4-1-2024

Matrix profile data mining for BGP anomaly detection

Ben A. Scott
Edith Cowan University

Michael N. Johnstone
Edith Cowan University

Patryk Szewczyk
Edith Cowan University

Steven Richardson
Edith Cowan University

10.1016/j.comnet.2024.110257
Scott, B. A., Johnstone, M. N., Szewczyk, P., & Richardson, S. (2024). Matrix profile data mining for BGP anomaly detection. Computer Networks, 242, article 110257. https://doi.org/10.1016/j.comnet.2024.110257

Computer Networks 242 (2024) 110257

Contents lists available at ScienceDirect

Computer Networks

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

Matrix Profile data mining for BGP anomaly detection

Ben A. Scott [*], Michael N. Johnstone, Patryk Szewczyk, Steven Richardson

Edith Cowan University, Perth, Australia

ARTICLE INFO

Keywords:
Anomaly detection
BGP
Cyber security
Internet security
Network security
Routing
Time series analysis

ABSTRACT

The Border Gateway Protocol (BGP), acting as the communication protocol that binds the Internet, remains vulnerable despite Internet security advancements. This is not surprising, as the Internet was not designed to be resilient to cyber-attacks, therefore the detection of anomalous activity was not of prime importance to the Internet creators. Detection of BGP anomalies can potentially provide network operators with an early warning system to focus on protecting networks, systems, and infrastructure from significant impact, improve security posture and resilience, while ultimately contributing to a secure global Internet environment. In this paper, we present a novel technique for the detection of BGP anomalies in different events. This research uses publicly available datasets of BGP messages collected from the repositories, Route Views and Réseaux IP Européens (RIPE). Our contribution is the application of a time series data mining approach, *Matrix Profile (MP)*, to detect BGP anomalies in all categories of BGP events. Advantages of the MP detection technique compared to extant approaches include that it is domain agnostic, is assumption-free, requires few parameters, does not require training data, and is scalable and storage efficient. The single hyper-parameter analyzed in MP shows it is robust to change. Our results indicate the MP detection scheme is competitive against existing detection schemes. A novel BGP anomaly detection scheme is also proposed for further research and validation.

1. Introduction

The Internet was devised decades ago, yet a communication system that binds this infrastructure, the Border Gateway Protocol (BGP), is reliant on trust and remains vulnerable to both malicious and non-malicious events [1–3]. Numerous cyber-attacks have affected internetworks with intense BGP traffic volume activity and ultimately overloaded the Internet [4–6]. For example, the Nimda incident caused a surge in BGP traffic approximately 30 times more than the normal volume, impacting hundreds of thousands of devices [7–9]. BGP insecurity places at risk the many businesses, transactions, devices, and global matters of state that depend on an operationally stable and secure Internet [10–12]. Recent BGP events have affected major Internet entities including Akamai, Apple, Amazon, Facebook, Google, Mastercard, and Microsoft [13–15], while from a nation-state perspective, BGP attacks have reportedly been used by Russia against Ukraine [16,17].

There exists a range of approaches to the detection of BGP anomalies. Comprehensive reviews of BGP anomaly detection techniques and BGP security proposals have been conducted; detection approaches have been categorized into machine learning (ML), reachability-based approaches, statistical pattern recognition, time series analysis, and validation studies based on historical BGP data, categorized by the type of BGP incident category a technique was applied to (e.g., direct, indirect, or outage events), and a review of BGP attacks and defenses [13,18]. Similar to communication protocol and intrusion detection scheme research more broadly, BGP anomaly detection studies are dominated by approaches that involve large numbers of features, parameters, domain-specific tuning, training and often contributing to unacceptable computational cost [19–21].

For example, the use of Support Vector Machine (SVM) [22–27], and Long Short-Term Memory (LSTM) approaches [6,20,22,28–31] have been successful in detection of a range of BGP incidents including Internet blackouts, leaks and worm attacks, though they often require training, extensive parameterization and tuning. Other non-ML approaches, such as the use of a Principal Component Analysis (PCA) based subspace method with BGP volume extraction, have been successful in detection, identification and differentiation of BGP anomalies, though router configuration requirements showed prohibitive factors to real-time detection [32].

Compared to other techniques, there are relatively fewer data mining applications for BGP anomaly detection. We propose the Matrix Profile (MP) family of algorithms as a candidate for BGP anomaly detection. The MP approach has been evaluated on hundreds of time series datasets [33,34], and has been shown to successfully detect anomalies in data with periodic characteristics, with minimal parameterization [35]. It has never been applied to the problem of BGP

[*] Corresponding author.
E-mail address: ben.scott@ecu.edu.au (B.A. Scott).

https://doi.org/10.1016/j.comnet.2024.110257
Received 1 May 2023; Received in revised form 12 February 2024; Accepted 15 February 2024
Available online 17 February 2024

anomaly detection. MP's established domain-agnostic nature [34,36], minimal parameter requirements [33,34], and handling of large, sparse datasets [37,38] position it as a possible BGP anomaly detection solution. Moreover, its scalability and storage efficiency, proven at least in other domains [36,38], suggest that it is suitable for the extensive data associated with BGP. We also propose and introduce the Matrix Profile BGP (MPBGP) anomaly detection scheme for scalable MP powered BGP anomaly detection.

Key contributions of our paper are summarized as:

- We show that Matrix Profile detects BGP anomalies in all categories and is competitive when evaluated against existing detection schemes. We directly compare the approach to Deep Learning techniques and one non-ML technique using the same BGP volume-centric incidents and features. In some incidents, MP can detect anomalous activity earlier than other techniques.
- We test the only parameter within the package (window size) and show it to be robust to change. We also test and validate the assertion that MP can discover anomalies in datasets with missing data, with no false negatives (FNs).
- We show that Matrix Profile provides a better understanding of some key BGP incidents that are consistent with previous research that showed potentially hidden anomalous behavior which may represent an early stage of BGP anomalies.

The remainder of this paper is structured as follows: Section 2 provides the relevant background, while Section 3 presents links to related work on BGP anomaly detection. In Section 4, we articulate the Matrix Profile (MP) algorithm, followed by an outline of BGP incidents in Section 5. Section 6 describes the methods, experiments, and metrics used to evaluate our approach, and we report on the findings in Section 7. The detection scheme is presented in Section 8, with the effectiveness of MP anomaly detection discussed in Section 9. The paper concludes in Section 10.

2. Background

BGP is the default inter-domain routing protocol for the Internet. The protocol has been revised multiple times since the first Request for Comment (RFC) proposal issued in 1989 [39–41]. RFCs exist as an Internet engineering and governance corpus [42].

Autonomous Systems (ASes) are internetworked routing domains administered by a single authority [43]. These structures are not simply physically or geographically bound but rather formed by corporate, organizational, and political relationships [44,45]; thus inferences about their operation based simply on physical topology can be misplaced [46,47]. The strategic objectives of any AS are reliant on connectivity and Network Reachability Information (NRI). The presence of NRI is required for connectivity and AS connectivity is provided by BGP. Two modes of BGP are available: Internal BGP (IBGP) and External BGP (EBGP). Where IBGP can provide connectivity between routers within a single AS, it is EBGP that connects BGP routers at different ASes [39].

All BGP messages are comprised of marker, length, and type fields that form a fixed header of 19 octets. The marker field represents the start of a message (16 octets, all set to 1), the length field (2 octets) identifies the total message length (including the header), and the type field represents one of four message types (OPEN, UPDATE, NOTIFICATION, and KEEPALIVE) [48–50]. Depending on the BGP router vendor, a fifth type of BGP message (ROUTE REFRESH) can be supported [51]. Once a Transmission Control Protocol (TCP) session is established an OPEN message is transmitted; an initial exchange of BGP messages is required to instantiate the ESTABLISHED state. Session termination information is found in a NOTIFICATION message and session maintenance information is provided for with the KEEPALIVE message.

Routes are stored in a BGP peer in a set of databases known as the Routing Information Base (RIB). The RIB consists of three distinct components: Adj-RIB-In, Adj-RIB-Out, and the local RIB (Loc-RIB). The Adj-RIB-In stores routes received from UPDATE messages (i.e., from other peers); in other words, Adj-RIB-In represents routes that have been learned from adjacent neighbors and are functional to the path decision process. The Adj-RIB-Out stores routes sent out from this peer via UPDATE messages, while the Loc-RIB contains the current optimal routes used by this peer determined by Adj-RIB-In information and path decision processes informed by its local policies. Incremental routing information changes are achieved via announcement, withdrawal, or existing attribute update messages following the RIB exchange.

It is from BGP anatomy and functionality that feature extraction for anomaly detection schemes can be conducted. There are many features that can be extracted from BGP traffic, and features can also be aggregated for analysis (e.g., number of announcements and withdrawals, NLRI prefix announced and withdrawn, can be aggregated for analysis) [6,9]. These features can be broadly grouped into two categories: BGP volume and AS-path [20,21]; previous research has shown that BGP changes are primarily observed in volume features [6,9]. MP is recognized in the literature for its efficiency in identifying repeating patterns (motifs) and anomalies (discords) within voluminous datasets [36,38]; this drives our investigation of a novel data mining solution for BGP anomaly detection, leveraging MP's domain-agnostic nature, minimal parameter requirements, and adeptness in handling large, sparse datasets efficiently [34,35,38].

3. Related work

BGP anomalies have been studied extensively in the literature, and have been defined as damaging BGP activity that exist on a spectrum of impact; the severity of which can range from the relatively innocuous (e.g., route-flapping) through to destructive (e.g., BGP 'hijacks', 'blackholing' and rerouting) and be driven by non-malicious or malicious intent [1,13,52]. Previous work produced a taxonomy of BGP anomalies with categories named direct, indirect and outages (link failure) [18]. We use six high-profile and well-studied incidents from the three known categories (direct, indirect and outages) for a direct comparison against other work that investigated the same well-studied incidents and showed BGP changes are primarily observed in volume features [6,9].

The direct anomaly category represents the range of BGP hijacks currently known in addition to route leak incidents (e.g., the Telekom Malaysia route leak). Indirect incidents include significant cyber-security events that impacted Internet operations, such as web servers. Cyber-attacks (e.g., Nimda, Code Red II, and Slammer worm attacks) are examples of indirect incidents that affected ASes with intensified BGP activity and ultimately overloaded the Internet. Outages caused by natural disasters and energy system failures (e.g., the Japanese Earthquake and Moscow Blackout) represent the final category of BGP incidents.

Anomaly detection categories including ML-based approaches, reachability-based methods, statistical pattern recognition, and validation studies based on historical BGP data, have been surveyed previously [9,18,53]. For example, numerous ML techniques for BGP anomaly detection have been described in the literature [1,18,54–56]. Supervised learning techniques have shown mixed results [1]. Other studies have used the same BGP incidents we outline in subsequent sections to investigate Hidden Markov Model (HMM), Support Vector Machine (SVM) and Deep Learning Recurrent Neural Networks (RNNs) approaches [6,56,57]. The use of graph feature for detection of IP prefix hijacking has been shown effective, such as the use of the PageRank algorithm to develop an Ontological Graph Identification (OGI) detection scheme [58]. Additionally, the use of graph features as ML inputs for BGP anomaly detection has also been conducted [26].

BGP data can be successfully analyzed as a time series [5,6,9,28]. There have been several time series techniques used to conduct BGP anomaly detection; these include Daubechies 5 (db5) Wavelet

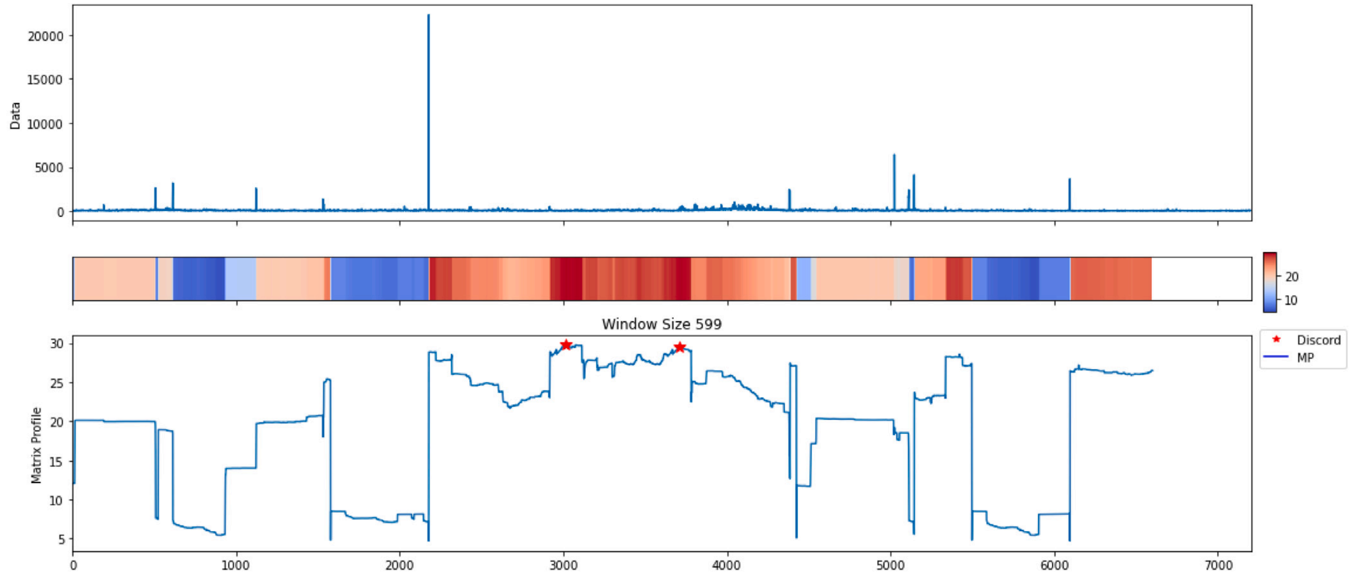

Fig. 1. Code Red I top-k discords (m_1=599).

transform, Fast Fourier Transform (FFT) based techniques, and Recurrence Quantification Analysis (RQA) [9,18]. BGP traffic has also been modeled as a dynamical system displaying determinism, nonlinear, periodic, and stable characteristics [9]. Previous work has employed Auto-Correlation Functions and FFTs to identify periodicity characteristics in unstable BGP traffic time series data [9,18].

Extant BGP anomaly detection approaches have varying strengths and limitations. For example, some ML approaches have been incapable of detecting AS-path spoofing from top-tier ASes [1], while time series approaches using Wavelet Transforms show promise in locating the source of an anomaly but are limited by time and inappropriate for real-time detection [18,59]. Alternatively, other techniques have succeeded with the identification and differentiation of BGP anomalies yet router configuration prohibited real-time detection [32].

Compared to other techniques, there are relatively few data mining applications for BGP anomaly detection. None of the extant examples of data mining techniques for BGP anomaly detection have been directly compared to predominant ML and statistical pattern recognition approaches [60,61]. Our work compares a novel BGP data mining anomaly detection technique to Deep Learning models and one non-ML statistical pattern recognition technique, using the same publicly available incident data [6,9]. MP is fundamentally different from traditional training in ML models. While training in ML can involve adjusting model parameters based on labeled data, MP computes a profile of the time series data without any labels. The profiling in MP is a process of understanding the inherent patterns in the data, not training on specific outcomes. In contrast to ML approaches, MP requires no training and has less parameters in comparison to the statistical pattern recognition technique.

4. Matrix profile

Matrix Profile (MP) is a technique that can be used to detect anomalies in time series data and has a sound theoretical basis (Euclidean distance between elements of time series) that has been evaluated on hundreds of time series datasets [33,34], although not in the domain of BGP anomaly detection. Additionally, the MP package utilized in our work has been used in a matrix profile-assisted LSTM model to forecast COVID-19 cases [62]. Previous research has shown the MP technique can successfully detect anomalies in data with periodic characteristics, with minimal parameterization [35]. Given previous research has illustrated that BGP traffic exhibits periodic characteristics, it is hypothesized that MP could provide for a novel BGP anomaly detection approach.

There are advantages using MP for analyzing time series data described in the literature [34]. MP is domain-agnostic as it does not rely on domain-specific features or patterns, instead, it captures intrinsic structures within the time series data itself, making it applicable across various domains (e.g., it has been applied to hundreds of different time series datasets) without any domain-specific tuning. In contrast to approaches that require extensive parameterization and tuning, MP can be parameter-free and does not require training data. Unlike a majority of time series detection algorithms, MP is unfazed by large, sparse datasets. It allows for anytime computation whilst being extremely scalable and storage efficient; massive datasets can be processed in main memory, for example, and MP is extremely parallelizable. Due to an exceptionally low parameter scope, MP discords minimize overfitting and are also free of data assumptions. MP has also shown it can discover anomalies in datasets with missing data, with no FNs [63,64].

Developed from similarity-join research, the MP algorithm has proven to be a useful application for similarity and anomaly detection in time series data [35]. The precursor data challenge originally proposed was: given a collection of data objects, retrieve the nearest neighbor for every object [34]. A full description of the mathematics underpinning MP and an efficient algorithm to calculate a MP of a time-series has been previously described in the literature [33,34,65], however a brief summary of the core concepts of MP is provided below.

Consider a time series $T = \left(t_1, t_2, t_3, \dots, t_n\right)$ of length n, where t_i are real numbers for $i = 1, 2, \dots, n$. A subsequence of length $m < n$ is any sequence of m consecutive values from T, and so a time series of length n has a total of n-m+1 subsequences. The set of all subsequences of T that have a length m will be denoted by $T_m = \left\{T_{1,m}, T_{2,m}, \dots, T_{n-m+1,m}\right\}$ where $T_{i,m}$ is the subsequence of length m starting from t_i. For the definitions that will follow it is useful to define the index sets $I_m = \{i \in \mathbb{Z} : 1 \le i \le n - m + 1\}$ and $I_{m,i} = \left\{j \in I_m : j \ne i\right\}$.

As an example, Fig. 1 is a BGP time series from the Code Red I event ($n = 7200$). The choice of m will determine the number of subsequneces. For example, in Fig. 2, there would be a total of 5761 subsequences (n-m+1) that can be generated by incrementally moving a window of $m = 1440$ data points from left to right across the time series and shows the time series with MP computed, with an example subsequence of length m=1440 illustrated by the red section.

The all-subsequence set T_m has a MP given by the vector $P_{T_m} = \left(p_1, p_2, \dots, p_{n-m+1}\right)$, with elements p_i given by:

$$p_i = \min_{j \in I_{m,i}} d\left(T_{i,m}, T_{j,m}\right), \quad i = 1, 2, \dots, n - m + 1 \tag{1}$$

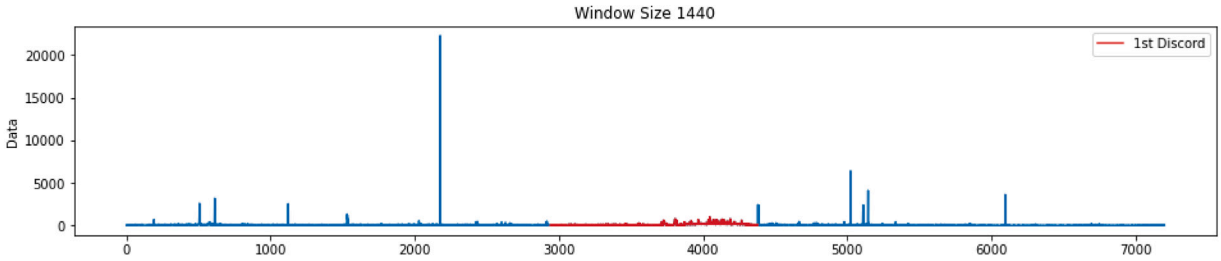

Fig. 2. Code Red I most significant discord subsequence range (m_2=1440).

where $d(a,b)$ denotes the Euclidean distance function:

$$d(a,b) = \sqrt{\sum_{i=1}^{m} \left(a_i - b_i\right)^2} \tag{2}$$

between vectors $a = \left(a_1, a_2, \dots, a_m\right)$ and $b = \left(b_1, b_2, \dots, b_m\right)$. That is, each element p_i of the MP gives the minimum Euclidean distance between the subsequence $T_{i,m}$ and all other subsequences of T_m. A relatively large value of p_i would indicate that subsequence $T_{i,m}$ is significantly different to the other subsequences of T_m, and therefore may correspond to an anomalous event. The MP index $Q_T = \left(q_1, q_2, \dots, q_{n-m+1}\right)$ is a vector of nearest neighbor indices corresponding to the MP vector P_{T_m}. That is, $q_i = j$ means that subsequence $T_{j,m}$ is the nearest neighbor of subsequence $T_{i,m}$.

A time series discord D_1 can be defined as $D_1 = T_{i_1,m}$ where $i_1 = \left\{i \in I_m : p_i > p_j \quad \forall j \in I_{m,i}\right\}$. That is, D_1 the subsequence of T_m that is most distant relative to its nearest neighbor [34]. The kth discord of a time series can be defined as $D_k = T_{i_k,m}$ where $i_k = \left\{i \in J_{m,k} : p_i > p_j \quad \forall j \in J_{m,k,i}\right\}$, $J_{m,k} = \left\{i \in I_m : T_{i,m} \cap D_j = \phi,\ j = 1, 2, \dots, k-1\right\}$ and $J_{m,k,i} = \left\{j \in J_{m,k} : j \neq i\right\}$. That is, the kth discord is the subsequence of T_m that is most distant relative to its nearest neighbor, excluding any subsequences that intersect/overlap a preceding discord [33]. We can also extract the *top-k* discords. In the example of Fig. 2, the red section is the k_1 discord.

There are several MP algorithms that exist, and when an underlying time series is large, the calculation of MP requires efficient algorithms; the STAMP, STOMP and SCRIMP algorithms are examples of efficient methods for calculating MP and time series discords that have been described in the literature [34,65]. Additionally, MP can also be deployed with zero parameters, using a pan-matrix-profile function; the package we use has this capability and is publicly available [66].

The efficiency and scalability of the MP approach are exemplified by a suite of algorithms tailored to various aspects of time series analysis. STAMP, for example, offers an ‘anytime’ capability, allowing for early insights into data patterns before the complete computation, making it particularly useful for large datasets where computational resources may be limited [34].

STOMP is an improvement over STAMP, and optimizes the matrix profile calculation by utilizing an ordered-search mechanism that significantly reduces the computational cost, making it suitable for even larger datasets [36,38]. SCRIMP++, on the other hand, combines optimal features from STAMP and STOMP, offering an efficient and incremental calculation of the matrix profile, and facilitating real-time analysis of streaming data [37]. These algorithms collectively contribute to MP’s scalability and storage efficiency.

A driving hypothesis for this work is that identification of discords in time series data might provide application for BGP anomaly detection, given BGP traffic exhibits characteristics that MP has been previously shown to detect. In addition, MP has been previously shown to detect anomalies in other environments within one second, which is an important indicator for the pursuit of near-real-time detection [35]. We seek to investigate the following questions:

- R_1: Can Matrix Profile detect BGP anomalies?
- R_2: Can Matrix Profile detect anomalous activity in different types of BGP incidents using BGP volume features?
- R_3: Is there any advantage with using Matrix Profile to detect anomalies in BGP?

5. Selected BGP incidents

We utilize publicly available data that has been previously used for work with both advanced nonlinear statistical analysis (RQA) and deep learning (RNNs) [6,57], to investigate the applicability of MP time series discords on the following BGP events. The MP software package used is publicly available [66]. The incidents include examples of all categories of BGP incident categories (direct, indirect and outage). The datasets were originally obtained from the RIPE and Route Views repositories, and the process of doing so has been documented in the literature for replication [6,20,57].

5.1. Code Red I

The Code Red I is a well-studied worm that is understood to have begun on July 15, 2001 and targeted web servers [6]. Studies have used data derived from the Code Red I incident and applied various machine learning approaches; specifically Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Broad Learning System (BLS) techniques [28,57]. The RNN results produced an accuracy of 90.69% for Code Red I.

5.2. Nimda

Nimda was a significant malware event that became public knowledge approximately September 18, 2001 that did not specifically target BGP but catalyzed BGP message overload across the Internet and thus is categorized an indirect BGP anomaly [9,67]. Nimda increased BGP volume traffic approximately 30 times from the normal volume of BGP updates [9]. Various machine-learning approaches have been used to analyze the Nimda incident, with LSTM detection achieving an accuracy of 92.00% [57,68]. Other approaches, such as RQA, have also used Nimda as a test-case, and successfully raised at least 10 True Positive alarms across the event [9].

5.3. Slammer

The Slammer malware was first observed in early 2003 and targeted servers [6]; while Slammer did not focus on BGP (indirect anomaly), the computer worm did significantly impact BGP. The impact of Slammer on BGP was shown to be rapid and devastating, contamination of at least 90 percent of vulnerable targets occurred in approximately 10 min [4,18]. Various anomaly detection techniques have used the incident as a test case. For example, RNN-based BGP anomaly detection approaches have previously utilized the Slammer incident as an experimental dataset, using GRU with an accuracy of 95.21% [6].

5.4. Moscow Blackout

The 2005 Moscow Blackout incident is an example of the anomalous category known as an outage (or 'link failure'). The incident resulted in hours of Internet 'blackout'; the outage impacted a Russian Internet exchange and several businesses. Approaches that have used the Moscow Blackout incident and associated data include machine-learning based approaches, such as RNNs, and statistical analysis techniques, such as RQA [6,18,57]. RQA has previously shown a high level of detection accuracy in this event (99.99%), raising at least 10 validated alarms across 597376 s of BGP updates [9].

5.5. Telekom Malaysia

Telekom Malaysia was an example of the Route Leak category of BGP incidents that occurred in 2015, whereby misconfiguration at the ISP resulted in a significant proportion of global routing table prefixes (179000) leaked to multinational telco Level 3 Communications. With Level 3 then accepting and propagating the IP prefixes. The incident has been previously described and investigated using RQA with reported accuracy of 100% [9].

5.6. WannaCrypt

The WannaCrypt incident of 2017 is the most recent of all the cases investigated in this paper. WannaCrypt was a devastating ransomware worm that obtained administrative privileges by deploying multiple exploits in systems running legacy Microsoft Windows operating systems. The WannaCrypt-related BGP data has been studied with deep learning RNNs (LSTM and GRU) and BLS, which were considered applicable due to their unique structure and capability to classify time series data [6]. Accuracy of 72.63% was achieved.

6. Methods

The BGP datasets we use for the evaluation of MP were drawn from established repositories (RIPE and Route Views) and are publicly available [6,20]. BGP update messages originally exist in multi-threaded routing toolkit (MRT) format within BGP repositories. As outlined in previous research [6], a Perl-based parsing tool is used to extract these datasets to ASCII, and a C# tool is used to extract features for the final datasets. Previous research has shown that BGP changes are primarily observed in volume features [6,9]. BGP anomalies caused by Slammer and WannaCrypt manifested noticeable changes in volume (e.g., BGP announcements and withdrawals) [6,18,21]. The research focus for these series of experiments was on BGP Volume features. We leave AS-PATH feature analysis for future research.

The MPA package used in our research is publicly available [66]. The use of the MPA package extends the functionality of conventional MP analysis by incorporating advanced algorithms such as STOMP, MASS, and SCRIMP++ [66], allowing for a comprehensive exploration of time series discords within BGP data. BGP Volume features are utilized, and we are interested in the presence of discords in the time series data. That is to say subsequences with the large (maximal) distances to their nearest neighbors. As described in Section 4, for a time series T, the MP of T will include a vector of subsequence pair distances (the distance profile) and a distance to nearest neighbor indexation (profile index).

While window selection in MP can be automated, effectively making MP parameter-free, the window size (m) has been described as robust such that effectively halving or tripling the window size has minimal impact on results [34,66]. It has also been stated that too small an m can increase false negatives and too large an m can produce false positives; with *a priori* knowledge of the anomaly length or duration shown as the optimal choice [34,69].

That is to say *reported* incident and attack knowledge is hypothesized as the ideal window sizes (e.g., m_1 and m_2). As such, we choose the anomaly duration as characterized by previous work with the same datasets as m_1 and the broadly reported period of the incident (eg., 24 h or 1440 min) as m_2, where possible. For direct comparison against the RQA technique and associated alarms, a smaller window size is required for the analysis of more than 10 alarms in some incidents. A comparable window size in these experiments is used ($m = 120$).

Having *a priori* knowledge of attack duration is not always possible, nor realistic, and we tested m robustness by examining a range of m values. Further evaluation of the MP detection scheme uses conventional combinations of false positive (FP), false negative (FN), true positive (TP) and true negative (TN), where the aforementioned have the usual definitions. As previous work has done with the same datasets [6,9], we utilize detection accuracy and F-score (F_1) metrics. Accuracy (A) is a measure of correctly classified anomalies:

$$\mathrm{A} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{3}$$

and F-score (specifically F_1, the harmonic mean of precision and recall) is a function of successful anomaly detection:

$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

The performance evaluation is based on the MP values of the time series data. For each sliding window, the MPA package computes its MP value, which represents its distance to its nearest neighbor in the time series, as described in Section 4. Windows with MP values exceeding a certain threshold are flagged as anomalies. For performance evaluation, the comparative works utilized A and F_1 score as evaluation metrics only, and we adopted the same for a consistent comparison. Use of additional performance metrics, such as Matthew's Correlation Coefficient (MCC), was constrained by the available data from the referenced studies we are directly comparing against, as not all the requisite information was available.

Our novel approach to BGP anomaly detection is compared to Deep Learning techniques (LSTM and GRU) and one non-ML statistical pattern recognition technique (RQA), using publicly available incident data [6,9]. Multiple RNN models (LSTM and GRU) are included due to their efficacy in handling sequential data and proven detection of BGP anomalies [6]. LSTM's design includes mechanisms to learn long-term dependencies, avoiding the vanishing gradient problem, while GRU models provide a simplified structure that can perform comparably with fewer parameters [6].

For a non-ML comparative analysis, the advanced nonlinear statistical analysis technique RQA for BGP anomaly detection is reported on [9], providing insights into the dynamical properties of the BGP time series data [70,71]. The RQA detection scheme reported on performance metrics (A and F_1) for Nimda, Moscow Blackout, and Telekom Malaysia using RIPE data [9]. As in the RQA study, we also complete manual investigation of source data to determine if a discord event is a TP or FP. While manual inspection is valuable and necessary for research and validation purposes, it is not scalable for a fully developed and operational detection scheme. In this stage of research, manual inspection is crucial to validate true positives, false positives, and other metrics, ensuring the robustness of our approach. The comparative work with machine learning anomaly detection approaches for Slammer and WannaCrypt, have described best A and F_1 results from using BGP update message data collected by Route Views, and for all remaining incidents we use the RIPE datasets [6].

7. Results and performance evaluation

We applied the MP technique to each of the BGP incident datasets previously described. We report on the results in chronological order thus: Code Red I (July 2001), Nimda (September 2001), Slammer (January 2003), Moscow blackout (May 2005), Telekom Malaysia (June 2015) and WannaCrypt (May 2017) (see Fig. 12).

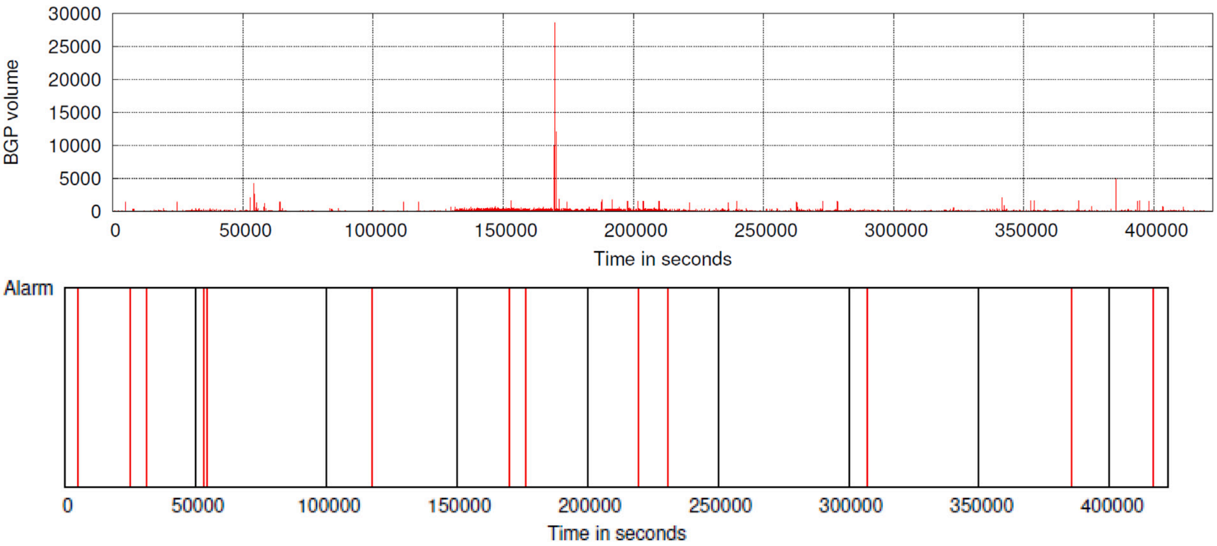

Fig. 3. RQA alarms raised in Nimda [9].

The RQA scheme has been previously applied to Nimda, Moscow Blackout, and Telekom Malaysia only. The RNN schemes have been applied to Code Red I, Nimda, Slammer, Moscow blackout, and WannaCrypt only. We compared the MP method against these detection schemes. As previously outlined, an ideal m parameter selection can be chosen with *a priori* knowledge of the reported incident, however, this is not always possible nor realistic and we therefore tested a range of m values with acceptable accuracy across 60 different m parameters ($A = 99.95$ and $F_1 = 92.57$).

The five-day Code Red I dataset ($n = 7200$) shown in Figs. 1 and 2 includes data for the day of the reported attack, coupled with data for two days prior and two days following the reported incident. A 599-minute period of traffic has been previously characterized by RNNs as anomalous ($A = 95.92$ and $F_1 = 73.96$) [6]. We utilize the previously characterized anomalous time period (599) as m_1 and one day of the reported incident period (1440 min) as m_2. We identify up to 10 discords and manual inspection of the data confirms that nine discords are TP alarms, while it is uncertain if a final discord is an FP or early detection of anomalous activity. We deem it to be FP until evidence suggests otherwise.

We conduct a top-k discord analysis (k=2) to analyze the top discords within the m_1 (599) and m_2 (1440) window sizes (see Figs. 1 and 2). As shown in Fig. 1, the k_1 discord of m_1 (599) signals an alarm at 3320 min, some 361 min earlier than the previously RNN characterized anomaly period [6], and a manual inspection confirms the discord subsequence range is encapsulated within the attack period. The k_2 discord of m_1 (599) shown in Fig. 1, is also validated as a TP within the attack period. The k_1 discord in the m_2 analysis shown in Fig. 2 also signals an early alarm (at 2900 min), and a manual inspection confirms the discord subsequence range is encapsulated within a previously classified attack period. While a number of TP late alarms were identified, they are excluded from further analysis.

The RQA scheme has previously reported 13 alarms signaled (10 TPs and 3 FPs) for the Nimda incident, with accuracy and F-score of 99.99% and 86.95% respectively (see Fig. 3 and Table 1). We compare the MP scheme with 13 discords across the same incident shown in Fig. 4. Similarly to the RQA technique [9], we establish that 10 alarms (including at least four alarms during the event date) are TP while it remains uncertain if three alarms constitute FP alarms or early anomaly detection. We deem these FPs until evidence proves otherwise. A period of 3674–4974 (1300 min) has been previously characterized by RNNs as anomalous. As previously described, incident and attack *a priori* knowledge has been hypothesized as the ideal window sizes (e.g., $m_1 = 1300$). As seen in Fig. 5, BGP volume traffic increased up to approximately 30 times from normal during Nimda, and the top-k analysis for window size m_1 shows one early alarm (at 1100 mins) and one alarm during the event (at 4360 mins). Additionally, analysis of the discord subsequence for the m_2 parameter produced an early alarm (k_2) at 1100 min, and extends into the previously RNN-classified anomalous traffic period.

The Slammer incident ($n = 7200$ mins) shown in Fig. 6 has been previously classified to have an 868-minute anomalous period (3200 - 4068 min) within a 24-hour period. As with previous experiments, previously classified anomalous activity serves as m_1 (868 min) and the 24-hour reported incident period serves as m_2 (1440 min). The k_2 discords for m_1 (868) and m_2 (1440) both signal early TP alarms, Fig. 6 illustrates the top-k discords for m_1. The k_2 discord shown in purple signals at 3030 min, extending into the previously RNN characterized anomalous period (3200-4068 min). One discord in this incident remained unclear whether it is a false alarm, and we label it FN until evidence can establish otherwise.

The RQA scheme has previously reported 12 alarms for the Moscow Blackout incident; 10 TPs and 2 FPs with accuracy and F-score of 99.99% and 90.90% respectively (Fig. 7). We compare the MP scheme with 12 discords analyzed across the same incident in Fig. 8. Similarly to the RQA study [9], we establish that 10 alarms (including at least four alarms during one established anomalous range between 2500–4500 mins shown in Fig. 8) are TP, while it remains uncertain if two of the alarms constitute FP alarms or early anomaly detection. An analysis of m_2 (1440) produced an early k_1 discord and the subsequence range (shown in red in Fig. 9) begins at 2800 mins and extends into the previously RNN classified anomalous period. In total, we validate that 10 alarms are TP while two remain unclear if they are FP or represent anomalous activity. We deem these two alarms FP until further evidence proves to the contrary.

With regards to the Malaysia Telekom incident, the RQA scheme has previously reported 8 alarms signaled found and all were TPs with accuracy and F-score of 100% and 100% respectively (Fig. 10). We compare the MP scheme with 8 discords across the same incident shown in Fig. 11 ($m = 120$). Via manual inspection of the dataset and reported event, we establish that 7 BGP anomaly alarms are TP, and we deem one early signal as an FP alarm. The earliest alarm was identified at approximately 150 min in the incident which is comparable to the signal identified at approximately 63 min with the RQA technique. As shown in Table 1, the MP has acceptable performance when directly compared with the RQA technique.

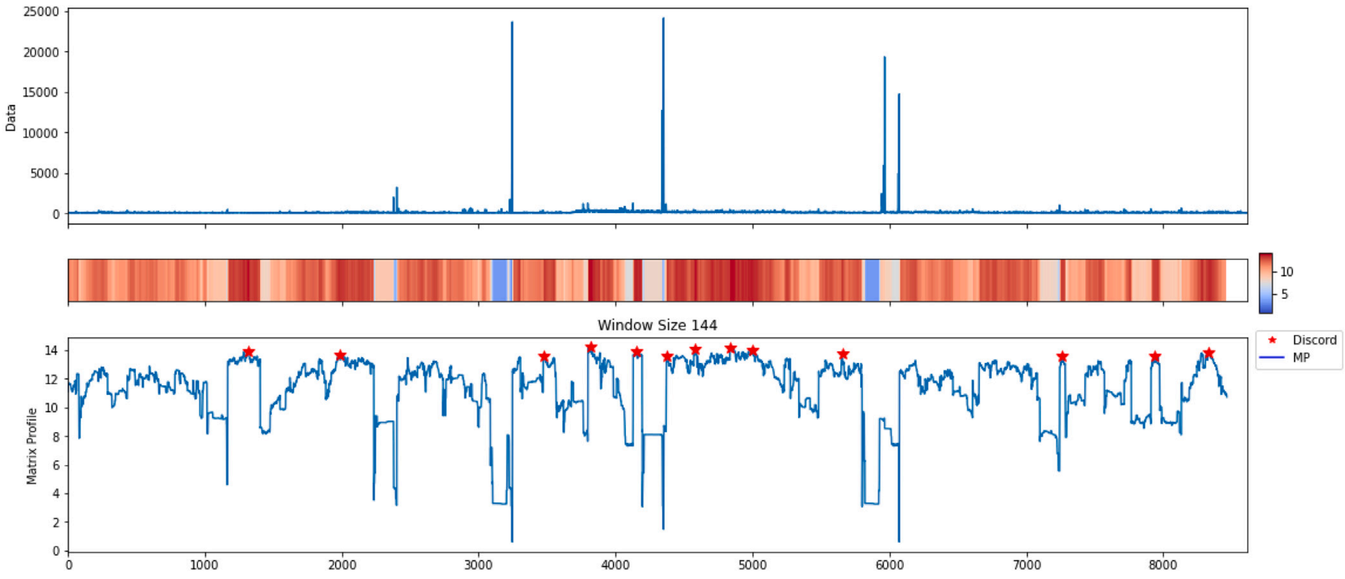

Fig. 4. MP alarms raised in Nimda.

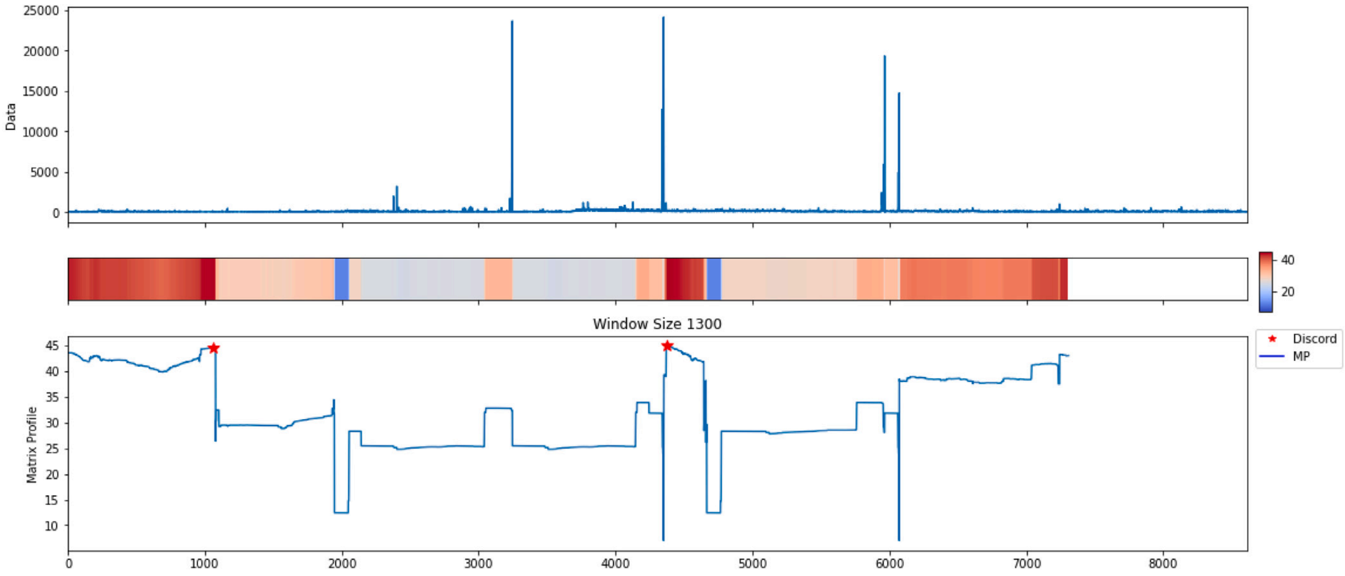

Fig. 5. Top-k discords for Nimda (m_1 = 1300).

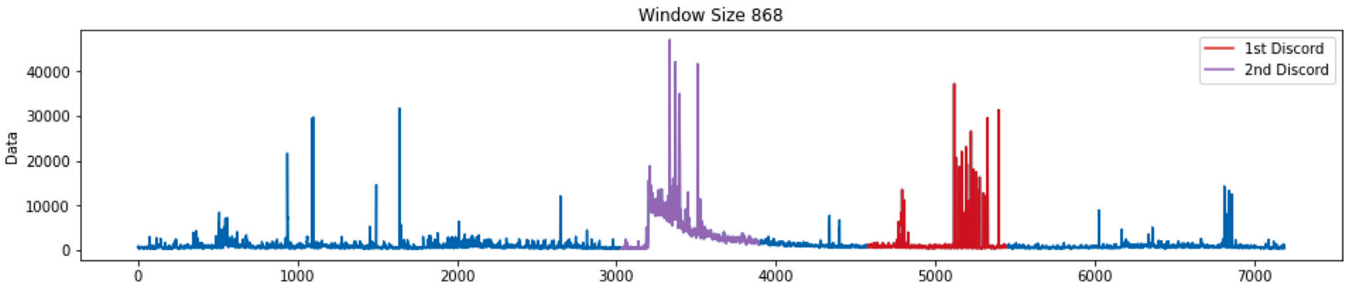

Fig. 6. Slammer top-k discord subsequences (m_1 = 868).

The WannaCrypt event has a 5759-minute period of traffic previously characterized by RNNs as anomalous. Due to the time series length (n), an m of 5759 will only produce a single significant discord and associated subsequence range, despite what top-k value is chosen. Therefore, for top-k analysis, we chose a 48-hour (2880) and 24-hour (1440) period to examine for m_1 and m_2 respectively. As is shown for m_1 (2880), the discords are both signaled during the previously RNN classified range (see Fig. 12). The k_1 discord begins at 3860 min and the k_2 at 6850 min. A top-k analysis of m_2 (1440) also identified the k_1 discord during the previously RNN classified range, whilst the k_2

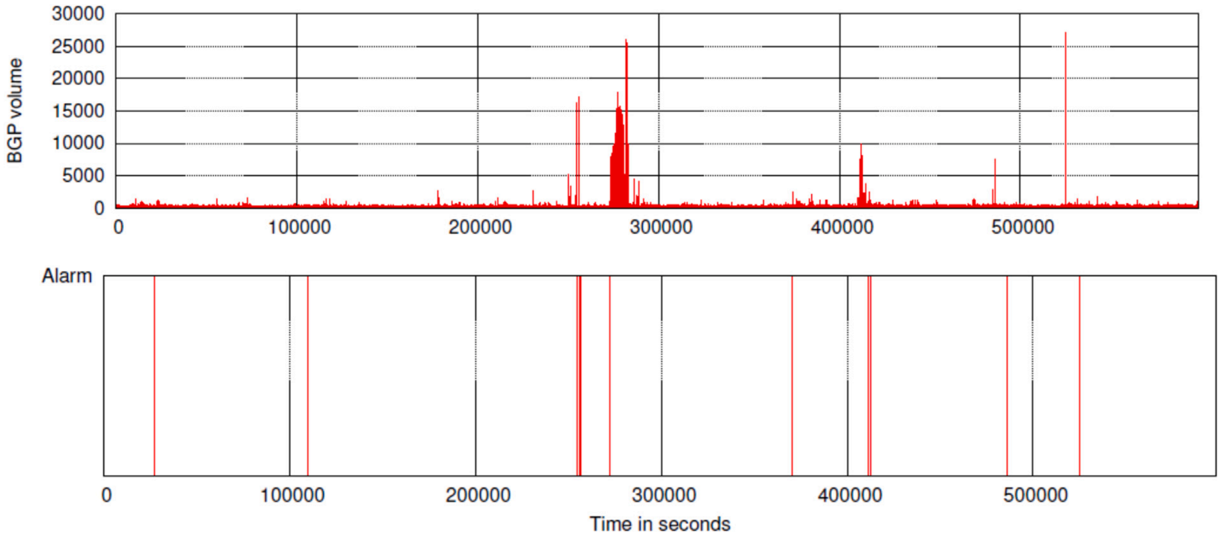

Fig. 7. RQA alarms raised in Moscow Blackout [9].

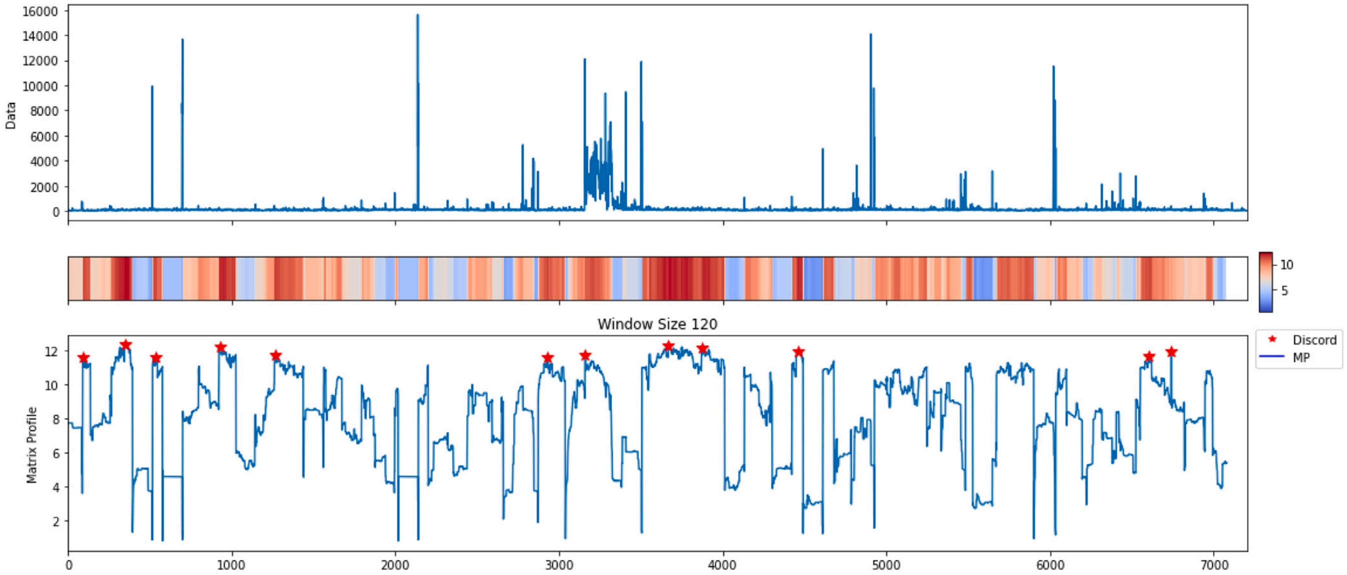

Fig. 8. MP alarms raised in Moscow Blackout.

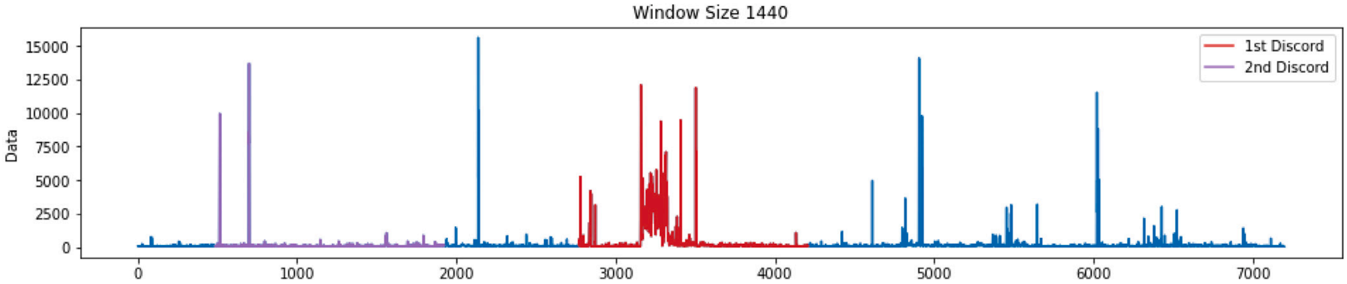

Fig. 9. Top-k subsequence for Moscow Blackout ($m_2 = 1440$).

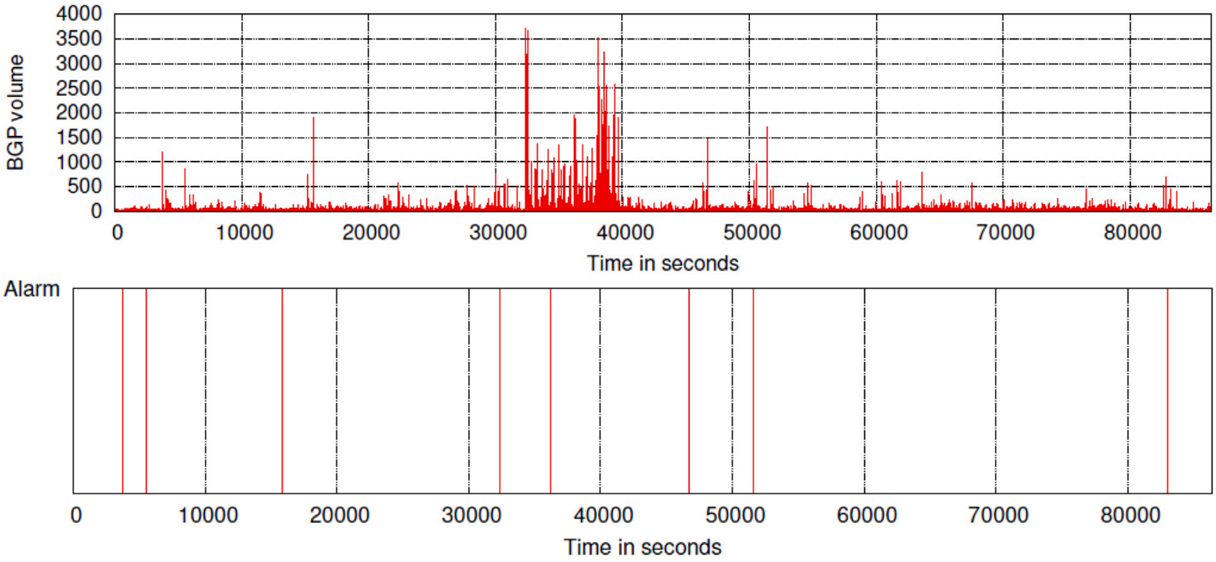

Fig. 10. RQA alarms raised in the Malaysia Telekom incident [9].

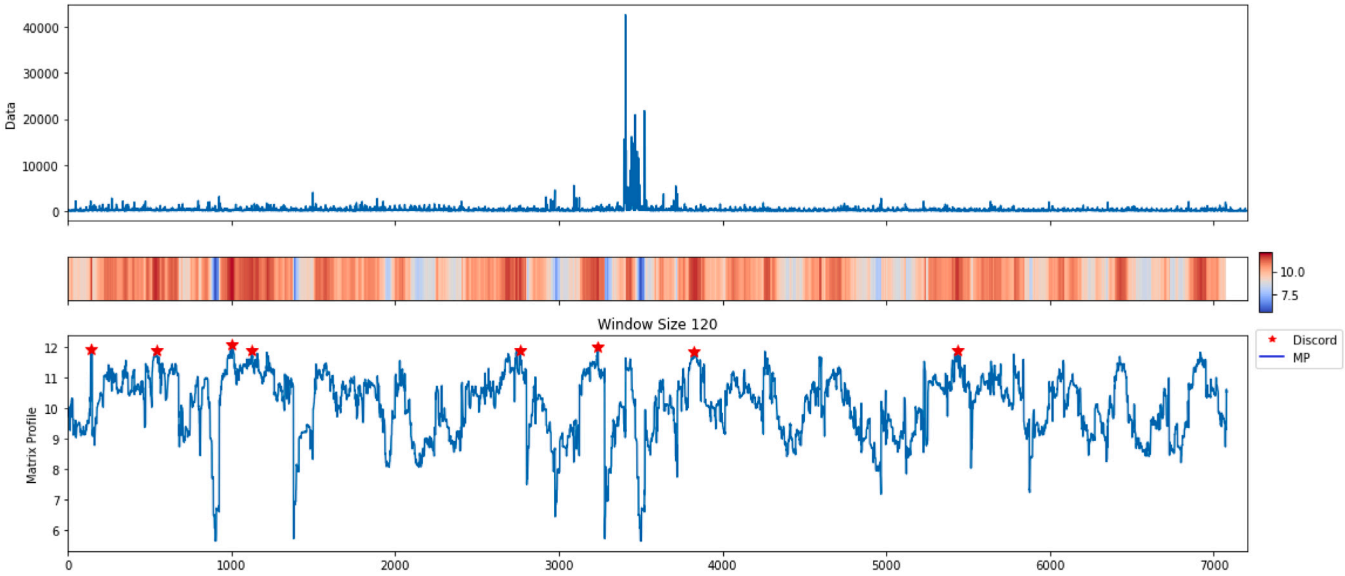

Fig. 11. MP alarms raised in the Malaysia Telekom incident.

Table 1
Performance evaluation.

Incident/technique	RQA (A)	RNNs (A)	MP (A)	RQA (F_1)	RNNs (F_1)	MP (F_1)	RQA (mins)	RNNs (mins)	MP (mins)
Code Red I	–	95.92	99.98	–	73.96	94.74	–	3681	2900
Nimda	99.99	92.00	99.85	86.95	95.83	86.96	83.3	3674	1100
Slammer	–	95.72	99.98	–	81.77	94.74	–	3200	2240
Moscow Blackout	99.99	98.30	99.97	90.90	35.15	90.91	417	3121	2770
Telekom Malaysia	100	–	99.99	100	–	93.33	3000	–	650
WannaCrypt	–	72.63	99.99	–	74.21	94.74	–	2881	2050

discord is an earlier alarm signaled at 2050 min. In total, we identified 9 TP alarms and 1 alarm that we deem an FP, as on manual inspection of the event data and reports, it remains unclear if the activity is anomalous.

8. Detection scheme

The MPBGP detection scheme is designed to be configured for both research purposes (e.g., networking testbeds and simulations) and real-world deployment (e.g., configured at BGP speakers or to obtain collector data). Preprocessing configuration is dependent on the use of the scheme. The specific configuration of the BGP speaker to enable BGP monitoring and exporting of data depends on the software being used (e.g., BIRD, Quagga, or ExaBGP). The MPBGP scheme contains configuration files and specific modules to communicate with BGP speakers and handle BGP protocol messages, for example, to provide the necessary functionality to establish a BGP session, send and receive BGP messages from a peer, collector or testbed environment. Regardless of the deployment environment, the feature extraction and preprocessing steps are the same. These steps are: extract the multi-threaded routing toolkit (MRT) format BGP data to ASCII using a BGP parser

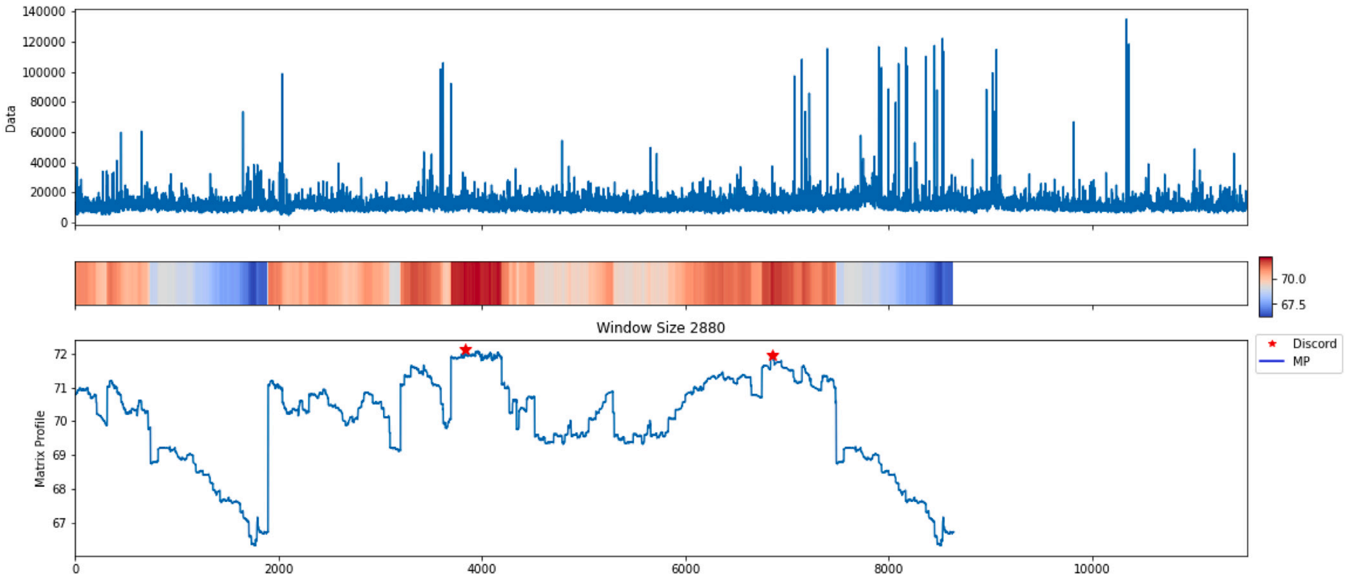

Fig. 12. WannaCrypt top-k analysis for m_2 (2880).

Algorithm 1: MPBGP Anomaly Detection Algorithm.

```
procedure MPBGPDetect($T$, $windowsize$, $k$)
    $dataset \leftarrow np.dataload(D)$
    $profile \leftarrow mp.compute(dataset, windowsize)$
    $discords \leftarrow mp.discover.discords(profile, k = k, exclusionzone = windowsize)$
    $discordnum \leftarrow list(range(1, k + 1))$
    for $i = 0$ to $k - 1$ do
        if $i == 0$ then
            $discordnum[i] \leftarrow str(discordnum[i]) + "st"$
        else if $i == 1$ then
            $discordnum[i] \leftarrow str(discordnum[i]) + "nd"$
        else if $i == 2$ then
            $discordnum[i] \leftarrow str(discordnum[i]) + "rd"$
        else
            $discordnum[i] \leftarrow str(discordnum[i]) + "th"$
        end if
    end for
    $mpadjusted \leftarrow$ np.append($profile['mp']$, [np.nan] $*$ $(profile['w'] - 1))$
    $fig, ax \leftarrow$ plt.subplots($1, 1, figsize = (16, 3)$)
    $ax.plot(profile['data']['ts'])$
    $ax.settitle(f"WindowSize(windowsize)")$
    $ax.setylabel('Data')$
    for $i, discord$ in enumerate(profile['discords']) do
        $x \leftarrow$ np.arange($discord, discord + profile['w']$)
        $y \leftarrow profile['data']['ts'][discord : discord + profile['w']]$
        $ax.plot(x, y, c =' C' + str(i + 3),$
$label = "Discord".format(discordnum[i]))$
    end for
    plt.legend()
    plt.show()
end procedure
```

and scanning tool whereby feature extraction is then completed and the output datasets are then used by the MPBGP algorithm.

The heart of the MP detection engine was described in Section 4, and is an algorithm that takes the preprocessed BGP feature extracted data and computes the matrix profile of the time series using a default or chosen window size, finding the top k discords (i.e., subsequences that are most different from the rest of the time series). The algorithm is shown in 1. The BGP discords are then plotted into an operator dashboard for visualization and analysis purposes. The implications for use of the detection scheme in future research are discussed in Section 9.

9. Discussion and future work

The advantages of MP as an anomaly detection scheme for BGP are due to the inherent advantages of the MP discords and the underlying algorithms themselves. MP is domain agnostic, free of data assumptions, and due to minimal parameters, there is a minimized risk of over-fitting. MP has also shown it can discover anomalies in datasets with missing data without producing FNs, as evidenced by the missing data points in the Nimda dataset that our work has utilized.

As one of the very few parameters in MP, the window size (m) has been previously described as robust to change, and we tested this assertion. Ultimately, our results are consistent with previous research that asserts an ideal m size for MP can be informed by *a priori* knowledge. Second, the assertion that m is robust in the absence of *a priori* knowledge is supported by the results. From the range of window sizes (m) tested, there were a total of 27% late alarms while 73% were signaled either early or during the reported incident, producing 55 TP alarms and 5 alarms we deem as FP, showing that the MP parameter is robust with respect to change. This emphasizes MP's capability for early detection, which is crucial for timely intervention before substantial impact.

Manual investigation of source data is not a trivial process, however it is necessary to determine if a discord event is a TP or FP, therefore we conduct this process with all discord alarms. MP is competitive, and in some cases, outperforms other detection schemes—the performance evaluation summary can be seen in Table 1.

In other environments, MP has been shown to be capable of detecting anomalies within the first–second. Previous work using advanced non-linear statistical analysis techniques, modeling BGP as a dynamical system, has also indicated that hidden anomalous behavior can represent an early stage of BGP anomalies. As such we compared MP against the RQA scheme that modeled BGP as a dynamical system—we identify similar early alarms and find MP to be competitive across the incidents directly compared: Nimda, Moscow Blackout, and Telekom Malaysia. Our work focused on BGP volume features, however we propose an investigation of MP using BGP path features for future work.

When compared against the RNN approach, the MP detection scheme signals earlier TP alarms in all the incidents investigated. The most significant early discord for Code Red I began some 781 min (approximately 13 h) before the previously deep learning RNN detected anomalous traffic, underscoring MP's early detection capabilities. Similar early alarms were identified in the Nimda, Slammer, Moscow Blackout, and WannaCrypt incidents, with MP signaling significant early discords well before RNNs, highlighting the importance of early detection for effective BGP anomaly mitigation.

The ability to detect BGP anomalies early, as demonstrated by MP, is akin to having a highly accurate fire alarm that alerts well in advance of an emergency, rather than too late to be of any use. A highly accurate but late fire alarm is of limited use. This early warning capability is the most crucial aspect of our findings. It is important to also highlight that this work desired to evaluate an under-represented area of BGP anomaly detection technique (data mining) with a direct comparative evaluation to well established ML and statistical pattern recognition techniques. As outlined in previous sections, the driving motivation of this paper was to first validate if MP can detect BGP anomalies, if there is any advantage using it, and how it performs with some of the most well studied volume-centric incidents and compared directly against other methods that used the same incidents using the same performance evaluation metrics.

Development and improvement of MP algorithms is also an active area of research whereby multi-dimensional variants may support improvements in MPBGP and is left for future research. Nevertheless, the use of the same publicly available datasets with publicly available packages allows for our results to be tested transparently against other approaches.

10. Conclusion

The suitability of using Matrix Profile (MP) for BGP anomaly detection has been evaluated in the context of all categories of BGP incidents. The advantages of MP that have been explored in the literature, discussed and evaluated in the context of BGP within this work, represent a promising novel BGP anomaly detection scheme. For example, BGP speaker configuration requirements have shown that some approaches can be prohibitive to real-time detection. The use of MP for anomaly detection in BGP research provides several advantages over existing techniques (as it has done in other domains). In contrast to many other approaches to BGP anomaly detection, MP is essentially parameter-free, MP is unimpeded by large, sparse datasets, whilst being extremely scalable and storage efficient–it is possible for massive datasets to be processed in main memory and handles missing data with no FNs. As mentioned in Section 9, MP discords minimize over-fitting and are also free of data assumptions. When considering the requirement for configuration of any anomaly detection scheme on BGP speakers, the advantages of MP are apparent. Our contribution is the application of a time series approach to detect BGP anomalies in all categories of BGP events. The single parameter analyzed in MP shows it is robust to change. Our results indicate that the MP detection scheme is competitive against existing detection schemes.

CRediT authorship contribution statement

Ben A. Scott: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Michael N. Johnstone:** Supervision, Writing – review & editing. **Patryk Szewczyk:** Supervision, Writing – review & editing. **Steven Richardson:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The work was supported by the Cyber Security Research Centre Limited whose activities are partially funded by the Australian Government's Cooperative Research Centres Programme.

References

[1] S. Cho, R. Fontugne, K. Cho, A. Dainotti, P. Gill, BGP hijacking classification, in: 2019 Network Traffic Measurement and Analysis Conference, TMA, IEEE, 2019, pp. 25–32, http://dx.doi.org/10.23919/TMA.2019.8784511, [Online]. Available: https://ieeexplore.ieee.org/document/8784511/.

[2] K. Kirkpatrick, Fixing the internet, Commun. ACM 64 (8) (2021) 16–17, http://dx.doi.org/10.1145/3469287, [Online]. Available: https://dl.acm.org/doi/10.1145/3469287.

[3] Q. Li, M. Xu, J. Wu, X. Zhang, P.P.C. Lee, K. Xu, Enhancing the trust of internet routing with lightweight route attestation, IEEE Trans. Inf. Forensics Secur. 7 (2) (2012) 691–703, http://dx.doi.org/10.1109/TIFS.2011.2177822.

[4] M. Lad, X. Zhao, B. Zhang, D. Massey, L. Zhang, Analysis of BGP Update Surge During Slammer Worm Attack, Springer, 2003, pp. 66–79.

[5] P. Moriano, R. Hill, L.J. Camp, Using bursty announcements for detecting BGP routing anomalies, Comput. Netw. 188 (2021) 107835, http://dx.doi.org/10.1016/j.comnet.2021.107835, [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1389128621000207.

[6] Z. Li, A.L.G. Rios, L. Trajkovic, Detecting Internet worms, ransomware, and blackouts using recurrent neural networks, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics, SMC, IEEE, Toronto, ON, Canada, 2020, pp. 2165–2172, http://dx.doi.org/10.1109/SMC42975.2020.9283472, [Online]. Available: https://ieeexplore.ieee.org/document/9283472/.

[7] M. Zhang, J. Li, S. Brooks, I-Seismograph: Observing, measuring, and analyzing internet earthquakes, IEEE/ACM Trans. Netw. 25 (6) (2017) 3411–3426, http://dx.doi.org/10.1109/TNET.2017.2748902.

[8] C. Zou, W. Gong, D. Towsley, L. Gao, The monitoring and early detection of Internet worms, IEEE/ACM Trans. Netw. 13 (5) (2005) 961–974, http://dx.doi.org/10.1109/TNET.2005.857113.

[9] B. Al-Musawi, Detecting BGP Anomalies Using Recurrence Quantification Analysis (Ph.D. thesis), Swinburne University of Technology, 2018.

[10] C.C. Demchak, Y. Shavitt, China's maxim–leave no access point unexploited: The hidden story of China Telecom's BGP Hijacking, Mil. Cyber Aff. 3 (1) (2018) 7.

[11] J.M. Smith, K. Birkeland, T. McDaniel, M. Schuchard, Withdrawing the BGP re-routing curtain: Understanding the security impact of BGP poisoning through real-world measurements, in: Proceedings 2020 Network and Distributed System Security Symposium, Internet Society, San Diego, CA, 2020, http://dx.doi.org/10.14722/ndss.2020.24240, [Online]. Available: https://www.ndss-symposium.org/wp-content/uploads/2020/02/24240.pdf.

[12] J. Sherman, The Politics of Internet Security: Private Industry and the Future of the Web, Tech. Rep., Atlantic Council, 2020, pp. 9–10, [Online]. Available: http://www.jstor.org/stable/resrep26661.

[13] A. Mitseva, A. Panchenko, T. Engel, The state of affairs in BGP security: A survey of attacks and defenses, Comput. Commun. 124 (2018) 45–60, http://dx.doi.org/10.1016/j.comcom.2018.04.013, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014036641731068X.

[14] C. Testart, P. Richter, A. King, A. Dainotti, D. Clark, Profiling BGP serial hijackers: Capturing persistent misbehavior in the global routing table, in: Proceedings of the Internet Measurement Conference, ACM, Amsterdam Netherlands, 2019, pp. 420–434, http://dx.doi.org/10.1145/3355369.3355581, [Online]. Available: https://dl.acm.org/doi/10.1145/3355369.3355581.

[15] P. Sermpezis, V. Kotronis, A. Dainotti, X. Dimitropoulos, A survey among network operators on BGP prefix hijacking, SIGCOMM Comput. Commun. Rev. 48 (1) (2018) 64–69, http://dx.doi.org/10.1145/3211852.3211862, [Online]. Available: https://dl.acm.org/doi/10.1145/3211852.3211862.

[16] F. Douzet, L. Pétiniaud, L. Salamatian, K. Limonier, K. Salamatian, T. Alchus, Measuring the fragmentation of the Internet: The case of the border gateway protocol (BGP) during the Ukrainian crisis, in: 2020 12th International Conference on Cyber Conflict, Vol. 1300, CyCon, 2020, pp. 157–182, http://dx.doi.org/10.23919/CyCon49761.2020.9131726.

[17] K. Limonier, F. Douzet, L. Pétiniaud, L. Salamatian, K. Salamatian, Mapping the routes of the Internet for geopolitics: The case of Eastern Ukraine, FM (2021) http://dx.doi.org/10.5210/fm.v26i5.11700, [Online]. Available: https://journals.uic.edu/ojs/index.php/fm/article/view/11700.

[18] B. Al-Musawi, P. Branch, G. Armitage, BGP anomaly detection techniques: A survey, IEEE Commun. Surv. Tutor. 19 (1) (2016) 377–396.

[19] A. Al-Bakaa, B. Al-Musawi, A new intrusion detection system based on using non-linear statistical analysis and features selection techniques, Comput. Secur. 122 (2022) 102906, http://dx.doi.org/10.1016/j.cose.2022.102906, [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167404822002991.

[20] P. Fonseca, E.S. Mota, R. Bennesby, A. Passito, BGP dataset generation and feature extraction for anomaly detection, in: 2019 IEEE Symposium on Computers and Communications, ISCC, IEEE, Barcelona, Spain, 2019, pp. 1–6, http://dx.doi.org/10.1109/ISCC47284.2019.8969619, [Online]. Available: https://ieeexplore.ieee.org/document/8969619/.

[21] N.H. Hammood, B. Al-Musawi, Using BGP features towards identifying type of BGP anomaly, in: 2021 International Congress of Advanced Technology and Engineering, ICOTEN, IEEE, Taiz, Yemen, 2021, pp. 1–10, http://dx.doi.org/10.1109/ICOTEN52080.2021.9493491, [Online]. Available: https://ieeexplore.ieee.org/document/9493491/.

[22] M. Hashem, A. Bashandy, S. Shaheen, Improving anomaly detection in BGP time-series data by new guide features and moderated feature selection algorithm, Turk. J. Electr. Eng. Comput. Sci. 27 (1) (2019) 392–406, http://dx.doi.org/10.3906/elk-1804-55, [Online]. Available: https://journals.tubitak.gov.tr/elektrik/vol27/iss1/29.

[23] A. Allahdadi, R. Morla, R. Prior, A framework for BGP abnormal events detection, 2017, http://dx.doi.org/10.48550/ARXIV.1708.03453, [Online]. Available: https://arxiv.org/abs/1708.03453 Publisher: arXiv Version Number: 1.

[24] N.M. Al-Rousan, L. Trajković, Machine learning models for classification of BGP anomalies, IEEE, 2012, pp. 103–108.

[25] X. Dai, N. Wang, W. Wang, Application of machine learning in BGP anomaly detection, J. Phys.: Conf. Ser. 1176 (2019) 032015, http://dx.doi.org/10.1088/1742-6596/1176/3/032015, Publisher: IOP Publishing.

[26] O.R. Sanchez, S. Ferlin, C. Pelsser, R. Bush, Comparing machine learning algorithms for BGP anomaly detection using graph features, in: Proceedings of the 3rd ACM CoNEXT Workshop on Big DAta, Machine Learning and Artificial Intelligence for Data Communication Networks, Big-DAMA '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 35–41, http://dx.doi.org/10.1145/3359992.3366640.

[27] K. Hoarau, P.U. Tournoux, T. Razafindralambo, Suitability of graph representation for BGP anomaly detection, in: 2021 IEEE 46th Conference on Local Computer Networks, LCN, IEEE, Edmonton, AB, Canada, 2021, pp. 305–310, http://dx.doi.org/10.1109/LCN52139.2021.9524941, [Online]. Available: https://ieeexplore.ieee.org/document/9524941/.

[28] M. Cheng, Q. Li, J. Lv, W. Liu, J. Wang, Multi-scale LSTM model for BGP anomaly classification, IEEE Trans. Serv. Comput. 14 (3) (2021) 765–778, http://dx.doi.org/10.1109/TSC.2018.2824809, [Online]. Available: https://ieeexplore.ieee.org/document/8334596/.

[29] M. Xu, X. Li, BGP anomaly detection based on automatic feature extraction by neural network, in: 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference, ITOEC, IEEE, Chongqing, China, 2020, pp. 46–50, http://dx.doi.org/10.1109/ITOEC49072.2020.9141762, [Online]. Available: https://ieeexplore.ieee.org/document/9141762/.

[30] T. Shapira, Y. Shavitt, A deep learning approach for IP Hijack detection based on ASN embedding, in: Proceedings of the Workshop on Network Meets AI & ML, NetAI '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 35–41, http://dx.doi.org/10.1145/3405671.3405814.

[31] P. Moriano, R. Hill, L.J. Camp, Using Bursty Announcements for Early Detection of BGP Routing Anomalies, 2019, arXiv preprint arXiv:1905.05835.

[32] Y. Huang, N. Feamster, A. Lakhina, J.J. Xu, Diagnosing network disruptions with network-wide analysis, SIGMETRICS Perform. Eval. Rev. 35 (1) (2007) 61–72, http://dx.doi.org/10.1145/1269899.1254890.

[33] E. Keogh, J. Lin, A. Fu, HOT SAX: efficiently finding the most unusual time series subsequence, in: Fifth IEEE International Conference on Data Mining, ICDM'05, 2005, p. 8, http://dx.doi.org/10.1109/ICDM.2005.79.

[34] C.-C.M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H.A. Dau, D.F. Silva, A. Mueen, E. Keogh, Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets, in: 2016 IEEE 16th International Conference on Data Mining, ICDM, IEEE, Barcelona, Spain, 2016, pp. 1317–1322, http://dx.doi.org/10.1109/ICDM.2016.0179, [Online]. Available: http://ieeexplore.ieee.org/document/7837992/.

[35] S. Duque Anton, L. Ahrens, D. Fraunholz, H.D. Schotten, Time is of the essence: Machine learning-based intrusion detection in industrial time series data, in: 2018 IEEE International Conference on Data Mining Workshops, ICDMW, IEEE, Singapore, Singapore, 2018, pp. 1–6, http://dx.doi.org/10.1109/ICDMW.2018.00008, [Online]. Available: https://ieeexplore.ieee.org/document/8637462/.

[36] Y. Zhu, Z. Zimmerman, N.S. Senobari, C.-C.M. Yeh, G. Funning, A. Mueen, P. Brisk, E. Keogh, Matrix profile II: Exploiting a novel algorithm and GPUs to break the one hundred million barrier for time series motifs and joins, in: 2016 IEEE 16th International Conference on Data Mining, ICDM, IEEE, Barcelona, Spain, 2016, pp. 739–748, http://dx.doi.org/10.1109/ICDM.2016.0085, [Online]. Available: http://ieeexplore.ieee.org/document/7837898/.

[37] Y. Zhu, C.-C.M. Yeh, Z. Zimmerman, K. Kamgar, E. Keogh, Matrix profile XI: SCRIMP++: Time series motif discovery at interactive speeds, in: 2018 IEEE International Conference on Data Mining, ICDM, IEEE, Singapore, 2018, pp. 837–846, http://dx.doi.org/10.1109/ICDM.2018.00099, [Online]. Available: https://ieeexplore.ieee.org/document/8594908/.

[38] C.-C.M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H.A. Dau, Z. Zimmerman, D.F. Silva, A. Mueen, E. Keogh, Time series joins, motifs, discords and shapelets: a unifying view that exploits the matrix profile, Data Min. Knowl. Disc. 32 (1) (2018) 83–123, http://dx.doi.org/10.1007/s10618-017-0519-9, [Online]. Available: http://link.springer.com/10.1007/s10618-017-0519-9.

[39] V. Jain, B. Edgeworth, Troubleshooting BGP: A Practical Guide to Understanding and Troubleshooting BGP, Cisco Press, 2016, Google-Books-ID LPLBDQAAQBAJ.

[40] K. Lougheed, Y. Rekhter, RFC1105: Border Gateway Protocol (BGP), RFC Editor, 1989.

[41] K. Lougheed, Y. Rekhter, A Border Gateway Protocol 3 (BGP-3), Tech. Rep, RFC 1267, Cisco Systems, TJ Watson Research Center, IBM Corp, 1991.

[42] S. Braman, Internet histories: the view from the design process, Internet Hist. 1 (1–2) (2017) 70–78, http://dx.doi.org/10.1080/24701475.2017.1305716, Publisher: Routledge.

[43] K. Boitmanis, U. Brandes, C. Pich, Visualizing Internet evolution on the autonomous systems level, in: S.-H. Hong, T. Nishizeki, W. Quan (Eds.), Graph Drawing, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2008, pp. 365–376, http://dx.doi.org/10.1007/978-3-540-77537-9_36.

[44] J. Ball, The Tangled Web We Weave: Inside the Shadow System that Shapes the Internet, Melville House Publishing, 2020.

[45] M. Roughan, W. Willinger, O. Maennel, D. Perouli, R. Bush, 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems, IEEE J. Sel. Areas Commun. 29 (9) (2011) 1810–1821, http://dx.doi.org/10.1109/JSAC.2011.111006.

[46] D.L. Alderson, J.C. Doyle, W. Willinger, Lessons from "a first-principles approach to understanding the Internet's router-level topology", SIGCOMM Comput. Commun. Rev. 49 (5) (2019) 96–103, http://dx.doi.org/10.1145/3371934.3371964, Place: New York, NY, USA Publisher: Association for Computing Machinery.

[47] R. Motamedi, B. Yeganeh, B. Chandrasekaran, R. Rejaie, B.M. Maggs, W. Willinger, On mapping the interconnections in today's Internet, IEEE/ACM Trans. Netw. 27 (5) (2019) 2056–2070, http://dx.doi.org/10.1109/TNET.2019.2940369, Conference Name: IEEE/ACM Transactions on Networking.

[48] Y. Rekhter, T. Li, A border gateway protocol 4 (BGP-4), 1995, http://dx.doi.org/10.17487/RFC1771, RFC 1771 (Draft Standard), RFC Editor, Fremont, CA, USA, Mar. 1995, obsoleted by RFC 4271. [Online]. Available: https://www.rfc-editor.org/rfc/rfc1771.txt.

[49] Y. Rekhter, T. Li, S. Hares (Eds.), A Border Gateway Protocol 4 (BGP-4), 2006, http://dx.doi.org/10.17487/RFC4271.

[50] E. Chen, Route Refresh Capability for BGP-4, 2000, http://dx.doi.org/10.17487/RFC2918, FC 2918 (ProposedStandard), RFC Editor, Fremont, CA, USA, Sep. 2000, updated by RFC7313. [Online]. Available: https://www.rfc-editor.org/rfc/rfc2918.txt.

[51] K. Patel, E. Chen, B. Venkatachalapathy, Enhanced Route Refresh Capability for BGP-4, 2014, http://dx.doi.org/10.17487/RFC7313, FC 7313 (Proposed Standard), RFC Editor, Fremont, CA, USA. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7313.txt.

[52] I.O. de Urbina Cazenave, E. Köşlük, M.C. Ganiz, An anomaly detection framework for BGP, IEEE, 2011, pp. 107–111.

[53] N.H. Hammood, B. Al-Musawi, A.H. Alhilali, A survey of BGP anomaly detection using machine learning techniques, in: S.R. Pokhrel, M. Yu, G. Li (Eds.), in: Applications and Techniques in Information Security, vol. 1554, Springer Singapore, Singapore, 2022, pp. 109–120, http://dx.doi.org/10.1007/978-981-19-1166-8_9, Series Title: Communications in Computer and Information Science. [Online]. Available: https://link.springer.com/10.1007/978-981-19-1166-8_9.

[54] A. Putina, S. Barth, A. Bifet, D. Pletcher, C. Precup, P. Nivaggioli, D. Rossi, Unsupervised real-time detection of BGP anomalies leveraging high-rate and fine-grained telemetry data, in: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS, 2018, pp. 1–2, http://dx.doi.org/10.1109/INFCOMW.2018.8406838.

[55] N. Al-Rousan, S. Haeri, L. Trajković, Feature selection for classification of BGP anomalies using Bayesian models, in: 2012 International Conference on Machine Learning and Cybernetics, Vol. 1, 2012, pp. 140–147, http://dx.doi.org/10.1109/ICMLC.2012.6358901.

[56] P. Batta, M. Singh, Z. Li, Q. Ding, L. Trajković, Evaluation of support vector machine kernels for detecting network anomalies, in: 2018 IEEE International Symposium on Circuits and Systems, ISCAS, 2018, pp. 1–4, http://dx.doi.org/10.1109/ISCAS.2018.8351647.

[57] Z. Li, A.L.G. Rios, G. Xu, L. Trajkovic, Machine learning techniques for classifying network anomalies and intrusions, in: 2019 IEEE International Symposium on Circuits and Systems, ISCAS, IEEE, Sapporo, Japan, 2019, pp. 1–5, http://dx.doi.org/10.1109/ISCAS.2019.8702583, [Online]. Available: https://ieeexplore.ieee.org/document/8702583/.

[58] O.S. Alkadi, N. Moustafa, B. Turnbull, K.-K.R. Choo, An ontological graph identification method for improving localization of IP prefix Hijacking in network systems, IEEE Trans. Inf. Forensics Secur. 15 (2020) 1164–1174, http://dx.doi.org/10.1109/TIFS.2019.2936975, Conference Name: IEEE Transactions on Information Forensics and Security.

[59] J. Mai, L. Yuan, C.-N. Chuah, Detecting BGP anomalies with wavelet, in: NOMS 2008 - 2008 IEEE Network Operations and Management Symposium, 2008, pp. 465–472, http://dx.doi.org/10.1109/NOMS.2008.4575169.

[60] S.T. Teoh, K. Zhang, S.-M. Tseng, K.-L. Ma, S.F. Wu, Combining visual and automated data mining for near-real-time anomaly detection and analysis in BGP, in: Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, VizSEC/DMSEC '04, ACM Press, Washington DC, USA, 2004, p. 35, http://dx.doi.org/10.1145/1029208.1029215, [Online]. Available: http://portal.acm.org/citation.cfm?doid=1029208.1029215.

[61] M.C. Ganiz, S. Kanitkar, M.C. Chuah, W.M. Pottenger, Detection of interdomain routing anomalies based on higher-order path analysis, in: Sixth International Conference on Data Mining, ICDM'06, 2006, pp. 874–879, http://dx.doi.org/10.1109/ICDM.2006.52.

[62] Q. Liu, D.L.X. Fung, L. Lac, P. Hu, A novel matrix profile-guided attention LSTM model for forecasting COVID-19 cases in USA, Front. Public Health 9 (2021) 741030, http://dx.doi.org/10.3389/fpubh.2021.741030, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2021.741030/full.

[63] Y. Zhu, A. Mueen, E. Keogh, Matrix profile IX: Admissible time series motif discovery with missing data, IEEE Trans. Knowl. Data Eng. 33 (6) (2021) 2616–2626, http://dx.doi.org/10.1109/TKDE.2019.2950623.

[64] R. Wankhedkar, S.K. Jain, Motif discovery and anomaly detection in an ECG using matrix profile, in: Progress in Advanced Computing and Intelligent Engineering, Springer, 2021, pp. 88–95.

[65] Y. Zhu, C.-C.M. Yeh, Z. Zimmerman, K. Kamgar, E. Keogh, Matrix profile XI: SCRIMP++: time series motif discovery at interactive speeds, in: 2018 IEEE International Conference on Data Mining, ICDM, IEEE, 2018, pp. 837–846.

[66] A.V. Benschoten, A. Ouyang, F. Bischoff, T. Marrs, MPA: a novel cross-language API for time series analysis, J. Open Source Softw. 5 (49) (2020) 2179, http://dx.doi.org/10.21105/joss.02179.

[67] M. Karimi, A. Jahanshahi, A. Mazloumi, H.Z. Sabzi, Border Gateway Protocol Anomaly Detection Using Neural Network, IEEE, 2019, pp. 6092–6094.

[68] J. Li, D. Dou, Z. Wu, S. Kim, V. Agarwal, An internet routing forensics framework for discovering rules of abnormal BGP events, SIGCOMM Comput. Commun. Rev. 35 (5) (2005) 55–66, http://dx.doi.org/10.1145/1096536.1096542, [Online]. Available: https://dl.acm.org/doi/10.1145/1096536.1096542.

[69] S.D.D. Antón, H.D. Schotten, Intrusion detection in binary process data: Introducing the hamming-distance to matrix profiles, 2020, CoRR [Online]. Available: https://arxiv.org/abs/2007.08813.

[70] N. Marwan, CRP Toolbox, 2013, [Online]. Available: https://tocsy.pik-potsdam.de/CRPtoolbox.

[71] N. Marwan, C.L. Webber, E.E.N. Macau, R.L. Viana, Introduction to focus issue: Recurrence quantification analysis for understanding complex systems, Chaos 28 (8) (2018) 085601, http://dx.doi.org/10.1063/1.5050929, Publisher: American Institute of Physics, [Online]. Available: https://aip.scitation.org/doi/full/10.1063/1.5050929.

Ben Scott completed a B.Sc. from the University of Queensland and a Masters (Cybersecurity) from Edith Cowan University (ECU). He is a Ph.D. Candidate at ECU and a Cybersecurity Cooperative Research Centre (CSCRC) Scholarship recipient. He researches the field of internet security engineering, mostly focusing on BGP anomaly detection.

Mike is an Associate Professor at the School of Science at Edith Cowan University where he teaches network security and mobile app development. As a member of the Security research Institute at ECU, his work on resilient systems covers secure development methodologies, wireless sensor networks and the security of IoT devices with a focus on critical infrastructure. With over 30 years of experience in ICT, he provides consultancy services in cyber security for private industry, government and research organizations and has held various IT roles including programmer, systems analyst, project manager and network manager before moving to academia.

Dr Patryk Szewczyk is a senior cyber security lecturer at Edith Cowan University, Australia and a senior member of the Australian Computer Society. Patryk's research specializations include cyber security, digital forensics and digital privacy. He has served as a reviewer for numerous international journals and conferences. Patryk has attained national awards for his research and community service achievements towards addressing end-user cyber security challenges.

Dr Steven Richardson is a senior lecturer in mathematics at Edith Cowan University, Australia. Steven's primary area of research interest is in mathematical modeling. He has served as a reviewer for a number of international journals.''