Profiling IoT-based Botnet Traffic using DNS

Owen P. Dwyer[1], Angelos K. Marnerides[1], Vasileios Giotsas[1], Troy Mursch[2]
[1]Security Lancaster, School of Computing and Communications, Lancaster University, UK
[1][o.dwyer, angelos.marnerides, v.giotsas]@lancaster.ac.uk
[2]Bad Packets LLC, Chicago, IL, USA
[2]troy@badpackets.net

***Abstract*—Internet-wide security and resilience have traditionally been subject to large-scale DDoS attacks initiated by various types of botnets. Since the Mirai outbreak in 2016 myriads of Mirai-alike IoT-based botnets have emerged. Such botnets rely on Mirai's base malware code and they infiltrate vulnerable IoT devices on an Internet-wide scale such as to instrument them to perform large-scale attacks such as DDoS. As recently shown, DDoS attacks triggered by Mirai-alike IoT-based botnets go far beyond traditional pre-2016 DDoS attacks since they have a much higher amplification and their propagation is far more aggressive. Thus, it is of crucial importance to tailor botnet detection schemes accordingly. This work provides a novel DNS-based profiling scheme over real datasets of Mirai-alike botnet activity captured on honeypots that are globally distributed. We firstly discuss features used in profiling botnets in the past and indicate how profiling IoT-based botnets in particular can be improved by leveraging DNS information out of a single DNS record. We further conduct an evaluation of our developed feature set over various Machine Learning (ML) classifiers and demonstrate the applicability of our scheme. Our resulted outputs indicate that the proposed feature set can significantly reduce botnet detection time whilst simultaneously maintaining high levels of accuracy of 99% on average under the random forest formulation.**

I. Introduction

A botnet is a group of infected hosts controlled remotely over the Internet by commands from a *botmaster*, used to facilitate illegal activities including keylogging and identity theft, malware propagation, cryptocurrency mining and, most infamously, distributed denial-of-service (DDoS) attacks. The scale and ambition of botnet activities have increased in recent years, which is particularly concerning when considered alongside the recent phenomenon of the *Internet of Things* (IoT); the extension of internet connectivity beyond home computers and mobile devices, and into embedded systems that have traditionally functioned offline. The growth of the IoT, coupled with the widely reported lack of security in IoT devices [1], [2] is a game-changer for botnets, which can now leverage IoT devices and amplify their attacks to scales orders of magnitude larger than previously possible.

The majority of identified IoT-based botnets exploit variants of the Mirai malware that was firstly observed through the Mirai botnet in October 2016 [3]. The Mirai botnet consisted of $550k$ compromised IoT devices from over 164 countries and initiated a DDoS attack on Dyn's datacenter, one of the major global Domain Name System (DNS) service providers. The resulted attack was recorded as the largest in history and caused the Internet outage to most of the developed countries, including the UK, for more than two days [3]. Evidently, the Mirai attack acted as the cornerstone to a new era of DDoS attacks since a plethora of its malware variants (e.g., IoTroop, Satori, Okiru, Owari) have emerged in the last three years. For instance, the IoTroop malware composes the Reaper botnet that instrumented a series of DDoS attacks on critical financial services in the Netherlands during the first three months of 2018 and is expected to affect more networks in the future. [1]

Whilst a number of previous studies has identified the distinctive transport layer (e.g. TCP or UDP) traffic features of botnets and created methods for their identification, many of these require time-consuming analysis of large datasets or the complex monitoring of entire networks over long periods of time. In parallel, work that aimed to identify botnets using Domain Name Service (DNS) information has shown to be used complementary within traditional botnet classification frameworks (e.g., [4], [5], [6]).

Nonetheless, little has been done in the context of explicitly identifying IoT-based botnets [7]. Within this work, we argue that knowledge gathered in terms of generic features on the transport layer alongside DNS-based information can be extremely useful for composing statistical meta-features within a given Machine Learning (ML) classifier. In contrast with other pieces of work in which multiple DNS records have been used within a clustering or classification process, we show that statistical meta-features composed by a single DNS record are adequate for accurate classification of IoT-based botnets with reasonable computational cost. Therefore, such a scheme can be beneficial for future detection mechanisms as well as for rule-based IP address blacklisting tools as used broadly by network operators. Our work is conducted over real datasets gathered by our globally distributed honeypots that contain infected IP addresses that are already part of Mirai or a Mirai-alike honeypot. In general, the contributions of this work are:

- The first study to profile Mira-alike botnet activity using DNS-based properties over real datasets.
- Highlighting the usefulness of meta-features using a single DNS record from a given flow for Mirai-alike identification.

[1]NexusGuard DDoS Threat Report 2018: https://www.nexusguard.com/threat-report-q2-2018

- Accuracy of more than 99% of average accuracy on detecting Mirai-alike IoT-based botnets using the random forest classifier.
- Computational time of less than 0.1 seconds for composing a concrete model for classifying Mirai-alike botnets.
- Computational time of less than 0.2 seconds to detect if a live DNS domain is Mirai-alike or not.

The remainder of this paper is structured as follows: Section II discusses related work and Section III describes the honeypot datasets used in this work. Section IV discusses the methodology employed in this work whereas Section V presents the results obtained through our evaluation. Finally, section VI concludes and summarises this paper.

II. Related Work

The importance of identifying infected hosts as a component of, for example, firewalls and intrusion detection systems, means that many existing studies have attempted to address this. Many studies ([4], [5], [6]) use the TTL value in DNS records, since a short TTL is advantageous for rapidly changing domain names. Holz et al. [8] argue however that a low TTL is *not* a good indicator of botnet activity, since it carries a risk of misclassifying benign content delivery networks (CDNs) as malicious fast-flux service networks (FFSNs). The way in which botnets are distributed across anonymous systems (ASs) is another feature which has been previously examined [9], but this study suffered from a high false negative rate as a result of relying on too minimal a set of features. Hoang and Nguyen's 2018 study [10] is one of the few that exclusively examines features derived from the text of the domain name, aiming to identify malware through their use of domain generation algorithms (DGAs). By considering the bi- and trigrams present in a given domain, and their relative commonality in natural English text as well as the frequency of vowels in the name, a domain's level of entropy is calculated. Based on these features, classification accuracy as high as 90.2% was achieved, but the applications of this are obviously limited to botnets that use DGAs.

Huang, Mao and Lee [11] present a "snapshot" system for identifying fast-flux networks, based on the principle that they are distributed across many internet service providers, and aims to quantify this with two measures. The first is a measure of distribution relying on time zones to discretise location data, but is ineffective if all hosts are in the time zone and struggles to differentiate FFSNs from benign CDNs. Therefore a second measure is presented, which considers the distance between the different IP addresses in one DNS query. The method claims an accuracy of 98.16% and is not relying on analysing long-term traffic flows.

The prevalence of fast-flux and its characteristic rapid changing of IP addresses means that most existing studies use temporal features based on changes in a server's DNS response over time, which take several hours to observe since the system must wait at least the record's TTL time before it can check for any changes. This limits practical applications and potential for real-time detection. However it has also been proven that it is possible to classify botnets based on information from a single point in time, based on its spatial distribution. The remainder of this study will aim to build upon this and establish a novel snapshot-based system that goes beyond currently deployed or proposed solutions since it incorporates non-spatial features and examines their relevance to the Mirai botnet specifically. Moreover, the utilisation of minimal DNS-based features out of a single record is an approach that reduces significantly the feature gathering process, thus establishing promising paths for future deployments of anomaly-based IDSs as well as next generation IP blacklisting tools.

III. Data Description

We gather probes from IP addresses that are part of a Mirai-alike botnet by operating 11 SSH and Telnet honeypots located in three countries. Our honeypots are located in: the United States (3 in Las Vegas, Nevada, 1 in Minden, Nevada, 3 in Los Angeles, California), Russia (2 in Moscow), and Brazil (2 in Sao Paulo). Each honeypot is configured to capture logs of all incoming traffic, and logfiles are then aggregated and indexed using Splunk for analysis. We subsequently match Mirai-alike fingerprints by comparing the TCP sequence with the IP addresses. In particular, Mirai bots send TCP SYN packets with the TCP initial sequence number equal to the destination IP of the targeted host [3]. Given that the TCP sequence number is a 32-bit integer, the likelihood of an identified Mirai-alike fingerprint being set at random is only $\frac{1}{2^{32}}$. Based on this technique, our honeypots have detected 811,636 Mirai-alike probes between 2017/02/17 – 2019/03/07.

From this dataset, IP addresses were randomly sampled and reverse queried to obtain domain names, of which a subset of 25% returned responses. These domain names were then forward queried to give us a set of DNS records associated with Mirai hosts. Within our methodology, the responsive domains are compared against a set of benign DNS records, obtained by querying the API [2]. The Majestic Million list ranks websites with the highest number of referring subnets and we consider these entries to be benign.

We following provide an insightful assessment of various DNS-based features from our datasets such as to describe their statistical profiles and pinpoint Mirai-alike characteristics.

IV. Methodology

A. Feature composition

1) DNS response: The answer to a DNS query, i.e. the IP address mapped to a given domain name, is contained in the packet's "Answer" section. In addition to this there is also an "Authority" section, which contains records pointing to the domain's authoritative nameservers, and an "Additional" section which can contain any record to aid query resolution, but in practice almost exclusively contains full records for the nameservers in the Authority section. Thus, a single response message will contain information on a domain's nameservers

[2]Majestic Million list: https://majestic.com/reports/majestic-million

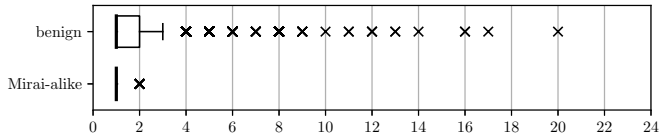

Fig. (1) Boxplot - number of records in the DNS Answer section

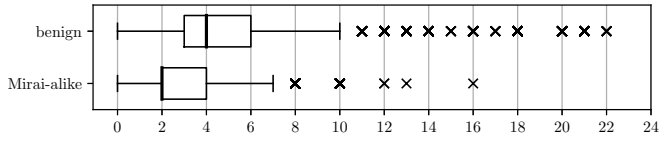

Fig. (2) Boxplot - number of records in the DNS Additional section

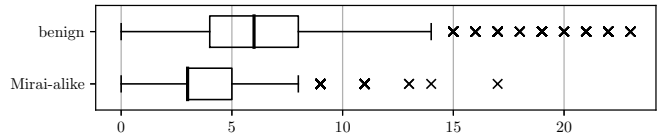

Fig. (3) Boxplot- number of ASes linked to a domain

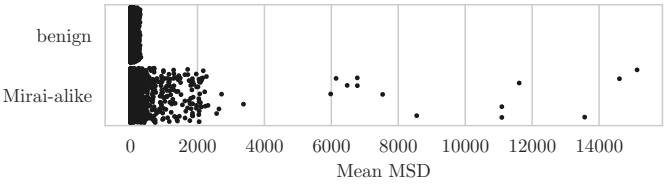

Fig. (4) Distribution of the mean MSD for each domain by classification

without the need to run additional queries. Previous studies [8], [6] have indicated that fast-flux domains on average return a higher number of records than a benign domain, making record count an easy to compute and useful indicator of botnet activity. However, this pattern was not present in Mirai-alike botnet activity, with Mirai-alike bots returning on average the same number of "Answer" records (fig. 1) and fewer additional records (fig. 2) than a benign domain. However, we did observe that Mirai-alike bots almost exclusively return one "Answer" record, forming a distinctive pattern which can be used in detection mechanisms and IP blacklisting tools.

2) TTL: Time to live (TTL) is an integer value included in every response record (RR), specifying the time interval for which that record should be cached in a server. After the period has passed, the server must pass any queries up the DNS hierarchy for resolution. For a botnet, especially one using fast-flux techniques, lower TTL values are preferable as this allows the botnet to evade detection and react more quickly to hosts being taken down. Stalmans and Irwin [6] found a significant difference in TTL between fast-flux and normal domains, with benign domains having an average TTL of 14885 and fast-flux domains only 595. In our datasetwe identified that the maximum TTL across all IP addresses in a single malicious record was on average 87,854, compared to 107,585 for benign domains. Thus, confirming that Mirai-alike bots tend to have lower TTLs.

As previously mentioned, TTL can be an unreliable feature for identifying botnets, since legitimate content delivery networks (CDNs) will also have low TTLs. Nevertheless, the classifiers discussed later incorporate TTL as one of the many features contributing to classification. Hence, we reduce the risk of CDN misclassification.

3) AS diversity: An autonomous system (AS) is a collection of IP routing prefixes under the control of a single administrative entity, sharing a common routing policy. A benign domain is likely to have most, if not all, of its servers within the same AS for practical reasons including the costs of using a particular ISP, or the convenience of keep the hosts geographically adjacent in one data centre. This pattern has been observed in a number of different botnets [6], [9]. Additionally Holz et al. [8] observe that legitimate domains and CDNs both tend to return only one ASN, making AS diversity a potentially useful feature in preventing misclassification of CDNs as botnets and countering any anomalies created by using TTL as a feature. However, the Mirai botnet does not share this pattern (Fig. 3), with hosts spanning a lower number of ASs than a benign domain on average.

4) Spatial distribution: Legitimate businesses will commonly cluster their servers closer together for cost reasons, whilst botnets by their nature will indiscriminately hijack anything with a vulnerability, resulting in a greater dispersion. The *spatial service relationship estimator* [11] takes advantage of this by measuring *service relationship*, i.e. the relationship between an IP address (i.e., the *consumer*) and a nameserver (i.e., the *provider*). For each address in the answer section, the minimum service distance (MSD) $d_{mm'}$ is calculated as the shortest distance between that address $q_m \in Q_{AS}$ and any address in the additional section $q_{m'} \in Q_{NS}$, based on the Euclidean distance between the two host's geographic coordinates: $d_{mm'} = \sqrt{(\text{lat}(q_m) - \text{lat}(q_{m'}))^2 + (\text{long}(q_m) - \text{long}(q_{m'}))^2}$ In this study however we use the Haversine distance formula , which is more suitable for data on a global scale, as it eliminates distortion at northern and southern extremities and allows latitudes to "wrap around" from -180° to +180°. For a malicious domain, the mean MSD was 355.5km, compared to 18.7 for a benign domain (fig. 4), confirming that Mirai-alike bots are more widely distributed than benign domains.

5) Vowel density: In natural English text, around 38.1% of letters are vowels [12], despite them only representing 19.2% of letters. Therefore the proportion of letters in a domain name may provide an indicator of how likely that domain is randomly generated. In our dataset, the domain names of Mirai-alike bots had an average vowel density (vowel count

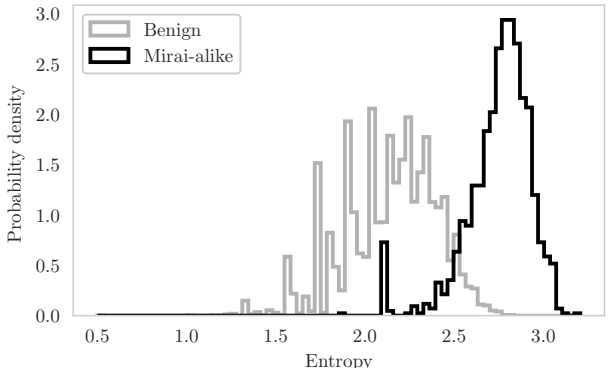

Fig. (5) Entropy distribution of domain names

/ length) of 0.16, compared to 0.32 for benign domains. This roughly corresponds to the percentages mentioned above, confirming that Mirai-alike botnets use easily detectable domain generation algorithms (DGAs).

6) Shannon entropy: We use the Shannon information entropy of a variable as a way to effectively quantify "randomness" being the negative logarithm of that value's probability mass function:

$$H(X) = -\sum_{i \in X} P(i)\ log_2\ P(i)$$

This is relevant when considering botnets since the domain names used by botnets are often randomly generated. Therefore botnet-related domains contain combinations of characters with entropy distinct from what would be expected from the English language. Mirai-alike domains were found to have on average a higher entropy than benign domains, as illustrated by the probability distribution in Fig. 5.

B. Feature engineering: system method

When optimising our system, input and output (I/O) is a concern, since this has the potential to slow down operation by orders of magnitude. DNS queries are particularly time-consuming, especially since a DNS server often has to query multiple other servers before it can respond. This is additionally limited by the bandwidth of the available bandwidth. Whilst control over this is limited, we can speed up database access time. To map IP addresses to ASNs and coordinates, we use the publicly available MaxMind GeoLite 2 databases[3]. The database file format is effectively a binary search tree, giving it $O(\log n)$ search complexity, and the API loads the entire database into memory once at the beginning of the process, minimising access times as far as possible.

C. IoT-based Botnet classification

In this study we investigated the performance of two supervised and one unsupervised classification algorithm:

[3]MaxMind GeoLite 2: https://dev.maxmind.com/geoip/geoip2/geolite2/

1) Naïve Bayes classifier: uses Bayes' theorem to calculate the most probable class of a variable under the "naïve" assumption that each of its attributes is conditionally independent. Whilst this is rarely true is most real-world scenarios, it significantly reduces the complexity of computation. Naïve Bayesian methods are expected to perform well on our dataset since we will be considering features from the domain name text, record count, and spatial distribution, all unrelated factors, meaning that the naïve assumption may actually be true and provide a more realistic model of the data, leading to more accurate classification. Within our evaluation we explore three different implementations of the Naïve Bayes classifier relating the the a-priori considered statistical distribution. Namely we assess the Gaussian, multinomial and Bernoulli distributions.

2) Random forest classifier: an ensemble classifier, i.e. it aggregates the results of multiple independent classifiers to create an overall model, increasing accuracy. Each individual decision tree classifier only uses a random subset of features as its input, and a majority vote determines the overall classification, meaning that over half need to be incorrect for the overall classification to be wrong. Random forests perform poorly where there is a large number of irrelevant variables, since this lowers the chance of a relevant variable being selected. A major advantage of this approach is the ability to see the effect of each individual feature on the final classification, and determine the importance of each feature.

3) k-Nearest Neighbours (k-NN) classifier: an example of unsupervised *instance-based learning*, since it makes no attempt to construct a model of the data or understand its underlying structure, instead simply storing each data point, thus exploiting a priori knowledge. Classification is based on a simple majority vote of a predefined number k of the closest pre-computed samples. The lack of any attempt to infer underlying structure has the potential to produce fast performance, and the majority voting mechanism suggests that accuracy will be similar to random forests, although it remains to be seen how well it will perform given the dimensionality of our data set.

4) Classification performance metrics: In order to assess the accuracy performance of the utilised classifiers we employ standard metrics in terms of the confusion matrix and operate over True Positives (TP), False Positives (FP), False Negatives (FN) and True Negatives (TN). Thus we employ:

- *Precision*, the proportion of detected malicious domains that actually were botnet domains.

$$Precision = \frac{TP}{TP + FP}$$

- *Recall*, representing the percentage of truly malicious domains correctly detected.

$$Recall = \frac{TP}{TP + FN}$$

- *F-score*, that is the harmonic mean of precision and recall such as to ensure greater capture of accuracy in our Mirai-alike classification process.

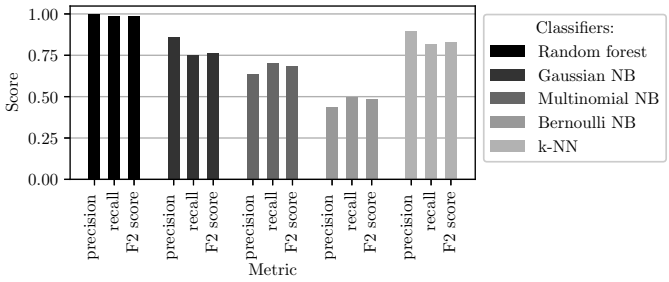

Fig. (6) Classification performance on detecting Mirai-alike hosts over 5 different classifiers.

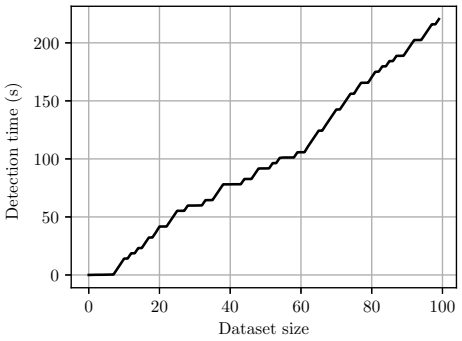

Fig. (7) Dataset size vs. detection time

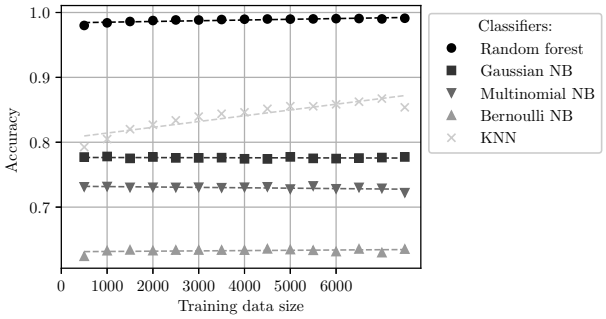

Fig. (8) Classification accuracy vs. training dataset size

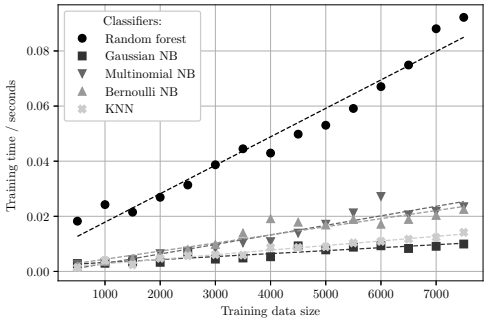

Fig. (9) Training time vs. training dataset size

$$F - score = 2.\frac{Recall.Precision}{Recall + Precision}$$

Alongside accuracy we also assess the importance of the features identified within ths section. In order to achieve that we employ the Gini Index. The Gini Index, G, is defined as:

$$G = \sum_{k=1}^{K} p_{mk}(1 - p_{mk})$$

where K is the total number of classes, and p_{mk} is the proportion of training data in the m^{th} region of assessment that belong to the k^{th} class.

V. Evaluation

As already presented earlier(Section IV, we consider a number of features that can be used to identify botnets, and how these apply to our Mirai-alike dataset. We subsequently assess the resulted feature vector with classification algorithms such as to evaluate the feasibility of an ML-based detection system. Although systems for botnet classification have been created in the past with success, they often require significant amounts of data and time-consuming analysis to complete. Whilst at least one previous study [11] has achieved detection times of half of a second, it relies entirely on geographic features with questionable applicability for IoT-based, Mirai-alike botnets[7]. Our aim therefore is to increase credibility using a wider range of features, whilst simultaneously maintaining rapid detection times explicitly for IoT-based botnets.

A. Classification performance

We assessed each of the classifiers over varying tuning phases within their algorithmic configuration and examined their performance in terms of accuracy, and computational cost. Figure 6 depicts how each classifier performed on average over the three classification accuracy metrics defined in section IV. As evident, we see the random forest formulation to be far more accurate than the 3 Bayesian-based classifiers and the K-means Nearest Neighbour (k-NN) clustering approach. Under varying model parameters (e.g., number of decision trees) the random forest achieved an average of 99% for precision and 98% for recall and F-score. We also witness that the best formulation for a Bayesian-based classifier is essentially the one that considers a Gaussian distribution with 80%, 75%, 77% for average precision, recall and F-score respectively. In parallel, the k-NN algorithm outperforms all three Bayesian-based classifiers with 90%, 78%, 80% for al the three aforementioned metrics. Therefore, the k-NN approach is much more capable of identifying precisely the proportion of detected malicious domains that are part of a greater botnet than any Bayesian-based classifier (i.e. precision).

Nonetheless, Figure 8 shows how accuracy varies according to the training set size. The three Bayesian methods demonstrate relatively consistent performance, k-NN improves noticeable with additional training data, whilst random forests improve slightly. The random forest method has a clear

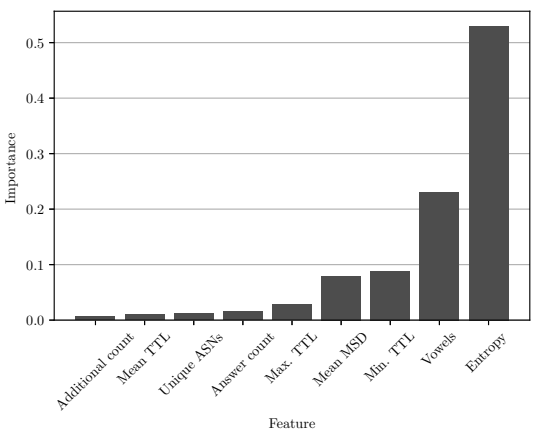

Fig. (10) Average feature importance in detecting Mirai-alike bots using the G Index.

advantage as it not only performs consistently well against all metrics, it performs well on smaller datasets. Thus, we consider random forests to be more practical for real-time implementations where limited amount of data is likely to be available particularly for medium-scale IP blacklisting tools or edge detection methods within small enterprise networks or medium-sized ISPs.

B. Computational performance

The relationship between a dataset's size and overall detection time (i.e. the entire process of reverse querying a given IP address, querying the domain name, extracting and preparing the features, and classifying the domain) is shown in figure 7 . A dataset of 100 domains takes around 200 seconds to analyse, giving an average detection time of 2 seconds. However, the historic nature of this dataset meant that a large proportion of the requests timed out since those hosts were no longer active. The average detection time excluding these timeouts was 0.17 seconds. In addition, a live implementation of such a scheme, for example in an intrusion detection system or an IP blacklisting tool, might extract a domain name from a packet flow in real time, necessitating only one DNS query and further reducing this time. Figure 9 shows the time taken to train different classifiers. Whilst the random forest algorithm was significantly slower than other approaches, training time was still less than 0.1 seconds, hence unlikely to be a limiting factor in any implementations of this scheme.

C. Feature importance

By utilising the Gini Index we assess the usefulness of the identified features used subsequently within our classification process. As evidenced by Fig. 10, vowel density and entropy were the most important features and demonstrates that the usage of DGAs by Mirai-alike bots, easily sets their profile apart from benign traffic.

VI. Conclusion

In this study we described a new and novel feature set for detecting Mirai-alike botnet activity through DNS, with the distinction that it can be entirely derived from the contents of a single DNS query. We conducted our evaluation over real honeypot datasets and witnessed that Mirai-alike activity tends to differ significantly from benign traffic. Mirai-alike bots tend to span fewer ASes and that an entropy-based description of domain names is a distinct property due to their use of DGAs. We assessed our proposed feature-set over a number of machine learning techniques, and witnessed that the random forest classifier provides the best accuracy performance with 99% of accuracy. In parallel, the exceptionally low computational cost of this technique places it as the best candidate for future implementation of intrusion detection systems as well as IP blacklisting tools.

VII. Acknowledgements

This work was supported by the EU's H2020 Enable Ancillary Services by Renewable Energy Sources project (EASY-RES - agreement No.764090). We also thank Bad Packets LLC for providing the datasets used as well as useful insights.

References

[1] M. A. Khan and K. Salah, "IoT security: Review, blockchain solutions, and open challenges," *Future Generation Computer Systems*, vol. 82, pp. 395 – 411, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X17315765

[2] M. Hypponen and L. Nyman, "The Internet of (Vulnerable) Things: On Hypponen's Law, Security Engineering, and IoT Legislation," *Technology Innovation Management Review*, vol. 7, pp. 5–11, Apr. 2017. [Online]. Available: http://timreview.ca/article/1066

[3] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the mirai botnet," in *Proceedings of the 26th USENIX Conference on Security Symposium*, ser. SEC'17, 2017, pp. 1093–1110.

[4] R. Perdisci, I. Corona, D. Dagon, and W. Lee, "Detecting Malicious Flux Service Networks through Passive Analysis of Recursive DNS Traces," in *2009 Annual Computer Security Applications Conference*, Dec. 2009, pp. 311–320.

[5] L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, "Exposure: A Passive DNS Analysis Service to Detect and Report Malicious Domains," *ACM Trans. Inf. Syst. Secur.*, vol. 16, no. 4, pp. 14:1–14:28, Apr. 2014. [Online]. Available: http://doi.acm.org/10.1145/2584679

[6] E. Stalmans and B. Irwin, "A framework for DNS based detection and mitigation of malware infections on a network," in *2011 Information Security for South Africa*, Johannesburg, South Africa, Aug. 2011, pp. 1–8.

[7] K. Angrishi, "Turning internet of things(iot) into internet of vulnerabilities (iov) : Iot botnets," *CoRR*, vol. abs/1702.03681, 2017. [Online]. Available: http://arxiv.org/abs/1702.03681

[8] T. Holz, C. Gorecki, K. Rieck, and F. C. Freiling, "Measuring and Detecting Fast-Flux Service Networks," in *Proceedings of the 15th Network & Distributed System Security Symposium*, 2008.

[9] M. Wielogorska and D. O'Brien, "DNS Traffic Analysis for Botnet Detection," in *Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science*, Dublin, Ireland, Dec. 2017, pp. 261–271.

[10] X. D. Hoang and Q. C. Nguyen, "Botnet Detection Based On Machine Learning Techniques Using DNS Query Data," *Future Internet*, vol. 10, p. 43, 2018.

[11] S.-Y. Huang, C.-H. Mao, and H.-M. Lee, "Fast-flux Service Network Detection Based on Spatial Snapshot Mechanism for Delay-free Detection," in *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '10. New York, NY, USA: ACM, 2010, pp. 101–111. [Online]. Available: http://doi.acm.org/10.1145/1755688.1755702

[12] P. Norvig, "English Letter Frequency Counts," 2012. [Online]. Available: http://norvig.com/mayzner.html