Digital Investigation 20 (2017) S47–S55

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

DFRWS 2017 Europe — Proceedings of the Fourth Annual DFRWS Europe

Behavioral Service Graphs: A formal data-driven approach for prompt investigation of enterprise and internet-wide infections

Elias Bou-Harb [a, *], Mark Scanlon [b]

[a] Cyber Threat Intelligence Laboratory, Florida Atlantic University, USA
[b] School of Computer Science, University College Dublin, Ireland

ARTICLE INFO

Article history:
Received 31 January 2017
Accepted 31 January 2017

Keywords:
Probing
Infections
Graphs
Threat modeling
Data analytics
Network forensics

ABSTRACT

The task of generating network-based evidence to support network forensic investigation is becoming increasingly prominent. Undoubtedly, such evidence is significantly imperative as it not only can be used to diagnose and respond to various network-related issues (i.e., performance bottlenecks, routing issues, etc.) but more importantly, can be leveraged to infer and further investigate network security intrusions and infections. In this context, this paper proposes a proactive approach that aims at generating accurate and actionable network-based evidence related to groups of compromised network machines (i.e., campaigns). The approach is envisioned to guide investigators to promptly pinpoint such malicious groups for possible immediate mitigation as well as empowering network and digital forensic specialists to further examine those machines using auxiliary collected data or extracted digital artifacts. On one hand, the promptness of the approach is successfully achieved by monitoring and correlating perceived probing activities, which are typically the very first signs of an infection or misdemeanors. On the other hand, the generated evidence is accurate as it is based on an anomaly inference that fuses data behavioral analytics in conjunction with formal graph theoretic concepts. We evaluate the proposed approach in two deployment scenarios, namely, as an enterprise edge engine and as a global capability in a security operations center model. The empirical evaluation that employs 10 GB of real botnet traffic and 80 GB of real darknet traffic indeed demonstrates the accuracy, effectiveness and simplicity of the generated network-based evidence.

Introduction

Undeniably, network forensics presents a rich problem space that typically deals with the collection, preservation, analysis and presentation of network-based knowledge. It is often exploited to generate actionable insights and intelligence that could be effectively leveraged by investigators. The latter is especially factual when attempting to fingerprint, assess and mitigate network security intrusions and misdemeanors. However, this attempt is recurrently hindered by various current technical challenges that face network forensics. First, network forensic analysts are significantly overwhelmed by huge amounts of low quality evidence. Such evidence is often generated from intrusion detection systems that are known to suffer from elevated levels of both false positives and negatives (Garcia-Teodoro et al., 2009), rendering the combined task of identifying relevant information and attributing the true malicious entity extremely challenging, if not impossible. Second, most network forensic approaches are passive or reactive, employ manual ad-hoc methods and are time consuming (Pilli et al., 2010; Adeyemi et al., 2013). This makes the generated evidence relatively obsolete to be acted upon in a timely manner and most certainly decreases its reliability and wastes valuable resources. Third, contemporary cyber attacks are getting more sophisticated than ever and continue to operate in an excessively coordinated and distributed manner. To this end, network forensic science is relatively lagging behind such advancement in the attacks. Further, most current network forensic practices do not support distributed inference, and if they do, they force the analysts to go through an error-prone, agonizing process of correlating dispersed unstructured evidence to infer a specific security incident.

Indeed, local and Internet-scale networks have been increasingly getting abused by various modernized attacks, including, distributed denial of service attacks (Fu et al., 2012), amplification

* Corresponding author.
E-mail address: ebouharb@fau.edu (E. Bou-Harb).

http://dx.doi.org/10.1016/j.diin.2017.02.002

attempts (Kührer et al., 2014), spamming (Xie et al., 2008) and advanced persistent threats (Daly, 2009). Such attacks are almost always being generated by groups of infected and distributed machines controlled by an external entity (Silva et al., 2013). In this paper, we refer to the latter orchestrated malicious groups as 'campaigns'. An effective approach to generate network forensic insights and inferences related to those campaigns is to analyze their generated probing activities. Such activities refer to reconnaissance techniques that are typically employed by those campaigns to obtain information about their targets prior to launching their targeted attacks (Allman et al., 2007). In fact, Panjwani et al. (2008) concluded that around 50% of attacks are indeed preceded by some form of probing activity. Additionally, such activity has been reported in numerous occasions as a concrete evidence of infection (Wang et al., 2014a; Whyte et al., 2006).

In essence, the presented research and development work attempts to answer the following question: *How can we design an approach that is able to effectively process, analyze and correlate large volumes of network traffic to generate, in a very prompt manner, formal, highly-accurate and actionable network forensic evidence that could be leveraged to infer infected campaigns?*

This paper attempts to answer this question. Specifically, the core contributions of this paper could be summarized in the following:

- Proposing a set of data behavioral analytics that scrutinize perceived probing activities to capture their various obscured features (i.e., machinery, strategies, natures, etc.). The analytics uniquely employ numerous statistical and entropy-based techniques to effectively generate feature vectors related to the infected probing sources or hosts.
- Presenting *Behavioral Service Graphs*, a novel approach that aims at providing investigators/analysts, network administrators and/or security operators with network forensic evidence related to infected machines within a constructed campaign. The approach models the probing sources that show evidence of infection as graphs. By exploiting ancillary graph theoretic concepts such as the maximum spanning tree and Erdős-Rényi random graphs, the approach is able to infer and correlate such distributed groups of infected machines. The approach is prompt since (1) it exploits probing activities to rapidly infer infections and (2) the inferred group of infected machines possesses the minimum number of members to formally claim that such group is indeed a malicious campaign. The latter idea is especially imperative as this will allow actionable thwarting of campaigns as soon as there exists evidence of their construction.
- Empirically evaluating the proposed approach using two real and significant datasets under two different deployments scenarios. The output concurs that the extracted inferences exhibit noteworthy accuracy and can generate significant, accurate and formal forensic insights that could be used for prompt mitigation and to facilitate further focused analysis.

The road-map of this paper is as follows. In the next section, we elaborate on the proposed approach. Specifically, we disclose the data preprocessing step, the employment of the data behavioral analytics, the rationale and construction of Behavioral Service Graphs and detail how they can be exploited to achieve the intended goals. In Section Empirical evaluation, we empirically evaluate the proposed approach and verity its accuracy and insights. We provide a discussion related to the approach, its limitations and possible improvements in Section Proposed approach: limitations & possible improvements. In Section Related work, we survey the related work on various concerned topics. Finally, Section Concluding remarks summarizes the goals, the methods and the results of the proposed approach and paves the way for future work that aims at providing extended network-based evidence to further support investigations.

Proposed approach

In this section, we describe and detail the rationale and employed steps of the proposed approach. In a nutshell, the proposed approach (1) fingerprints and extracts probing activities from perceived network traffic, (2) applies the proposed behavioral analytics to generate feature vectors related to the infected probing sources, (3) constructs Behavioral Service Graphs that model those probing machines, and (4) manipulates such graphs to infer distributed campaigns possessing minimum members of infected machines. The latter four steps are detailed next.

Fingerprinting probing activities

Motivated by the fact that probing activities precede a plethora of attacks (Allman et al., 2007; Panjwani et al., 2008) coupled with the rationale that such activities are the very first signs of any infection (Wang et al., 2014a; Whyte et al., 2006), the proposed approach leverages the latter to extract probing activities generated from infected machines. The intrusion detection system community provides extensive techniques on how to accomplish this task (Bhuyan et al., 2011). In this work, to successfully and accurately fingerprint probing activities, we leverage the work by Staniford et al. (2002) and cross match the output, for validation purposes, by using two open-source detection systems, namely, Snort (Roesch et al, 1999) and Bro (Paxson, 1999). We have selected to employ the latter three approaches as they are the de-facto standards when it comes to probing detection, possess the capability to operate in real-time, and have been extensively and repetitively evaluated and validated. The output of this step is probing traffic, generated from unique sources, coupled with their network sessions that have been saved in packet capture (.pcap) format for further analysis.

Data behavioral analytics

In order to capture the behaviors of the inferred probing sources, we propose to employ the following set of behavioral analytics. This aims at generating the feature vectors of the infected probing machines to be employed as input for the subsequent steps. The proposed approach takes as input the previously extracted probing sessions and outputs a series of behavioral characteristics related to the probing sources. In what follows, we pinpoint the concerned questions and subsequently present the undertaken approach in an attempt to answer those.

Is the probing traffic random or does it follow a certain pattern?

When sources execute their probing traffic, it is imperative to infer and capture the fashion in which they achieve their goal. To realize this task, we proceed as follows. For each unique pair of hosts extracted from the probing sessions (probing source to target), we test for randomness of their inter-arrival times in the traffic using the non-parametric Wald-Wolfowitz statistic test. If the outcome is positive, we record it for that precise probing source and apply the test for the remaining probing sessions. If the result is negative, we conclude that the generated traffic follows a certain pattern. To deduce the particular employed pattern, we model the probing traffic as a Poisson distribution and capture the maximum likelihood estimates for the Poisson parameter λ that corresponds to that traffic, at a 95% confidence level. The choice to model the traffic as a Poisson process is motivated by our previous work (Bou-

Harb et al., 2016), where we have noticed that probe arrivals is consistent with that distribution. Please note that we are not particularly interested in the derived pattern values; we only employ them to characterize the probing traffic to build the feature vectors of the probing sources.

How are the targets being probed?

As shown in Dainotti et al. (2012), coordinated probing sources adopt numerous strategies when probing their targets. These strategies could include IP-sequential, reverse IP-sequential, uniform permutation or other types of permutations. In an effort to infer the probing strategies, we execute the following. For each probing source, we retrieve its corresponding distribution of target IP addresses. To distinguish between sequential and permutation probing, we leverage the Mann-Kendall statistic test, a non-parametric hypothesis testing approach, to check for monotonicity in those distributions. The rationale behind the monotonicity test is that sequential probing will indeed induce a monotonic signal in the distribution of target IP addresses while permutation probing will not. Moreover, in this work, we set the significance level to 0.5% since an elevated value could introduce false positives. To discriminate between (forward) IP-sequential and reverse IP-sequential, for those distributions that tested positive for monotonicity, we also take note of the slope of the distribution; a positive slope renders a forward IP-sequential strategy while a negative one defines a reverse IP-sequential strategy. For those distributions that tested negative for monotonicity (i.e., not a sequential strategy), we apply the chi-square goodness-of-fit statistic test (Anderson and Darling, 1954). The latter insight will notify us whether or not the employed strategy is a uniform permutation; if the test returns a negative output, then the employed strategy will be deemed as a permutation; uniform permutation otherwise.

What is the nature of the probing source?

It is of momentous importance as well to infer the nature of the probing source; is it a probing tool or a probing bot. In this work, we are predominantly interested when the sources are bots as this will provide more concrete evidence of infection. From the two preceding questions, we can deduce those probing events that are random and monotonic. It is known that monotonic probing is a behavior of probing tools in which the latter sequentially probe their targets (IP addresses and ports). Additionally, for random events (i.e., events that do not disclose the use of certain patterns in their *inter-arrival times*), the monotonic trend checking would aid in filtering out traffic caused by non-bot scanners (Li et al, 2011). Hence, we deem a probing source as employing a probing bot *only* if their traffic possesses pattern usage *and* if they adopt a probing approach other than sequential probing (i.e., including reverse IP-sequential); a probing tool otherwise. To this end, we acknowledge that this problem of classifying the nature of the probing source is indeed challenging. Future work will attempt to further fortify the extracted evidence from our employed heuristic method by investigating the correlation between the perceived probing traffic and probing traffic extracted from malware samples.

Is the probing targeted or dispersed?

When sources probe their targets, it would be also beneficial to infer whether their probing traffic is targeted towards a small set of IP addresses or dispersed. In an attempt to answer this, for each probing source b, we denote GF(b) as the collection of flows generated by that particular source. The destination target IP addresses used by the flows in GF(b) induce an empirical distribution. Consequently, we adopt the concept of relative uncertainty (Xu et al., 2005), an information theoretic metric and execute it on those distributions. The latter index is a conclusive metric of variety, randomness or uniformity in a distribution, regardless of the sample size. A result that is close to 0 points out that the probing source is employing a targeted approach while an outcome value close to 1 means that its corresponding probing traffic is dispersed.

Miscellaneous inferences

For each probing bot, we also record its rate (packets/second), its ratio of destination overlaps defined as $r = nc/nt$ where nc defines the number of common sessions between all the sources and nt is total number of all probing sessions, and its target ports.

It is evident that the latter set of behavioral analytics significantly rely on various statistical tests and methods to uncover the behavior of the probing sources. We emphasize that such approach is arguably more sound than heuristics or randomly set thresholds. Further, it is noteworthy to indicate that all the employed statistical tests assume that the data is drawn from the same distribution. Since the approach operates on one type of data, namely, network data extracted from a certain network topology, we can safely presume that the values follow and are in fact drawn from the same distribution.

Behavioral Service Graphs

We model the probing machines that show signs of infection (i.e., those inferred as bots using the behavioral analytics) coupled with their feature vectors using what we refer to as Behavioral Service Graphs. Such graphs are of the form $G = (N, E)$ where N represents the set of infected probing sources/machines (i.e., nodes) and E characterizes the edges between such nodes. It is worthy to mention that G is an undirected complete graph (Díaz et al., 2002), with weights on the edges representing the probability of behavioral similarity (P_{bs}) computed by piecewise comparisons between the previously inferred feature vectors of each of the nodes.

$$\text{Bot 1:} \begin{bmatrix} \texttt{Random Traffic} \\ \texttt{Sequential Probing} \\ \texttt{Dispersed Probing} \\ \texttt{Rate: 60 pps} \\ \texttt{Destinations Overlap: 100} \\ \texttt{Port Number: 80} \end{bmatrix}$$

$$\text{Bot 2:} \begin{bmatrix} \texttt{Pattern in Traffic} \\ \texttt{Sequential Probing} \\ \texttt{Targeted Probing} \\ \texttt{Rate: 55 pps} \\ \texttt{Destinations Overlap: 200} \\ \texttt{Port Number: 80} \end{bmatrix}$$

To clarify how P_{bs} is computed, consider the above two feature vectors that capture the behavior of two distinct bots. By performing binary comparisons between each corresponding pair of

features of those bots, one can note that the similarity is 3/6 or 50%. Please note that for the rate and destinations overlap features, we consider a conservative 15% as being similar. It is also important to mention that we generate different complete graphs for different targeted port numbers that cluster a number of inferred bots. This will indeed provide the capability to identify infected machines participating in each unique campaign, given that there are multiple, simultaneously active different campaigns. Therefore, in essence, each constructed graph is actually modeling infected machines, their behavioral similarity and what specific network service is being probed. This aims at providing the investigator with additional inference about the activities of the current infections and to warn about possible future attacks that could specifically abuse that service.

In summary, Behavioral Service Graphs allow the prompt inference of bot infected machines by solely analyzing their probing activities. Further, they extend such inferences to automate the amalgamation of evidence from distributed entities, as well as providing auxiliary valuable insights related to the behaviors of the infected machines and their possible intended actions.

Friends of the enemy stay closely connected: inferring infected campaigns

Campaigns of infected machines could be distinguished from other security incidents as (1) the population of the participating bots is several orders of magnitude larger, (2) the target scope is generally the entire Internet Protocol (IP) address space, and (3) the bots adopt well orchestrated, often botmaster-coordinated, stealth strategies that maximize targets coverage while minimizing redundancy and overlap (Dainotti et al., 2015). In the context of an enterprise network, such campaigns not only hinder the legitimate users' overall experience and productivity, but also jeopardize the entire cyber security of the enterprise (i.e., causing vulnerabilities or opening backdoors in the internal network). Further, they significantly degrade the provided quality of service since the compromised machines will most often cause an excessive increase in bandwidth utilization that could be rendered by extreme peer to peer usage, spamming, command-and-control communications and malicious Internet downloads. Additionally, if the enterprise network would be used to trigger, for instance, a malware-orchestrated spamming campaign, then such enterprise could as well encounter serious legal issues for misusing its infrastructure (i.e., for example, under the US CAN SPAM Act (Parliament of Canada; Governement of the USA)). Consequently, this will immensely adversely affect the enterprise's business, reliability and reputation.

To this end, forensic investigators of enterprise networks are interested in possessing a capability that aims at inferring such campaigns of infected machines. However, one crucial requirement of such capability would be its promptness in deeming a group of infected machines as a campaign. In other words, the undertaken approach would be required to provide tangible evidence related to the *minimum* number of infected machines that compose a campaign. Indeed, this would generate actionable evidence that could be exploited to promptly thwart the expansion of the campaign and thus would significantly limit the sustained possible collateral damage and any symptoms of infection. We next elaborate on such an approach.

Previous work (Rajab et al., 2006) demonstrated that coordinated bots within a campaign probe their targets in a similar fashion. Indeed, Behavioral Service Graphs were initially engineered to naturally and intuitively support the latter; they cluster the infected machines targeting the same service and they combine their feature vectors (and their similarly probability) for further analysis. The proposed approach executes two steps to retrieve the minimum number of infected machines to deem a group of such machines as a campaign.

First, given a complete Behavioral Service Graph $G = (N, E)$, the approach extracts a subgraph $G' = (N', E')$ where $N' = N$ and $E' \subseteq E$. This aims at reducing the number of edges while maximizing the behavior probability between the infected machines (i.e., nodes). To achieve this task, we employ the graph theoretic concept of a maximum spanning tree (Ozeki and Yamashita, 2011) by implementing a slightly modified version of Kruskal's algorithm (Kruskal, 1956). Although there exists a plethora of approaches for the creation of maximum spanning trees, the latter algorithm was the basis of many and is abundantly available in numerous tool sets. Second, to understand the structure of the subgraph formed by members of a botnet on the complete graph, suppose that there are m bots, thus forming a graph with m corresponding nodes. Let the set $X = \{X_1, X_2, \cdots, X_m\}$ denote these nodes and P_e denote the probability of having an edge between any given X_i and X_j, for $i \neq j$ where $1 \leq i \leq m$ and $1 \leq j \leq m$. Since P_e would exist with an equal and a random probability given any pair of X_i and X_j, the subgraph formed by the nodes $X_1, X_2, \cdots, X_m$ on a complete graph is indeed an Erdős-Rényi random graph, where each possible edge in the graph possesses an equal probability of being created.

One interesting property shown by Erdős and Rényi is that such graphs have a sharp threshold of edge probability for graph connectivity (Milo et al., 2002). Simplified, if the edge-probability is greater than such threshold, then all of the nodes produced by such a model will be strongly connected. Erdős and Rényi have shown that the sharp connectivity threshold is $th_s = ln\theta/\theta$, where θ is the number of nodes in the graph. The proposed approach exploits this neat graph theoretic property; given the previously extracted maximum spanning tree subgraph, the approach eliminates all nodes/edges whose bot-edge probability (i.e., behavioral similarity P_{bs}) is less than th_s, deeming the rest of the bots, given such formal forensic evidence, as the *niche* of the botnet.

In conclusion, according to the random peer selection model, the niche members of the same infected campaign are expected to be closely connected to each other on a subgraph extracted from Behavioral Service Graphs.

Empirical evaluation

We evaluate the proposed approach in two different deployment scenarios using two real datasets. This aims at validating the accuracy, effectiveness and simplicity of the generated network-based evidence as well as demonstrating the portability of the proposed approach.

Scenario 1: enterprise capability

In this first scenario, Behavioral Service Graphs are employed to infer infected machines within an enterprise network. Although the notion of an enterprise network could extend to an Internet service provider or even a backbone network, in this scenario, for simplicity purposes, we depict a small department within an organization having a deployment setting similar to what is illustrated in Fig. 1. Such department includes 26 machines that are connected to the Internet via an enterprise commodity edge server. The proposed approach is deployed on that server.

Building the ground truth

In order to systematically assess the accuracy of the proposed scheme, one needs to know the IP addresses/hosts of the members of the malicious campaign in a given network. Otherwise, nothing can be said about the true positive or false alarm rate.

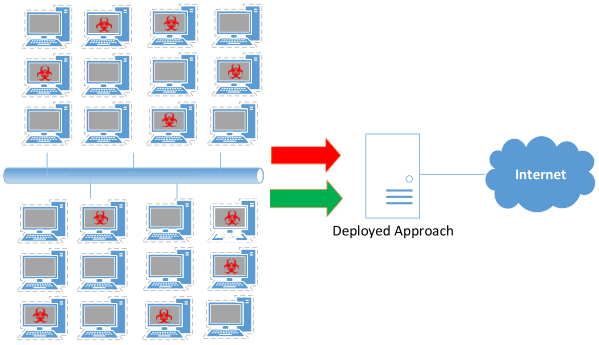
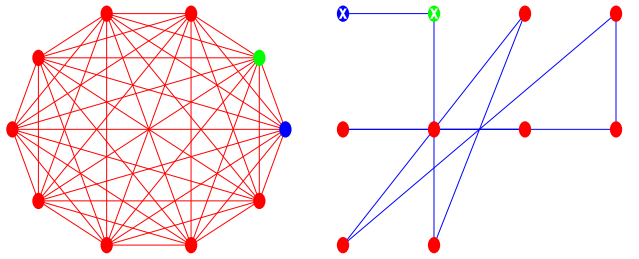


Fig. 1. The proposed approach deployed as an enterprise edge engine.



(a) Enterprise Complete Graph (Inferring the members of the campaign)

(b) Enterprise SubGraph (Inferring the niche members of the campaign)

Fig. 2. The creation of the Enterprise Complete and Sub Graphs.

In order to establish the ground truth for our experiment, we obtained 10 GB of real probing traffic retrieved from the Carna botnet[1] The latter orchestrated campaign is rendered as one of the largest and most comprehensive IPv4 probing census ever. Subsequently, we presumed, as shown in Fig. 1, that 10 out of the 24 machines are infected and thus are generating their malicious probing traffic towards the web service using TCP as the transport protocol and 80 as the destination port number. We successfully achieved this by substituting the IP addresses of the assumed infected departmental machines with 10 IP addresses belonging to 10 unique sources of the Carna botnet that are probing that service. To provide a realistic evaluation scenario, we now assume that we have no knowledge about the infected departmental machines. Subsequently, we generated a legitimate background traffic dataset of 15 GB by leveraging the Security Experimentation EnviRonment (SEER) tool set[2] and randomly merged it (using tcpslice[3]) with the malicious probing traffic dataset generated by those 10 IP addresses ($\approx$8 GB). The newly created merged dataset (of 23 GB) could be thought of as the network data generated by the departmental machines and received by the edge server, where the proposed approach has been deployed for inference and analysis.

Evaluation

By invoking the proposed approach on the merged dataset, the Behavioral Service Graph and its corresponding maximum spanning tree were inferred as depicted in Fig. 2. From a performance perspective, it is worthy to note that processing the 23 GB dataset to generate the feature vectors as well as model the infected machines in complete and subgraphs and infer the niche of the campaign required 8 min and 16 s using a commodity machine with an Intel 3.4 GHz i7 processor with 16 GB of RAM. We believe that this iteration of the implementation of the proposed approaches, which exploit the `C libcpap` library for network processing and the `C Boost Graph Library (BGL)` for graph manipulations, is quite efficient.

A number of observations could be extracted from the complete graph (Fig. 2a). First, the number of assembled Behavioral Service Graphs is accurate; the approach generated one complete graph which is correct as the infected machines in the illustrated scenario are probing only one service, namely, the web service. Second, the number of nodes in this Behavioral Service Graph is also precise; the approach inferred and correlated 10 infected machines which is consistent with the number of infected departmental machines. Third, after a semi-automated analysis and comparison that was based on the logged probing IP traffic flows, we identified that all the 10 machines that the proposed approach has identified are indeed the same IP addresses of the infected departmental machines (i.e., the IP addresses of the Carna botnet). Therefore, based on the latter three observations, we can claim that the proposed approach yielded no false negative or false positive from a threat modeling perspective.

However, to further fortify the latter claim in an attempt to generalize it as it applies to various network scenarios, we performed several other experiments. Specifically, we were interested in evaluating the accuracy of the proposed approach as (1) the number of probed services increase in diversity and (2) as the number of infected machines scale up in a given network. Thus, we first augmented the number of probed services, one at a time, up to 100 various probed TCP and UDP services. The results disclosed that the number of generated Behavioral Service Graphs remained accurately reflecting the number of probed services. Moreover, to verify the scalability of the proposed approach, we increased the number of ground truth infected machines, by slots of 100, up to 1000 machines. The outcome disclosed consistent accuracy in terms of constructed number of nodes and the positive infection state of such nodes. Such experiments relatively validate the accuracy and the scalability of the proposed scheme. Nevertheless, it is important to mention that by further executing scalability experiments exceeding the 1000 threshold, the proposed approach started to suffer from false positives and negatives in terms of infection status of those created nodes. Specifically, we quantified that as the number of machines increased by slots of 500 machines beyond that 1000 threshold, the false positive rate increased by around 2% and the false negative rate increased by around 1.5%, on average. While we did not have a chance to fully analyze this issue, our initial investigation showed that it could be an implementation issue with the employed graph library. Future work will attempt to address this scalability finding and will also investigate a distributed implementation of the proposed graph-theoretic notions.

We were further concerned about the quality of the formed cluster provided by the complete Behavioral Service Graph. Since such a graph is supposed to correlate the nodes based on their infection state as well as their behaviors, we thought it would be significantly beneficial to assert such grouping of nodes by employing another approach. To achieve this, we relied on an unsupervised, machine learning data clustering technique, namely, the *k*-means algorithm (Hartigan and Wong, 1979). Typically, the standard *k*-means algorithm requires, as apriori knowledge, the number of clusters *k*. However, since our aim is to provide a robust evaluation methodology, we relied on an approach to automatically determine the optimal number of clusters. In particular, we leveraged the Calinski-Harabasz criterion (Caliński and Harabasz, 1974)

[1] http://internetcensus2012.bitbucket.org/download.html.
[2] http://seer.deterlab.net/trac.
[3] http://sourceforge.net/projects/tcpslice/.

that operates by systematically verifying various number of clusters and subsequently recording the variances between and within the formed clusters. To determine the optimal number of clusters, the metric should be maximized with respect to k; the optimal number of clusters is the solution with the highest Calinski-Harabasz index value. To apply the k-means on the infected 10 nodes as previously inferred by the complete Behavioral Service Graph, we (1) retrieved their probing traffic from the merged dataset using a simplistic tcpdump[4] filter, (2) extracted their packet features[5] using the open source jNetPcap API[6] and (3) compiled the extracted features into a unified data file and then applied the k-means algorithm in conjunction with the Calinski-Harabasz metric on such file. The outcome of the k-means execution is illustrated in Fig. 3.

Motivated by Ding and He (2004) that asserted 1) that the relaxed solution of the k-means clustering, specified by the cluster indicators, could be given by the principal components from the Principal Component Analysis (PCA) technique (Shlens, 2014) and 2) that the PCA subspace spanned by the principal directions is identical to the cluster centroid subspace, Fig. 3 reveals the formed cluster on the first two principal axes of the PCA. One can notice the formation of one, and only one, relatively strongly correlated cluster. This result strongly advocates that the nodes possess strong similarity characteristics. In summary, such an outcome that was generated by approaching the clustering problem from another perspective, indeed validates the grouping capability of the infected machines (i.e., nodes) that is provided by the constructed complete Behavioral Service Graph of Fig. 2a.

Thus, up to this stage, an enterprise forensic investigator can leverage such precise and actionable evidence for prompt detection, containment and mitigation of such infected machines from the concerned network. Intuitively, an investigator can also isolate such machines to collect additional evidence by monitoring their network and system activities for auxiliary data analysis or extraction of digital artifacts. The latter evidence could be exploited to generate signatures of attacks or for thoroughly analyzing a specific security phenomenon of interest.

Campaign niche inference

Extracted from the Behavioral Service Graph, the enterprise subgraph that is depicted in Fig. 2b also provides noteworthy inferences. First, it was able to generate formal forensic evidence that clustered the machines into a well-defined orchestrated infected campaign. Recall, that the formality arises from the fusion of the bots behavioral similarly as previously extracted by the data analytics of Section Data behavioral analytics coupled with the graph theoretical notion of a maximum spanning tree. Second, and more importantly, the approach was able, by leveraging Erdős-Rényi random graphs, to infer the niche members of that infected campaign. Fig. 2b shows those nodes with marked 'Xs'. In fact, the proposed approach revealed that these two nodes render the creation of the campaign. Thus, in theory, these nodes should have caused the creation of the campaign in the first place. To validate this, we manually investigated those IP addresses by going back to the Carna botnet dataset. Our investigation showed that these two IP addresses are used as root nodes in the botnet to infect other machines for propagation purposes. The latter fact was validated as these two nodes were among the top 3 nodes to generate most of the probing traffic in the dataset. The latter formal forensic evidence could be promptly exploited by investigators to prioritize the eradication of those two nodes in order to seize the expansion of the campaign on their networks. This would indeed significantly limit any present or future possible sustained collateral damage and any symptoms of infection that could be caused by the infected bots.

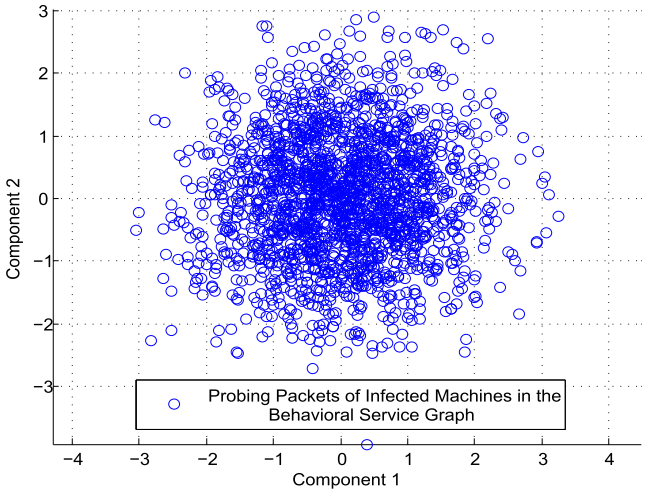


Fig. 3. Validating the clustering capability of the complete Behavioral Service Graph.

Scenario 2: global capability

In the previous scenario, we have demonstrated how the proposed scheme can be exploited to operate within the context of an enterprise network. In this section, we port the approach to a global scale and elaborate on how it can be employed to monitor, infer and distribute Internet-scale forensic intelligence. Thus, in this scenario, the approach is envisioned to operate in a model similar to what is dubbed as a global Security Operation Center (SOC). Typically, such operational centers have access to significant various real-time and raw data streams from around the globe. They often exploit such data for analysis, correlation and generation of intelligence that would be distributed to concerned parties for alert and mitigation purposes. Such centers were initially formed as global independent entities to combat an increasing trend of external (in contrary to internal) threats and attacks.

Thus, in this second scenario, Behavioral Service Graphs are postulated to be deployed as an additional forensic capability in one of those SOC centers. In this context, we operate the scheme by investigating darknet data. In a nutshell, a darknet (also commonly referred to as a network telescope) is a set of routable and allocated yet unused IP addresses (Moore et al., 2004). It represents a partial view of the entire Internet address space. From a design perspective, a darknet is transparent and indistinguishable compared with the rest of the Internet space. From a deployment perspective, it is rendered by network sensors that are implemented and dispersed on numerous strategic points throughout the Internet. Such sensors are often distributed and are typically hosted by various global entities, including Internet Service Provides (ISPs), academic and research facilities, and backbone networks. The aim of a darknet is to provide a lens on Internet-wide malicious traffic; since darknet IP addresses are unused, any traffic targeting them represents anomalous unsolicited traffic. Such traffic (i.e., darknet data) could be leveraged to generate various cyber threat intelligence, including inferences and insights related to probing activities; some of the probes of an infected machine, while probing the Internet space, will also hit the darknet and thus will be subsequently captured. Recall, that the probing machine, while spraying its probes, can not avoid the darknet as it does not have any knowledge about its existence. Further, it

[4] http://www.tcpdump.org/.

[5] Adopted from Alshammari and Zincir-Heywood (2011), where they have been shown to produce distinguishing characteristics when applied on network data.

[6] http://jnetpcap.com/.

is extremely rare if not impossible for a probing source to have any capability dedicated to such avoidance (Cooke et al., 2006).

To this end, we have access to real and logged darknet data provided by the Center for Applied Internet Data Analysis (CAIDA)[7] We leverage a sample of such data to evaluate the proposed approach as it is deployed in the SOC model.

The ground truth

Similar to the previous scenario, there exists a need to have a concrete knowledge about a ground truth to properly evaluate the proposed scheme. For this purpose, in this scenario, we rely on a reported Internet-scale event related to a large-scale probing campaign. Particularly, on October 10, 2012, the Internet Storm Center (ISC) received a report of a probing campaign targeting Internet-wide SQL servers[8] This incident was also interestingly corroborated by Dshield. Dshield data comprises of millions of intrusion detection log entries gathered daily from sensors covering more than 500,000 IP addresses in over 50 countries. Further, the ISC report noted that the probing campaign involved more than 9000 distributed sources, which aim at exploiting that service. We rely on the occurrence of such disclosed incident as the ground truth as we proceed in our evaluation.

Evaluation

From our darknet data repository, we extract one week of data ($\approx$80 GB) pertaining to the period of October 4th to October 11th, 2012. The aim is to employ the proposed approach on such data to evaluate the scheme's capability and effectiveness in disclosing insights related to that reported campaign. By executing the proposed approach on the extracted probing traffic from our darknet dataset, the outcome demonstrated that one of the Behavioral Service Graphs was indeed able to infer and correlate around 800 unique sources[9] targeting the SQL service. Further, the behavioral analytics (1) showed strong behavioral similarity between those sources and (2) inferred that those sources were indeed bots, thus providing strong evidence that such campaign was triggered from Internet-wide infected machines. The latter inferred bots could be visualized as in Fig. 4, where nodes that are close to each other represent a maximized behavioral similarity (P_{bs}) (recall Section Behavioral Service Graphs). Additionally, it might be interesting to mention that the proposed approach deemed 84 bots as the niche of the campaign by leveraging the approach of Section Friends of the enemy stay closely connected: inferring infected campaigns.

Thus, provided with such forensic evidence, SOC analysts can demand an immediate take-down of those 84 bots to limit the expansion of such campaign on the global Internet. In addition, they can promptly notify concerned parties to employ mitigation approaches against the abuse of SQL servers. From a performance perceptive, using the same commodity machine as in the previous experiment, this SOC experiment required approximately 39 min to process the 80 GB darknet dataset in order to build the Behavioral Service Graph coupled with its corresponding subgraph, and to infer the niche of the infected campaign.

Proposed approach: limitations & possible improvements

In this section, we discuss several limitations of the proposed approach and attempt to provide some remediation strategies.

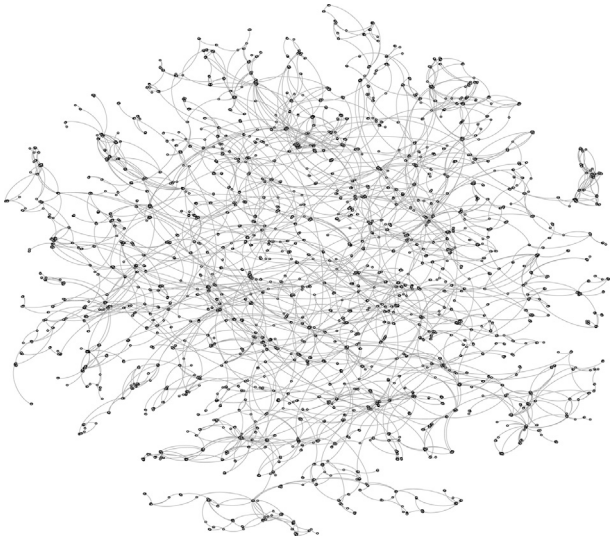


Fig. 4. The proposed approach revealing the bots of the SQL probing campaign.

First, the approach analyzes probing activities (for promptness reasons) by leveraging a number of behavioral analytics to infer enterprise and Internet-wide bots. Intuitively here, there is a need to further fortify the infection evidence. To this end, as mentioned in Section Data behavioral analytics, we are attempting to devise an approach, which endeavors to correlate perceived probing activities from such probing sources with malware traffic samples to corroborate the bot infection evidence as well as to attribute the inferred infected machines to a specific malware family. Another related issue is that bots or infected machines that do not generate probing activities will not be captured by the proposed approach. In this context, literature approaches which offer host-based solutions to infer maliciousness could be employed as a (preprocessing) complementary detection capability. Second, the proposed approach of Section Friends of the enemy stay closely connected: inferring infected campaigns infers the niche of the campaign by heuristically selecting the nodes/edges that possess a similarity behavior above a threshold indicated by the Erdős-Rényi random graphs. To this end, it would be very interesting to find a formal mathematical relation to infer and extrapolate the number of such nodes (beyond just the analyzed dataset) as a function of the amount of generated traffic, the geographic distribution of the inferred nodes and the similarity behavioral metrics. Finally, the proposed approach is still experimental. We are currently continuing its development to render it operational in an automated fashion in both of the experimented scenarios. In this context, the pinpointed scalability issue will also be thoroughly investigated.

Related work

In this section, we briefly review the related work on two topics, namely, anomaly detection using graphs and data forensic approaches.

Anomaly detection using graphs

Wang et al. (2014b) approached the problem of anomaly detection as a change-point hypothesis constructed on a time series of graphs. The authors proposed a stochastic model that is based on the use of scan statistics; metrics that can extract normal traffic and compare it to anomalous traffic. Their model was evaluated and

[7] http://www.caida.org/data/overview/.

[8] https://isc.sans.edu/forums/diary/Reports+of+a+Distributed+Injection+Scan/14251/.

[9] Since the monitored darknet space is a/8, we are only able to see a portion of the campaign using that dataset.

validated on real email data. In another work, Brdiczka et al. (2012) proposed an approach for proactive detection of insider threats by combining structural anomaly detection from social and information networks, and psychological profiling of individuals. Their approach is specifically tailored to detect anomalies in multi-player online games. In a different work, Hassanzadeh et al. (2012) proposed a framework for analyzing the effectiveness of various graph theoretic properties in detecting anomalous users on online social networks. Their empirical evaluations demonstrated that such derived properties are indeed accurate in modeling anomalous behaviors. Further, Ding et al. (2012) employed bipartite graph representation of network flow traffic coupled with community detection techniques in an attempt to infer malicious sources. To achieve such a task, the authors further employed hard thresholds and heuristics derived from empirical evaluations. Furthermore, Wang and Daniels (2008) proposed a graph-based approach to correlate and reason about generic network security incidents.

Data forensic approaches

The author in Guarino et al. (2013) explored the challenges of data forensics as applied to digital investigation. He elaborated on how techniques and algorithms that are typically used in data analysis could possibly be adapted to the unique context of digital forensics. The author discussed various approaches ranging from managing evidence to machine learning techniques, and analysis of forensic disk images and network traffic dumps. In an alternate work, Zhu (2011) proposed a data-driven iterative algorithm for discovering network attack patterns via a feedback mechanism. The author claimed that the algorithm is fully unsupervised as it does not require user-defined thresholds. Simulations were conducted to validate the accuracy of the proposed approach. Moreover, in Garfinkel (2012), Garfinkel presented numerous lessons learned from writing digital forensic tools and managing a 30 TB digital evidence corpus. Specifically, the author elaborated on the technical difficulties analyzing such data, the possible hardware and software issues that could be faced, and how to accurately retain the extracted evidence. The author concluded by stating some present issues related to data forensic approaches, namely, diversity of data that needs to be analyzed, the size of the datasets, and the mismatch between the technical skills of investigators and the difficulty level of the work.

The proposed approach of this work is unlike the above two categories as (1) it tackles a different problem rendered by inferring enterprise and Internet-wide infections, (2) uniquely scrutinizes probing activities using a set of behavioral analytics to promptly infer infections, (3) employs a new concept of similarity service graphs to infer malicious campaigns, (4) exclusively exploits graph theoretic notions such as the maximum spanning tree and Erdős-Rényi random graphs to infer the niche of the infected campaign and (5) it has been empirically evaluated using two real and significant datasets in two diverse deployment scenarios.

Concluding remarks

Network forensic approaches that endeavor to contribute, both, scientifically and operationally, are particularly rare. Motivated by this, in this work, we have devised Behavioral Service Graphs, a data-driven approach that is able to effectively process, analyze and correlate large volumes of network traffic to promptly generate formal, highly-accurate and actionable network forensic evidence. Such evidence could indeed be leveraged by investigators to infer enterprise and Internet-wide infected machines, which operate within the context of malicious campaigns. Empirical evaluations with real data under two different deployment scenarios have verified the accuracy and effectiveness of the proposed approach in terms of inferring the infections as well as pinpointing the niche of such campaigns. We hope that the forensic community could consider the approach as a building block for complementary analysis and investigation.

As for future work, other than tackling the issues mentioned in Section Proposed approach: limitations & possible improvements, we are currently exploring the problem of campaign analysis; the ability to infer what the probing infected bots will eventually execute after finalizing their initial probing activities. We aim to achieve the latter by correlating the generated inferences from this work with other data sources, including but not limited to, passive DNS, and public intrusion and firewall logs. Additionally, we will be leveraging our proposed approach in the near future to conduct a large-scale Internet measurement study using CAIDA's darknet data to report and analyze on simultaneously active Internet-scale malicious campaigns.

References

Adeyemi, Ikuesan R., Razak, Shukor Abd, Azhan, Nor Amira Nor, 2013. A review of current research in network forensic analysis. Int. J. Digital Crime Forensics (IJDCF) 5 (1), 1–26.

Allman, Mark, Paxson, Vern, Terrell, Jeff, 2007. A brief history of scanning. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. ACM, pp. 77–82.

Alshammari, Riyad, Zincir-Heywood, A Nur, 2011. Can encrypted traffic be identified without port numbers, ip addresses and payload inspection? Comput. Netw. 55 (6), 1326–1350.

Anderson, Theodore W., Darling, Donald A., 1954. A test of goodness of fit. J. Am. Stat. Assoc. 49 (268), 765–769.

Bhuyan, Monowar H., Bhattacharyya, D.K., Kalita, Jugal K., 2011. Surveying port scans and their detection methodologies. Comput. J. bxr035.

Bou-Harb, Elias, Debbabi, Mourad, Assi, Chadi, 2016. A novel cyber security capability: inferring internet-scale infections by correlating malware and probing activities. Comput. Netw. 94, 327–343.

Brdiczka, O., et al., May 2012. Proactive insider threat detection through graph learning. In: IEEE Symposium on Security and Privacy Workshops, pp. 142–149.

Caliński, Tadeusz, Harabasz, Jerzy, 1974. A dendrite method for cluster analysis. Commun. Stat-theory Methods 3 (1), 1–27.

Cooke, Evan, Bailey, Michael, Jahanian, Farnam, Mortier, Richard, 2006. The dark oracle: perspective-aware unused and unreachable address discovery. In: NSDI, vol. 6 pages 8–8.

Dainotti, Alberto, et al., 2012. Analysis of a "/0" stealth scan from a botnet. In: The 2012 ACM Conference on Internet Measurement Conference, IMC'12. ACM.

Dainotti, A., King, A., Claffy, K., Papale, F., Pescapé, A., 2015 Apr. Analysis of a "/0" stealth scan from a botnet. IEEE/ACM Trans. Netw. 23 (2), 341–354.

Daly, Michael K., Nov, 4, 2009. Advanced Persistent Threat. Usenix.

Díaz, Josep, Petit, Jordi, Serna, Maria, 2002. A survey of graph layout problems. ACM Comput. Surv. (CSUR) 34 (3), 313–356.

Ding, Chris, He, Xiaofeng, 2004. K-means clustering via principal component analysis. In: Proceedings of the Twenty-first International Conference on Machine Learning. ACM, p. 29.

Ding, Qi, et al., 2012. Intrusion as (anti)social communication: characterization and detection. In: 18th ACM SIGKDD, pp. 886–894.

Fu, Zhang, Papatriantafilou, Marina, Tsigas, Philippas, 2012. Mitigating distributed denial of service attacks in multiparty applications in the presence of clock drifts. Dependable Secure Comput. IEEE Trans. 9 (3), 401–413.

Garcia-Teodoro, Pedro, Diaz-Verdejo, J., Maciá-Fernández, Gabriel, Vázquez, Enrique, 2009. Anomaly-based network intrusion detection: techniques, systems and challenges. Comput. Secur. 28 (1), 18–28.

Garfinkel, Simson, 2012. Lessons learned writing digital forensics tools and managing a 30tb digital evidence corpus. Digit. Investig. 9 (Suppl. (0)), S80–S89. The Proceedings of the Twelfth Annual {DFRWS} Conference 12th Annual Digital Forensics Research Conference.

Governement of the USA. CAN-SPAM Act: A Compliance Guide for Business. http://tinyurl.com/nu85mzc.

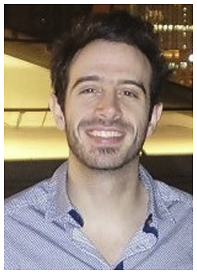
Guarino, Alessandro, 2013. Digital forensics as a big data challenge. In: Reimer, Helmut, Pohlmann, Norbert, Schneider, Wolfgang (Eds.), ISSE 2013 Securing Electronic Business Processes. Springer, Fachmedien Wiesbaden, pp. 197–203.

Hartigan, John A., Wong, Manchek A., 1979. Algorithm as 136: a k-means clustering algorithm. Appl. Stat. 100–108.

Hassanzadeh, Reza, et al., 2012. Analyzing the effectiveness of graph metrics for anomaly detection in online social networks. In: Web Information Systems Engineering – WISE 2012, vol. 7651. Springer, Berlin Heidelberg, pp. 624–630.

Kruskal, Joseph B., 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Am. Math. Soc. 7 (1), 48–50.

Kührer, Marc, Hupperich, Thomas, Rossow, Christian, Holz, Thorsten, 2014. Exit from hell? reducing the impact of amplification ddos attacks. In: USENIX Security Symposium.
Li, Zhichun, et al., 2011. Towards situational awareness of large-scale botnet probing events. IEEE Trans. Inf. Forensics Secur.
Milo, Ron, Shen-Orr, Shai, Itzkovitz, Shalev, Kashtan, Nadav, Chklovskii, Dmitri, Alon, Uri, 2002. Network motifs: simple building blocks of complex networks. Science 298 (5594), 824–827.
Moore, David, Shannon, Colleen, Voelker, Geoffrey M., Savage, Stefan, 2004. Network Telescopes: Technical Report. Department of Computer Science and Engineering, University of California, San Diego.
Ozeki, Kenta, Yamashita, Tomoki, 2011. Spanning trees: a survey. Graphs Comb. 27 (1), 1–26.
Panjwani, Susmit, Tan, Stephanie, Jarrin, Keith M., Cukier, Michel, 2008. An experimental evaluation to determine if port scans are precursors to an attack. In: Dependable Systems and Networks, 2005. DSN 2005. Proceedings. International Conference on. IEEE, pp. 602–611.
Parliament of Canada. BILL C-28. http://tinyurl.com/avh9vzv.
Paxson, Vern, 1999. Bro: a system for detecting network intruders in real-time. Comput. Netw. 31 (23), 2435–2463.
Pilli, Emmanuel S., Joshi, Ramesh C., Niyogi, Rajdeep, 2010. Network forensic frameworks: survey and research challenges. Digit. Investig. 7 (1), 14–27.
Rajab, Moheeb Abu, Zarfoss, Jay, Monrose, Fabian, Terzis, Andreas, 2006. A multifaceted approach to understanding the botnet phenomenon. In: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement. ACM, pp. 41–52.
Roesch, Martin, et al., 1999. Snort: lightweight intrusion detection for networks. In: LISA, vol. 99, pp. 229–238.
Shlens, Jonathon, 2014. A Tutorial on Principal Component Analysis arXiv preprint arXiv:1404.1100.
Silva, Sérgio SC., Silva, Rodrigo MP., Pinto, Raquel CG., Salles, Ronaldo M., 2013. Botnets: a survey. Comput. Netw. 57 (2), 378–403.
Staniford, Stuart, Hoagland, James A., McAlerney, Joseph M., 2002. Practical automated detection of stealthy portscans. J. Comput. Secur. 10 (1), 105–136.
Wang, Wei, Daniels, Thomas E., October 2008. A graph based approach toward network forensics analysis. ACM Trans. Inf. Syst. Secur. 12 (1), 4:1–4:33.
Wang, Yini, Wen, Sheng, Xiang, Yang, Zhou, Wanlei, 2014. Modeling the propagation of worms in networks: a survey. Commun. Surv. Tutor. IEEE 16 (2), 942–960.
Wang, Heng, Tang, Minh, Park, Y., Priebe, C.E., Feb 2014. Locality statistics for anomaly detection in time series of graphs. Signal Process. IEEE Trans. 62 (3), 703–717.
Whyte, David, Kranakis, Evangelos, van Oorschot, Paul C., 2006. Dns-based detection of scanning worms in an enterprise network. In: NDSS.
Xie, Yinglian, Yu, Fang, Achan, Kannan, Panigrahy, Rina, Hulten, Geoff, Osipkov, Ivan, 2008. Spamming botnets: signatures and characteristics. ACM SIGCOMM Comput. Commun. Rev. 38 (4), 171–182.
Xu, Kuai, Zhang, Zhi-Li, Bhatt, Supratik, 2005. Profiling internet backbone traffic: behavior models and applications. In: ACM SIGCOMM Comp. Comm Rev, pp. 169–180.
Zhu, Ying, August 2011. Attack pattern discovery in forensic investigation of network attacks. Sel. Areas Commun. IEEE J. 29 (7), 1349–1357.



Dr. Elias Bou-Harb is currently an Assistant Professor at the computer science department at Florida Atlantic University, where he directs the Cyber Threat Intelligence Laboratory. Previously, he was a visiting research scientist at Carnegie Mellon University. Elias is also a research scientist at the National Cyber Forensic and Training Alliance (NCFTA) of Canada. Elias holds a Ph.D. degree in computer science from Concordia University, Montreal, Canada. His research and development activities and interests focus on the broad area of operational cyber security, including, attacks detection and characterization, Internet measurements, cyber security for critical infrastructure and big data analytics.

Dr. Mark Scanlon is an Assistant Professor in Forensic Computing in the School of Computer Science, University College Dublin, Ireland and a Fulbright Scholar in Cybersecurity and Cybercrime Investigation. His research interests include Remote Evidence Acquisition, Evidence Whitelisting, Data Deduplication, Cloud Forensics, and Digital Forensics Education. Dr. Scanlon is an active member of the digital forensics research community and is a keen associate editor, keynote speaker, reviewer and conference organizer across a range of key journals and conferences in the field.