ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Traffic Identification Method Engine: An Open Platform for Traffic Classification

A.R. Arunachalam, K.G.S. Venkatesan, Abdul Basith.K.V, M. Sriram

Dept. of C.S.E., Bharath University, Chennai, India

ABSTRACT: The availability of open supply traffic classification systems designed for each experimental and operational use will facilitate collaboration, convergence on normal definitions and procedures, and reliable analysis of techniques. during this article, we tend to describe Traffic Identification Engine (TIE), associate open supply tool for net-work traffic classification , that we tend to started developing in 2008 to market sharing common implementations and information during this field. In the traffic identification engine we use to special path or send messeges. in this algorithm we have the lowest path of width to send messages. Through scientific collaborations, and thanks to the support of the open source community, this platform gradually evolved over the past five years, supporting an increasing number of functionalities, some of which we highlight in this articles.

KEYWORDS: TIE, Functions,Web Architecture.

I. INTRODUCTION

we initial offer a short summary of the evolution of traffic classification and also the challenges addressed within the last years during this analysis field [2]. we tend to then describe the most parts and functionalities of TIE by particularization a number of our style decisions, conjointly driven by such analysis of the state of the art. we tend to finally illustrate some representative use cases of applying TIE to specific analysis problems: Comparing the accuracy of various classifiers
Comparing their classification performance. Investigating multi-classification and combination methods
Despite the big body of literature printed on traffic classification, and analysis efforts of the many educational and trade analysis teams, there square measure few open supply implementations of network traffic classifiers on the market. additionally, one in all the most problems once novel classification approaches square measure given is that the inability to properly valuate and compare them [3]. within the thought of traffic categories (specific applications, application classes, etc.), still as within the metrics accustomed valuate classification accuracy [3] .

II. TRAFFIC CLASSIFICATION AND RELATED CHALLENGES

The evolution of web applications has created ancient ways for classifying network traffic more and more less effective. Port-based approaches will simply misclassify traffic flows, principally attributable to new applications reusing port numbers registered at IANA with different applications, arbitrarily choosing port numbers, or rental users select a most popular port. Payload-based approaches — that examine packets con-tent to spot peculiar patterns — square measure thought of a lot of reliable, however create privacy, technological, and economic challenges, and can't be applied to encrypted and obfuscated traffic. additionally, the increasing use of protocol encapsulation and multi-channel applications (i.e., victimization totally different communication channels for heterogeneous services) has more hindered the power to classify web traffic.

To address such limitations, many different ways are planned within the literature, generally applying techniques and algorithms originally developed in different analysis fields (signal process ,applied mathematics models, characterization of network traffic, machine learning) to numerous traffic properties. samples of properties of network traffic used for such functions square measure (at flow level) period, volume, mean packet size, and (at packet level) size and inter-packet time of the primary n packets of a flow. Machine-learning techniques [5] and heuristic approaches [6] haveestablished significantly promis-ing once handling obfuscated and encrypted traffic.

However, progress during this space quickly found obstacles to assessing the state of the art and reaching consensus (in terms of methodologies, definitions, and best practices) within the analysis and operational community. the variety within the terminology and definitions adopted once describing approaches and metrics [2], similarly because the wide selection of granularities in process flows and traffic categories across approaches [1], created it troublesome to check completely different studies. as an example, completely different approaches assign flows to traffic categories of various coarseness (e.g., distinguishing application classes like peer-to-peer vs. specific applications like Kazaa, Bittorrent). Agreeing on shared procedures, benchmarking metrics, flow definitions, traffic categories, similarly as mapping between different classification granularities (e.g., mapping the IMAP, POP, and SMTP application protocol categories to the mail category class) would instead yield a lot of rigorous results and facilitate the assessment (and so the progress) of the state of the art during this field [1]. per this philosophy, we have a tendency to designed TIE to simply compare classifiers, as we have a tendency to show within the 1st use case.

Another issue (mostly thanks to privacy concerns) is in accessing traffic traces representative of various eventualities, to be used as take a look at knowledge or as reference for validation
[1].galvanized by solutions planned by the community [3], we have a tendency to value-added in TIE support for sharing traffic traces while not payload together with per-flow reference labels (sometimes known as ground truth).

Most classification approaches weren't designed to figure in real-world eventualities (i.e., online), as an example, for live reportage or triggering of actions per classification results is predicted. Many compromises have then been proposed to search out the correct trade-offs among accuracy, performance, and cost: reducing the quantity of traffic knowledge analyzed (e.g., limiting the amount of packets inspected for every flow [7, 8]); reducing the procedure overhead (e.g., shrinking the set of options [9]); exploiting the high similarity of recent laptop architectures (e.g., all-purpose graphical processing units, GPGPUs [10]). we have a tendency to describe however we have a tendency to designed TIE to support on-line classification and performance evaluation, 2 key functionalities within the second use case mentioned during this article.

III. LITERATURE SURVEY

Traffic classification technology has magnified in connectedness this decade, because it is now utilized in the definition and implementation of mechanisms for service differentiation ,network style and engineering, security, accounting, advertising, and analysis. Over the past ten years the analysis community and also the networking business have investigated, projected and developed many classification approaches. we have a tendency to define the persistently unsolved challenges within the field over the last decade, and counsel many methods for endeavor these challenges to market progress within the science of net traffic classification [11].

Comparing traffic classifiers, several well-thought-of analysis teams have printed many fascinatingpapers on traffic classification, proposing mechanisms of various nature. We see a minimum of 2hurdles before this may happen. a serious issue is that we want to search out ways in which to share full-payload information sets, or, if that doesn't persuade be possible, a minimum of anonymized traces with complete application layer meta-data. a comparatively minor issue refers to finding AN agreement on that metric ought to be wont to appraise the performance of the classifiers .during this note we have a tendency to argue that these are necessary problems that the
community ought to address, and sketch many solutions to foster the discussion on these topics [15].

Tie a community-oriented traffic classification platform, throughout the last years the analysis on network traffic classification has become terribly active. The analysis community, emotional by increasing difficulties within the machine-driven identification of network traffic and by considerations associated with user privacy, began to investigate and propose classification approaches different to port-based and payload-based techniques. Despite the massive amount of works printed within the past few years on this subject, only a few implementations are getting different approaches were created accessible to the community. Moreover, most approaches projected in literature suffer of issues associated with the power of evaluating and examination them [17].

Blinc multi level traffic classification within the dark, we have a tendency to gift a basically completely different approach to classifying traffic flows in line with the applications that generate them. In distinction to previous ways, our approach relies on observant and characteristic patterns of host behavior at the transport layer. we have a tendency to analyze these patterns at 3 levels of skyrocketing detail (i) the social, (ii) the purposeful and (iii) the appliance level. This structure approach of watching traffic flow is perhaps the foremost necessary contribution . Our results show that we have a tendency to ar ready to classify 80%- ninetieth of the traffic with quite ninety fifth accuracy [19].

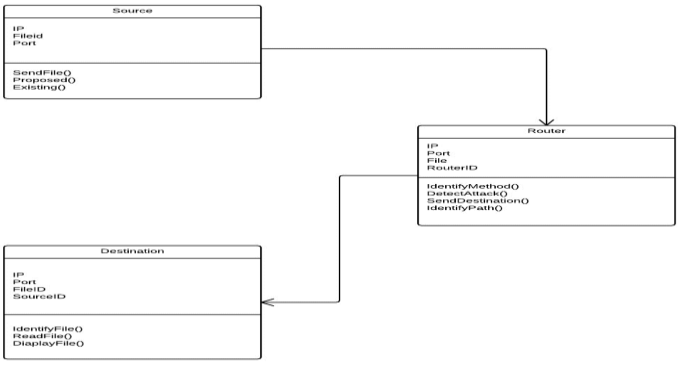
Early application identification, the automated detection of applications related to network traffic is an important step for network security and traffic engineering. The feasibleness of application identification at the start of a protocol affiliation. supported AN analysis of packet traces collected on eight completely different networks, we discover that it's doable to tell apart the behaviour of AN application from the observation of the scale and also the direction of the primary few packets of the protocol affiliation. we have a tendency to apply 3 techniques to cluster protocol connections: K-Means, mathematician Mixture Model and spectral bunch. ensuing clusters are used along side assignment and labelling heuristics to style classifiers [20].

Traffic classification through joint distributions of packet-level statistics, Interest in traffic classification, in each business and world, has dramatically full-grown within the past few years. analysis is devoting nice efforts to applied mathematics approaches exploitation sturdy options. during this paper we have a tendency to propose a classification approach supported the joint distribution of Packet Size (PS) and put down Packet Time (IPT) and on machine-learning algorithms. Provided results, obtained exploitation completely different real traffic traces, demonstrate however the projected approach is in a position to attain high (byte) accuracy (till 98%) and the way the new discriminating options have properties of hardiness, that counsel their use within the style of classification/identification approaches sturdy to traffic secret writing and protocol obfuscation [12].

IV. DEFINITIONS AND OPERATING MODES

DEFINITIONS - so as to match completely different classification approaches, TIE proposes a unified illustration of classification results. It defines IDs for application categories (applications) and associates them with cluster categories (groups), that embrace applications providing similar services. Such mapping allows the comparison of techniques functioning at completely different granularities (e.g., applications vs groups) or, for example, the comparison of traffic classifiers that have application-level protocol categories employing a coarser coarseness. Moreover, many application sub-classes (sub applications) are related to every application, so as to discriminate connected traffic flows serving completely different functions (signaling vs. data, Skype voice v.
OPERATING MODES - TIE are often run in 3 in operation modes, every cherish a special overall behavior: Offline mode: A flow is assessed only if it expires or at the top of TIE execution. This mode is helpful for evaluating classification techniques once no temporal arrangement constraints apply, or once a classifier needs perceptive flows for his or her entire period of time.

V.CLASS DIAGRAM



10.15680/ijircce.2015.0303163

V. COMPARING MEMORY AND COMPUTATIONAL OVERHEAD

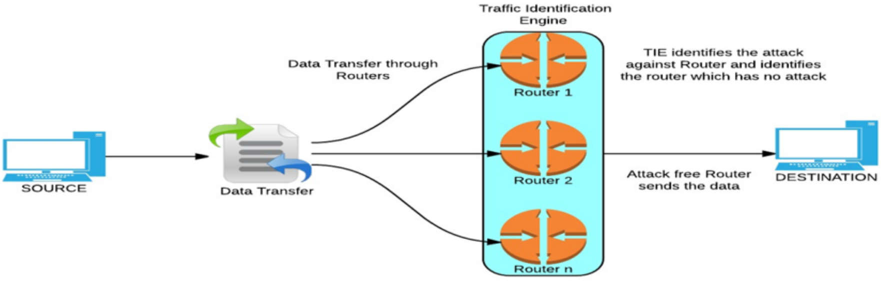
TIE may also be accustomed compare the performance of traffic classification approaches in terms of classification time and CPU/memory usage. In [8], by mistreatment identical traffic trace and classifiers of the previous use case, we tend to conjointly compared their memory and computer hardware overhead mistreatment TIE's extensions for performance analysis. we tend to conducted our tests on GNU/Linux (kernel two.6.27), when corroboratory that no different user processes were intense important computer hardware time, and none of the operating system [21].

To measure classification time, we tend to compiled TIE with the temporal order possibility enabled, and used the values measured for the Port plugin as a reference, since this can be the quickest classification technique (it solely needs one operation on a hash table for the primary packet of every session). Figure 2a summarizes the general results with per-session average values. Such results are in step with Port Load requiring, by implementing a set string comparison, a lot of less procedure resources than pattern matching supported regular expression.

We monitored resource usage by sanctioning the memory dump practicality, through that heap memory allocations is logged to file, and grouping the central processor usage of TIE with a shell script supported the annotation command. Figure 2b shows that the central processor usage of TIE, once running solely L7, reached the most price among the primary sixty s once decreasing to a gradual price. The slow decay is perhaps attributable to the very fact that the central processor %reportable by annotation may be a moving average. On the opposite hand, the Port and Port Load plugins showed similar qualitative behaviors, shortly reaching steady levels [25].

VI. TIE AND THE RESEARCH COMMUNITY

Starting from the primary unharness of TIE in 2009 (at that point offered upon request by email), the platform has been cited in additional than forty publications and, through many collaborations, has been extended to support new classification options and schemes — as well as the mix of multiple classifiers — also on run techniques already offered in rail (third use case). TIE has been downloaded quite a hundred and fifty times consistent with statistics collected at the official web site (unique downloads of distinct users United Nations agency filed missive of invitation through an online form). The transfer requests originated from universities (62 percent), corporations (30 percent), and people (8 percent) [27].


Traffic Identification Engine
Data Transfer through Routers
TIE identifies the attack against Router and identifies the router which has no attack
Router 1
Router 2
Router n
SOURCE
Data Transfer
Attack free Router sends the data
DESTINATION

VII. ARCHITECTURE OF TIE

In order to optimize process potency, advanced features square measure collected as long as nominal by a instruction possibility and if a skip-session flag isn't set (this flag avoids process extra packets once enough packets have already been inspected). whereas we tend to enclosed support for options supported the foremost common classification techniques (port-based, flow-based,
payload-based, etc.), TIE will simply be extended to extract new options supported definitions already printed within the literature or to support new techniques.

In order to speedily experiment with techniques implemented by external tools, this stage will optionally dump for every session the corresponding classification options in conjunction with the label assigned by a classifier (e.g., a payload-based classifier will wont to establish ground truth). TIE supports selling options directly in some common formats, like the arff format utilized by maori hen,10 one in every of the foremost used tools within the field of machine-learning Associate Professor classification.

The fourth stage of the TIE engine consists of a multi-decisional engine fabricated from a choice combiner (hereinafter DC) and one or additional classification plugins (hereinafter classifiers) implementing completely different classification techniques.

The DC is chargeable for classifying sessions by combining multiple classifiers per completely different algorithms, as report-ed in Table 1b. Whenever a brand new packet related to associate degree unclassified session is processed by the feature extractor, if all the classifiers square measureable to be invoked on it session, the DC combines their results per the designed algorithmic program so as to create the ultimate call. A confidence worth between zero and one hundred represents the general reliableness of such a choice [30].

IX. PROPOSED SYSTEM

The development of associate degree open supply tool for traffic classification referred to as Traffic Identification Engine (TIE). TIE has been designed as a community-oriented tool, to produce researchers and practitioners a platform to simply implement (and share) traffic classification techniques, and change their comparison and combination. we tend to finally illustrate some representative use cases of applying TIE to specific analysis problems: scrutiny the accuracy of various classifiers. scrutiny their classification performance. investigation multi-classification and combination methods [29].

A.CORRELATIONANALYSIS

The related flows sharing identical three-tuple area unit generated by identical application. The three-tuple heuristic concerning flow correlation has been thought of in many sensible traffic classification schemes planned a payload primarily based cluster methodology for protocol logical thinking, within which they classified flows into equivalence clusters exploitation the heuristic. Tested the correctness of the three-tuple heuristic with real-world traces.

B.OUTPUTGENERATOR

When TIE is employed to classify traffic, the last stage of the TIE engine is answerable for generating output files containing data concerning the sessions processed and their classification. The output format is exclusive, counters and timestamps linguistics rely on each the operative mode and session kind. once operating in cyclic mode; such output will simply be processed to get live visual reports [30].

C.PRECLASSIFIER

The TIE is employed to coach classifiers, the fourth stage of the TIE engine pre-loads the labels related to every session from a ground-truth file. It are often obtained as output by running TIE on identical traffic trace with a ground-truth classifier [31].

D. call COMBINER

TIE is employed to classify traffic, the fourth stage of the TIE engine consists of a multi-decisional engine product of a call combiner (hereinafter DC) and one or additional classification plugins (hereinafter classifiers) implementing completely different classification techniques .The DC is an swerable for classifying sessions by combining multiple classifiers in keeping with completely different algorithms, as reportable in .Whenever a brand new packet related to associate degree unclassified session is processed by the feature extractor, if all the classifiers area unit able to be invoked thereon session. The DC combines their results in keeping with the organized algorithmic program so as to form the ultimate call [32].

VIII. Conclusion and Future Work

In this article we tend to describe TIE, a platform we tend to started developing in 2008 to assist researchers to tackle unresolved challenges in traffic classification. because of the support of the open supply community and scientific collaborations, the platform has gradually evolved throughout the past 5 years, facultative the production of serious scientific results. within the half-moon of 2014, we tend to attempt to unharness a replacement version of the platform supported feedback and contributions from users collected within the past 2 years. Thereafter, we tend to attempt to extend TIE by:
Investigating the best combination strategy and set of classifiers to come up with reliable ground truth whereas preserving privacy. Extending the support for sharing labeled traffic with anonymized traces.

VIII Acknowledgement

The author would like to thank the Vice Chancellor, Dean-Engineering, Director, Secretary, Correspondent, HOD of Computer Science & Engineering, Dr. K.P. Kaliyamurthie, Bharath University, Chennai for their motivation and constant encouragement. The author would like to specially thank **Dr. A. Kumaravel** for his guidance and for critical review of this manuscript and for his valuable input and fruitful discussions in completing the work and the Faculty Members of Department of Computer Science &Engineering. Also, he takes privilege in extending gratitude to his parents and family members who rendered their support throughout this Research work.

REFERENCES

1. L. Salgarelli, F. Gringoli, and T. Karagiannis, “Comparing Traffic Classi-fiers,” *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 3, 2007, pp. 65–68.
2. A. Dainotti, W. de Donato, and A. Pescapé, “TIE: A Community-Oriented Traffic Classification Platform,” *Traffic Monitoring and Analysis*, Springer, 2009, pp. 64–74.
3. T. T. Nguyen and G. Armitage, “A Survey of Techniques for Internet Traf-fic Classification using Machine Learning,” *IEEE Commun. Surveys & Tuto-rials*, vol. 10, no. 4, 2008, pp. 56–76.
4. T. Karagiannis, K. Papagiannaki, and M. Faloutsos, “Blinc: Multilevel Traf-fic Classification in the Dark,” *ACM SIGCOMM Comp. Commun. Review*, vol. 35, no. 4, 2005, pp. 229–40.
5. L. Bernaille, R. Teixeira, and K. Salamatian, “Early Application Identifica-tion,” *Proc. 2006 ACM CoNEXT Conf.*, 2006, p. 6.
6. G. Aceto *et al.*, “Portload: Taking the Best of Two Worlds in Traffic Classi-fication,” *IEEE INFOCOM Wksps.*, 2010, pp. 1–5.
7. N. Williams, S. Zander, and G. Armitage, “A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow Classification,” *ACM SIGCOMM CCR*, vol. 36, no. 5, Oct. 2006, pp. 7–15.
8. G. Szabó *et al.*, “Traffic Classification over Gbit Speed with Commodity Hardware,” *IEEE J. Commun. Software and Systems*, vol. 5, 2010.
9. A. Callado *et al.*, “Better Network Traffic Identification Through the Inde-pendent Combination of Techniques,” *J. Network and Computer Applica-tions*, vol. 33, no. 4, 2010, pp. 433–46
10. A. Finamore *et al.*, “Experiences of Internet Traffic Monitoring with tstat,” *IEEE Network*, vol. 25, no. 3, 2011, pp. 8–14.
11. F. Gringoli *et al.*, “Mtclass: Enabling Statistical Traffic Classification of Multi-Gigabit Aggregates on Inexpensive Hardware,” *2012 8th Int’l. Wireless Commun. and Mobile Computing Conf.*, 2012, pp. 450–55.
12. A. Este, F. Gringoli, and L. Salgarelli, “On-Line SVM Traffic Classifica-tion,” *2011 7th Int’l. Wireless Communications and Mobile Computing Conf.*, 2011, pp. 1778–83.
13. S. Lee *et al.*, “Netramark: A Network Traffic Classification Benchmark,” *SIGCOMM Comp. Commun. Rev.*, vol. 41, no. 1, Jan. 2011, pp. 22–30.
14. M. Canini *et al.*, “Gtvs: Boosting the Collection of Application Traffic Ground Truth,” *Traffic Monitoring and Analysis*, Springer, 2009, pp. 54–63.
15. F. Gringoli *et al.*, “Gt: Picking Up the Truth from the Ground for Internet Traffic,” *Comp. Commun. Rev.*, vol. 39, no. 5, 2009, pp. 12–18.
16. Dainotti *et al.*, “Classification of Network Traffic via Packet-Level Hid-den Markov Models,” *IEEE GLOBECOM ’08*, 2008, pp. 1–5.
17. M. Crotti *et al.*, “Traffic Classification Through Simple Statistical Finger-printing,” *ACM SIGCOMM CCR*, vol. 37, no. 1, Jan. 2007, pp. 7–16.
18. K.G.S. Venkatesan. Dr. V. Khanna, Dr. A. Chandrasekar, “Autonomous System(AS) for mesh network by using packet transmission & failure detection”, Inter. Journal of Innovative Research in computer & comm. Engineering, Vol. 2, Issue 12, PP. 7289 – 7296, December -2014.
19. K.G.S. Venkatesan and M. Elamurugaselvam, “Design based object oriented Metrics to measure coupling & cohesion”, International journal of Advanced & Innovative Research, Vol. 2, Issue 5, PP. 778 – 785, 2013.

20. Teerawat Issariyakul • Ekram Hoss, "Introduction to Network Simulator NS2".
21. S. Sathish Raja and K.G.S. Venkatesan, "Email spam zombies scrutinizer in email sending network Infrastructures", International journal of Scientific & Engineering Research, Vol. 4, Issue 4, PP. 366 – 373, April 2013.
22. K.G.S. Venkatesan, G. Julin Leeya, G. Dayalin Leena, "Efficient colour image watermarking using factor Entrenching method", International Journal of Advanced Research in computer science & software Engg., Vol. 4, Issue 3, PP. 529 – 538, March – 2014.
23. K.G.S. Venkatesan. Kausik Mondal, Abhishek Kumar, "Enhancement of social network security by Third party application", International Journal of Advanced Research in computer science & software Engg., Vol. 3, Issue 3, PP. 230 – 237, March – 2013.
24. Annapurna Vemparala, Venkatesan.K.G., "Routing Misbehavior detection in MANET'S using an ACK based scheme", International Journal of Advanced & Innovative Research, Vol. 2, Issue 5, PP. 261 – 268, 2013.
25. K.G.S. Venkatesan. Kishore, Mukthar Hussain, "SAT : A Security Architecture in wireless mesh networks", International Journal of Advanced Research in computer science & software Engineering, Vol. 3, Issue 3, PP. 325 – 331, April – 2013.
26. Annapurna Vemparala, Venkatesan.K.G., "A Reputation based scheme for routing misbehavior detection in MANET"S ", International Journal of computer science & Management Research, Vol. 2, Issue 6, June - 2013.
27. K.G.S. Venkatesan, "Planning in FARS by dynamic multipath reconfiguration system failure recovery in wireless mesh network", International Journal of Innovative Research in computer & comm. Engineering, Vol. 2, Issue 8, August - 2014.
28. K.G.S. Venkatesan, AR. Arunachalam, S. Vijayalakshmi, V. Vinotha, "Implementation of optimized cost, Load & service monitoring for grid computing", International Journal of Innovative Research in computer & comm. Engineering, Vol. 3, Issue 2, PP. 864 – 870, February - 2015.
29. R. Karthikeyan, K.G.S. Venkatesan, M.L. Ambikha, S. Asha, "Assist Autism spectrum, Data Acquisition method using Spatio-temporal Model", International Journal of Innovative Research in computer & communication Engineering, Vol. 3, Issue 2, PP. 871 – 877, February - 2015.
30. K.G.S. Venkatesan, B. Sundar Raj, V. Keerthiga, M. Aishwarya, "Transmission of data between sensors by devolved Recognition", International Journal of Innovative Research in computer & comm. Engineering, Vol. 3, Issue 2, PP. 878 – 886, February - 2015.
31. K.G.S. Venkatesan, N.G. Vijitha, R. Karthikeyan, "Secure data transaction in Multi cloud using Two-phase validation", International Journal of Innovative Research in computer & comm. Engineering, Vol. 3, Issue 2, PP. 845 – 853, February - 2015.
32. K.G.S. Venkatesan, "Automatic Detection and control of Malware spread in decentralized peer to peer network", International Journal of Innovative Research in computer & comm. Engineering, Vol. 1, Issue 7, PP. 15157 – 15159, September - 2013.