Near Real-time Intrusion Alert Aggregation Using Concept-based Learning

Gordon Werner
gxw9834@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

Shanchieh Jay Yang
sjyeec@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

Katie McConky
ktmeie@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

ABSTRACT

Intrusion detection systems generate a large number of streaming alerts. It can be overwhelming for analysts to quickly and effectively find related alerts stemmed from correlated attack actions. What if fast arriving alerts could be automatically processed with no prior knowledge to find related actions in near real-time? The Concept Learning for Intrusion Event Aggregation in Realtime (CLEAR) system aims to learn and update an evolving set of temporal 'concepts,' each consisting of aggregates of related alerts that exhibit similar statistical arrival patterns. With no training data, the system constructs the concepts in near real-time from statistically similar alert aggregates. Tracked concepts are then applied to incoming alerts for fast and high-fidelity aggregation. The concepts learned by CLEAR are significantly more unique and invariant when compared to those learned by alternative drift detection methods. Furthermore, it provides insights for how specific individual, or co-occuring, alerts arrive with distinct and consistent temporal patterns.

CCS CONCEPTS

• **Security and privacy → Intrusion detection systems**; Usability in security and privacy; • **Information systems →** *Data stream mining*.

KEYWORDS

cyber intrusion analysis, alert aggregation, concept learning

ACM Reference Format:
Gordon Werner, Shanchieh Jay Yang, and Katie McConky. 2021. Near Real-time Intrusion Alert Aggregation Using Concept-based Learning. In *Computing Frontiers Conference (CF '21), May 11–13, 2021, Virtual Conference, Italy.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3457388.3458663

1 INTRODUCTION

Intrusion detection systems (IDSs) are commonly used within networks to monitor traffic and detect potentially anomalous or malicious behavior [12]. While these systems are necessary for maintaining security, they can quickly generate an overwhelming number of alerts making it difficult or even impossible for an analyst to deduce insights in a reasonable amount of time [6]. Alert aggregation [11] is an emerging field of research that aims to reduce the overall number of alerts by removing redundant ones, but provides no further insight into the alerts themselves. IDS alert processing systems such as attack graphs attempt to correlate alerts to provide a deeper understanding of the threats facing a network [9]; however, these usually depend on expert knowledge and network scanning. Finding intrusion alerts that are related or correlated is a difficult task as there is no direct labeling of alerts to attacker actions. Mappings such as MITRE ATT&CK [15] are based on expert knowledge and are not always applicable given how attacker tactics change over time [25]. There is a need for fast, automated processing that gives analysts a deeper understanding of related alerts and their characteristics.

In an attempt to meet this need, research has attempted to automatically summarize alerts as attack models, e.g., [18]. Note that network traffic and cyber alert arrivals are non-stationary processes [26]; alert arrival behavior changes over time, sometimes drastically. Figure 1 illustrates cyber alerts generated over time within a network. This toy example shows that the arrival patterns are changing and potentially repeating over time. It would be difficult for an analyst to detect and track such changes, but an automated system could potentially process and group alerts in near real-time without contextual knowledge. By learning and tracking these invariant and unique temporal arrival patterns, 'concepts' could be learned to aggregate correlated but not necessarily identical alerts. Furthermore, it will be of great value if specific alerts, reflecting specific attack actions, exhibit consistent temporal characteristics. Such information on the arrival timing of critical alerts can potentially enable proactive decision making for cyber defense.

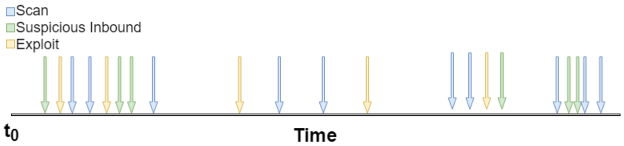

Figure 1: An example of intrusion alerts arriving over time

This work introduces the Concept Learning for Intrusion Event Aggregation in Realtime (CLEAR) system, which is driven by the intuition that alerts exhibiting similar temporal characteristics are most likely related to one another in some way. The system consumes alerts as they are produced by IDS. CLEAR adds a dimension to alert aggregation by grouping alerts based on the stationarity of their inter arrival times (IATs). Rather than simply reducing 'redundant' alerts, aggregates now represent a consistent and continual arrival behavior captured in near real-time.

CF'21, May 11-13, 2021, Virtual
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8404-9/21/05…$15.00
https://doi.org/10.1145/3457388.3458663

Aggregates with statistically similar distributions are collected into 'concepts' by the system in an unsupervised manner with no training. This novel system learns, maintains and updates historic concepts as new alerts arrive. Tracked concept's statistics are applied to incoming alerts to provide faster and more confident aggregation. CLEAR's learning was designed to generate 'tight,' invariant concepts reflecting specific and unique temporal behaviors. CLEAR's continual aggregation and concept learning set it apart from similar drift detection methods such as the Two stage shift detection based on EWMA (TSSD-EWMA) [21].

The main contributions of this work are:

- Develop a novel CLEAR system that continually learns to aggregate alerts in near real-time by statistically matching the distribution of arrival times to a learned 'concept.'
- Demonstrate the performance improvements of CLEAR over the baseline approach, TSSD-EWMA, in terms of concept uniqueness and tightness, using a real-world cyber intrusion alert dataset.
- Derive insights of intrusion activities using CLEAR to discover alerts that exhibit consistent temporal behaviors and co-occurrence in a real-world penetration testing setting.

The rest of this paper is organized as follows. Section 2 details the related work in alert aggregation, correlation of cyber events and concept drift detection. Section 3 describe the CLEAR system design. Section 4 provides the design of experiments and Section 5 discusses the results of testing. Section 6 concludes the paper.

2 RELATED WORK

2.1 Alert Aggregation

IDS systems work to detect anomalous or malicious network activity and raise alerts to network administrators to allow them to combat threats to a network [12]. Given the prevalence, size and complexity of modern networks most IDS systems produce an extremely large number of alerts [6]. One of the greatest challenges facing the security field is processing these alerts effectively in order to construct a clear and unified knowledge of a network's security status [13].

Aggregating IDS alert data is an emerging field of study that aims to reduce the overall number of alerts with minimal information loss [11]. Alert reduction is accomplished by removing successive alerts of the same type caused by ongoing activity, e.g, scanning, or alerts that are generated from the same activity by multiple scanners [10]. Traditional aggregation aims to remove 'redundant' alerts [23] but provides no deeper insight into the alerts presented to an analyst.

Researchers in [18] aggregated alerts to train Bayesian models around individual attributes in an attempt to learn attack behaviors. A naive timing threshold was used that ended an aggregate after a sufficiently long time with no new arrivals. This approach is not ideal as it is unclear what a "good" threshold is, or how it should change depending on overall network patterns. CLEAR's statistics driven approach to aggregation based on the stationarity of alert arrival statistics improves the value of aggregates by ensuring they represent unique and invariant temporal behaviors.

2.2 Temporal Correlation of Cyber Events

The idea that alerts can be related to temporally near ones is not an inherently novel idea. Time series modeling has been applied to network traffic, alert counts and cyber intrusion events to model temporal relationships in the data. ARIMA models effectively forecasting cyber event counts in [29] suggest that cyber event occurrences are temporally correlated. Further work found that malicious activity levels can change and repeat over time [30]. Time series modeling of hourly counts of individual signatures was conducted in [27] to detect abnormalities in occurrence. The findings of these related works indicate it is worthwhile to consider temporal relationships across alerts. This helped inform the intuition that alerts exhibiting similar temporal characteristics are related. To our knowledge, this is the first work to process cyber alert inter arrival times to determine such relationships.

2.3 Concept Drift Detection with EWMA

It cannot be assumed that the distributions of and relationships between features and labels do not change over time [21]. This phenomena is commonly referred to as concept drift [16] and is very common in network and human generated traffic online [26]. Drift can happen abruptly or gradually depending on the data and its context [7]. If unaccounted for drift will cause a degradation in model [2].

Concept drift adaptation is an emerging field of study that has been explored under various names in research [16]. Handling concept drift is a necessary component for processing cyber alerts as the relevance of features can change over time [19]. Most applications of drift detection schemes are to classification problems [7] and it is common to use classifier error for drift detection [1]. Feature or covariate drift measures changes in the distributions of only the input features of the system [16].

Data streams are increasingly common and the assumption of stationarity can rarely be made in such a context [28]. While many are multi-variate, univariate streams exist and can exhibit drift [5]. A univariate model uses historic measurements to forecast future ones. Should the relationship between historic and future measurements change, a feature drift has occurred in the system [5]. Some drift detection tests, such as change detection mechanisms, process the statistics of individual features to determine if a drift has occurred [4].

Raza et. al proposed a two stage feature drift detection system for univariate and multivariate series built around the exponentially weighted moving average based control chart [21]. A control chart is a graphical representation of a series used in statistical process control theory [22]. EWMA Charts measure the moving average of a series and construct a control limit (CL) based on the standard deviation of the one step ahead prediction error. It is a two stage system; in stage one the control chart raises a warning when a measurement falls outside the CL. In stage two more measurements are collected and the two-sample Kolmogorov Smirnov test (KS-Test) [14] is used to determine if drift has occurred.

Recently, research in the area has expanded to address the potential for historic concepts to re-appear. In such a scenario, a system that leverages historic concepts can reduce the impact of a drift by more quickly detecting and adapting to drift [3]. As with standard drift detection most methods focus on classification problems. Solutions maintain multiple models and make a classification with an ensemble approach [17]. Little focus has been made towards learning the statistics of the various concepts exhibited within a system.

3 CLEAR: DESIGN METHODOLOGY

3.1 Definition: Concepts and Aggregates

CLEAR processes a series of cyber alert event arrivals X, with timestamps $T = \{t_0, t_1, ...t_n\}$ and interarrival times $\Delta_i = t_i - t_{i-1}$. As new alerts arrive their IAT statistics are processed to group them based on arrival characteristics in an attempt to capture an attacker action. An aggregate is defined as a consecutive set of alerts $A_j = \{x_i, x_{i+1}, ..., x_{i+n}\}$ that exhibit similar inter-arrival times. A concept C_k is a collection of aggregates whose IAT distributions are statistically similar to one another.

There are two main components within CLEAR. The first conducts aggregation in near real-time using EWMA control charts in terms of alert IATs. The second continually learns and maintains concepts from aggregates and uses the two-sample KS-Test to optimally match the current aggregate A_{cur} to a known concept C_{best}. Completed aggregates are used to update existing concept statistics while tracked concepts are used to assist in ongoing real-time aggregation. The current aggregate remains open so long as the distribution of the alert IATs remains stationary. When a change in stationarity is detected, the current aggregate is ended and passed to the concept learning engine to find and update the concept that best reflects its temporal patterns C_{best}.

3.2 Control Charts

CLEAR captures aggregates using an EWMA control chart constructed around alert IATs. Control charts are designed to measure the stationarity of a system [22], and have been applied to general concept drift detection in the past [21]. Control charts are an ideal candidate for alert processing as they update with each new arrival and can detect potential changes in stationarity in at most one additional arrival from the change point.

The control chart maintains a moving average of incoming alert IATs as shown in (1). The moving average is used as a one step ahead forecast for the next IAT measurement $\hat{\Delta}_{i+1} = z_i$. The variance of this prediction error (2) is incorporated into a second moving average (3) used to construct control limits for the chart (4). The Control Limit can be widened or narrowed by changing the parameter L; it is normally 1.96 [22]. A measurement that falls outside of the chart's CL is refereed to as a point of drift [21].

$$z_i = \alpha\Delta_i + (1 - \alpha)z_{i-1} \tag{1}$$

$$\epsilon = x_i - z_{i-1} \tag{2}$$

$$\sigma^2_{\epsilon_i} = \alpha_\epsilon \epsilon^2_i + (1 - \alpha_\epsilon)\sigma^2_{\epsilon_{i-1}} \tag{3}$$

$$CL_{i+1} = z_i \pm L\sigma_{\epsilon_i} \tag{4}$$

3.3 Two-Sample KS-Test

The two-sample KS-Test measures the maximum distance between the cumulative distribution function (CDF) of two sample distributions and is described in (5) where sup_x is the supremum, and n_i is the number of measurements in a sample.

$$D_{KS} = sup_x|CDF_{1,n_1}(x) - CDF_{2,n_2}(x)| \tag{5}$$

The test operates under the null hypothesis that both samples are from the same distribution. To test this hypothesis, a critical distance can be estimated based on the sizes of the samples and the level of confidence required. This approximation can be found in (6), with the note that a 95% confidence was used to determine the scaling coefficient applied to the function [8]. Should the measured distance of the KS-Test exceed the critical value, the null hypothesis must be rejected; the two samples *must* be from unique distributions.

$$D_{crit} = 1.36\sqrt{\frac{n_1 + n_2}{n_1 n_2}} \tag{6}$$

3.4 CLEAR's Approach to Aggregation

A flowchart detailing how CLEAR processes a new alert arrival is described in Figure 2. A current aggregate A_{cur} is maintained that intakes new alert arrivals in near real-time. When a new alert arrives it is added to the current aggregate and the most recent IAT is computed. The IAT is compared with the control limits of the aggregate's EWMA chart. If it falls within the chart's control limit the KS-Test is used to find the concept C_{best} whose IAT distribution is most statistically similar to the current aggregate's. If C_{best} has not changed since the last alert the current aggregate's EWMA is updated with the most recent IAT. Should a new concept be determined to be the best match the current aggregate copies that concept's EWMA statistics over its own and updates them with *all* IATs in the current aggregate. By leveraging the two-sample KS-Test and historic concept control charts in such a way CLEAR provides near immediate insight into the expected characteristics of the current aggregate. Additionally, the confidence in the aggregate's control chart is increased as it is built on a larger pool of historic measurements and not just the limited ones contained in A_{cur}.

Under the TSSD-EWMA system, when a new measurement falls outside of the CL a warning is raised and the system stops processing until a predefined number of new measurements arrive. These new measurements are then compared with the measurements made prior to the warning using the two-sample KS-Test. If the distance is above the critical value, the measurements made since the warning are considered a new concept. A new control chart is generated using these measurements and processing continues with the prior concept being forgotten.

In CLEAR, a new IAT that falls outside of the CL immediately ends the current aggregate. Should the concept remain the same after the drift point the new aggregate will be matched with the same concept. Functionally, this will result in the same behavior and control chart without the need to wait for a number of new measurements. This increases speed without compromising the effectiveness of concept learning. The most recent alert is removed from the current aggregate and saved as it will be placed into the next current aggregate. If the aggregate ended due to small IAT, the two most recent alerts are removed instead. Figure 3 illustrates why this is done. Small IATs indicate that the previous alert occurred more closely to the current alert than to the third most recent. Although the previous IAT did fall within the control limit it is more intuitive to treat the two latest arrivals as the beginning of a new aggregate as highlighted in Figure 3.

Before beginning a new aggregate, the now completed aggregate is matched with a tracked concept using the KS-Test. The concept

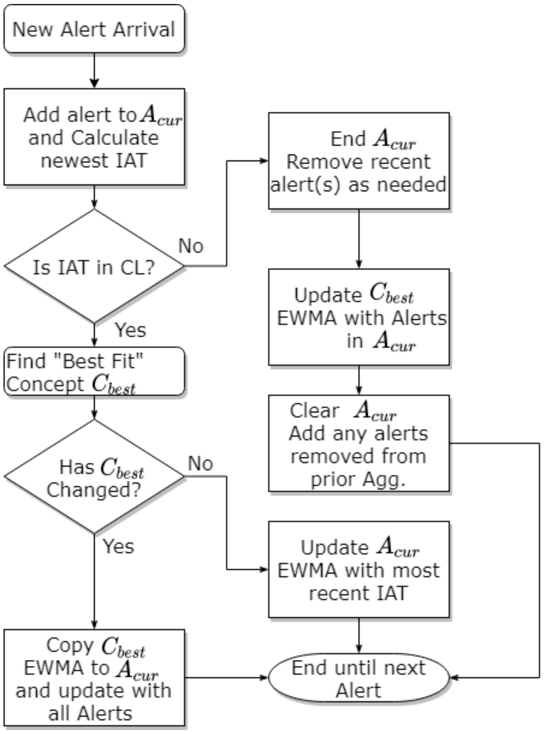
New Alert Arrival
Add alert to A_{cur} and Calculate newest IAT
Is IAT in CL?
No
Yes
End A_{cur} Remove recent alert(s) as needed
Update C_{best} EWMA with Alerts in A_{cur}
Clear A_{cur} Add any alerts removed from prior Agg.
Find "Best Fit" Concept C_{best}
Has C_{best} Changed?
No
Yes
Update A_{cur} EWMA with most recent IAT
Copy C_{best} EWMA to A_{cur} and update with all Alerts
End until next Alert

Figure 2: Flowchart of CLEAR alert processing

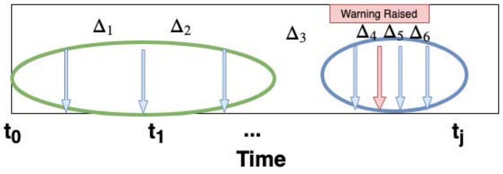

Figure 3: New aggregate caused by small IAT

is then updated with all of the aggregate's alert timestamps, IATs, and EWMA statistics. After this a new aggregate is created and populated with the saved alerts. By removing the need to wait for further measurements seen in the TSSD-EWMA approach CLEAR is able to process alerts as they arrive and detect changes in behavior rapidly and confidently. An aggregate is ended with a delay of at most one additional alert arrival, and learned concepts are tracked and updated with the end of each aggregate.

4 DESIGN OF EXPERIMENTS

Our experiments aim to assess CLEAR's ability to effectively correlate temporally near alerts through aggregation. The TSSD-EWMA drift detection approach is used as the baseline for comparison. We will assert three hypotheses:

- Concepts learned by CLEAR are consistently "tight" within a stream, with a concept's IATs exhibiting a minimal coefficient of variation;
- Concepts learned by CLEAR are unique within a stream, resulting in high KS Distances between them;
- Analysis of alert within concepts reveal certain signatures that exhibit consistent temporal patterns across streams;

4.1 Dataset and Experimental Setup

The dataset used for this research consists of Suricata alerts collected during the 2018 National Collegiate Penetration Testing Competition (CPTC) [20]. Eight teams were provided an identical network and were tasked with conducting penetration testing. A similar competition is the National Collegiate Cyber Defense Competition [31] however there is not the same level of alert generation, making it difficult to obtain as a full dataset.

The CPTC took place over the course of a single day from 8 AM until 6 PM with eight teams competing. All traffic within each team's network was passed through the Suricata [24] IDS to generate alerts. Data streams were created by filtering alerts first by team and then by source IP. By parsing the data in this way each stream can be interpreted as a single attacker's behavior as each team member is provided a single machine.

There were 127 total streams of data spread across the eight teams. Each stream was processed independently by both CLEAR and TSSD-EWMA using a separate instantiation of each system. Concepts learned in one stream were not applied to others. While it was not designed with aggregation in mind, to mimic the aggregation of CLEAR the warnings raised by TSSD-EWMA were interpreted as the end of aggregates. A window size of 15 alerts was used for testing if a drift had occurred with the KS-Test in the TSSD-EWMA system.

4.2 Experimental Hypothesis

4.2.1 Concept Coefficient of Variation. The learned concepts represent a specific temporal pattern exhibited by the alerts contained within. While the non-temporal attributes of alerts within a concept may not be identical, it is expected that the temporal characteristics are similar across all alerts. Therefore, concepts should exhibit relatively 'tight' or consistent behavior throughout, represented by a low coefficient of variation, $CV = \frac{\sigma_c}{\mu_c}$. If a concept captures consistent temporal patterns then the overall deviation of its measurements should be minimal relative to their mean. A high measurement variance implies that the concept captured too "vague" of a behavior, leading to unrelated alert groupings. The tightness of all concepts learned by CLEAR across all streams was recorded and is presented in the following section.

4.2.2 Concept Uniqueness. CLEAR aims to learn distinct and unique concepts within a single stream of intrusion alerts, and therefore there should be minimal overlap in the IAT statistics of concepts. Each concept should be significantly unique from all others within the same data stream. To measure this the Two Sample KS-Test was applied to all concept pairs within a data stream; for each concept the minimum distance to another concept in the stream was recorded. Whether or not the null hypothesis of the Two Sample KS-Test could be rejected was also recorded for all concept pairs. Ideally no two concepts should exist in a data stream that cannot reject the hypothesis as they should represent different and unique arrival statistics.

4.2.3 Alert Signature's Temporal Characteristics. The purpose of aggregation in CLEAR is to produce groups of temporally correlated alerts to gain a better understanding of attack patterns within a network. After processing alerts and learning the various concepts, the alert signatures were analyzed to find any that exhibited consistent temporal patterns. In total there were 193 unique signatures contained within all streams analyzed by CLEAR. To determine a potentially consistent signature, a probabilistic approach was taken in analyzing the distribution of alerts within the concepts learned for a stream. While no explicit ranking metric was used, a number of statistics were collected for each signature such as:

- The number of concepts containing a specific signature relative to the number of streams it appeared in;
- The total number of alerts with that signature;
- The mean and standard deviation of all IATs within concepts containing the signature;

The rationale for using these statistics is that a signature with consistent temporal characteristics should be found in a limited number of concepts per stream. Looking at the overall concept statistics allows for discovery of consistent signatures should CLEAR create temporally near or overlapping concepts in a given stream.

5 RESULTS AND DISCUSSIONS

Across all streams CLEAR generated 558 concepts while the TSSD-EWMA drift scheme only generated 282. There were 69 streams in which the TSSD-EWMA system did not detect a new concept, instead classifying all alerts in the stream into a single concept. This did not occur for any of the streams processed by CLEAR. When analyzing concept uniqueness for TSSD-EWMA, the streams with only 1 concept were not included as there was no other concept to compare with using the two sample KS Test.

5.1 Concept Tightness with CLEAR

Figure 4 plots the coefficient of variation for all concepts learned by both approaches. CLEAR's concepts exhibit a much lower ratio on average indicating significantly "tighter" IAT statistics. CLEAR concepts averaged a ratio of .864 while TSSD-EWMA resulted in an average ratio of 6.14. The concepts, and by extension the aggregates, learned by our system show much more uniformity in their arrival patterns with each other compared to the other aggregation method. As CLEAR learns from historic concepts it can quickly match recent observations with existing concepts or detect novel concepts more quickly than TSSD-EWMA can. Low coefficient of variation highlights CLEAR's effectiveness at quickly and accurately detecting the end of aggregates; if there were significant outliers the CV would be higher.

CLEAR's ability to generate such 'tight' concepts shows that it more effectively captures uniform alert behavior patterns. This provides confidence that any two aggregates from the same concept represent extremely similar temporal behaviors. Aggregating alerts based on their arrival behavior leads to more clearly defined concepts, and provides a strong temporal correlation. Further relationships found between alerts within a concept give a better understanding of the potential actions an attacker can take given the temporal characteristics of intrusion alerts.

The TSSD-EWMA constructs a new concept's control chart around the first m measurements regardless of their similarity to one another allowing for significantly wider concepts. CLEAR instead is able to apply historic control chart statistics to a new aggregate as early as the second alert, limiting the likelihood of highly variant IATs within a concept. In such a potentially dynamic environment as cyber intrusion detection this weakness of TSSD-EWMA may be exacerbated as the data is generated from human driven actions which quickly change and adapt [26]. When analyzing individual alert signatures their temporal behaviors are better detailed by our system. If a signature only occurs within a fixed arrival pattern it will be found in specific CLEAR concepts. It is more difficult to isolate a specific signature's temporal characteristics from the TSSD-EWMA concepts as they are significantly broader.

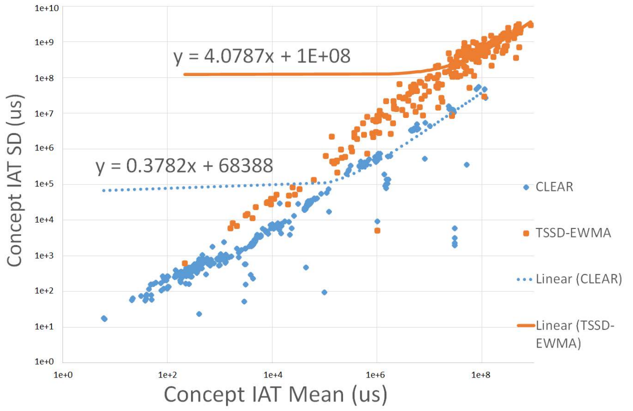

Figure 4: Concept means v. SD for both methods

5.2 Separation Across Concepts

For each stream, a scatter plot was generated matching the number of concepts in the stream to the average of the minimum KS-Distances measured for each concept within the stream and can be found in Figure 5. The figure summarizes the uniqueness of concept distributions within each individual stream. Two non-unique concepts would have very similar statistical distributions, which would lead to a smaller KS-Distance between them.

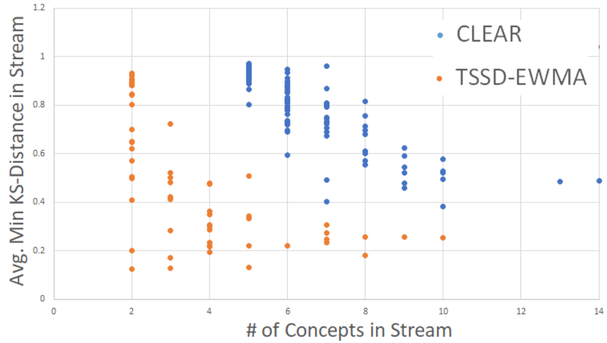

Figure 5: Total concepts v. avg. min. KS-Dist. in streams

Streams processed by CLEAR resulted in more statistically unique concepts overall than those handled by TSSD-EWMA in spite of the fact that CLEAR on average produces more concepts per stream. For streams with an equivalent number of concepts across the two methods in no cases does TSSD-EWMA produce concepts with higher average minimum KS-Distance than CLEAR. Our system is able to learn more temporal concepts that are also more distinct from one another. Knowing concepts are unique gives stronger context to additional relationships found between alerts. Correlations found across concepts indicate a behavior independent of temporal patterns, while those found within a concept can be confidently said to have a high dependence with the temporal arrival patterns of alerts. Concept uniqueness increases the concentration of behaviors to single concepts within a stream.

Figure 6 expands on the uniqueness of concepts by comparing the overall distribution of minimal KS-Distances between concepts. Each data point used to generate the histograms represent the smallest KS-Distance between a single concept and all other concepts in the same stream. A majority of CLEAR concepts have a minimum KS-Distance of .8 or greater from the nearest concept, indicating nearly no overlap between most concepts. TSSD-EWMA concepts show much more overlap in behaviors captured, with a majority measuring less than .4 from their nearest concept. These results show the ability of CLEAR to learn distinct and unique temporal behaviors within a data stream.

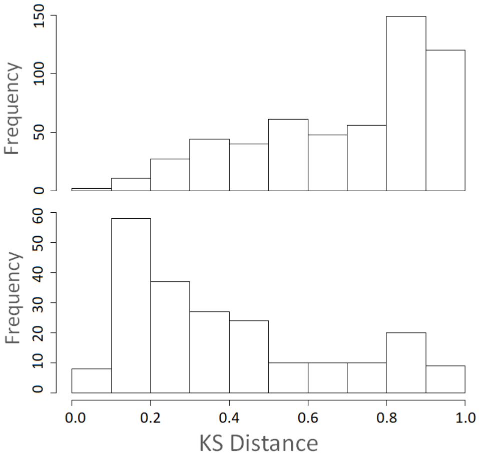

Figure 6: Min KS-Dist. between concepts for CLEAR (top) and TSSD-EWMA (bot)

Even when looking at individual data streams CLEAR generates more unique concepts when compared to the TSSD-EWMA method. For a majority of streams the TSSD-EWMA could not detect a second concept, while CLEAR always produced at least two concepts. Considering that alerts are derived from network traffic and that network traffic is inherently bursty and non-stationary [26], it is unrealistic to expect a stream should have only one concept. For streams where the standard drift approach detected more than one concepts, 28.6% of concepts were close enough to another that the null hypothesis could not be rejected; meaning it could not be concluded that the two concepts were generated by different distributions. CLEAR concepts saw only 14.3% of concepts unable to reject the null hypothesis, half the rate with more than double the concepts when compared to the TSSD-EWMA approach.

Unlike in the TSSD-EWMA's case however, the concepts that make up CLEAR's 14.3% statistic are either two small concepts or one large and one small concept. In the case of very small sample sizes, it is nearly impossible to reject the null hypothesis for the KS-test as the critical distance is very large. For CLEAR concepts that failed to be further from another concept than the critical distance, marking them "similar;" the average critical value for the KS-test was .63. In the case of TSSD-EWMA, the average critical value was .26; a substantially lower value. This means that it was much easier on average for the concepts learned with the standard drift method to be statistically different from each other, and yet similar concepts are still generated by TSSD-EWMA at twice the rate of CLEAR.

To better illustrate the uniqueness of concepts learned by the two systems box plots were generated over the IATs contained in the concepts of a single stream and are shown in Figure 7. As shown TSSD-EWMA produces fairly homogeneous concepts over time with large amounts of overlap while CLEAR produces distinct concepts with much lower deviation in the arrival times contained.

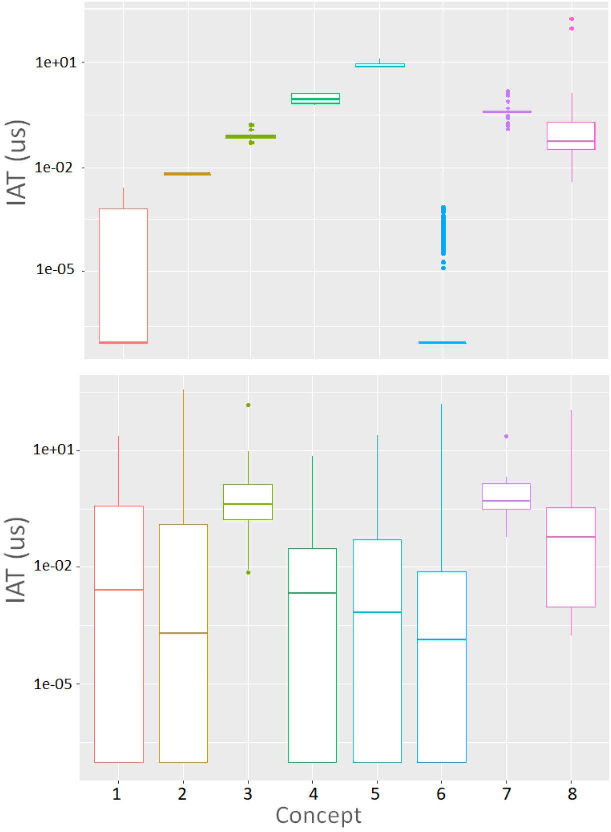

Figure 7: CLEAR (top) & TSSD-EWMA (bottom) concept IATs

TSSD-EWMA also lacks the ability to learn and update historic concepts, leading to seemingly identical concepts being generated at different times in the stream. CLEAR's concepts are instead learned over time to ensure that they are representative of temporal behaviors captured within alert aggregates. When analyzing signatures, it is important to know that individual concepts within a stream are unique. If a signature only occurs in a single stream it exhibits a very consistent temporal behavior. Such findings are strengthened when the same signature is found in statistically similar concepts in multiple streams. Given the broad and overlapping TSSD-EWMA concepts, it is difficult to draw any concrete conclusions from signature analysis.

5.3 Single Stream Results

To better visualize the performance of the aggregation methods, a single stream of data will be used for the remaining discussion to better highlight the behavior of the various approaches. Employing the TSSD-EWMA drift detection scheme detailed in [21] results in a smaller number of Aggregates split over 2 concepts as shown in Figure 8. The graph shows the box plots of the IATs in each of the aggregates generated by the system. In the figure there are multiple very similar aggregates covering the same ranges of IATs for most of the stream, with little deviation over time. Looking at these results, it seems as though the IATs for this specific dataset are relatively uniform over time.

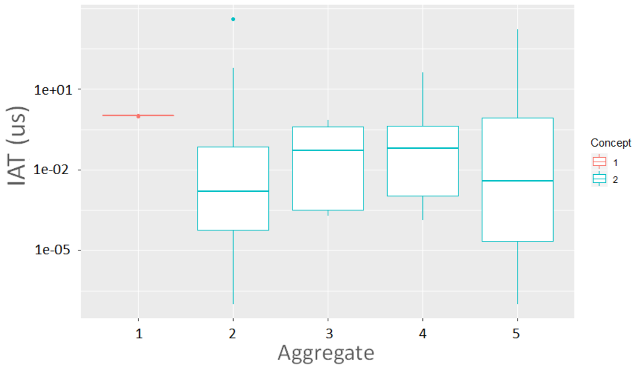

Figure 8: Aggregates generated by TSSD-EWMA

Arrival patters are not in fact uniform however as shown by a short period of aggregation obtained when applying the CLEAR system to the stream in Figure 9. When analyzing new data using historic learned concepts, the CLEAR system is able to quickly adapt to changes in arrival patterns, separating arrivals into three unique concepts.

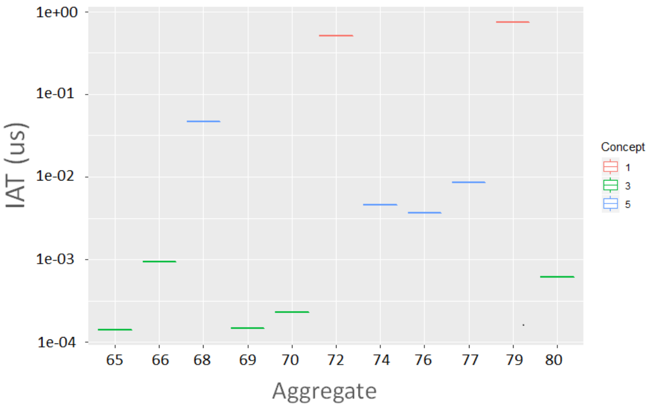

Figure 9: Aggregates generated by CLEAR

The aggregates in the figure averaged around 3 alerts each, further highlighting CLEAR's reactivity. By applying known concept statistics to aggregation CLEAR is more sensitive to sudden changes in arrival behavior. The CLEAR system is able to produce "tighter" and more unique concepts that may otherwise be overlooked by traditional streaming drift detection methods.

In such a fast changing environment this reactivity is key, and is a distinct weakness of the TSSD-EWMA approach by comparison. Requiring a fixed number of new measurements after a drift point means that it is possible that two or more unique concepts are merged together, creating broad non-descriptive aggregates like those seen in Figure 8.

5.4 Temporally Correlated Alert Signatures

CLEAR's unique and 'tight' concepts ensure alerts are aggregated based on their temporal arrival patterns accurately. Further relationships and consistent characteristics of the alerts contained can be found by analyzing concepts. This section investigates the alert signatures contained in concepts learned by CLEAR and discusses some findings.

In the data streams processed by CLEAR there were 161 unique alert signatures. Individual signatures appeared in an average of 7.68 streams and 14 concepts resulting in an average ratio of 1.83 concepts per stream. Signatures were analyzed based on their overall concentration of alerts across streams. A selection of signatures and their corresponding statistics are presented in Table 1. The means and standard deviation of IATs for the concepts containing the signatures as well as the ratio of concepts containing the signature to streams containing the signature were used to analyze how consistent a signature's temporal characteristics were in the data. Since CLEAR learns both "tight" and unique concepts, should a signature behave in a relatively consistent temporal manner it should only appear in a small number of concepts in the streams it is found in. While the ratio of concepts per streams was the primary factor in determining concentration, the concept statistics and total number of alerts was also used to accommodate the potential for a signature's temporal range to be wider than that of a single concept learned by CLEAR.

Table 1: Signature rankings based on various statistics

Sig. Abbr.(ID)	Conc./Stream	Alerts	μ IAT	σ IAT
SQLAISA(170)	1	13	0.3 ms	N/A
PHPENV(91)	1	34	6.9 ms	0.38 ms
ETWSH(33)	1.5	122	4.38 ms	2.33 ms
ETWS(34)	1.5	122	4.38 ms	2.33 ms
MONGOVR(14)	1.25	45	3.94 ms	14.9 ms
MONGODER(15)	1.25	46	3.94 ms	14.9 ms
PHPINJ(80)	6	111	27.5 ms	39.6 ms
PSSQLSCAN(6)	2.3	966	15.72 s	40.3 s
OSSCAN(9)	1.6	441	0.33 s	0.44 s
CURL(22)	1	116	0.5 s	4.9 s
RFI ATT.(116)	1	1	56.4 ms	N/A

5.4.1 *Alert Signatures with Consistent Arrival Patterns.*

Figure 10 shows the box plots of the concepts containing the signature "ET WEB_SERVER PHP ENV SuperGlobal in URI"(PHPENV(91)) across the data streams that contain it. This signature appeared in a single stream of each of two teams and is most likely used as part of a vulnerability injection. Both concepts exhibit nearly identical IAT statistics and contain the same 12 signatures. This is strong evidence of the same action being taken by both teams, most likely an identical script or exploit.

Another signature that exhibits consistent temporal characteristics is "ET WEB_SERVER Possible MySQL SQLi Attempt Information Schema Access" (SQLAISA(170)). This signature is usually part of a SQL injection attempt, and was used by one team multiple

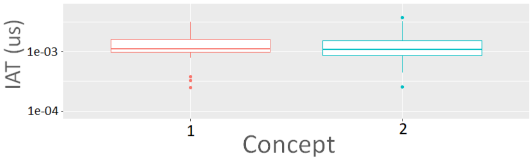

Figure 10: Concepts containing "ET WEB_SERVER PHP ENV SuperGlobal in URI" (PHPENV(91))

times over the course of the competition. Although the injection attempts were made hours apart from one another, CLEAR classified all alerts with the signature into the same concept. This result highlights the effectiveness of CLEAR's concept tracking and matching with aggregates.

5.4.2 Co-occurring Alert Signatures. It was not uncommon to find groups of signatures where all alerts were aggregated into the same concepts. Usually these signatures were very closely related (GPL WEB_SERVER service.cnf access, GPL WEB_SERVER services.cnf access and GPL WEB_SERVER writeto.cnf access) and only occurred once or twice across all datasets.

A pair of alerts warning of Mongo database version and database enumeration requests were seen in streams of members from seven teams. The signatures "ETPRO ATTACK_RESPONSE MongoDB Version Request" (MONGOVR(14)) & "ETPRO ATTACK_RESPONSE MongoDB Database Enumeration Request" (MONGODER(15)) were found together in the same 20 concepts within 16 streams. The signatures detail two unique requests to a Mongo database and occur in near equal numbers (45 vs. 46 alerts respectively) and with similar and mostly consistent timing patterns. Nearly all concepts containing the signatures exhibited microsecond IAT statistics; though one concept was in the sub-second range.

Similar behavior was observed between the signatures "ET WEB_ SERVER Possible CVE-2014-6271 Attempt" (ETWS(34)) and "ET WEB _SERVER Possible CVE-2014-6271 Attempt in Headers" (ETWSH (33)). Both signatures occurred in pairs of alerts throughout the two data streams that they appeared in across two unique teams. This behavior is unique, as there were other streams that contained only the signature ETWSH(33) without the other. All concept means fall between 1 and 6 milliseconds which is very consistent across streams and teams.

5.4.3 Alerts with Multiple Distinct Arrival Patterns. Scanning is an extremely common, near constant passive action taken by many entities. Given its nature it is not surprising to find that most scanning signatures, such as "ET SCAN Suspicious inbound to PostgreSQL port 5432" (PSSQLSCAN(6)), are found in a number of concepts with varying timing characteristics. Interestingly however another type of scanning, "ET SCAN NMAP OS Detection Probe" (OSSCAN(9)) was used by 6 teams in one of two distinct "modes" with unique timing characteristics. Figure 11 shows the concepts containing the signature; no single stream contained concepts from both modes.

The specific operating mode likely depends on surrounding attacker actions. This could be due to difference in target OS, the parameters used by the attacker or a complementary action being executed in parallel.

6 CONCLUDING REMARKS

The CLEAR system is able to intelligently aggregate non-stationary streaming alerts through efficient concept learning and matching of arrivals. It learns concepts that exhibit significantly less variation and more unique distributions when compared to the TSSD-EWMA process. Intrusion alert IATs grouped into concepts exhibit a standard deviation that is on average less than the concept's mean. TSSD-EWMA concepts on average have a tightness ratio that is 7.1 times greater than CLEAR. Furthermore, the concepts learned by CLEAR are significantly more unique than those learned by TSSD-EWMA. On average, CLEAR concepts are twice as distant from others within the same stream than those learned by TSSD-EWMA when measured by the two-sample KS-Test. CLEAR is not limited in its change point detection and can quickly detect changes in stationarity correctly without waiting for additional arrivals. By quickly matching alert aggregates to the optimally matched concept in near real-time, CLEAR reveals attack behaviors exhibited by temporally related intrusion alerts.

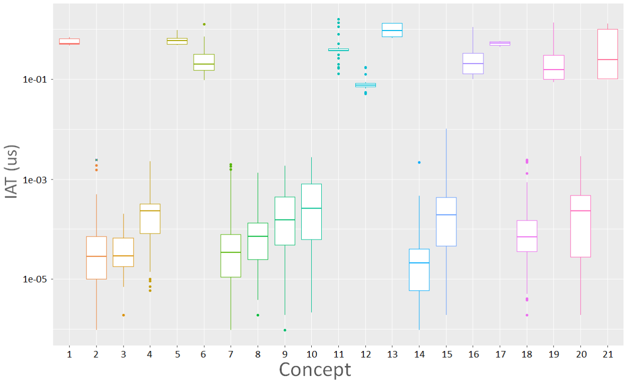

Figure 11: Concepts containing signature "ET SCAN NMAP OS Detection Probe" (OSSCAN(9))

As demonstrated in the experiments conducted using CPTC 2018 intrusion alerts, CLEAR is able to identify specific attack action signatures with similar temporal characteristics. A number of signatures were found in concepts with consistent arrival patterns across multiple streams and teams. Groups of signatures such as ETWS(34) and ETWSH (33) always appeared in concepts together, also across multiple team's data streams. Some signatures exhibited multiple 'modes' of operation, i.e., different alert arriving speeds. These results highlight the value of CLEAR, which produces the temporal context of intrusion activities. Extracting related alert signatures from temporal proximity and arrival patterns helps to provide security analysts broader insights on attack behavior: which attack actions are likely to co-occur together and at what inter-arrival time. Such insights can be helpful to effectively determine when, where and how to interrupt a cyberattack campaign.

ACKNOWLEDGMENTS

This work was partially funded by NSF awards #1526383 and #1742789, as well as the RIT GCI Seed Fund.

REFERENCES

[1] M. Baena-Garcıa, del Campo-Á., R. Fidalgo, A. Bifet, R. Gavalda, and R. Morales-Bueno. 2006. Early drift detection method. In *Fourth international workshop on knowledge discovery from data streams*, Vol. 6. 77–86.

[2] J. Barddal, H. Gomes, and F. Enembreck. 2015. Analyzing the impact of feature drifts in streaming learning. In *International Conference on Neural Information Processing*. Springer, 21–28.

[3] J. Barddal, H. Gomes, F. Enembreck, and B. Pfahringer. 2017. A survey on feature drift adaptation: Definition, benchmark, challenges and future directions. *Journal of Systems and Software* 127 (2017), 278–294.

[4] A. Bifet and R. Gavalda. 2007. Learning from time-changing data with adaptive windowing. In *Proceedings of the 2007 SIAM international conference on data mining*. SIAM, 443–448.

[5] R. Cavalcante, L. Minku, and A. Oliveira. 2016. Fedd: Feature extraction for explicit concept drift detection in time series. In *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 740–747.

[6] H. Debar and A. Wespi. 2001. Aggregation and correlation of intrusion-detection alerts. In *International Workshop on Recent Advances in Intrusion Detection*. Springer, 85–103.

[7] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar. 2015. Learning in nonstationary environments: A survey. *IEEE Computational Intelligence Magazine* 10, 4 (2015), 12–25.

[8] P. Hartigan. 2019. Critical Values for the Two-Sample Kolmogorov-Smirnov Test. http://sparky.rice.edu/astr360/kstest.pdf.

[9] J. Holsopple, S. Yang, and M. Sudit. 2015. Mission Impact Assessment for Cyber Warfare. In *Intelligent Methods for Cyber Warfare*. Springer, 239–266.

[10] M. Husák and M. Čermák. 2017. A graph-based representation of relations in network security alert sharing platforms. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 891–892.

[11] M. Husák, M. Čermák, M. Laštovička, and J. Vykopal. 2017. Exchanging security events: Which and how many alerts can we aggregate?. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 604–607.

[12] H. Liao, Y. Lin, C.and Lin, and K. Tung. 2013. Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications* 36, 1 (2013), 16–24.

[13] F. Maggi, M. Matteucci, and S. Zanero. 2009. Reducing false positives in anomaly detectors through fuzzy alert aggregation. *Information Fusion* 10, 4 (2009), 300–311.

[14] F. Massey Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* 46, 253 (1951), 68–78.

[15] MITRE. 2016. The MITRE ATT&CK Framework. https://attack.mitre.org/.

[16] J. Moreno-Torres, T. Raeder, R. Alaiz-RodríGuez, N. Chawla, and F. Herrera. 2012. A unifying view on dataset shift in classification. *Pattern Recognition* 45, 1 (2012), 521–530.

[17] H. Nguyen, Y. Woon, W. Ng, and L. Wan. 2012. Heterogeneous ensemble for feature drifts in data streams. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 1–12.

[18] A. Okutan and S. Yang. 2019. ASSERT: attack synthesis and separation with entropy redistribution towards predictive cyber defense. *Cybersecurity* 2, 1 (2019), 15.

[19] A. Okutan, S. Yang, K. McConky, and G. Werner. 2019. CAPTURE: Cyberattack Forecasting using Non-Stationary Features with Time Lags. In *Proceedings of the 7th Annual Conference on Communications and Network Security (CNS '19)*. IEEE.

[20] J. Pelletier. 2018. Collegiate Penetration Testing Competition. https://nationalcptc.org/.

[21] H. Raza, G. Prasad, and Y. Li. 2015. EWMA model based shift-detection methods for detecting covariate shifts in non-stationary environments. *Pattern Recognition* 48, 3 (2015), 659–669.

[22] S. Roberts. 2000. Control chart tests based on geometric moving averages. *Technometrics* 42, 1 (2000), 97–101.

[23] J. Sun, L. Gu, et al. 2020. An Efficient Alert Aggregation Method Based on Conditional Rough Entropy and Knowledge Granularity. *Entropy* 22, 3 (2020), 324.

[24] Suricata. 2020. Suricata Open Source IDS. https://suricata-ids.org/.

[25] Symantec. 2017. Symantec 2017 Internet Security Threat Report. https://docs.broadcom.com/doc/istr-22-2017-en.

[26] P. Vaz de Melo, C. Faloutsos, R. Assunção, and A. Loureiro. 2013. The self-feeding process: a unifying model for communication dynamics in the web. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1319–1330.

[27] J. Viinikka, H. Debar, L. Me, A. Lehikoinen, and M. Tarvainen. 2009. Processing intrusion detection alert aggregates with time series modeling. *Information Fusion* 10, 4 (2009), 312–324.

[28] H. Wang and Z. Abraham. 2015. Concept drift detection for streaming data. In *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–9.

[29] G. Werner, S. Yang, and K. McConky. 2017. Time Series Forecasting of Cyber Attack Intensity. In *Proceedings of the 12th Annual Conference on Cyber and Information Security Research (CISRC '17)*. ACM, New York, NY, USA, Article 18, 3 pages. https://doi.org/10.1145/3064814.3064831

[30] G. Werner, S. Yang, and K. McConky. 2018. Leveraging Intra-Day Temporal Variations to Predict Daily Cyberattack Activity. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*. IEEE, 58–63.

[31] D. Williams. 2018. Collegiate Cyber Defense Competition. https://www.nationalccdc.org/.