Computer Networks 73 (2014) 224–243

Contents lists available at ScienceDirect

Computer Networks

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

Survey Paper

Traffic anomaly diagnosis in Internet backbone networks: A survey

A.K. Marnerides [a,c,*], A. Schaeffer-Filho [b], A. Mauthe [a]

[a] InfoLab21, Department of Computing and Communications, Lancaster University, Lancaster, UK
[b] Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[c] School of Computing & Mathematical Sciences, Liverpool John Moores University, Liverpool, UK

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 16 July 2014
Accepted 9 August 2014
Available online 23 August 2014

Keywords:
Internet traffic anomalies
Anomaly detection
Feature selection
Digital signal processing
Information theory
Statistical methods

ABSTRACT

Computer networks are becoming increasingly important in supporting business and everyday activities. In particular, the Internet has become part of the critical infrastructure and has a strategic importance in our society and in the digital economy. These developments have led to a highly dynamic network utilization, where traffic fluctuations and seemingly random and anomalous traffic patterns are commonly manifested and hard to diagnose. In order to ensure the protection and resilience of such networks, it is necessary to better analyze and observe network traffic. Thus, anomaly diagnosis aims to discover and characterize critical anomalies affecting the network infrastructure, where the source of these anomalies may be deliberately malicious (e.g. attacks) or unintentional (e.g. failures, misconfigurations or legitimate but abnormal use of the network such as in flash crowds). However, although there is a multitude of algorithms and techniques looking at different elements of the analysis of network traffic anomalies, most research typically focuses on a specific aspect or methodology and there is very little regard for the overall context. This survey aims to present a comprehensive investigation of the current state of the art within the network anomaly diagnosis domain, in particular for Internet backbone networks. We decompose the overall anomaly diagnosis problem spectrum into four main dimensions, namely, processing costs, diagnosis granularity, theoretical methodologies and traffic features. Subsequently the anomaly diagnosis research area is structured further and an overview of the most relevant research is provided by individually reviewing each component of the problem spectrum and proposed solutions with a deeper focus on methodologies and features. Further, we also present and review seminal pieces of work that are considered cornerstones of the anomaly diagnosis research domain.

1. Introduction

As part of the critical infrastructure computer networks are of curial importance to societies and the global economies. They need to be well managed and protected from threats to the infrastructure as well as the actual communication. Hence, with constantly increasing traffic volumes and persistently changing traffic patterns, *anomaly diagnosis* has received considerable attention by Internet Service Providers (ISPs) and the networking research community in recent years [1–13] in order to prove better network management and more resilient networks.

Traffic anomalies occur frequently in the Internet, causing atypical changes in network traffic levels that possibly compromise the efficiency and reliability of the network.

* Corresponding author at: InfoLab21, Department of Computing and Communications, Lancaster University, Lancaster, UK.
E-mail addresses: a.marnerides2@lancaster.ac.uk (A.K. Marnerides), alberto@inf.ufrgs.br (A. Schaeffer-Filho), a.mauthe@lancaster.ac.uk (A. Mauthe).

http://dx.doi.org/10.1016/j.comnet.2014.08.007
1389-1286/

The source of these anomalies may be deliberately malicious (e.g. attacks) or unintentional (e.g. failures, misconfigurations or legitimate but abnormal use of the network such as in flash crowds). Two mainstream approaches are being used to detect and identify network anomalies, i.e. *signature-based* anomaly detection as part of Intrusion Detection Systems (i.e. IDS) (e.g. [12,14–18]) and *statistical* anomaly detection (e.g. [1–6,8,12,13]). The former aims to extract and identify on-going anomalies using pattern-based pre-knowledge gathered in the past, whereas the latter is trying to isolate and characterize patterns that substantially deviate from the normal behavior of the network. A number of systems and products use a signature based anomaly detection approach. However, according to [12] signature-based or rule-based anomaly detection methods are resource-intensive and have detection rates lower than 70%. As a result, the research community is focusing its attention on methods based on statistical anomaly detection. Consequently *anomaly diagnosis* has developed into a research domain. Traffic anomaly diagnosis constitutes the primary aspect in the control and management loop that is required to make communication networks more resilient since it enables the identification, understanding and classification of abnormal behavior in the network. One of the key problems in this context is how to discriminate between *operational overload* due to legitimate service requests as in a flash crowd or *malicious attacks* (e.g. a DDoS attack) and then apply adequate countermeasures to mitigate or remediate the problem [19]. Hence, anomaly diagnosis is a multi-stage process with *anomaly detection* identifying the onset of the abnormal behavior in the network and anomaly *classification/clustering* determining and categorizing the root cause of the anomaly.

Due to this multifaceted nature of anomaly diagnosis there has been lots of research on specific aspects but no common picture has emerged. In this survey we systematically look at the anomaly diagnosis domain in backbone networks by elaborating upon the dimensions of *Processing Cost, Diagnosis Granularity, Theoretical Methodologies* and *Traffic Features* as introduced in [20]. We advocate that in order to understand the anomaly diagnosis problem as a whole one cannot ignore any of the aspects above. Instead, it is necessary to look at the entire problem space and further explain how the different dimensions fit and coexist together. Therefore, the purpose of this paper is twofold: firstly, it provides an overview of the network anomaly diagnosis domain by presenting the problem spectrum as a whole in an accessible form for those unfamiliar with this research area. Secondly, it individually categorizes and reviews the most influential research work. Furthermore, this work places a stronger emphasis on the dimensions of *Theoretical Methodologies* and *Traffic Features* by discussing and presenting seminal pieces of work in the hope of providing a comprehensive understanding to any general audience.

This paper is structured as follows: Section 2 introduces the multiple dimensions of the anomaly diagnosis problem. Section 3 describes in depth the theoretical foundations of methodologies for anomaly diagnosis, while Section 4 presents aspects related to traffic features and feature preprocessing required by these methodologies. Section 5 discusses important examples found in the literature that apply the methods and techniques surveyed in this paper. Concluding remarks are presented in Section 6.

2. Anomaly diagnosis in a nutshell

The notion of anomalous behavior has been a heavily discussed topic within a range of disciplines and to best to our knowledge it has not yet been comprehensively defined capturing its use throughout all disciplines. However an atypical event within a backbone network is considered as anomalous if it exhibits an outlier characteristic with respect to the overall normal profiling that is initially constructed via a statistical anomaly detection methodology. Thus, we clarify that a network anomaly corresponds to a pattern of observable network traffic (e.g. packet/flow/byte counts) or feature (e.g. distribution source/destination IP address pairs) that does not conform to an expected notion of normal behavior as defined by a given statistical model. At the same time, we state that despite the fact that a large portion of anomalies is caused under malicious intent (e.g. DDoS attacks, worms, viruses) and have an immediate effect on the performance of a network, there are several instances where a detected anomaly does not necessarily cause a degradation of the network's overall performance (e.g. port scan) and is not always malicious (e.g. flash crowd events). Traditionally, an anomaly diagnosis technique for backbone networks has the ultimate goal of differentiating between normal and anomalous behavior in the network. In particular, it is a process performed under a range of requirements and is mainly accomplished by combining several techniques [21]. In any situation where anomaly detection is required there are two conflicting goals, i.e. optimal detection and protection of networks and systems, and efficient and cost effective operations. On the one hand, the best service which should protect from all malicious intents carried within the network is desired. On the other hand, the economical and feasibility aspects have to be considered, which might lead to prioritization of specific service users (e.g. business vs. domestic) or accepting certain weaknesses within the detection infrastructure. Hence, to evaluate diagnosis mechanism one needs to look at the entire problem spectrum and its different dimensions.

2.1. Background and related work

The reliable identification and categorization of the various network anomaly types is not straightforward and a number of issues have to be looked at. Firstly, network anomaly diagnosis (particularly in backbone networks) becomes complex since anomaly patterns have to be identified and interpreted from large amounts of multi-dimensional traffic data. Such data contain a substantial amount of noise either caused by the measurement process itself or the cross-traffic burstiness which is aggregated from all the measurement links. Secondly, anomaly diagnosis in such networks entails a number of technical difficulties in terms of both high monitoring and processing costs of

network traffic measurements accompanied by the lack of automated tools for real-time detection [1,7,22]. Furthermore, due to the multidimensionality of the problem, no overall classification scheme or framework has so far been established. Thirdly, the literature on the subject is fragmented and fails to provide a comprehensive picture of the problem spectrum. Fourthly, due to business-related issues that involve user privacy as well as the cost for maintaining resources for the gathering of network data, the applicability of any proposed anomaly diagnosis technique is limited. Even when network traces are provided by a service provider, it is likely that this data has been captured at particular observational points (e.g. links between basic Points of Presence) within a backbone which may not be sufficient to construct an adequate training model that complies with the overall varying dynamics manifested in a given backbone network. Naturally, this constraint has a direct relationship with the diversity of measurement techniques applied on links or edge routers that affect the view of an anomaly detection component with respect to the overall characterization of the backbone network. For instance, cases where operational data is acquired on a single backbone link require a completely different post-processing compared to cases where traces are captured by multiple links on Points of Presence (PoPs). As pointed out in [23], the strong asymmetric traffic profile presented in single-point backbone link measurements does not allow the use of the well known Origin–Destination (i.e. OD) flow aggregation and the analysis of bidirectional network flows.

Nevertheless, previous studies have taken different approaches to compare existing techniques for anomaly detection [24,25]. There are studies that, under a metric-based perspective, distinguish those techniques for anomaly detection that rely on traffic volume changes [13,26–28], from techniques that observe unusual changes in the distribution of certain features of the network traffic [1,6,12,29]. Overall, when evaluating the quality of anomaly detection and classification techniques, a number of standard performance metrics can be taken into account [30]. The most common performance metrics are true positive rate (TPR), specificity (SPC), precision (p), negative predictive value (NPV), false positive rate (FPR), false discovery rate (FDR), accuracy (ACC) and F1-score. In particular, false positive flows or packets are those that should have been marked as non-malicious, but were flagged as malicious. In contrast, false negatives are flows or packets that should have been marked as malicious, but were flagged as non-malicious. Despite these generic metrics there are still some fundamental constraints towards the adequate evaluation of anomaly detection schemes. Apart from the different approaches for defining the network ground truth (i.e. normal behavior) that might lead to conflicting results when comparing state-of-the-art anomaly detectors, the authors in [31,32] also indicate that an additional difficulty in the evaluation of anomaly detectors relates to the reproducibility of experiments. Moreover, relying on fixed-sized network traffic traces also prevents researchers from performing sensitivity analysis in anomaly detection schemes, which could further measure the size of a given anomaly in order to get it detected [32].

Throughout past and recent literature, the comparison of anomaly detection techniques can also be based on distinguishing those techniques that depend on an underlying model of normal traffic data [25,33], from techniques that do not require such historical data to identify anomalies [34]. Techniques for anomaly detection can also be classified based on the domain in which network anomalies are observed, typically in terms of unusual variations in the time-domain [3,4,29], in the space-domain [12] or in the frequency-domain [26] of the network traffic. There are numerous efforts assessing the aforementioned variations and they are widely known and categorized as *temporal*, *spatial* and *spatio-temporal* solutions [12]. Nevertheless, despite the hybrid mathematical formulations embodied in such solutions, we strictly dedicate ourselves at presenting the fundamental theoretical foundations that underly such approaches, with the intention to introduce the problem to a wider audience. Therefore, we avoid comparing such methods as commonly discussed in [12,35,36], and rather illustrate the most influential methods for anomaly diagnosis proposed in the last decade.

2.2. Anomaly types

There is a wide variety of causes that initiate anomalies, ranging from malicious targeted network attacks to unintentional misconfigurations due to human mistakes. Misconfigurations can cause the network to operate sub-optimally or make it not able to operate at all (e.g., routing misconfiguration), or leave the network vulnerable to well-known operational threats (e.g., firewall misconfigurations). Also, the provision of new services and equipment in the network, or the failure of existing ones are also likely to cause divergence from the expected normal traffic profile, and thus indicate the occurrence of an anomaly. For example, the BBC iPlayer [37] is understood to have caused an excessive growth in the volume of streaming traffic which affected a large number of ISP providers. Likewise, anomalies may be caused by an unexpected increase in the number of accesses to a service, e.g. a news website, driven by factors outside the realm of the network, in a phenomenon known as flash crowds. This might affect the performance of the network or even prevent it from providing a service (e.g., by bringing down a Web server). Finally, due to the rapid development of Internet technologies and operating systems there is a large number of new anomalies (e.g. application-specific threats) that are not likely to be detected by a network-level detection and diagnosis component. Even though the greatest majority of detection strategies are aware of *lightweight*[1] anomalies (e.g. port scans, worm propagation, phishing attacks), it is infeasible to develop a single technique that is adaptive to all types of anomalies.

Network anomalies have been surveyed and categorized in previous works [7,8]. However, in order to better structure the problem space in this work we adopt the

[1] We refer to lightweight anomalies as anomalous network behavior that does not consume enormous amounts of traffic volume with respect to bytes, packets and flows such as stealthy worms [38], viruses, phishing attacks and port-scans.

taxonomy of network anomalies presented by Barford et al. [39]:

- *Malicious attacks*: attempts to make the infrastructure and services unavailable to legitimate users, such as a Distributed Denial of Service (DDoS) attack.
- *Legitimate but abnormal use*: legitimate demand for a service that is unusually greater than what the system is prepared to handle, e.g. flash crowds.
- *Measurement anomalies*: due to problems with the data collection, for example, loss of flow data due to router overload.
- *Misconfigurations/failures*: component faults due to software bugs, hardware failures as well as human mistakes leading to improper resource configuration.

The effects of a network anomaly can be typically observed in two manners: (i) through sudden changes in the expected volume of overall traffic carried in the network, or (ii) through sudden changes in the distribution of certain characteristics of the network traffic, such as addresses and ports (i.e., traffic features). Broadly, the anomaly detection techniques discussed in this paper fall in one of these categories.

2.3. Anomaly diagnosis problem spectrum

Due to the multidimensionality of the problem space, the construction of an anomaly diagnosis framework is complex. The goal of providing one unifying model to fully resolve all abnormal traffic phenomena efficiently and accurately leads to an ill-posed problem. This is mainly due to the numerous and sometimes conflicting requirements posed by a range of domains where such a framework should be used. The focus of this paper is on *backbone network anomaly diagnosis*, which we define as a composite, multi-dimensional problem that involves a range of heterogeneous domains. In this work we follow the descriptive scheme in [20] that enabled the decomposition of the anomaly diagnosis problem into four distinct dimensions, namely.

2.3.1. Processing cost

Costs in general are associated with different parts of the network and network operation. The cost dimension can be decomposed into two main sub-dimensions; i.e. system and network costs. The cost dimension also implies several other subdivisions such as economic and social aspects. Though, these are outside the scope of this paper.

There is a number of cost factors, which are generally divided into *fixed costs* (i.e. not directly depended on the operational level within a given period such as labor costs or equipment depreciation) and *variable costs* (i.e. costs directly linked to the level of operation). Within the anomaly detection domain, costs are often associated with the granularity and accuracy of the processing [40] and as such are directly related to the operation since they represent the additional overhead incurred through anomaly detection. Such costs are for instance related to monitoring, analyzing, detecting, and responding to anomalies. *Information costs* are defined as the time required for processing information and comparing them against the detection model [41]. Hence there is a direct performance-cost-trade-off between the effort of processing information and the increased accuracy that can be achieved. In [42], Zhanga et al. contrast the operational costs to the failure costs incurred by system failure due to attacks that are not detected and responded to in time (called response costs). Additionally, other cost factors such as the communication overhead have to be considered.

Within the scope of this paper we focus on *operational* (respectively processing) costs. The cost factors in this context are the costs for gathering information at a certain degree of granularity (monitoring costs), the costs for processing this information (analyzing costs), the costs for applying a certain detection model (detection cost). Additionally there are costs for reacting to anomalies (responds costs). For example we can consider the costs related to the sampling factor, which in practice determines the performance of the measurement process within an ISP. Consequently, an increased sampling factor increases the monitoring costs as well as the analyzing costs (since there will be higher processing costs). However, it will increase the detection accuracy and can therefore reduce the response costs or the subsequent costs associated with system failure. We discuss issues related to traffic sampling in Section 2.4.

Another aspect to consider is the location of the detection since there is an additional trade-off between local-system performance (e.g., a mechanism used only at one or two points within a network) and network-wide performance (e.g., in the case of a distributed instrumentation). Considering the latter it is clear that a network-wide mechanism also includes a performance trade-off between a single system within the network and the entire network. The relevant cost aspects in relationship with the other dimensions as depicted in the model in Fig. 1 are variable costs factors related to the operational costs. A higher *granularity* has a direct impact on the monitoring and analyzing costs. Naturally, a selected *methodology* has impact on the detection costs as expressed in the computational complexity of the different models. Moreover, the selection of *features* has impact on the monitoring and analyzing costs, and to some extend on the detection cost.

2.3.2. Diagnosis granularity

Granularity can refer to the detection of abnormal behavior only, or the classification of the abnormal behavior into known classes of network anomalies, or both. The level of granularity defined within an anomaly diagnosis component depends on the network performance costs as determined by a given organization or ISP. Anomaly diagnosis may exist in the form of a detection scheme that is only able to *detect* abnormal deviations from a "normal behavior" without being aware of the exact type of anomaly class (e.g. DDoS, alpha flow, etc.). A diagnosis component may also be intelligent enough and apply supervised, semi-supervised or unsupervised *classification* without the need of initially detecting an anomalous deviation. However, as it has been shown in past work [43],

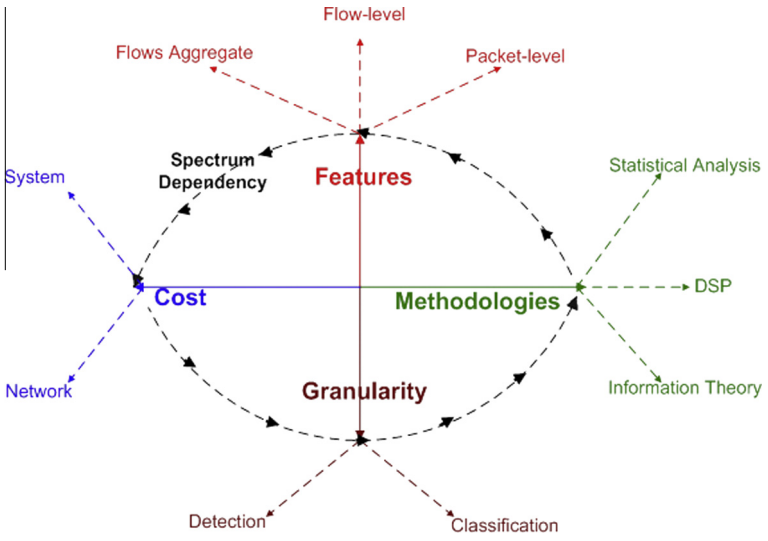

Fig. 1. Anomaly diagnosis problem spectrum.

this approach of direct classification results in higher processing costs since all traffic (anomalous or not) has to be passed to a supervised classification component. The two granularity sub-dimensions may also co-exist and provide a fine-grained diagnosis scheme. As shown in several studies (e.g. [1,7,8,39,44]) a mechanism involving a hybrid mathematical model using selected features may successfully employ detection and further classification.

2.3.3. Theoretical methodologies

Different anomaly diagnosis approaches commonly involve a blend of theoretical foundations, whose main properties are discussed in this paper. Important results such as [1,2,7,8,39,44–46] would never have been found without applying a combination of various mathematical domains. The three most used theoretical domains that compose the methodologies dimension are: *Statistics*, *Digital Signal Processing* (DSP) and *Information Theory* (e.g. entropy/mutual information estimations). In practice, multiple methodologies have been combined in order to better detect and distinguish anomalies. For instance, from a pure mathematics point of view, a signal processing technique (i.e. treating network data as a signal) for identifying high-frequency patterns (e.g. by using wavelets) also requires the use of standard statistical approaches such as timeseries analysis. On the other hand several statistical timeseries models (e.g. ARIMA) are dependent on DSP techniques (e.g. Fourier transforms) for representing their values over the frequency domain in order to achieve a clearer understanding of traffic trends within a time–frequency domain. Finally, there are also approaches that use all these sub-dimensions together, as in [44], where entropy timeseries of certain flow-features are used as an input to a Principal Component Analysis (PCA) method for further detailed classification.

2.3.4. Traffic features

A challenging issue in the design of an anomaly diagnosis component is the feature selection. The feature dimension can be decomposed into three subdivisions that represent the network-related traffic characteristics, in terms of *Raw Packet Information*, *IP Flows*, and *Flow Aggregates*. In general, by defining the diagnosis granularity and the theoretical methodology to be applied, there is an implicit decision on which level of aggregation network data should be collected. For instance, if a detection scheme is used to observe large deviations without classification, a collection of flow aggregates may be used as input to a threshold-based process for comparing volumes of traffic. On the other hand, a sophisticated component that includes both detection and classification would probably require certain statistical properties of particular network flow features (e.g. packet inter-arrival timeseries, src/dst port/address pair distribution) as used in [1,2,6,29,44].

2.4. Sampled network data in anomaly diagnosis

An important element within the formulation of anomaly diagnosis schemes relates to the sampling factor. Nevertheless, in our anomaly diagnosis problem decomposition we do not explicitly state the sampling factor as an independent dimension since it is a component that affects the other dimensions.

Packet and flow sampling techniques are widely used for the purposes of reducing the overhead in network monitoring and traffic engineering applications and they mostly rely upon methods embedded within network routers (e.g. Cisco's Netflow [47]). A number of studies [48,49] have shown that the level of granularity used for sampling affects the performance as well as the processing costs. Regardless of its simple implementation and low memory requirements, the authors in [50] illustrate that traditional and adaptive packet sampling schemes result in inaccurate interpretations of flow statistics such as the original flow size distribution [51]. Furthermore, the study in [50] exhibits the pitfalls of flow sampling with respect to high CPU usage and prohibitive memory requirements. Thus, from a pure network monitoring viewpoint, the sampling

factor has a direct relationship with the *processing cost* and the *traffic features* dimensions.

From an anomaly diagnosis perspective the authors in [51–53] show that packet sampling methods engage several trade-offs regarding the detection of anomalies. The thorough analysis performed on real un-sampled backbone flow data in [52] stated that several volume-based flow metrics (e.g. packet, byte counts per flow) seem to be an insufficient input for an anomaly detection method, thus biasing the *granularity* of accurately detecting and categorizing a given anomaly. On the other hand, meta-statistics composed by some non-volume metrics such as the entropy of the destination IP addresses distribution are not affected by the sampling scheme, hence they can adequately identify both volume and lightweight anomalies. Moreover, the work in [51] shows that several flow-based sampling schemes (e.g. random flow sampling, smart sampling) degrade the detection performance of anomaly diagnosis methods for either volume-based anomalies (e.g. DDoS) or lightweight anomalies such as port scans. Hence, sampling performed either on a flow or packet basis surely affects the dimensions of *granularity* as well as the *cost* invoked by the algorithmic aspect of the *methodology* used for an anomaly diagnosis component. When raw features are flow aggregates there is only a minimal processing cost since the dataset to be pre-processed is already summarized. The source of intensive processing in this case arises from the basic flow statistics (e.g. mean flow interarrival time) that need to be extracted from the overall aggregate. In general, aggregate records exist in smaller amount compared to individual flows.

2.5. Discussion

In Fig. 1 the dependency between the dimensions of *Processing Cost, Diagnosis Granularity, Theoretical Methodologies* and *Traffic Features* have been depicted in an anti-clockwise direction. Starting from the processing cost, the argument is that this dimension is directly dependent on the diagnosis granularity that a possible framework is desired to achieve. Subsequently, techniques that aims to achieve only detection or classification or both are required to follow certain theoretical methodologies. There is a number of hybrid methodologies employing various mathematical principles from different areas. For simplicity reasons the diagram includes as sub-dimensions the de facto areas that have been frequently used in the literature, namely *Statistics*, *Digital Signal Processing* (DSP) and *Information Theory*. Apart from the granularity and cost requirements, methodology selection is also dependent on which features are used. Again, feature selection can be related back to the cost requirements since based on the type of data to be processed, network or system-wide costs change. Although we do not consider sampling as an independent dimension, as described in the previous section the dimensions that we identified also relate to the sampling factor.

In the remainder of this paper we concentrate on the dimensions of theoretical methodologies and traffic features. These form the core of a conceptual framework for anomaly diagnosis and for this reason we discuss the most influential work in these research areas. Our aim is to provide the reader with an understanding of the most fundamental concepts related to methods and techniques for anomaly diagnosis. We also provide an extensive set of references to the most influential work in this area in the last decade.

3. Theoretical methodologies

Existing solutions for anomaly detection and classification on IP networks rely on techniques derived from the three theoretical domains discussed earlier. Despite the various commonalities embodied within those domains, this section makes an effort to discuss them separately.

We provide a brief description of each domain before discussing them in depth:

- *Statistical Methods:* mathematical schemes that capture characteristics, occurrences and temporal trends in order to profile a process and capture specific dynamics (e.g. network anomalies) heavily rely on statistical methods. In fact, other methodologies (such as signal processing and information theory) have at their foundation theoretical elements from statistics (e.g. mean, variance, autocorrelation sequence, etc.). However, there are many pure statistical techniques and models solely used to detect anomalous characteristics and further forecast the behavior of the processes examined. Pure statistical methods for anomaly diagnosis are mostly used for determining the accuracy of an anomaly detection framework (e.g. histogram-based, maximum likelihood estimation, residual analysis). Nevertheless, in most cases statistical methods are employed in a combined manner for characterizing and forecasting network behavior in order to reveal anomalous trends (e.g. Auto Regressive Integrated Moving Average – ARIMA, Analysis of Variance – ANOVA) as in [1,20,64].
- *Digital Signal Processing:* is used to represent network traffic as signal components that can be processed independently. Typically, a signal is converted from the time (or space) domain into the frequency domain, e.g., by means of a Fourier transform. This step transforms the original signal data into magnitude and phase components for each individual frequency. In the frequency spectrum, the signal can then be divided into its high, medium and low frequency parts, with each part being further divided as needed. Short term anomalies (for example, equipment failures and malicious attacks) can be identified by filtering out the low frequency components that represent the predictable part of the signal. This permits the isolation of anomalous signals in the traffic data through the analysis of signal properties, e.g., through the detection of a sharp increase in the variance of the data [10,39,55].
- *Information Theory:* is based on the fields of probability theory and statistics, and involves the *quantification of information*. One of the key measures of information is known as *entropy* representing the

Table 1
Exemplar uses of the three main theoretical methodologies through various techniques: statistics, digital signal processing and information theory.

Statistics	Digital signal processing	Information theory
Flow aggregate		
Linear modeling [45]	Kalman filters [45]	Compressive sensing [24,33]
Subspace/PCA [22,44,54,55,36]	Wavelets [39]	Shannon entropy [22,24,33,44,54]
Sketches [54]	Fourier [36]	
Multivariate quantiles [35]		
ARIMA [33,36,56]		
Flow		
PCA [1]	Wavelets [1,13,36,39]	Renyi entropy [20]
ARIMA [1,20]	Bispectrum [20]	Shannon entropy [22,54,57,58]
Point processes [34]	Fourier [36]	Tsallis entropy [59]
Histograms [6,60,61]	Change detection [62,63]	
Markov processes [12]		
ANOVA [64]		
Packet		
AR/ARIMA/ARFIMA [65–67]	Wavelets [64,65]	Maximum entropy [68]
Inverse distribution [69]	Power spectral density [53,70]	Shannon entropy [53,58,71,72]
		Conditional entropy [72]

amount of uncertainty associated with a value of a discrete random variable [73]. The entropy is maximized when all values have the same probability of occurring, i.e., the greatest degree of unpredictability. The use of entropy estimation for anomaly detection relies on the principle that certain types of network anomalies will (meaningfully) disturb the distribution of traffic features (e.g. source/destination ports, source/destination IP addresses) [6,22,44,53]. In this manner, significant changes to the entropy of destination ports observed in the traffic data might for instance indicate the occurrence of a port scan attack, whereas changes to the entropy of source/destination IP addresses might indicate an ongoing DDoS attack.

Table 1 outlines significant efforts in these areas. It indicates the use of each methodology, applied to different granularities of data, particularly in the context of packets, flows and flow aggregates. The following sections describe the three main theoretical methodologies in greater detail, accompanied by exemplar techniques and theoretical models. The interested reader can also find further information on the use of specific techniques for anomaly detection in the referenced literature.

3.1. Statistical analysis

Statistical data analysis is the basis for the theoretical foundations introduced earlier. The explicit task of anomaly diagnosis that includes both detection and classification schemes involves the use of both descriptive (i.e. mean, standard deviation and frequency-descriptive) and inferential (i.e. hypothesis testing, regression, timeseries) statistics. For instance, traffic classification and attack-related traffic identification with the use of Support Vector Machines as presented in [43] initially involves the frequency representation (i.e. how many times a feature occurred within a particular time-range) under a hypothesis on whether a particular feature (e.g. byte counts, payload size) belongs to a known application. Furthermore, anomaly-specific classification techniques involving unsupervised clustering as presented in [44] involved the clustering of unlabeled time-measured PCA metadata in order to determine the most suitable clusters for several types of anomalies. From a general viewpoint, the majority of statistical approaches require a level of hypothesis testing in order to establish a ground truth with respect to the normal behavior of the network and to further determine decision thresholds [45]. Moreover, there have been several propositions that involved a range of statistical hypothesis tests such as the goodness of fit Kolmogorov—Smirnov test [12] in order to validate *de facto* statistical assumptions (e.g. stationarity) for a given distribution that was describing a selected network feature (e.g. packet count). The list of detection/classification methodologies is extensive, therefore this subsection concentrates on *de facto* methodologies mainly used in inferential statistics.

Timeseries analysis used in the context of stochastic processes (i.e. random processes) may take several forms and is considered as the main component for *temporal* anomaly detection [7,19,45,20,27,63,66]. Obviously each case differs and there is a broad direction in analysis using timeseries ranging from examination of serial dependence between data elements (i.e. autocorrelation) up to spectral analysis determining cyclic models (i.e. seasonality) on a particular data pattern within time. In the past years, several models were introduced, targeting specific domains such as prediction and forecasting. Particularly for IP networks such models were used with high rates of detection accuracy of over 90% in general [66,63].[2] In particular the work in [63] exhibits a reduction of false negatives and false positive ratios with respect to their anomaly detection rate performance under Autoregressive Integrated Moving Average (ARIMA) models when compared with the Exponential-Weighted Moving Average (i.e. EWMA) technique that they assessed. Also, the work in [66] has demonstrated the benefits at accurately detecting and forecasting anomalous

[2] As also discussed throughout this paper, we clarify that the level of accuracy actually depends on the use case, specific methodology and techniques employed within a given study.

events on a backbone link. The authors in [66] achieved to identify and further forecast a number of 37 anomalies out of the 39 which were pre-labeled by the network operators who provided the dataset used in the experimentation process. Thus, their proposed scheme seemed to achieved a relatively high forecasting accuracy of approximately 95% on detecting various types of anomalies and as also used in subsequent studies (e.g. [7,19,20,27]) it has proved to be a robust, reliable and efficient method.

Modern characterization, prediction and forecasting models are based on the use of Autoregression models (AR) and Moving Average (MA) models, also known as ARMA models. In general, an AR process can fill in the gaps of the MA process and vice versa, thus their joint use is common. The granularity of the anomaly detection is dependent on the raw features used in this particular scheme. However, ARMA and ARIMA models perform better on *temporal* approaches that target the forecasting of heavy volume-based anomalies [27,63] and they were mostly employed to operate a close-to real-time mode. According to [27] the ARIMA formulations aid the initial detection and further forecasting of volume-based attacks (e.g. DDoS). However, as indicated in [63] a critical aspect for the efficiency of ARIMA models lies with the adequate selection of appropriate packet header meta-features (e.g. distribution of source/destination pairs or combinations as "sketches" of traffic as in [63]). Hence, the aspect of granularity that relates with the detection and prediction of lightweight anomalies has a direct relationship to the appropriate selection of traffic features as we have already discussed in Section 2.3. Moreover, both [27,63] managed to implement ARIMA models with minimal processing costs and overhead when deployed on a close-to real-time operation.

Nonetheless, since we intend to present the basic formulation of either the ARMA of ARIMA methods, we begin with their first component, the AR process, represented as:

$$\chi_1 = \xi + \Phi_1 * \chi_{(t-1)} + \Phi_2 * \chi_{(t-2)} + \Phi_3 * \chi_{(t-3)} + \ldots + \varepsilon \quad (1)$$

where ξ is a constant, $\Phi_1,\ \Phi_2,\ \Phi_3$ are the AR model parameters, and ε is a random error component also known as random shock.

The above formula shows that under the assumptions of stationarity in most network packet or flow timeseries consist of serially dependent measurements and there is the flexibility of estimating a coefficient or a set of coefficients describing those consecutive measurements. In simple terms, it shows that every observation is made up of a random error caused by the measurement process and a linear combination of prior observations of that process [19,27].

A process that fills the gaps of AR is the Moving Average (MA) model, which is independent of the autoregressive process. In practice the MA factor denotes that each packet or flow measurement in the series may also be affected by the past measurement error (in our case ε) that is neglected by the AR component. Therefore an MA process may be represented as:

$$\chi_1 = \mu + \varepsilon_t - \theta_1 * \varepsilon_{(t-1)} - \theta_2 * \varepsilon_{(t-2)} - \theta_3 * \varepsilon_{(t-3)} - \ldots \quad (2)$$

where again μ is a constant and $\theta_1,\ \theta_2,\ \theta_3$ are the model parameters. As we have already mentioned both models are essential to operate on timeseries that satisfy certain stationarity requirements. For further details refer to [74].

Another joint scheme of Autoregression (AR) and Moving Average (MA) models is the ARIMA (AutoRegressive Integrated Moving Average) model. In particular, an ARIMA model is a forecasting technique that captures the dependency of future network packet or flow values based on current and past measurements as described in [7,20,27,63]. ARIMA models in conjunction with single and cross-spectrum analysis using Fourier transforms as well as with certain techniques involving smoothing (i.e. Holt-Winters, Exponentially Weighted Moving Average, Simple Smoothing) are specifically designed to detect deviations and observe cyclic patterns in timeseries of random network or packet-level measurements. An ARIMA model is defined in terms of three parameters: the autoregressive parameter (p), the number of differentiation passes (d) and the moving average parameter (q).

Hence, a general ARIMA(p, d, q) model is defined as:

$$z_k - \sum_{i=1}^{p} \Phi_i \cdot z_{k-1} = \varepsilon_k - \sum_{i=1}^{q} \theta_j \cdot \varepsilon_{k-1} \quad (3)$$

where z_k is obtained through the differencing of the original timeseries d times if $d \geqslant 1$, or by subtracting the mean from the original timeseries if $d = 0$, ε_k is the forecast error at time k, $\Phi_i (1 \leqslant i \leqslant p)$ is the autoregressive coefficient, and $\theta_i (1 \leqslant j \leqslant q)$ is the moving-average coefficient. ARIMA models can be applied to obtain a prediction of z_k. The error ε_k is obtained by subtracting the prediction of z_k from the actual z_k. In the end, this error represents the anomalous traffic, indicating that any traffic behavior that does not conform to the model is considered anomalous [19,27]. However, the identification of the error from the initial graphical representation is hard and usually requires the use of certain explanatory techniques involving the Autocorrelation Function (ACF) as well as the Partial Autocorrelation Function (PACF) [74] as indicated in [21].

A popular formulation within the realms of probability theory and inferential statistics is the representation of a network as a Hidden Markov Model (HMM) [7,8,75]. The intuition behind this approach is to map the abnormal traffic state as one of the possible states that a network (modeled as a Markov process) may be at a particular time unit. A Markov process assumes that the system under test satisfies the *memoryless* Markov property, i.e., its future state is not dependent on its past but only on its present state. For instance if we let an IP network flow be defined as a stochastic process X where its state changes in time t (i.e. $X(t),\ t \geqslant 0$) for every $n \geqslant 0$ time points $0 \leqslant t_1 \langle t_2 \langle \ldots \langle t_n \langle t_{n+1}$ and holds visible states as $i_0, i_1, \ldots, i_{n+1}$ it fulfils the following property also known as Markovian property:

$$P(X(t_{n+1}) = i_{n+1} | X(t_n) = i_n,\ X(t_{n-1}) = i_{n-1}, \ldots, X(t_0) = i_0) \quad (4)$$

and (4) is simplified as

$$= P(X(t_{n+1}) = i_{n+1} | X(t_n = i_n)) \quad (5)$$

This formulation is complimentary to what we referred to earlier about the fact that the future network flow state is not dependent on the past flows (i.e. events) but only on

the current state. Therefore, knowledge of the network's history does not add any valuable information.

In the specific case of detecting hidden anomalies a HMM employs probability measurements throughout each transition from one state to another involving a detection rate. The main difference with a traditional Markov process is that the observable state $X(t_n)$ is also dependent on a hidden, unobservable state $Y(t_n)$ also satisfying the Markovian property. In addition, a HMM enables the estimation of a range of state transition probabilities (usually a multi-dimensional matrix) that is sensitive enough for extracting at exact timing a hidden unobservable state. This unobservable state may be considered as a possible network anomaly but it however requires great processing time as stated in [7,8,75]. The pitfall with respect to the processing time is derived due to the computation of the state transition matrix in the case where a full hidden Markov model has to estimate the values for multiple flows, hence, the greatest number of flows the larger the processing time. The work in [7] emphasizes this constraint by discussing the outcomes on several pieces of work that had to employ HMM modeling for anomaly detection in computer systems [76] and networks [77].

3.2. Digital Signal Processing (DSP)

The use of DSP in anomaly detection provides major advantages with respect to the analysis of discrete time-signals. Mapping several DSP techniques on IP networks has produced several solutions in the context of applications related with traffic characterization and anomaly diagnosis. For instance, work in [20] demonstrated that the highly non-stationary properties of Internet's traffic volume (i.e. bytes/packets timeseries) on backbone and access links could be adequately characterized on the Time–Frequency (TF) plane by the bispectrum tool [78] where its explicit ability at properly mapping signal phase transition peaks could precisely determine and forecast the exact timing and amplitute of volume peaks and also pinpoint volume-wise "lightweight" anomalies such as malware. Moreover, the work in [79] provided a traffic classification scheme that apart from classifying application-layer protocols (e.g. HTTP, FTP, etc.) it could also classify up to a 96% of overall accuracy anomalies on known ports and also packets with particular payload information (.e.g malicious packets) with the usage of the Cohen-based energy TF distributions [80]. DSP is a multi-disciplinary domain merging statistics, mathematics and physics. Techniques derived from DSP have proven to adequately confront the challenges of anomaly detection and be in a position to pinpoint both volume-based anomalies (e.g. DDoS attacks [10,20,39,45]), as well as lightweight anomalies (e.g. port scans, worms, malware [7,8,20,45]).

DSP is used to represent network packets or flow measurements as a discrete signal within the time, frequency, spatial, autocorrelation and time–frequency domains. Critical issues involved in all these domains relate to the correct choice of the measurement's sampling and the selection of appropriate filters. However, we will not delve into the specifics of these issues but rather provide an overview of how DSP is applicable to IP networks for the tasks of traffic modeling and anomaly diagnosis.

A fundamental property of all the DSP domains is the continuous Fourier Transforms (FTs) and their algorithms for fast computation. Since FTs were firstly introduced for analog signals, the Discrete Fourier Transform (DFT) which is a special case of the Z-transforms represents a network flow or packet measurement as a complex number of length N within only the frequency domain (and not the time domain), defined as:

$$X(k) = \sum_{k=0}^{N-1} x(n) e^{-j2\pi kn/N} \quad k = 0, 1, 2, \ldots, N-1 \tag{6}$$

DFTs are not able to indicate clearly frequency shifts over time due to the fact that a signal's (in our case network measurement) phase is hidden by the plot of the DFT magnitude spectrum. For this reason, a more flexible approach is considered to be the Short Time Fourier Transforms (STFT). The intuition behind the STFT is to enable a time–frequency distribution for a given network measurement by using a specific sliding time window function $g(t)$ resulting in a two-dimensional time–frequency function $X(t,f)$ as shown in Eq. (7):

$$X_{STFT} = X(t,f) = \int_{-\infty}^{+\infty} x(t') g(t - t') e^{-2\pi i f\ t} dt \tag{7}$$

Even though the STFT overcomes the limitations of the DFT, it is still inadequate regarding the resolution provided for each time–frequency segment under evaluation. This is mainly due to size-limitations of the windowing function employed. From an anomaly detection perspective, such a limitation leads to an inaccuracy with respect to the exact timing in which an anomaly occurs based on its frequency distribution. That is, it might be observed but cannot be precisely localized, neither in the time nor frequency spectrums of its representative spectrogram. According to [81] the STFT in the context of anomaly detection is limited due to its strong dependency in its windowing function that can either give a good resolution on identifying the frequency of the anomaly or provide a good timing resolution and indicate its exact initiation on the time plane. This is the main reason why wavelet-based approaches (which in theory are considered as an extension of STFTs) have been proposed and are nowadays widely used in the domain of anomaly diagnosis [1,13,39,65].

Techniques based on wavelet analysis [1,13,39,65] are applicable for decomposing the traffic measurement signal into a hierarchy of component signals, formed by low, mid and high-frequencies. Low-frequency signals represent the slow-varying portions of the original signal, whereas high-frequency signals represent more spontaneous variations in the signal. This particular methodology has demonstrated promising results with respect to real-time operation due to its intrinsic properties at quickly decomposing the traffic signal under small processing cost [1,13,39,65]. In particular, the work in [39] employed wavelets in conjunction with a seminal deviation score algorithm on semi real-time scenario in order to determine the thresholds related with normal activity and further distinguish the various anomalous traffic patterns that seemed to have close to similar traffic characteristics. As the authors claim,

this proposed synergistic approach did not pose high levels of processing cost.

As described in [39], wavelet processing can be divided into two complementary steps:

- *Analysis:* iteratively extracts the hierarchy of derived signals by applying filters to the original network measurement. For each iteration, the input is a signal x of length N and the output will be two or more derived signals of length $N = 2$. Filter L has a smoothing effect and $L(x)$ corresponds to the low-frequency output. A number of filters H_1 to H_r with $r \geqslant 1$ are used to obtain the high-frequency parts of the signal. In the following iteration, $L(x)$ is further decomposed into $L^2(x), H_1L(x), \ldots, H_rL(x)$, and the algorithm continues in a similar manner for the subsequent iterations. Each iteration j will thus increment the number of times the low-pass filter is applied to the signal, i.e. decomposing the signal into $L^j(x), H_1L^{j-1}(x), \ldots, H_rL^{j-1}(x)$, and therefore recording each time a smoother part of the signal (lower frequency).
- *Synthesis:* performs the inverse of the analysis, where the input of each iteration is $L^j(x), H_1L^{j-1}(x), \ldots, H_rL^{j-1}(x)$ and the output is the signal $L^{j-1}(x)$. Typically, some of the derived signals are altered before reconstructing the original network measurement again, for example, to eliminate low-frequency components that may not be relevant for the analysis.

Techniques based on distributed signal processing promote the idea of decomposing network measurement in a distributed fashion and then re-synthesizing it for further analysis on a centralized facility [65]. In general such approaches involve PCA-based techniques that are the discrete implementations of the Karhunen–Loeve Transform [88] and Singular Value Decomposition (SVD) as in [22]. Additionally, the broad area of DSP offers tools related to signal filtering such as the Kalman filter [89], which has been successfully used in the context of filtering normal behavior and extracting the abnormal spectrum. Further discussion on the use of PCA and Kalman filters for anomaly detection (as in [22,44,45]) is presented in Section 5.4.

3.3. Information theory

By combining the disciplines of applied mathematics and engineering, information theoretical approaches tend to model traffic behavior based on the information content of a complex system such as a networked environment. The information content is produced as the opposite of the uncertainty measure (i.e. entropy estimation) assigned to a random variable (in our case network flows or packets/bytes of a flow) within that system.

Traditionally the concept of entropy has its roots in thermodynamics. However, Shannon in his pioneering work in [73] has provided a different entropy formulation that was initially dealing with symbols and alphabetical characters in an information theoretical manner. The newly proposed Shannon's entropy was the first building block of the domain of information theory. Based on Shannon, we assume that X corresponds to a finite number of possible values $[x_1, x_2, \ldots, x_n]$ where each value possesses a probability value denoted similarly as $P = [p_1, p_2, \ldots, p_n]$. The goal is to find a number that will measure the size of uncertainty. Hence, the function is expressed as the uncertainty related to an event $X = x_i$ where $i = 1, 2, 3, \ldots, n$. For each n there is a corresponding uncertainty function $a(p)$ and therefore the function $A_n(p_1, p_2, \ldots, p_n)$ states the average uncertainty for the range of all the finite random values in X.

Given the definitions in [73], the well-known Shannon entropy expression is defined by:

$$A_n(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{8}$$

With respect to anomaly diagnosis, the traditional logarithmic based Shannon entropy, as well as some of its variations (i.e., mutual entropy, sample entropy, Rényi entropy [90]) have contributed to a great extent on characterizing the evolution and intrinsic behavior of particular flow features (e.g. src/dst IP address, payload) and networks faults [57–59,68,91]. In fact, the work in [44] has successfully indicated that certain packet features compose distinct entropy distributional values on different types of anomalies (i.e. for a Denial of Service anomaly, src/dst IP addresses and src/dst port) resulting as an input for a more accurate classification. In most of the proposed solutions, the actual entropy computation was undertaken in an offline mode by using a sample of traffic data and it was feasible to identify a plethora of lightweight anomalies that could have not been revealed by simply observing the distribution of the traffic volume (i.e. the distribution of packet and/or byte counts on the backbone links). Overall, the cost at producing entropy metrics relates with the selection and mapping of a given probability density function (i.e. pdf) on the examined network feature which in practise is the building block for the computation of entropy as we following illustrate.

Information theory is an evolving field that reaches beyond the standardized Shannon foundations and also involves non-logarithmic entropies. For instance, recently the use of entropy spectrums containing distributional behavior of flow features has been introduced [59]. In particular this is based on the Tsal lis Entropy [92], which is a generalization of the well-known Boltzmann–Gibbs entropy and is defined as:

$$S_q(p) = \frac{1}{q-1}\left(1 - \sum_x (p(x))^q\right) \tag{9}$$

where $q \to 1$ recovers the original Boltzmann–Gibbs entropy.

As presented in [59], this particular type of entropy in conjunction with generalized entropy metrics shows greater sensitivity to lightweight anomalies, such as worms and port-scans, in comparison to traditional Shannon-based entropy metrics.

3.4. Towards spatial, temporal and spatio-temporal approaches

By definition, *temporal* analysis of anomalies aims to identify *when* an anomaly has occurred within an examined

traffic trace, by either investigating deviations of volume statistics (e.g. counts of bytes/packets) or by analyzing the distributional behavior of selected packet features (e.g. src/dst IP addresses). On the other hand, *spatial* approaches aim to identify *where* an anomaly has occurred. Finally, *spatio-temporal* techniques enable the identification of the exact time and location instant of an anomaly within a network. Table 2 illustrates the relationship between existing solutions and the theoretical methodologies discussed earlier, and their applicability either for a *spatial*, *temporal* or *spatio-temporal* solution.

According to Table 2, it is clear that the majority of techniques for anomaly detection are dependent on the foundational domain of statistics. For instance, the Hidden Markov Models used in [75] are based on probability theory, which is considered as a sub-domain of statistical analysis. Moreover, several formulations such as Support Vector Machines (SVMs) and Bayesian Networks used in [43,83,85,82], respectively, are representatives of supervised Machine Learning (ML), but also belong to the realms of statistical inference analysis.

In summary, Table 2 provides a snapshot of the hybrid techniques used within the anomaly detection community throughout the years. Furthermore, it indicates that anomaly detection itself is a complex issue that requires sophisticated techniques that are formulated by a combined use of different theoretical foundations. Undoubtedly, each formulation corresponds to a particular solution path either by observing at a temporal, spatial or spatio-temporal level of the traffic process.

3.5. Discussion

Considering the anomaly diagnosis problem spectrum, this section briefly presented the three theoretical domains that can be employed in the construction of approaches for the detection and further classification of network anomalies. These are summarized in Table 3. Neither of the these theoretical domains should be remarked as superior to any other, but rather they should be considered complementary, and Section 3.4 acknowledged that typically anomaly diagnosis requires their combined use. Due to this interdependency, it is crucial for an anomaly diagnosis approach to blend these core techniques together. Thus, the use of a pure statistical technique will often require the distinct properties of a DSP-based approach under the intention of eliminating noise components that are anticipated to be present in the raw data. Furthermore, an information theoretic approach should consider the statistical properties of the initial traffic distribution in order to correctly approximate entropy representations.

Table 2
Relationship between methods and the foundational theoretical domains in some exemplar anomaly diagnosis studies; ST = Statistics, DSP = Digital Signal Processing, IT = Information Theory.

Method	Theoretical domain
Temporal	
EWMA/Fourier/Wavelets [1,13,36,65]	ST, DSP
ARIMA/Holt-winters [27,39,66]	ST, DSP
Bayesian networks [82–84]	ST
Support vector machines [83,85,43]	ST
Spectral analysis [86,87]	ST, DSP
Hidden Markov models [7,75]	ST
Wavelets [13,39,65]	ST, DSP
Spatial	
Entropy/PCA [22,44]	ST, IT, DSP
PCA [22,44,55]	ST, DSP
GQS (ST) [35]	ST
Spatio-temporal	
Sketch/change detection [63]	ST
Markov modulated process [12]	ST, DSP
Graph-wavelets [64]	ST, DSP
Kalman filter, ROC curves[45]	ST, DSP

4. Traffic features

Feature composition is subject to several sub-processes, as presented in Fig. 2. Typically any detection mechanism interacts with a measurement facility that collects sampled network monitoring information. Subsequently (and normally offline) data mining is then used in the transformation of network monitoring information into statistical features that are then used to generate desired statistical metadata. There is a plethora of techniques that focus on the selection of network features that cover a large area in statistics as well as graph theory. For instance, work in [93] aims to detect botnets using Internet connection graphs and by strictly considering a graph-theoretic approach for mapping the behavior of source/destination pairs. Nevertheless, for the purpose of this paper we present the pre-processing aspect from a statistical approach perspective.

As already mentioned, the pre-processing stage includes the selection of the statistical methods to be used to process this information. The optimal set of traffic features must be identified. A larger number of features not necessarily ensures better anomaly detection, as some additional features might introduce noise in the analysis process. Typically, a recursive "trial-and-error" validation scheme is used to identify the usefulness of a set of features. Features extracted from raw data that do not meet the algorithmic requirements are discarded, and in this case the pre-processing step is restarted. Statistical feature selection is then re-considered based on the selected detection/classification algorithms. This step is purely dependent on the rest of cost and granularity requirements (as discussed in Section 2.4). Most commonly, metadata used by existing anomaly detection techniques is referred to as traffic features in the literature, and initially consists of simple volume measurements (e.g. counts of bytes/packets per second) captured at dedicated points on either access, peering or transit backbone links. Unfortunately and as reported by several studies (e.g. [22,25,44]) these measurements cannot be used as a direct observation of the traffic rate between Origin–Destination pairs (i.e. ODs) or Points of Presence (i.e. PoPs), thus for efficient anomaly diagnosis it is essential to follow sophisticated pre-processing schemes as well as fine grained solutions for optimizing the ill-posed traffic matrix problem [24,25,33,94].

This section describes in detail how packets, flows and aggregates are transformed into meaningful numbers for

Table 3
Summary of the three theoretical domains used in anomaly diagnosis approaches.

Statistics	Digital signal processing	Information theory
Description		
Looks for a network anomaly based on the variance or deviation of particular samples within discrete time bins, with respect to monitored traffic trends	Represents traffic data as discrete signals with different frequencies, and looks for an anomalous signal frequency by initially identifying the frequency-wise normal patterns	Calculates mutual information or entropy values and identifies anomalous distributions of certain traffic features in the observed traffic samples
Usage examples		
Traffic spikes representing a sudden deviation from the expected amount of traffic seen in a particular link might indicate heavy volume anomalies, such as DDoS or flash crowds	Typically low-frequency signals represent the normal behavior which varies very slowly, whereas high-frequency signals might indicate the temporary effects of a DDoS attack	Abrupt changes in the distribution of destination port addresses in the observed traffic might indicate, for example, a port scan being carried as part of a worm attack
Limitations		
Under the intention of profiling the normal patterns of the traffic process their filtering capabilities are limited regarding the extraction of lightweight anomalies (e.g. port scans)	Has several constraints with respect to the identification of the exact instant that an anomaly occurred due to its dependence on representing the traffic signal mainly on the frequency domain rather than on the time domain	Assumes statistical properties for the timeseries of a traffic-related feature (e.g. Gaussianity), which may result in inaccuracies

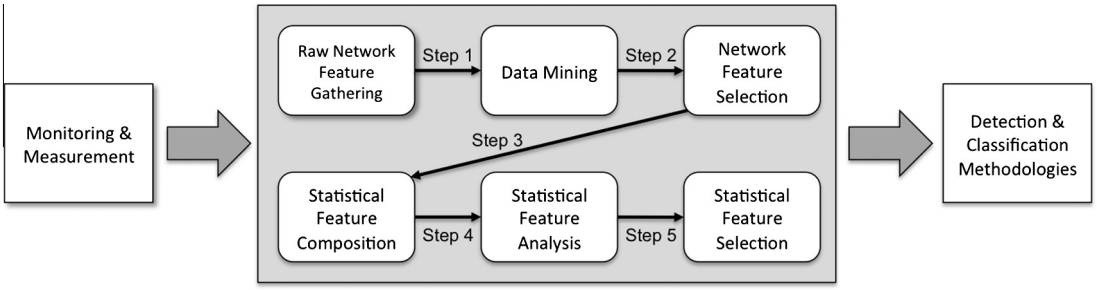

Fig. 2. Sub-processes included in the overall process of feature composition.

a given mathematical process. Most of the literature in the area largely avoids this topic (with a few exceptions, e.g. [62]). Typically, the technical difficulties and issues in the pre-processing stage are not explicitly described, although authors usually tend to refer to the type of raw network data used. For instance, the work reported in [39] used flow-level information of both IP flow data as well as SNMP. However, the signal analysis that has been performed on both types of data has shown differences, sometimes resulting in contradictory outputs on detection and further anomaly categorization. Similarly, the work presented in [22] used link measurements under SNMP, whereas in their follow-up work [29] the same authors emphasized some of the limitations of SNMP (e.g., loss of information due to its UDP dependency) and therefore suggested the use of Cisco's Netflow data. Additionally, the work presented in [56,63] favored data collection in a Netflow format. Furthermore, research efforts reported in [86,87] have performed data collection and further deep packet inspection using well-known monitoring tools such as tcpdump [95], and also using DAG cards [96] for hardware-based monitoring. Apart from these standard techniques, there are several other types of specialized tools for application-specific (e.g., P2P, VoIP traffic) as well as protocol-specific (e.g., BGP) monitoring. Anomaly detection approaches as presented in [56,84] have included information gathered from such specialized monitoring modules in order to take into account traffic features from higher layers in the protocol stack. However, for the purposes of this document, we describe strictly pre-processing approaches for standard monitoring and measurement, mainly dealing with SNMP, Netflow and pcap records.

4.1. De facto raw information

Raw information pre-processing is an important factor towards the initial composition of basic statistics regarding the network overall behavior. Next, the pre-processing aspects in different granularities of raw information are described in detail. The next section will described how statistics either on a packet, flow or flow aggregate level can be used to establish basic datasets for a given diagnosis methodology.

4.1.1. Flow aggregates

In general, the decomposition and pre-processing stage of SNMP information is fairly different from how Cisco's Netflow [47] or pcap records are processed. As described in [97], SNMP records are initially dependent on raw pcap

records that are mostly captured by tcpdump [95]. Even though the aggregation of pcap records into a more coherent SNMP representation enables better standardized traffic visualization, anomaly detection may become more difficult explicitly in the case of high-speed networks. For instance, the Multi Router Traffic Grapher (MRTG) [98] allows the visualization of collected network information using Object Identifiers (OIDs), as provided by SNMP protocol's Management Information Bases (MIBs). However, this procedure conceals certain protocol specific behaviors. Apart from this and from the UDP-dependence posed by SNMP, there is another shortcoming related to the sampled monitored traffic: in tools such as MRTG, the sampling time interval is standardized and hard to re-configure if needed (i.e., fixed five minute time bins). Consequently, such standardized sampling is likely to lead to post-processing limitations. Given these reasons, the authors in [39] complemented their use of SNMP link measurements with IP-flow information, in order to extract protocol and application-specific timeseries on bytes and packet counts.

4.1.2. Flow records

Cisco's Netflow permits packet and flows aggregation, but it is also dependent on UDP or the SCTP protocol and it poses certain disadvantages on enforcing in-depth anomaly detection due to the unreliable state of UDP and the limitation of SCTP at confronting scenarios where multiple routers need to interact with multiple Netflow data collectors. In general, Netflow is enabled on an interface-by-interface basis meaning that products that analyze Netflow data primarily do so by presenting per-interface reports. Often, during the pre-processing stage, it is needed to understand how the network actually carries the traffic between the two end points of the flow reported. However, since traffic may pass through many hops in its routed path from one end-point to another, and since routing paths may change dynamically over time, understanding the routed network's role in delivering the end-to-end traffic reported by Netflow is often difficult. This is due to the possibility that Netflow may not be enabled broadly enough to know which flows crossed which interfaces.

4.1.3. Packet records

Being the most basic monitoring and measurement record, pcap is the resulting format of directly monitoring full packets on a network interface. Through the use of a dedicated API (libcap [95], winpCap [99]), such records are most commonly manipulated and further processed by tools such as tcdump, ipsumdump [100], CoralReef [101], Wireshark [102] and several other open-source and commercial products. Even though these post-processing pcap tools provide deep inspection and sophisticated sniffing for every packet transmitted within a network, they pose high computational requirements. Since pcap provides a packet-by-packet representation without any grouping into flows, its pre-processing involves more steps when compared to Netflow or SNMP-based approaches. Network engineers and researchers must specify feature-related requirements via several data mining techniques (i.e., filtering, flow grouping, extraction of particular basic statistics) either manually (e.g., under shell scripts) or with the use of pcap processing tools as mentioned earlier. By considering the fine-grain information present in pcap records, a concern is the potentially large size of the files it generates. Thus, whilst monitoring a large backbone network or even a single high-speed link, the processing of pcap records may require a long time and result in high processing costs.

4.2. Raw features vs. statistical interpretation

Anomalies display certain characteristics which can be represented via meaningful statistical metadata resulting from the pre-processing stage as discussed previously. Within the pre-processing stage unimportant raw data is filtered out. Further, it also provides general data statistics such as the average packet size per flow, packet inter-arrival times and flow inter-arrivals. Anomaly identification requires sophisticated processing to estimate by how much a particular statistical feature deviates from the data collected. A mathematical process is used to determine which model the traffic may approximately fit in order to distinguish normal from abnormal characteristics. Unfortunately, there is no universal formulation that can be employed in all networks. Due to this diversity, often initial statistical features (e.g., mean, standard deviation, variance, covariance) need to be transformed into different functions in order to be suitable to a given model.

Signal processing techniques [25,45] assume that traffic may be generally modeled as a linear state space model having Gaussian behavior. In practice, most of the initial statistics produced from large networks promote non-stationarity and non-linearity [20]. Such an assumption requires the transformation of datasets after the pre-processing stage in order to deal with residual values of the basic statistics after performing forecasting statistical techniques (e.g. ARIMA and Holt-Winter models as in [39]). Often researchers refer to byte or packet counts, but implicitly they mean the outcomes of such raw features after transforming them in a completely different conceptual and numerical representation. Another example is the work presented in [26,39], where traffic was initially gathered and pre-processed based on byte and packet counts but in order to enforce a particular detection scheme, these were converted to a time–frequency (TF) domain based on their initial Fourier transforms. Subsequently, those values were in a position to be compatible with their wavelet-based framework for recognizing low, mid and high frequency spectra each denoting particular anomalies. The same also happened in [44] where sample entropy values on top of basic statistics had the core feature role in the subspace-based framework.

The features used by a particular algorithm for the separation of abnormal from normal traffic (i.e. detection) do not need to be compatible with a classification mechanism. In fact, often a classification scheme relies on its own mathematical principles to provide meaningful clusters in the anomalous spectrum identified by the initial detection phase. For example, in [44] unsupervised learning using two clustering algorithms (k-means and hierarchical agglomeration) was applied. In this case, the initial values obtained by the detection scheme were transformed. In

particular, the authors defined their distinct anomaly clusters based on distance metrics as produced by the k-means algorithm. The initial anomalous entropy-based components as extracted by a PCA-based scheme were difficult to cluster due to their mathematical incompatibility and their different probabilistic foundational schemes. The k-means clustering algorithm was chosen since it enables the transformation of probabilistic values (such as entropy estimations) into the Euclidian space and it further relates them based on their distance. There are several examples such as in [43,84] as well as in [20,85] that propose certain theoretical re-approximations, where detection-related statistics follow an adjustable and mathematically valid scheme compatible with a given classification scheme.

4.3. Discussion

Through a brief presentation of measurement techniques, this section illustrated the inter-dependency between the traffic features with the other dimensions in Fig. 1. In reality, any anomaly diagnosis component is required to operate under a basic input provided by link measurements in the form of either packets, flows or flow aggregates. However, due to the nature of such link measurements, it is difficult to have a complete view of the traffic rates between distinct IP flows.

Each of the raw data formats described in this section have their own constraints with respect to computational cost alongside the network-wide view mentioned earlier. Moreover, the composition of traffic features in any of the three main measurement infrastructures requires different levels of meta-mapping on any mathematical formulation that intends to extract distinct anomalies on the source/destination viewpoint (Section 4.2). As discussed in this section, records derived from SNMP and Netflow measurements include noisy and also include a level of opaqueness regarding protocol-specific dynamics. Although SNMP may be considered a "cheaper" solution for ISPs, Cisco's Netflow may give a better insight on source/destination traffic dynamics. On the other hand, Netflow is not used widely over a network but rather on specific points. Hence, inaccuracies on network-wide anomaly diagnosis can occur. In contrast to SNMP and Netflow records, packet records provided by pcap can express even further the behavior of traffic flows but its high processing cost on the pre-processing stage as well as on the composition of traffic features can heavily impact on the performance of an anomaly detection component.

5. Examples of methods and techniques

In order to better characterize the current state of network traffic anomaly diagnosis, this section discusses important examples found in the literature that apply the methods and techniques surveyed in this paper. The list of examples is not meant to be exhaustive but instead it reflects the most influential contributions to the field in our opinion. The discussed techniques present some of the de facto methodologies that have been used in recent years, which demonstrate their application and have been acknowledged as seminal work. Other examples can also be found in recent work (e.g. [1,3,4,6,8,12,13,20,53]).

5.1. Principal component analysis in backbone networks

Principal Component Analysis (PCA) is a statistical technique that has been broadly used in the area of data mining that aims to reduce the dimensionality of high-dimensional datasets and further indicate the data points with the largest possible variance. Given the representation of the data points into new orthogonal axes (i.e. principal components) it is feasible to identify outlier points that hold a high probability at being related with anomalous activity. This technique has seen a great level of success and has been used by several studies as reported in [1,7,8,36]. It was firstly introduced in [22] therefore we following provide a brief description of how its was used and presented in that novel piece of work.

Lakhina et al. [22] describe the use of *Principal Component Analysis* (PCA) to separate network traffic measurements into normal and anomalous subspaces. This method focuses on the detection of volume-based anomalies in Origin–Destination flow aggregates in backbone networks and it is a core component within several IDS systems today (e.g. SnortAD [15,103]). The intuition behind this method is that a volume anomaly propagates through the network and it thus should be observable on all links involved. In the PCA approach, instead of identifying anomalous traffic volumes on a link by comparison with past values (as in ARIMA timeseries models, see Section 3.1), anomalies are detected by comparing the values in a link with other values in the network at the same time (relying on a *spatial analysis* rather than on a *temporal analysis*). Ultimately, PCA separates the space of link traffic measurements into subspaces representing normal and anomalous traffic behavior.

Routing traffic is represented in a matrix A of size (#links) $\times$ (#OD-flows), where $A_{ij} = 1$ if OD flow j passes through link i, and 0 otherwise. Additionally, link traffic measurements are denoted in a matrix Y of size $t \times m$, where t is the number of time intervals and m is the number of links in the network. Any column i in matrix Y represents the time series of link i, whereas any row j represents the measurements in all links at a single timestep j. The PCA method can be used to map a given set of data points onto new axes, which are called the principal components, and point in the direction of maximum variance remaining in the data. The principal components are ordered by the amount of variance they capture, with the first principal component capturing the greatest variance in the data. PCA is applied to the measurement matrix Y, resulting in m principal components, $\{v_i\}_{i=1}^{m}$.

Principal components are calculated iteratively, such that the kth principal component describes the maximum variance of the difference between the original data and the data mapped in the first $k-1$ principal components. Each principal component v_k is computed as follows:

$$v_k = \arg\max_{\|v\|=1} \left\| (Y - \sum_{i=1}^{k-1} Y v_i v_i^T) v \right\|$$

After calculating the k principal components, it is possible to observe the amount of variability captured by each principal component. Typically, most of the variability can be captured in a small number of principal components. Subsequently the original dataset is mapped into these new axes. The mapping of the data to principal axis i is given by Yv_i, which can be normalized to unit length by dividing it by $\|Yv_i\|$:

$$u_i = \frac{Yv_i}{\|Yv_i\|} \quad i = 1, \ldots, m$$

The measurement matrix Y, when weighted by the principal component v_i, will produce one dimension of the transformed data. Vector u_i captures the temporal variation along principal axis i, and since the axes are ordered by overall variance, u_1 captures the strongest temporal trend in all link traffic (the most normal behavior), followed by u_2, and so on. The set $\{u_i\}_{i=1}^{m}$ is divided into normal and anomalous subspaces, where u_x $(1 \leqslant x \leqslant m)$ and all subsequent dimensions will belong to the anomalous space if u_x traffic projection indicates unusual network conditions. A simple threshold-based method is used to separate between normal and anomalous projections.

Anomaly detection is achieved by projecting the original link traffic at any timestep onto these two subspaces, thereby obtaining the *modeled* and the *residual* traffic. In general, a volume anomaly will result in a large change to the residual traffic, which can be statistically measured using *squared prediction error* (SPE).

5.2. Signal analysis of traffic anomalies

As already introduced throughout this paper, the broad domain of signal processing has seen a considerable level of success with respect to its applicability in anomaly detection in IP networks. One of the most established and widely used technique derived by this domain is wavelet analysis. This particular technique has been employed in various past and recent studies as in [1,7,8,13,20,36,81] and it was firstly introduced in [39]. Hence we following describe how this technique was firstly introduced in that seminal work.

In [39], Barford et al. use *wavelet filters* (see Section 3.2) to isolate anomalous signals within the data via the detection of a sharp increase in the variance of the filtered data (both from SNMP – byte and packet counts – and from IP flows). Anomalies are identified by firstly filtering low frequency components and therefore removing from the signal its predictable part (using wavelets), and only then applying statistical methods. The technique is able to distinguish long-lived anomalies, e.g. flash crowds (exposed using the low frequencies) as well as short-lived anomalies, e.g. failures and attacks (exposed by a combination of data from mid- and high-frequencies) in the normal daily and weekly traffic cycles.

Many wavelet transforms are possible and selecting the most adequate method requires an expert understanding of the performance of the decompositions and the choice of the most suitable transform for a given application. This selection is based on a balance between its *time–frequency localization* characteristics. Time localization relates to the length of the filters used in the transform. Simply enough the longer the filter, the more blurring will be present in the time domain, and the more difficult it will be to determine when the anomaly has happened. Frequency localization is related to two characteristics of the wavelet system, namely vanishing moments and approximation order. The decision to measure frequency localization using either of these depends on the nature of the algorithm that is employed [39].

A *deviation score* algorithm is used to detect anomalies in the measured data, where the high and medium parts of the signal are normalized to have variance 1, and then computing the variance of the data falling within a moving window (whose length depends on the duration of the anomalies that are intended to be captured). The variability of the high and medium parts of the signal is then combined by using weighted sum, and then statistically compared against a threshold. However, the wavelet parameters and length of the detection window are highly dependent on the particular application and anomaly types that one intends to detect.

5.3. Entropy estimation based on feature distributions

Overall, entropy-based detection formulations have been proposed in numerous past and recent studies as in [3,4]. The credibility and success of these approaches were firstly demonstrated by Lakhina et al. in [44], thus we dedicate this subsection to briefly describing the basic outcomes of this approach.

In fact, the work presented by Lakhina et al. [44] describes the use of *entropy estimation* (see Section 3.3) in the analysis of traffic features distributions in both single-link and network-wide (backbone) traffic. Entropy estimation identifies changes that can be used to detect anomalies that do not cause significant disruptions in the traffic volume. Instead of relying on the volume as the main metric this method assumes that anomalies will disturb the distribution of traffic features in a detectable manner. It also enables the automatic classification of network anomalies using *unsupervised learning*, in which similar anomalies are grouped into specific clusters according to patterns discovered in the network traffic.

Taking into account the properties of entropy, the authors then use the *multiway subspace method* to detect anomalies across multiple traffic features and multiple Origin–Destination flows. Initially, the *subspace method* is used to identify unusual deviations in the variation of a set of correlated metrics. Observation of features are represented in a matrix X of size $t \times p$, where columns represent features and rows represent values observed. PCA is used to select a new set of features m $(m \ll p)$ that defines the dimensionality of the subspace. This is achieved under the assumption that features are correlated and that variations can be expressed as a linear combination of less than p variables. Normal variation is defined as a projection of data in this new subspace, and anomalous variation is defined as a deviation from this subspace.

The *multiway subspace method* is an extension which allows the analysis over multiple traffic features (source

and destination addresses and source and destination ports). This extension relies on four matrixes, each of size $t \times p$, containing the entropy timeseries of length t for p OD flows for one particular traffic feature. This allows the extraction of anomalies across multiple traffic features and multiple OD flows. Compared to the PCA method described in Section 5.1, this work relies on the feature distribution variation as the main metric instead of volume-based variation of the traffic.

5.4. Filtering and statistical methods

In [45], the authors present a hybrid approach combining Kalman filtering and statistical methods for detecting volume anomalies in large-scale backbone networks. Based on the assumption that an anomaly will traverse multiple links along its flow path, this approach relies on the monitoring of multiple links simultaneously, which would thus allow more evidence about a potential anomaly to be collected. Instead of performing anomaly detection directly on monitored data, this technique performs anomaly detection on origin–destination (OD) flows, whose data is inferred from other measurements (link statistics). Initially, a Kalman filter [89,104] is applied to filter out the normal-looking traffic and to isolate the prediction error, followed by a statistical analysis of the residual traffic. Differently from the work in [22], which also focuses on network-wide anomaly detection, Kalman filters were used to process the incoming link data rather than PCA analysis.

The authors use traffic matrix estimation [25] to provide predictions of future values of the traffic matrix, and then compare future predictions to an inference of the actual traffic matrix using new link-level SNMP measurements. The Kalman filter estimates the network state X_t, representing the OD flows, by using a two step approach for each time t. $\widehat{X}_{t|i}$ denotes the estimation of X_t based on time $i, t \geqslant i$:

- *Prediction step*: $\widehat{X}_{t|t}$ denotes the estimate of the network state at time t based on all observations up to time t. This term has an associated variance denoted by $P_{t|t}$. In the prediction step, given $\widehat{X}_{t|t}$ and $P_{t|t}$, the one step predictor denoted by $\widehat{X}_{t+1|t}$ and its associated variance $P_{t+1|t}$ are computed.
- *Estimation step*: in this step the Kalman filter updates the state estimates $\widehat{X}_{t+1|t+1}$ and its variance $P_{t+1|t+1}$ through the combination of its predicted values and the new observation of link-level SNMP measurements. The new observation is thus used to correct the previous prediction. By using the predictive ability of the filter it is therefore possible to estimate the future values of the traffic matrix [45].

If the difference between the prediction and the estimation (using the most recent measurements) becomes larger than a given threshold, then the residual is further analyzed to determine whether an anomaly alert should be raised. Four methods for the statistical analysis of the residual were compared, each focusing on different aspects of the traffic pattern change: (a) simple comparison of the instantaneous residual traffic to a threshold, (b) comparing a local variance calculation to a global variance on the filtered residual signal and triggering an alarm whenever the ratio between the two exceeds a threshold, (c) applying wavelet analysis on the filtered signal and then raising an alert when the signal exceeds a threshold at a sufficient number of timescales, and (d) through the identification of a change in the mean rate of the residual process.

5.5. Discussion

Network anomaly diagnosis is most commonly used as a compilation of several mathematical and data mining techniques, which initially assess the statistical normality of a dataset in order to identify the anomalous behavior. In contrast to the studies described in Sections 5.1, 5.3 and 5.4, the work described in Section 5.2 is capable of identifying volume-based anomalies that are only visible by observing simple link measurements such as the byte and packet counts. Furthermore, the techniques described in Sections 5.1, 5.3 and 5.4 had the objective of revealing even more hidden anomalies, which could not have been extracted without observing the distributional behavior of selected packet features. Moreover, all techniques were heavily dependent on the traffic matrix estimation that could provide estimates regarding the traffic rate between OD pairs. Consequently, estimating the aggregate traffic traversing OD pairs or PoPs rather than examining only the observable link measurements allows the identification of new anomaly types that may not be visible by a volume-based approach.

6. Concluding remarks

Traffic anomalies occur frequently in the Internet due to a number of reasons, ranging from failures and misconfigurations of networked devices to operational overload due to legitimate demand or malicious attacks. Internet traffic anomaly diagnosis as a research domain is essentially composed of two parts, i.e. anomaly detection and anomaly classification. Its main objective is the identification of abnormal traffic patterns and the identification of their causes. As such it forms part of a wider resilience strategy with remediation as the subsequent step. In order to remediate against identified threats successfully it is necessary to reliably distinguish between the different types of anomalies so that a targeted remediation approach can be taken. At its core, network anomaly diagnosis is a complex and multi-dimensional problem that comprises a range of heterogeneous domains. Substantial research has gone into identifying specific aspects, and devising methods and techniques for various (sub-)problems but no complete picture of the problem space (let alone the solution space) has so far emerged. The overwhelming majority of the literature on the subject is fragmented and focuses on specific aspects only. Thus, it fails to provide a comprehensive view of the problem spectrum.

This survey presented a comprehensive discussion on the anomaly diagnosis problem. In order to structure the problem space, we divided it into four interdependent dimensions, namely *Processing Cost*, *Diagnosis Granularity*,

Theoretical Methodologies and *Traffic Features* and elaborated upon their distinct properties as well as their interdependency. In particular, we focused on the methodology and feature dimensions, where we discussed the most influential research work. The explicit interest on the latter two dimensions has been done under the intention to construct tangible examples via seminal literature and enable a clearer understanding for readers coming from a general audience. Moreover, it was aimed to illustrate how the different approaches and techniques relate to each other through a conceptual framework. Hence, we discussed the relationship between each dimension and explained how these building blocks are typically combined in anomaly detection approaches. Overall, the problem space is presented in a more systematic manner, providing an overview and guide to any interested reader.

Furthermore, this survey addressed the practical aspects related to anomaly diagnosis schemes. In particular, we have discussed in detail the process of feature composition and how it is achieved. It was briefly illustrated that the process of constructing meaningful and suitable features for an anomaly diagnosis methodology is commonly decomposed into several sub-tasks that involve data mining. Apart from the initial filtering of the raw network data features (either on a packet, flow or flow aggregate level) there exists a further analysis that depends on the cost and granularity requirements that need to be suitable for a given mathematical methodology. This paper has also presented an extensive discussion on the advantages and shortcomings of each type of raw information, and emphasized the importance of the selection of the best statistical meta-features to be used for anomaly detection and classification. With the aim of illustrating practical examples of anomaly diagnosis components, this work selected some of the most influential efforts in this area and elaborated upon their techniques with respect to the features and the methodologies employed.

Finally, it has to be acknowledged that the anomaly diagnosis problem space is vast and not even all of its aspects have so far been explored by research. This paper provides a comprehensive view of the anomaly diagnosis problem. In order to organize the research field and place the related research efforts and techniques within a structure that captures the different core-elements relevant within this discussion we outlined the anomaly diagnosis problem spectrum in Section 2.3. As discussed within the paper, this problem space poses a dynamic persona and with organizing it this way the intention was to provide the basis for a clearer comprehension of the explicit domain of anomaly diagnosis and place it within the overall picture provided within this paper.

References

[1] S. Novakov, C.-H. Lung, I. Lambadaris, N. Seddigh, Studies in applying pca and wavelet algorithms for network traffic anomaly detection, in: 2013 IEEE 14th International Conference on High Performance Switching and Routing (HPSR), 2013, pp. 185–190. http://dx.doi.org/10.1109/HPSR.2013.6602310.

[2] J. Wang, D. Rossell, C.G. Cassandras, I.C. Paschalidis, Network anomaly detection: a survey and comparative analysis of stochastic and deterministic methods, CoRR abs/1309.4844.

[3] B. Tellenbach, M. Burkhart, D. Schatzmann, D. Gugelmann, D. Sornette, Accurate network anomaly classification with generalized entropy metrics, Comput. Netw. 55 (15) (2011) 3485–3502, http://dx.doi.org/10.1016/j.comnet.2011.07.008.

[4] A. Coluccia, A. D'alconzo, F. Ricciato, Distribution-based anomaly detection via generalized likelihood ratio test: a general maximum entropy approach, Comput. Netw. 57 (17) (2013) 3446–3462. http://dx.doi.org/10.1016/j.comnet.2013.07.028.

[5] T. Gamer, Collaborative anomaly-based detection of large-scale internet attacks, Comput. Netw. 56 (1) (2012) 169–185, http://dx.doi.org/10.1016/j.comnet.2011.08.015.

[6] D. Brauckhoff, X. Dimitropoulos, A. Wagner, K. Salamatian, Anomaly extraction in backbone networks using association rules, IEEE/ACM Trans. Network. 20 (6) (2012) 1788–1799, http://dx.doi.org/10.1109/TNET.2012.2187306.

[7] A. Patcha, J.-M. Park, An overview of anomaly detection techniques: existing solutions and latest technological trends, Comput. Netw. 51 (12) (2007) 3448–3470, http://dx.doi.org/10.1016/j.comnet.2007.02.001.

[8] T.M., G. Liu, C. Ji, Anomaly detection approaches for communication networks, Algorithms for Next Generation Networks, 2010.

[9] J.M. Estvez-Tapiador, P. Garcia-Teodoro, J.E. Daz-Verdejo, Anomaly detection methods in wired networks: a survey and taxonomy, Comput. Commun. 27 (16) (2004) 1569–1584.

[10] P. Barford, N. Duffield, A. Ron, J. Sommers, Network performance anomaly detection and localization, in: INFOCOM 2009, IEEE, 2009, pp. 1377–1385. http://dx.doi.org/10.1109/INFCOM.2009.5062053.

[11] F. Soldo, A. Metwally, Traffic anomaly detection based on the ip size distribution, in: INFOCOM, 2012 Proceedings IEEE, 2012, pp. 2005–2013. http://dx.doi.org/10.1109/INFCOM.2012.6195581.

[12] I.C. Paschalidis, G. Smaragdakis, Spatio-temporal network anomaly detection by assessing deviations of empirical measures, IEEE/ACM Trans. Netw. 17 (3) (2009) 685–697, http://dx.doi.org/10.1109/TNET.2008.2001468.

[13] M. Salagean, Real network traffic anomaly detection based on analytical discrete wavelet transform, in: 2010 12th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), 2010, pp. 926–931. http://dx.doi.org/10.1109/OPTIM.2010.5510445.

[14] M. Roesch, Snort – lightweight intrusion detection for networks, in: Proceedings of the 13th USENIX Conference on System Administration, USENIX Association, Berkeley, CA, USA, 1999, pp. 229–238.

[15] Snort <http://www.snort.org>.

[16] Bro ids <http://www.bro.org>.

[17] V. Paxson, Bro: a system for detecting network intruders in real-time, in: Computer Networks, 1999, pp. 2435–2463.

[18] H. Debar, M. Dacier, A. Wespi, A revised taxonomy for intrusion-detection systems, Ann. Télécommun. 55 (7–8) (2000) 361–378.

[19] T. Peng, C. Leckie, K. Ramamohanarao, Survey of network-based defense mechanisms countering the dos and ddos problems, ACM Comput. Surv. 39 (1) (2007) 3. http://dx.doi.org/10.1145/1216370.1216373.

[20] A.K. Marnerides, On Characterization & Decomposition of Internet Traffic Dynamics, Ph.D. thesis, Dept. of Computing & Communications, Lancaster University, 2011.

[21] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (3) (2009) 15:1–15:58, http://dx.doi.org/10.1145/1541880.1541882.

[22] A. Lakhina, M. Crovella, C. Diot, Diagnosing network-wide traffic anomalies, in: SIGCOMM '04: Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, ACM, New York, NY, USA, 2004, pp. 219–230. http://dx.doi.org/10.1145/1015467.1015492.

[23] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, K. Cho, Extracting hidden anomalies using sketch and non Gaussian multiresolution statistical detection procedures, in: Proceedings of the 2007 Workshop on Large Scale Attack Defense, LSAD'07, ACM, New York, NY, USA, 2007, pp. 145–152. http://dx.doi.org/10.1145/1352664.1352675.

[24] Y. Zhang, M. Roughan, W. Willinger, L. Qiu, Spatio-temporal compressive sensing and internet traffic matrices, in: Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, SIGCOMM'09, ACM, New York, NY, USA, 2009, pp. 267–278. http://dx.doi.org/10.1145/1592568.1592600.

[25] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, C. Diot, Traffic matrices: balancing measurements, inference and modeling, SIGMETRICS Perform. Eval. Rev. 33 (1) (2005) 362–373. http://dx.doi.org/10.1145/1071690.1064259.

[26] P. Barford, D. Plonka, Characteristics of network traffic flow anomalies, in: IMW'01: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement, ACM, New York, NY, USA, 2001, pp. 69–73. http://dx.doi.org/10.1145/505202.505211.

[27] A.H. Yaacob, I.K. Tan, S.F. Chien, H.K. Tan, Arima based network anomaly detection, Int. Conf. Commun. Software Netw. 0 (2010) 205–209. <http://doi.ieeecomputersociety.org/10.1109/ICCSN.2010.55>.

[28] V. Siris, F. Papagalou, Application of anomaly detection algorithms for detecting syn flooding attacks, in: Global Telecommunications Conference, 2004, GLOBECOM '04, vol. 4, IEEE, 2004, pp. 2050–2054. http://dx.doi.org/10.1109/GLOCOM.2004.1378372.

[29] A. Lakhina, M. Crovella, C. Diot, Characterization of network-wide anomalies in traffic flows, in: IMC'04: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, ACM, New York, NY, USA, 2004, pp. 201–206. http://dx.doi.org/10.1145/1028788.1028813.

[30] E. Alpaydin, Introduction to Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2004.

[31] C. Gates, C. Taylor, Challenging the anomaly detection paradigm: a provocative discussion, in: Proceedings of the 006 Workshop on New Security Paradigms, NSPW '06, ACM, New York, NY, USA, 2007, pp. 21–29. http://dx.doi.org/10.1145/1278940.1278945.

[32] H. Ringberg, M. Roughan, J. Rexford, The need for simulation in evaluating anomaly detectors, SIGCOMM Comput. Commun. Rev. 38 (1) (2008) 55–59, http://dx.doi.org/10.1145/1341431.1341443.

[33] Y. Zhang, M. Roughan, C. Lund, D. Donoho, An information-theoretic approach to traffic matrix estimation, in: Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM'03, ACM, New York, NY, USA, 2003, pp. 301–312. http://dx.doi.org/10.1145/863955.863990.

[34] F. Silveira, C. Diot, N. Taft, R. Govindan, Astute: detecting a different class of traffic anomalies, in: SIGCOMM'10: Proceedings of the ACM SIGCOMM 2010 Conference on SIGCOMM, ACM, New York, NY, USA, 2010, pp. 267–278. http://dx.doi.org/10.1145/1851182.1851215.

[35] P. Chhabra, C. Scott, E.D. Kolaczyk, M. Crovella, Distributed spatial anomaly detection, in: Proceedings of Infocom 2008, 2008.

[36] Y. Zhang, Z. Ge, A. Greenberg, M. Roughan, Network anomography, in: IMC'05: Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement, USENIX Association, Berkeley, CA, USA, 2005, p. 30.

[37] C. Arthur, Why isps loathe the bbc iplayer, The Guardian, April 2008. <http://www.guardian.co.uk/technology/2008/apr/07/iplayer.isps>.

[38] N. Weaver, V. Paxson, S. Staniford, R. Cunningham, A taxonomy of computer worms, in: Proceedings of the 2003 ACM Workshop on Rapid Malcode, WORM '03, ACM, New York, NY, USA, 2003, pp. 11–18. http://dx.doi.org/10.1145/948187.948190.

[39] P. Barford, J. Kline, D. Plonka, A. Ron, A signal analysis of network traffic anomalies, in: IMW'02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement, ACM, New York, NY, USA, 2002, pp. 71–82. http://dx.doi.org/10.1145/637201.637210.

[40] D. Gao, M.K. Reiter, D. Song, On gray-box program tracking for anomaly detection, in: Proceedings of the 13th Conference on USENIX Security Symposium, SSYM'04, vol. 13, USENIX Association, Berkeley, CA, USA, 2004, p. 8. <http://dl.acm.org/citation.cfm?id=1251375.1251383>.

[41] W. Lee, D. Xiang, Information-theoretic measures for anomaly detection, in: Proceedings. 2001 IEEE Symposium on Security and Privacy, 2001, SP 2001, 2001, pp. 130–143. http://dx.doi.org/10.1109/SECPRI.2001.924294.

[42] Z. Zhang, F. Nat-Abdesselam, P.-H. Ho, Y. Kadobayashi, Toward cost-sensitive self-optimizing anomaly detection and response in autonomic networks, Comput. Secur. 30 (6–7) (2011) 525–537.

[43] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, K. Lee, Internet traffic classification demystified: myths, caveats, and the best practices, in: CoNEXT'08: Proceedings of the 2008 ACM CoNEXT Conference, ACM, New York, NY, USA, 2008, pp. 1–12.

[44] A. Lakhina, M. Crovella, C. Diot, Mining anomalies using traffic feature distributions, SIGCOMM Comput. Commun. Rev. 35 (4) (2005) 217–228. http://dx.doi.org/10.1145/1090191.1080118.

[45] A. Soule, K. Salamatian, N. Taft, Combining filtering and statistical methods for anomaly detection, in: IMC'05: Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement, USENIX Association, Berkeley, CA, USA, 2005, p. 31.

[46] D. Zuev, A.W. Moore, Traffic classification using a statistical approach, in: Passive and Active Measurement, 2005.

[47] Cisco Ios Netflow <www.cisco.com/web/go/netflow>.

[48] N. Duffield, Sampling for passive internet measurement: a review, Stat. Sci. 19 (2004) 472–498.

[49] C. Duffield, N. Lund, M. Thorup, Properties and prediction of flow statistics from sampled packet streams, in: Proceedings of the 2Nd ACM SIGCOMM Workshop on Internet Measurement, IMW'02, ACM, New York, NY, USA, 2002, pp. 159–171. http://dx.doi.org/10.1145/637201.637225.

[50] N. Hohn, D. Veitch, Inverting sampled traffic, IEEE/ACM Trans. Network. 14 (1) (2006) 68–80, http://dx.doi.org/10.1109/TNET.2005.863456.

[51] J. Mai, C.-N. Chuah, A. Sridharan, T. Ye, H. Zang, Is sampled data sufficient for anomaly detection?, in: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, IMC'06, ACM, New York, NY, USA, 2006, pp 165–176. http://dx.doi.org/10.1145/1177080.1177102.

[52] D. Brauckhoff, B. Tellenbach, A. Wagner, M. May, A. Lakhina, Impact of packet sampling on anomaly detection metrics, in: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, IMC'06, ACM, New York, NY, USA, 2006, pp. 159–164. http://dx.doi.org/10.1145/1177080.1177101.

[53] D. Brauckhoff, K. Salamatian, M. May, A signal processing view on packet sampling and anomaly detection, in: INFOCOM, 2010 Proceedings IEEE, 2010, pp. 1–9. http://dx.doi.org/10.1109/INFCOM.2010.5462154.

[54] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, A. Lakhina, Detection and identification of network anomalies using sketch subspaces, in: IMC'06: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, ACM, New York, NY, USA, 2006, pp. 147–152. http://dx.doi.org/10.1145/1177080.1177099.

[55] L. Huang, X.L. Nguyen, M. Garofalakis, A. Joseph, M. Jordan, N. Taft, In-network pca and anomaly detection, in: Advances in Neural Information Processing Systems (NIPS), Vancouver, BC, 2006.

[56] Y. Zhang, S. Singh, S. Sen, N. Duffield, C. Lund, Online identification of hierarchical heavy hitters: algorithms, evaluation, and applications, in: IMC'04: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, ACM, New York, NY, USA, 2004, pp. 101–114. http://dx.doi.org/10.1145/1028788.1028802.

[57] G. Nychis, V. Sekar, D.G. Andersen, H. Kim, H. Zhang, An empirical evaluation of entropy-based traffic anomaly detection, in: IMC'08: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, ACM, New York, NY, USA, 2008, pp. 151–156. http://dx.doi.org/10.1145/1452520.1452539.

[58] A. Wagner, B. Plattner, Entropy based worm and anomaly detection in fast ip networks, in: 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprise, 2005, pp. 172–177. http://dx.doi.org/10.1109/WETICE.2005.35.

[59] B. Tellenbach, M. Burkhart, D. Sornette, T. Maillart, Beyond shannon: characterizing internet traffic with generalized entropy metrics, in: PAM'09: Proceedings of the 10th International Conference on Passive and Active Network Measurement, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 239–248. http://dx.doi.org/10.1007/978-3-642-00975-4_24.

[60] A. Kind, M. Stoecklin, X. Dimitropoulos, Histogram-based traffic anomaly detection, IEEE Trans. Network Serv. Manage. 6 (2) (2009) 110–121, http://dx.doi.org/10.1109/TNSM.2009.090604.

[61] D. Brauckhoff, X. Dimitropoulos, A. Wagner, K. Salamatian, Anomaly extraction in backbone networks using association rules, in: IMC'09: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, ACM, New York, NY, USA, 2009, pp. 28–34. http://dx.doi.org/10.1145/1644893.1644897.

[62] M. Thottan, C. Ji, Anomaly detection in ip networks, IEEE Trans. Signal Process. 51 (8) (2003) 2191–2204.

[63] B. Krishnamurthy, S. Sen, Y. Zhang, Y. Chen, Sketch-based change detection: methods, evaluation, and applications, in: IMC'03: Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement, ACM, New York, NY, USA, 2003, pp. 234–247. http://dx.doi.org/10.1145/948205.948236.

[64] V. Bandara, A. Pezeshki, P. Anura, Modeling spatial and temporal behavior of internet traffic anomalies, in: 2010 IEEE 35th Conference on Local Computer Networks (LCN), 2010, pp. 384–391. doi: 10.1109/LCN.2010.5735749.

[65] W. Lu, A.A. Ghorbani, Network anomaly detection based on wavelet analysis, EURASIP J. Adv. Signal Process 2009 (2009) 4:1–4:16, http://dx.doi.org/10.1155/2009/837601.

[66] J.D. Brutlag, Aberrant behavior detection in time series for network monitoring, in: LISA'00: Proceedings of the 14th USENIX Conference on System Administration, USENIX Association, Berkeley, CA, USA, 2000, pp. 139–146.
[67] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, P. Abry, Non-Gaussian and long memory statistical characterizations for internet traffic with anomalies, IEEE Trans. Depend. Secur. Comput. 4 (1) (2007) 56–70. http://dx.doi.org/10.1109/TDSC.2007.12.
[68] Y. Gu, A. McCallum, D. Towsley, Detecting anomalies in network traffic using maximum entropy estimation, in: Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement, IMC'05, USENIX Association, Berkeley, CA, USA, 2005, p. 32. <http://dl.acm.org/citation.cfm?id=1251086.1251118>.
[69] V. Karamcheti, D. Geiger, Z. Kedem, S. Muthukrishnan, Detecting malicious network traffic using inverse distributions of packet contents, in: MineNet'05: Proceedings of the 2005 ACM SIGCOMM Workshop on Mining Network Data, ACM, New York, NY, USA, 2005, pp. 165–170. http://dx.doi.org/10.1145/1080173.1080176.
[70] C.-M. Cheng, H. Kung, K.-S. Tan, Use of spectral analysis in defense against dos attacks, in: Global Telecommunications Conference, 2002, GLOBECOM'02, vol. 3, IEEE, 2002, pp. 2143–2148. http://dx.doi.org/10.1109/GLOCOM.2002.1189011.
[71] L. Feinstein, D. Schnackenberg, R. Balupari, D. Kindred, Statistical approaches to ddos attack detection and response, in: Proceedings of the DARPA Information Survivability Conference and Exposition, vol. 1, 2003, pp. 303–314. http://dx.doi.org/10.1109/DISCEX.2003.1194894.
[72] W. Lee, D. Xiang, Information-theoretic measures for anomaly detection, in: SP'01: Proceedings of the 2001 IEEE Symposium on Security and Privacy, IEEE Computer Society, Washington, DC, USA, 2001, p. 130.
[73] C.E. Shannon, Prediction and entropy of printed english, Bell Syst. Tech. J. 30 (1951) 50–64.
[74] G. Box, G.M. Jenkins, G. Reinsel, Time Series Analysis: Forecasting and Control, third ed., Prentice Hall, 1994.
[75] S.-J. Han, S.-B. Cho, Detecting intrusion with rule-based integration of multiple models, Comput. Secur. 22 (7) (2003) 613–623, http://dx.doi.org/10.1016/S0167-4048(03)00711-9.
[76] C. Warrender, S. Forrest, B. Pearlmutter, Detecting intrusions using system calls: alternative data models, in: Proceedings of the 1999 IEEE Symposium on Security and Privacy, 1999, 1999, pp. 133–145. http://dx.doi.org/10.1109/SECPRI.1999.766910.
[77] D.-Y. Yeung, Y. Ding, Host-based intrusion detection using dynamic and static behavioral models, Pattern Recogn. 36 (2003) 229–243.
[78] C. Nikias, M.R. Raghuveer, Bispectrum estimation: a digital signal processing framework, Proc. IEEE 75 (7) (1987) 869–891, http://dx.doi.org/10.1109/PROC.1987.13824.
[79] A. Marnerides, D. Pezaros, H. chul Kim, D. Hutchison, Internet traffic classification using energy time–frequency distributions, in: 2013 IEEE International Conference on Communications (ICC), 2013, pp. 2513–2518. http://dx.doi.org/10.1109/ICC.2013.6654911.
[80] L. Cohen, Time–frequency distributions – a review, Proc. IEEE 77 (7) (1989) 941–981, http://dx.doi.org/10.1109/5.30749.
[81] W. Lu, A.A. Ghorbani, Network anomaly detection based on wavelet analysis, EURASIP J. Adv. Signal Process 2009 (2009) 4:1–4:16, http://dx.doi.org/10.1155/2009/837601.
[82] C. Hood, C. Ji, Proactive network-fault detection [telecommunications], IEEE Trans. Reliab. 46 (3) (1997) 333–341, http://dx.doi.org/10.1109/24.664004.
[83] V. Sotiris, P. Tse, M. Pecht, Anomaly detection through a bayesian support vector machine, IEEE Trans. Reliab. 59 (2) (2010) 277–286, http://dx.doi.org/10.1109/TR.2010.2048740.
[84] A.W. Moore, D. Zuev, Internet traffic classification using bayesian analysis techniques, in: SIGMETRICS'05: Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, ACM, New York, NY, USA, 2005, pp. 50–60. http://dx.doi.org/10.1145/1064212.1064220.
[85] A.K. Marnerides, D. Pezaros, H. Kim, D. Hutchison, Unsupervised two-class and multi-class support vector machines for abnormal traffic characterization, in: 10th International Passive and Active Measurements Conference, PAM Conference Student Workshop, Seoul, South Korea, 2009.
[86] A. Hussain, J. Heidemann, C. Papadopoulos, A framework for classifying denial of service attacks, in: SIGCOMM'03: Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, ACM, New York, NY, USA, 2003, pp. 99–110. http://dx.doi.org/10.1145/863955.863968.
[87] A. Hussain, J. Heidemann, C. Papadopoulos, Identification of repeated denial of service attacks, in: Proceedings of the IEEE Infocom, IEEE, Barcelona, Spain, 2006. <http://www.isi.edu/johnh/PAPERS/Hussain06a.html>.
[88] M. Loève, Probability theory, Graduate Texts in Mathematics, fourth ed., vol. 46, Springer-Verlag, 1978. vol. II.
[89] R. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 1 (82) (1960) 35–45.
[90] A. Rényi, On measures of information and entropy, in: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, 1960, pp. 547—561.
[91] A. Marnerides, S. Malinowski, R. Morla, M. Rodrigues, H. Kim, Towards the improvement of diagnostic metrics fault diagnosis for dsl-based iptv networks using the Renyi entropy, in: Global Communications Conference (GLOBECOM), 2012 IEEE, 2012, pp. 2779–2784. http://dx.doi.org/10.1109/GLOCOM.2012.6503537.
[92] C. Tsallis, Nonextensive statistics: theoretical, experimental and computational evidences and connections, Braz. J. Phys. 29 (cond-mat/9903356) (1999) 1–35.
[93] S. Ruehrup, P. Urbano, A. Berger, A. D'Alconzo, Botnet detection revisited: theory and practice of finding malicious p2p networks via internet connection graphs, in: 2013 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2013, pp. 435–440. http://dx.doi.org/10.1109/INFCOMW.2013.6562902.
[94] Y. Zhang, M. Roughan, N. Duffield, A. Greenberg, Fast accurate computation of large-scale ip traffic matrices from link loads, in: Proceedings of the 2003 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS'03, ACM, New York, NY, USA, 2003, pp. 206–217. http://dx.doi.org/10.1145/781027.781053.
[95] Tcpdump/Libcap Public Repository <http://www.tcpdump.org>.
[96] The Dag Project <http://dag.cs.waikato.ac.nz>.
[97] J. Schönwälder, A. Pras, M. Harvan, J. Schippers, R.M. van de, SNMP traffic analysis: approaches, tools, and first results, in: Proceedings of the 10th IFIP/IEEE International Symposium on Integrated Network Management, IM'07, IEEE Computer Society Press, Piscataway, 2007, pp. 324–332. <http://doc.utwente.nl/64390/>.
[98] T. Oetiker, Mrtg – the multi router traffic grapher, in: LISA'98: Proceedings of the 12th USENIX Conference on System Administration, USENIX Association, Berkeley, CA, USA, 1998, pp. 141–148.
[99] WinpCap: The Packet Capture and Monitoring for Windows <http://www.winpcap.org>.
[100] Ipsumdump Tool <http://www.cs.ucla.edu/kohler/ipsumdump/>.
[101] Caida's Coralreef Tool <http://www.caida.org/tools/measurement/coralreef/>.
[102] Wireshark <http://www.wireshark.org>.
[103] Snort ad Tool <http://anomalydetection.info/>.
[104] T. Kailath, A.H. Sayed, B. Hassibi, Linear Estimation, Prentice Hall, 2000.

Angelos K. Marnerides is a Lecturer (Assistant Professor) in the School of Computing and Mathematical Sciences at Liverpool John Moores University, UK. His research interests include anomaly detection, network security, resilience and cloud computing. Prior to that he was a Research Associate in the Department of Computing and Communications at Lancaster University (2012—2014), a Post-doctoral Research Fellow in the Carnegie Mellon University – Portugal postdoctoral scheme at IT, University of Porto (2011–2012) and an Honorary Research Associate with the Department of Electronic and Electrical Engineering at University College London (UCL) (2012–2013). He obtained his M.Sc. and Ph.D. in Computer Science from Lancaster University in 2007 and 2011 respectively.

Alberto Schaeffer-Filho is an Associate Professor in the Institute of Informatics at Federal University of Rio Grande do Sul (UFRGS). His research interests include network management, security and resilience of next generation networks. Prior to that he was a Research Associate in Lancaster University for three years. He obtained his Ph.D. in Computing from Imperial College London in 2009. He is a member of the IEEE. See http://www.inf.ufrgs.br/alberto for more details and selected papers.

Andreas Mauthe is a Senior Lecturer at the School of Computing Communications at Lancaster University. He has been working in the area of distributed and multimedia systems for more than 20 years. His particular interests are in the area of content management systems and content networks, and QoE as well as Cyber Security and resilient networks with an explicit interest in anomaly detection. Prior to joining Lancaster University, Andreas headed a research group at the Multimedia Communications Lab (KOM), at the TU Darmstadt. After completing his Ph.D., Andreas worked for more than four years in different positions in industry in the area of content management in content production and media archives.