monitoring is key to a wide range of network applications such as traffic engineering, anomaly detection, and security analysis.

A number of critical network management decisions, such as blocking traffic to a victim destination, require extraction and analysis of real-time spatio-temporal patterns in network traffic. The large traffic volumes seen in today's high-speed networks pose enormous computational and storage requirements for accurate traffic measurements. Traditionally, traffic measurements are performed by configuring conservative sampling factors[1] at the routers with very limited local storage. The collected samples are periodically sent to high-end servers where they are post-processed to answer some higher-level user queries (e.g., traffic volume from a customer domain) or to perform network troubleshooting and anomaly detection.

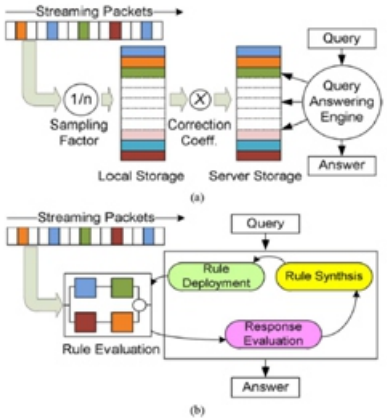

Fig. 1. Network measurement paradigms. (a) Traditional sampling-based methods. (b) Iterative measurement framework.

Existing System:

Sampling solutions, though straightforward, often introduce inaccuracies in estimating various flow statistics or in preserving traffic features critical for anomaly detection [2]. To bridge the gap between accuracy and detection latency, the concept of programmable measurements was proposed [3] to configure measurement rules that are representative of user requirements.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Such measurement specifications may not readily be available, especially in situations where the adaptation is based on network behavior rather than a fixed pattern. Forinstance, during the search of a volumetric anomaly such as a heavy-hitter, the measurements need to quickly adapt and track the evolving anomaly instead of being updated periodically with static sampling ratios.Recently, iterative measurements have gained attention as alternates to sampling-based solutions [4], [5]. The idea behind iterative measurements is to perform multiple sequential measurements and analysis of progressively finer resolutions. The contention is that repetitive measurements, analysis, and automated refinement of measurement goals smartly prunes away uninteresting data in a manner that is tied with the user requirements.

Proposed System:

The contributions of this paper are summarized as follows.

• We propose three traffic streaming algorithms, Equilibrium Rollback, Flow Momentum, and Directed Momentum, to guide the iterative configuration of measurement rule-sets by taking into account resource constraints, detection latency, and measurement goals. We demonstrate how our proposed algorithms address the shortcomings of MRT.
• With the recent trend toward software-defined networking paradigm and the adoption of OpenFlow [8] in various switches, we extend the evaluation of the iterative measurement framework on both hardware and software platforms.
We examine the actual rule-processing costs associated with different algorithms and examine the tradeoffs between cost and accuracy of our algorithms when implemented on these two different rule-processing platforms.
• We extend our algorithmic analysis to a distributed framework and show the effectiveness of our algorithms in detecting hard-to-isolate global icebergs. Our results demonstrate 100% detection accuracy of our algorithms across the platform choices, with low to moderate utilization of computation and storage budgets.

RELATED WORK:

Network traffic measurement fundamentally involves collecting information about a subset of traffic that satisfies some criteria.

Traffic is generally grouped in terms of flows, where a flow refers to a set of packets that have the same -tuple values in their header fields. Typical definitions of the flow include 6-tuple: , where is the protocol field, is type of service, and are the source and destination IP addresses, and and are the source and destination ports, respectively. We define a flowset to be an aggregation of flows. For instance, the Classless Inter-Domain Routing (CIDR) prefix is a particular type of a flowset that aggregates all the flows that have matching significant bitscorresponding to the size of the prefix. Traditional measurement schemes typically maintain unique "per-flow"-based statistics. The collected information is post-processed offline for answering higher-level user queries [11] such as detecting an anomalous behavior. The per-flow schemes, however, require storing information about potentially huge number of flows, straining the limited SRAM budgets of measurement hardware. The scalability issues of the per-flow scheme have traditionally been addressed using packet- or flow-based sampling approaches [12]. Studies have shown, however, that sampling leads to inaccuracies in answering the user queries [2]. Recently, smart sampling approaches, such as cSamp [13] and FlexSample [3], are proposed to balance the monitoring goals with resource constraints through smarter provisioning of resources based on application requirements. However, these schemes require measurement goals to be defined a priori, which can be challenging withhighly dynamic network or traffic conditions.

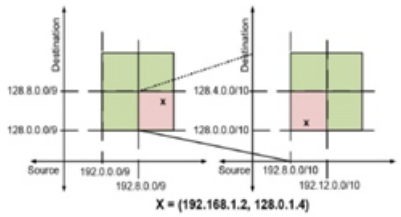

Fig. 2. MRT with zoom ratio of four.

MOTIVATION AND PROBLEM STATEMENT:

The key idea of the MRT is that one can, by observing a flowset, infer the characteristics of its subsets or objects (the flows). Therefore, one can selectively zoom into flowsets that might contain anomalies, such as heavy-hitters while ignoring others. As the algorithm explores the traffic landscape, it logs explored regions in a tree structure where nodes represent monitored regions in the IP-space.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The parent nodes represent regions in IP-space that are supersets of the region covered by their children nodes, with the root node of the tree covering the entire IP-space. The expansion ratio corresponds to the number of children, or arcs, originating from a parent node. As shown earlier, the MRT algorithm helps guide traffic measurements in the vast -tuple search space. However, the limited visibility, on which the iterative guided measurements have to base their decisions, can lead to false negatives and positives in detecting an anomaly. For instance, a brief spike in activity may lead the MRT to incorrectly declare the presence of a heavy flow when it may only be a transient Flash crowd [19]. Similarly, a brief absence of an anomaly can lead the MRT to discard a region from future consideration. Thus, when the anomalous behavior returns, the MRT will have to restart its tracking process from the top level with coarse granularities, resulting in false negatives and wastage of measurement resources as well as increased detection latencies.

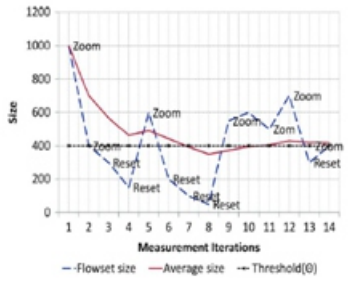

Fig. 3. Sensitivity of MRT algorithm to dynamic traffic fluctuations.

STREAMING ALGORITHMS FOR SMART GUIDED MEASUREMENTS:

We now present our streaming solutions that aim to address the challenges associated with application-aware rule-based online traffic measurements, with the goal of providing accurate response despite highly volatile traffic. As discussed earlier, a key design challenge is in maintaining computation and storage scalabilities while offering high accuracies and minimum latencies in answering the user query. In this context, we target volumetric anomalies, such as tracking of HHs and GIs, as our main query.

Flow Momentum:

The Flow Momentum (FM) algorithm addresses the issue of MRT resets by taking into account the average bit rate that is encountered over a hierarchical path reaching a flowset. This is in contrast to the original MRT, where the expansion/rejection decision is solely based on the flowset's current size. The FM algorithm thus effectively gives the leaf nodes a grace period in the active rule-set to cope with the temporal variations in the anomaly. The durations of the grace period are proportional to the intensity, or momentum, of the anomalous flows that guided the measurements toward the leaf node in the first place. Thus, in the case of the HH, a leaf node may be more active if the anomalous flow has larger volume.

Equilibrium Rollback:

The FM algorithm increases the duration where a given flowset remains covered by active rule-set until its effective momentum also falls below the thresholds. In the scenario where the momentum goes below the thresholds, the FM also resets its tracking process from the top level, similar to MRT. The Equilibrium Rollback (ER) algorithm addresses the problem of such MRT and FM resets through gracefully rolling back, or zooming out, of the expanded flowsets, instead of discarding off the flowsets. In doing so, the ER algorithm effectively tries to filter moderate traffic variations in the tracked flowsets, thereby achieving an equilibrium point over the zoomed hierarchy that just passes the threshold requirements.

Directed Momentum:

The streaming algorithms discussed so far are quite generic in nature, that is, they do not take into account any opportunity or constraints presented by application or available computational platform. They are thus best suited for scenarioswhere the knowledge of the anomaly or the environment is limited. However, such a separation between the application/platform and the algorithm may lead to suboptimal use of the computational resources.A very large rule-set can throttle the system by consuming scarce resources to process redundant or unnecessary rules. An intelligent hacker could actually use this deficiency to outsmart the detection process in real time by injecting a huge number of false flowsets (or leads) to be tracked. A smart algorithm therefore needs a mechanism to filter out irrelevant leads from the active rule-set.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Algorithm 1: Directed Momentum

input : R: Active rule-set
input : $Q\{\}$: Set of Rule Processors
input : Φ: The expansion/zoom ratio
input : δ: Measurement Interval
output : E_f: set of elephant flows
$Stretch \leftarrow |R| - |Q|$
$Pull \leftarrow |\min(Stretch, 0)|$

/* Measurement Phase */
1 **while** $t \leq \delta$ **do**
2 **for** $R_i \in R$ **do**
3 **if** $R_i = P_t$ **then**
4 $R_i.Size \leftarrow R_i.Size + P_t.Size$
5 $R_i.M \leftarrow R_i.M + P_t.Size$

/* Decision Phase */
6 **for** $R_i \in R$ **do**
7 **if** $(R_i.Size > Size_{Th})$ **then**
8 **if** Granularity$(R_i) = MAX$ **then**
9 $E_f \leftarrow E_f + R_i$
10 **else**
11 R.replace$\{R_i$, Expand$(R_i, \Phi)\}$
12 $R_{expanded}.M \leftarrow R_{parent}.M$
13 $R_{expanded}.Static \leftarrow 0$
14 **else if** $pForce(R_i, Pull)/\lambda \geq \theta$ **then**
 $R_i.Static \leftarrow R_i.Static + 1$
 R_i.Hold
15 **else** R_i.Drop

/* Calculates Directed Momentum */
procedure pForce $(R_i, Pull)$
$\lambda_i \leftarrow (R_i.M/k * \delta)$
$d_i^n \leftarrow \exp^{(Pull * R_i.Static/Granularity(R_i))}$ /* pForce*/
return (λ_i/d_i^n)
end procedure

Algorithm 2: Equilibrium Rollback

/* Decision Phase */
1 **for** $R_i \in R$ **do**
2 **if** $(R_i.Size > Size_{Th})$ **then**
3 **if** Granularity$(R_i) = MAX$ **then**
4 $E_f \leftarrow E_f + R_i$
5 **else** R.replace$\{R_i$, Expand$(R_i, \Phi)\}$
6 **else if** Granularity$(R_i) > 1$**then**
7 R.replace$\{R_i$, Collapse$(R_i)\}$
8 **else** R_i.Drop

EMPIRICAL EVALUATION:

In this section, we present our empirical evaluation of the proposed algorithms utilizing BURAQ hardware and Radix- Trie software-based rule-processing platforms.We begin by discussing the rule-set sizes created by the algorithms that have a direct implication on running costs of the algorithms. We thenmap the algorithms on BURAQ platform and demonstrate their accuracies in its context.

This is followed by a discussion on the effects of tuning Zoom-Ratio parameter toward the accuracy and detection latencies of the solutions. We next evaluate the algorithms on software platform and analyze the latencies that contribute in both hardware- and software-mapped solutions. Finally, we apply our algorithms to a distributed measurement framework and demonstrate its effectiveness in isolatingdistributed anomalies.

CONCLUSIONS:

In this paper, we have proposed the k-zero day safety as a novel network security metric, discussed its computation and application, and demonstrated its power in practical scenarios. Specifically, we formally defined the k-zero day safety model and showed that the metric satisfied the required algebraic properties of a metric function. We then studied the complexity of computing the metric and proposed efficient algorithms for determining the metric value. Next, we applied the proposed metric to the practical issue of network hardening and extended the metric to characterize various hardening options; we also discussed in details how the abstract model may be instantiated for given networks in practice. Finally, we demonstrated how applying the proposed metric may lead to interesting and sometimes surprising results through a series of case studies; we also discussed how the metric may potentially be applicable to SCADA security.

REFERENCES:

[1] Cisco, San Jose, CA, USA, “Cisco IOS NetFlow,” [Online]. Available: http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html.

[2] J. Mai, C.-N. Chuah, A. Sridharan, T. Ye, and H. Zang, “Is sampled data sufficient for anomaly detection?,” in Proc. 6th ACM SIGCOMM IMC, 2006, pp. 165–176.

[3] A. Ramachandran, S. Seetharaman, and N. Feamster, “Fast monitoring for traffic subpopulations,” in Proc. 8th ACM SIGCOMM IMC, 2008, pp. 257–270.

[4] C. Estan, S. Savage, and G. Varghese, “Automatically inferring patterns of resource consumption in network traffic,” in Proc. SIGCOMM, 2003, pp. 137–148.

[5] L. Jose, M. Yu, and J. Rexford, “Online measurement of large traffic aggregates on commodity switches,” in Proc. , Hot-ICE, 2011, p. 13.

[6] L. Yuan, C.-N. Chuah, and P. Mohapatra, “ProgME: towards programmable network measurement,” in Proc. SIGCOMM, 2007, pp. 97–108.

[7] F. Khan, N. Hosein, C.-N. Chuah, and S. Ghiasi, “Streaming solutions for fine-grained network traffic measurements and analysis,” in Proc. 7th ACM/IEEE ANCS, 2011, pp. 227–238.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[8] N. McKeown, T. Anderson,H.Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: enabling innovation in campus networks," Comput. Commun. Rev., vol. 38, no. 2, pp. 69–74, Mar. 2008.

[9] F. Khan, S. Ghiasi, and C.-N. Chuah, "A dynamically reconfigurable system for closed-loop measurements of network traffic," IEEE Trans. Comput., 2012, to be published.

[10] F. Khan,N.Hosein, S.Ghiasi, C.-N. Chuah, and P. Sharma, "Streaming solutions for fine-grained network traffic measurements and analysis," UC Davis, Davis, CA, USA, Tech. Rep. ECE-CE-2013-2, 2013 [Online]. Available: http://www.ece.ucdavis.edu/cerl/techreports/2013-2/

[11] N. Brownlee, C. Mills, and G. Ruth, "Traffic flow measurement: Architecture," RFC 2722, 1999 [Online]. Available: http://www.ietf.org/ rfc/rfc2722.txt

[12] N. G. Duffield, "Sampling for passive Internet measurement: A review," Statist. Sci., vol. 19, no. 3, pp. 472–498, 2004.

[13] V. Sekar, M. K. Reiter, W. Willinger, H. Zhang, R. R. Kompella, and D. G. Andersen, "CSAMP: A system for network-wide flow monitoring," in Proc. 5th USENIX NSDI, San Francisco, CA, Apr. 2008, pp. 233–246.

[14] C. Estan and G. Varghese, "New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice," Trans. Comput. Syst., vol. 21, no. 3, pp. 270–313, 2003.

[15] G. Cormode and S. Muthukrishnan, "An improved data stream summary:the count-min sketch and its applications," J. Algor., vol. 55, pp. 58–75, Apr. 2005.

[16] Q. G. Zhao, M. Ogihara, H. Wang, and J. J. Xu, "Finding global icebergs over distributed data sets," in Proc. 25th ACM SIGMOD-SIGACT-SIGART PODS, 2006, pp. 298–307.

[17] H. Zhao, A. Lall, M. Ogihara, and J. Xu, "Global iceberg detection over distributed data streams," in Proc. IEEE ICDE, Mar. 26, 2010, pp. 557–568.

Author's:

T.Durga Prasad is a student of Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, TS,India.Presently he is Pursuing his M.Tech (CSE) from this college. His area of interests includes Information Security, Cloud Computing, Data Communication & Networks.

Ms. P.Spoorthi is an efficient teacher, received M.Tech from JNTU Hyderabad is working as an Assistant Professor in Department of C.S.E, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, AP,India. She has published many papers in both National & International Journals. Her area of Interest includes Data Communications & Networks, Information security, Database Management Systems, Computer Organization, C Programming and other advances in Computer Applications