This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Multi-Dimensional Anomalous Entity Detection via Poisson Tensor Factorization

Maksim E. Eren*†, Juston S. Moore†, Boian S. Alexandrov‡
*Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County
†Advanced Research in Cyber Systems, Los Alamos National Laboratory
‡Theoretical Division, Los Alamos National Laboratory

***Abstract*—As the attack surfaces of large enterprise networks grow, anomaly detection systems based on statistical user behavior analysis play a crucial role in identifying malicious activities. Previous work has shown that link prediction algorithms based on non-negative matrix factorization learn highly accurate predictive models of user actions. However, most statistical link prediction models have been constructed on bipartite graphs, and fail to capture the nuanced, multi-faceted details of a user's activity profile. This paper establishes a new benchmark for red team event detection on the Los Alamos National Laboratory Unified Host and Network Dataset by applying a tensor factorization model that exploits the multi-dimensional and sparse structure of user authentication logs. We show that learning patterns of normal activity across multiple dimensions in one unified statistical framework yields improved detection of penetration testing events. We further show operational value by developing fusion methods that can identify anomalous users, source devices, and destination devices in the network.**

***Index Terms*—anomaly detection, Poisson tensor factorization, cyber security, canonical polyadic decomposition**

I. Introduction

Detection of compromised accounts and insider threats continues to be a significant challenge for cyber defenders. In 2016, when Turcotte et al. introduced the Poisson matrix factorization model for cyber anomaly detection, 63% of confirmed data breaches involved stolen user credentials [1]. This figure has climbed to 80% in 2020 [2], and the average number of yearly security breaches has increased by 67% within the past 5 years [3]. At the same time, the cost of malicious insider attacks increased by 15% in 2019, and continues to be one of the threat types that takes the longest to resolve [3]. When hunting for intruders or malicious insiders on their networks, incident response teams primarily rely on rule-based indicators such as hand-crafted signatures or open-source threat intelligence feeds. Although rule-based indicators perform well when detecting known attacks, they require immense manual work to tune for each enterprise network, and often fail to detect patient and persistent attackers. Currently, alerts are generated only for 9% of attacks [4], and the average cost of a security breach is $3.86 million [5]; therefore, there is an urgent need to improve statistical anomaly detection methods and their associated operational workflows in order to drive increased adoption.

This manuscript has been approved for unlimited release and has been assigned LA-UR-20-26304.

U.S. Government work not protected by U.S. copyright

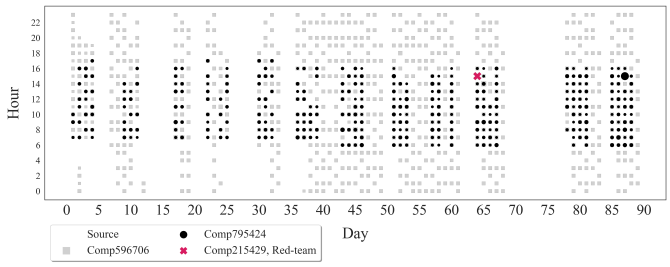

Fig. 1. Hourly authentication events from multiple source computers over 90 days for one compromised user in the LANL Unified Host and Network Dataset. The user's activity reveals time- and device-based predictable patterns that deviate from the single anomalous log-on.

Generating actionable alerts with anomaly detection systems requires identifying unusual events that correspond to malicious activity. New events happen continually on a network, and no labelled datasets exist with enough detail to build reliable detection systems using supervised learning alone. Because the number of daily events on a corporate network can easily reach into the millions or billions, deployable anomaly detectors must achieve extremely low false alarm rates. Eliminating rare, but benign, events from these alerts is challenging, since human activities are difficult to predict; for example, users authenticate to new network resources on a continual basis. This work builds upon state-of-the-art algorithms for user behavior analysis to build more nuanced methods of normal user behavior over time. We show that our model can accurately detect actions that deviate from the norm by demonstrating its performance detecting stolen user credentials during a penetration testing event on the Los Alamos National Laboratory (LANL) network.

Previous work has shown that recommendation system models, based on matrix factorization, can identify "peer groups" of users and devices, which allow data-driven predictions of future user actions [1], [6], [7]. Users tend to exhibit a seasonal behavior in the network, where patterns of activities are correlated in time. For example, a simple predictable user behavior is an employee initiating a logon from a desktop computer every weekday, except Friday, at approximately 7:00am. Figure 1 shows activities from such a real, anonymized LANL user (User087542) over 90 days.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on January 04,2021 at 21:34:21 UTC from IEEE Xplore. Restrictions apply.

In this work, we extend peer-based models to include multiple dimensions of an activity profile, including users, source devices, destination devices, authentication status, and time. We apply tensor factorization to extract high-dimensional latent activity profiles that capture correlations across observed dimension. Our analysis subsequently leverages these learned profiles to improve the sensitivity and specificity of peer-based anomaly detection.

Our contributions include:

- Generalizing existing statistical models to *jointly* learn multi-dimensional activity profiles.
- Demonstrating that jointly-learned activity profiles improve the detection of *anomalous events*.
- Presenting state-of-the-art results for the task of *anomalous entity* detection, for example, identifying a penetration testing team's source device within the top 3 most anomalous devices during the month-long test period.
- Establishing a new benchmark for red team activity detection on the LANL Unified Host and Network Dataset.

II. Background

Our work draws on prior advances in the areas of statistical anomaly detection and tensor factorization. Here we present a brief summary of related work in both fields.

A. Anomaly Detection

A variety of classical factorization techniques, such as Principle Component Analysis (PCA) [8] and Non-negative Matrix Factorization (NMF) [9], have been applied to detect anomalies using reconstruction error as a metric. These techniques extract "normal" patterns hidden in the data to perform dimensionality reduction. However, these models don't offer a direct statistical interpretation, and thus don't directly produce a p-value for anomalies. Statistical models have stronger mathematical guarantees, and provide more direct methods for fusing analytic outputs. Prior studies have explored statistical Poisson matrix factorization methods, based on recommendation systems, to perform anomaly detection [1], [6], [10].

Sanna Passino et al. used two bipartite graphs with the dimensions *User - Destination* and *User - Source*, applied Poisson Matrix Factorization (PMF) to detect anomalies, and extended their statistical model to incorporate covariates about users and computers [6]. Similarly, Volkovs et al. showed that a deep neural network can learn to augment user preference data and alleviate the problems of cold start [11]. Price-Williams et al. developed a model that detected anomalous users via their historic authentication times [10]. Turcotte et al. demonstrated that Fisher p-value fusion can combine independent p-value scores of user's logon and process start events for anomaly detection [1].

B. Tensor Factorization

Tensors are higher order extensions of matrices [12]. Cyber event logs can naturally be encoded as tensors. For example, users authenticating between two devices can be represented by a 3rd-order tensor with dimensions *User*, *Source*, and *Destination* where an index $\mathcal{X}_{u,s,d}$ in this tensor represents the number of authentications that user u performs going from source device s to destination device d. In this work, we use *binary tensors*, where the element $\mathcal{X}_{u,s,d}$ is set to 1 if user u authenticates from source device s to destination device d at least once, and is set to 0 otherwise.

Tensor factorization is a cutting-edge method for factor analysis. It decomposes high-dimensional data into factor matrices where the factor matrices carry the latent features in each tensor dimension. Specifically, we use a form of non-negative tensor factorization called Poisson tensor factorization. We choose non-negative factorization because our dataset is inherently non-negative (i.e., our dataset involves counts of event types, and a negative event count would be impossible). Importantly, non-negativity requires the extracted latent features to be additive components of the original data, which improves interpretability [13].

Bruns-Smith et al. originally applied Poisson tensor factorization in the cyber security domain [14], but they manually analyzed their resulting factors to find malicious activity. While the authors successfully identified indicators of malicious incidents, their manual analysis does not scale to large data. Our work leverages Canonical Polyadic Decomposition (CPD) to extend existing Poisson matrix factorization models and thus automate ranking and scoring.

CPD [15] is an important tool for unsupervised learning, feature extraction, and dimensionality reduction. By definition, if a tensor can be written as a single outer product of vectors, it has rank 1. Any arbitrary tensor can be decomposed as a weighted sum of rank-1 tensors, which is called Polyadic Decomposition. If the number R of rank-1 tensors is minimal, then the decomposition is a CPD. Importantly, in the non-negative case, a best rank-R approximation always exists [16], and it is almost always unique [17]. Usually, each factor is normalized to sum to 1 and weight is absorbed by γ_r to achieve a unique solution. For example, an order 3 tensor $\mathcal{X}$ with dimensions u, s, d and shape N_u, N_s, N_d can be approximated by a sum of R rank-1 tensors, each called a *component*. Each component is encoded as the outer product of 3 factor vectors, $\theta_r^{(u)}, \theta_r^{(s)}, \theta_r^{(d)}$, with lengths N_1, N_2, and N_3, respectively. Equation (1) shows the CPD tensor approximation, where $\circ$ represents vector outer product.

$$\mathcal{X} \approx \hat{\mathcal{X}} \equiv \sum_{r=1}^{R} \gamma_r \cdot \theta_r^{(u)} \circ \theta_r^{(s)} \circ \theta_r^{(d)} \tag{1}$$

The problem of selecting the optimal rank R for a specific application is essential in finding low-dimensional latent tensor representations. If we perfectly reconstruct the input tensor, our tensor factorization carries little information about peer groups or other shared structure; if our rank is too low, we lose vital information. Truong et al. discussed this problem in [18], and introduced the *Non-Negative RESCAL* method. Minimal multirank in RESCAL is chosen to be the rank with low relative error and high silhouette score. Similarly, our model attempts to find the best rank to deviate from arbitrary rank

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on January 04,2021 at 21:34:21 UTC from IEEE Xplore. Restrictions apply.

selection. Differently than RESCAL, we use log-likelihood on held-out validation data to find the optimal tensor rank for link prediction and anomaly detection.

III. Multidimensional Anomaly Detection

Simultaneous analysis and extraction of latent features by tensor factorization enhances the detection of unusual activities by making the system sensitive to different correlations between the dimensions. For example, we can train our models to learn user patterns and daily/hourly periodicity jointly.

Our model is based on Poisson CPD. For a D-dimensional tensor with shape $N_1, \ldots, N_D$, we model each element as an independent draw from a Poisson distribution, where the rate $\lambda_{i_1,\ldots,i_D}$ ($1 \leq i_1 \leq N_1$, $\ldots$, $1 \leq i_D \leq N_D$) is determined by a CPD of rank R:

$$\mathcal{X}_{i_1,\ldots,i_D} \sim \text{Poisson}\left(\lambda_{i_1,\ldots,i_D}\right) \tag{2}$$

$$\lambda_{i_1,\ldots,i_D} = \sum_{r=1}^{R} \gamma_r \prod_{d=1}^{D} \theta_{r,i_d}^{(d)} \tag{3}$$

where $\theta_r^{(d)}$ is r-th component in the d-th dimension (or factor).

During training, we learn latent factors to maximize the *joint log-likelihood* of all observed counts:

$$\begin{aligned}\log P(\mathcal{X}) = \sum_{i_1=1}^{N_1} \cdots \sum_{i_D=1}^{N_D} ((\mathcal{X}_{i_1,\ldots,i_D} \cdot \log \lambda_{i_1,\ldots,i_D}) \\ - \log \Gamma(x_{i_1,\ldots,i_D} + 1)) - \Lambda\end{aligned} \tag{4}$$

where

$$\Lambda = \sum_{r=1}^{R} \gamma_r \left[\left(\sum_{i_1=1}^{D_1} \theta_{r,i_1}^{(d)} \right) \cdot \ldots \cdot \left(\sum_{i_D=1}^{N_D} \theta_{r,i_D}^{(d)} \right) \right] \tag{5}$$

Note that this log-likelihood function is efficient to compute on sparse data because the first two terms are 0 whenever the count $\mathcal{X}_{i_1,\ldots,i_D}$ is 0. Therefore, the sum can be implemented efficiently by summing only over non-zero (i.e. observed) counts.

The most efficient solver for this likelihood function is CANDECOMP-PARAFAC Alternating Poisson Regression (CP-APR) [19], which is equivalent to minimizing the Kullback–Leibler divergence with a non-negativity constraint, via a modified multiplicative update (MU) algorithm[1].

A. Rank Selection

CPD is a non-convex problem where it is assumed that the tensor rank is known [19]. We use log-likelihood (Equation 4) evaluation on held-out time periods (i.e. validation data) to find the rank that best predicts future user actions. We split training data into two separate time periods: *validation-train* and *validation-test*. We fit the tensor factorization using *validation-train* on all ranks from 1 to 100 (with a step size of 5 between 10-100), and evaluate log-likelihood on *validation-test*. The rank with highest log-likelihood is chosen as R during our subsequent training and testing procedures.

[1]CP-APR is available at https://www.tensortoolbox.org/

B. Poisson Rate Smoothing

Because tensors representing cyber security logs are extremely sparse, we encounter numerical underflow when estimating the tensor factorization. In order to alleviate this problem, we multiplicatively inflate our counts such that the mean value in the tensor is approximately 1. Additionally, because of the sparse structure of these tensors, many of the estimated factors are sparse (i.e. have a large quantity of zero values). Zero values in the factors result in estimated Poisson rates of 0 during the testing phase; thus we need to regularize our estimation procedure. We do this by estimating a rank-1 factorization and a rank-R factorization of the training tensor, where the optimal R is computed by maximizing validation log-likelihood. Since the sum of counts across any axis of our tensor is non-zero, we are guaranteed to have non-zero factors in our rank-1 factorization. We use this fact to regularize our estimation of the Poisson rate $\lambda_{i_1,\ldots i_D}$:

$$\lambda_{i_1,\ldots,i_D} = 0.1 \cdot \lambda_{i_1,\ldots,i_D}^{1} + 0.9 \cdot \lambda_{i_1,\ldots,i_D}^{R} \tag{6}$$

C. Anomalous Event Scoring

We perform anomaly detection by computing the *p-value* of each observed count during our test period. The p-value is the probability of observing a count *at least as extreme* as the observed value, under the model learned during training time[2]: $P(X_{i_1,\ldots,i_D} \geq x \mid \lambda_{i_1,\ldots,i_D})$. That is, our null hypothesis is that a user's behavior will follow our previously learned activity profile. A lower p-value is an indication of an anomalous event.

D. Score Fusion

To make operational use of the anomaly scores produced by our system, we need to summarize these results into the detection of malicious entities, such as stolen user credentials or compromised bastion hosts. We achieve this summarization using p-value fusion of *dependent* p-values [20]. This fusion is accomplished by taking the harmonic mean over all p-values, including the p-values for unobserved events (which are, by definition, 1). Note that fusion can either produce a ranked list of entities (e.g. users, source devices, destinations, days, etc.) or reduce to a lower-dimensional set of events (e.g. user-source and user-destination interactions, etc.).

Finally, we find that fusing the ranked lists produced by multiple multi-dimensional tensors allows us to identify multiple complementary aspects of anomalous behavior, and thus achieve better results than identifying anomalies with any one tensor alone. For fusing ranked lists, we use mean reciprocal rank (MRR) [21].

IV. Los Alamos National Laboratory (LANL) Authentication Dataset

Detailed and diverse datasets at the enterprise scale are rare in the cyber security domain due to privacy and security concerns. Turcotte et al. introduced the publicly available

[2]This p-value can be computed by the Poisson survival function.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on January 04,2021 at 21:34:21 UTC from IEEE Xplore. Restrictions apply.

TABLE I
TENSOR DETAILS AND TEST SET P-VALUE STATISTICS FOR RED TEAM AND BENIGN EVENTS

	Tensor Details			Red Team p-value			Benign p-value		
	Dimensions Size	% Non-Zero	Optimal Rank	Mean	Std	Count	Mean	Std	Count
USDs	11260 x 15055 x 4796 x 2	1.02×10^{-7}	4	.2721	.4090	119	.9575	.1677	125,285
USDHs	11260 x 15055 x 4796 x 24 x 2	3.04×10^{-8}	5	.1062	.2621	137	.9801	.1215	955,945
USDHDs	11260 x 15055 x 4796 x 24 x 7 x 2	1.60×10^{-8}	45	.0175	.0765	138	.9946	.0664	3,513,665

TABLE II
ATTRIBUTE COUNTS AND SELECTED DAYS FOR DATASET SPLITS

Set	User	Source	Destination	Events	Fail %	Days
Train	11,260	15,055	4,796	194,841,640	1.82	1-56
Validation-Train	11,118	14,705	4,698	166,712,680	2.13	1-48
Validation-Test	9,181	10,778	3,508	28,013,171	12.68	49-56
Test	10,165	12,526	4,176	91,547,561	3.88	57-82

Unified Host and Network Dataset[3] to address this critical need [22]. The dataset contains host event and netflow logs collected over a 90-day period at LANL, including red team activity occurring from days 57 to 82. This red team activity provides ground truth information for evaluation of anomaly detection techniques. Attributes in the dataset are anonymized, but the dataset curators took care to ensure the collection remained meaningful for research.

While the LANL dataset contains both network and host data, our work focuses on a subset of the host data. We base our work on the 3.5 million daily average user authentication events collected by the Windows Logging Service (WLS) at endpoint devices in the LANL dataset. We filter the dataset to include only *EventID* 4624 and 4625 which are collections of various types of successful and failed logon records. In particular, we limit *LogonType* to *Interactive*, *Network*, *Batch*, *Service*, *Unlock*, *NewCredentials*, *RemoteInteractive*, and *CachedInteractive*[4]. We disregard events performed by local and system processes (instances where the *UserName* ends with "$") to minimize the presence of automated activity. We extract the following attributes from the remaining authentication data, to be used as dimensions in our tensors:

- *UserName*, user which initiates the log-on.
- *Source*, device where the authentication originates.
- *LogHost*, destination device to be authenticated to.
- *EventID*, fail or success status of the authentication.
- *Time*, timestamp of the event.

Additionally, daylight savings time occurs at day 42 in the LANL dataset, shown in Figure 1. As a result, we increment hours by 1 after day 42 at 2:00 am to normalize the time. Finally, we drop any data instances with missing values.

We split all extracted data instances by time into training, test, validation-train, and validation-test sets. The attribute sizes and day distributions for each split are shown in Table II.

[3]Dataset is available at https://csr.lanl.gov/data/2017/

[4]Detailed attribute descriptions can be found in [22].

A. Tensor Construction

We build three separate binary tensors with dimensions *User - Source - Destination - status* (USDs), *User - Source - Destination - Hour - status* (USDHs), and *User - Source - Destination - Hour - Day - status* (USDHDs). The *status* indicates failed or succeeded logon activity. The *Day* dimension is day of the week (Monday through Sunday), and *Hour* is hour of the day (0 through 23). *User* represents the account which initiates the authentication event (e.g., *UserName*), *Source* and *Destination* are the origin and target devices of the log-on event (*Source* and *LogHost*, respectively, in our data). A tensor entry of 1 indicates the presence of at least one event along the specified dimensions. Table I shows statistics for each tensor.

Authentication events result in immensely sparse problems, where only a small fraction of the large tensor is made up of non-zero elements. Zero values that comprise the majority of the tensor do not need to be stored in memory, allowing us to deviate from dense tensor representation. Instead, sparse tensors can be stored as list of coordinates and a corresponding list of non-zero values.

We store coordinates of the non-zero values with element-to-index mappings of the categorical dimensions. Entities that do not exist in the training data are not mapped in the corresponding test sets.

V. EXPERIMENTS AND RESULTS

We conduct experiments targeting two tasks: (1) detecting anomalous events, and (2) detecting anomalous entities. Anomalous events are analogous to single log messages, for example, a single anomalous *User*, *Source*, *Destination*, *Hour*, and *Day* combination. Anomalous entities are higher-level abstractions, discovered by finding commonalities between multiple anomalous events, for example a single malicious user or a single malicious device.

Our model *does not see any labels for malicious activity*. Following common practice in user behavior analysis, we assume that the vast majority of activity during the training period is benign, and observations during the training time period are used to establish a baseline activity profile. Incident response teams can use our model as a streaming detector, where daily activities are scored against an existing model, and a batch re-training procedure is undertaken on a recurring basis to refine the model.

We quantitatively evaluate our detections using the area under the receiver operating characteristic (ROC) curve (ROC-AUC), which evaluates the extent to which the model assigns

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on January 04,2021 at 21:34:21 UTC from IEEE Xplore. Restrictions apply.

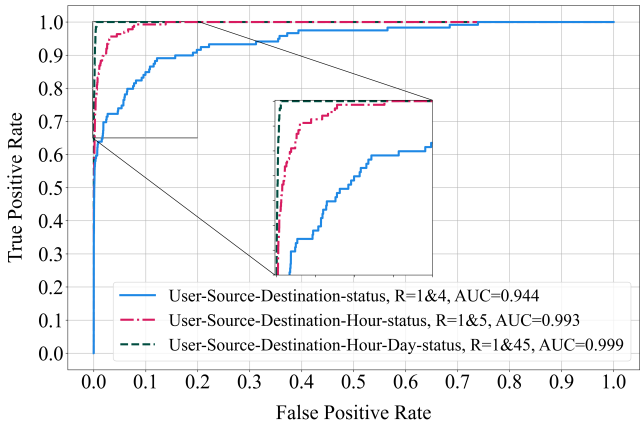

Fig. 2. ROC curve for the authentication tensors with increasing dimension, showing that increasing dimensions improves overall ranking of red team events.

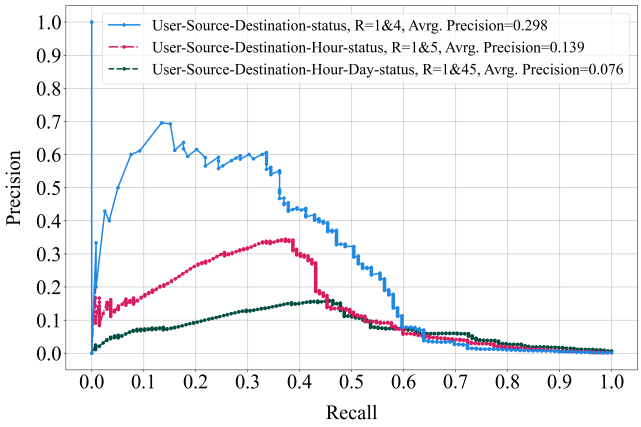

Fig. 3. Precision-recall plot for authentication tensors with increasing dimension, highlighting the cost of higher false positive rates on red team event detection.

lower p-values to red team events than benign events, and average precision (AP)[5], which essentially measures the model's sensitivity to false positives.

A. Detecting Anomalous Events

We detect anomalous events by calculating a p-value for each element in the observed tensor. Our experiments demonstrate that our tensor factorization model can identify anomalies across multiple modalities, and that adding dimensions to the analysis improves the learning of detailed activity profiles.

Table I shows statistics for the p-values inferred across red team and benign events. Red team events have substantially lower average p-values than benign events, which shows that our model discovers meaningful anomalies in an unsupervised manner. As we add dimensions representing the hour of the day and the day of the week to consideration, average p-values decrease for red team events and increase for benign events. Simultaneously, the standard deviation of the p-values drops when the new dimensions are added to the tensor. This result indicates that learning the temporal characteristics of the connections jointly with the peer structure connecting users and their devices enhances detectability. This time-based anomaly detection is novel within a joint statistical framework, and greatly enhances capabilities in applications such as insider threat detection.

Previous work that applied non-negative matrix factorization (NMF) detected anomalies using only two dimensions at a time, such as a *User-Destination* pair [6]. 2-dimensional link prediction methods cannot extract the multi-faceted details of a user's activity profile; for instance, detection of anomalies via *User-Destination* dimensions alone will miss a malicious activity if the only abnormal characteristic of the connection is the *Source* of the authentication event. Using tensors, we jointly learn activity profiles that include all dimensions of a user's behavior, and our experimental results show detection improves, with minimal increase in the computational cost, for tensors with up to 6 dimensions.

[5]Average precision is roughly the area under the precision-recall curve

TABLE III
ENTITY COUNTS FOR FUSION DIMENSION(S)

Target Dimensions	Total Entities	Red Team Entities
User	10,129	76
Source	12,519	1
Destination	4,167	93
User-Source	31,287	76
User-Destination	69,045	117
Source-Destination	70,533	93

Figure 2 shows that adding dimensions improves the *ranking* of red team events within the full list of events observed on the network. However, ROC curves for each tensor cannot be compared directly due to the differing number of benign events. Figure 3 shows that our detections are sensitive to *class imbalance*. For example, with the p-value threshold of 0.001 the USDs tensor identifies 56 of the 119 anomalies while falsely classifying 104 out of 125,285 events. With the same p-value threshold, the USDHDs tensor can detect 108 of the 138 red team events while falsely classifying 3,483 out of nearly 3.5 million events. The insight that lower-dimensional tensors yield better performance, when evaluated in terms of false positives, leads us to develop our anomalous entity detection method to reduce the workload for analysts.

B. Detecting Anomalous Entities

We detect anomalous entities by fusing p-values over tensor dimensions to assign anomaly scores to one or more target dimension(s). In our results, an entity can be a single physical object in the network, such as a *User*, or a combination of physical objects, such as a *User-Source*. Table III shows the total number of entities. Fusing down to more than one dimension allows us to compare our results directly to previous performance benchmarks with matrix factorization.

Figure 4 shows ROC-AUC and AP scores for each tensor when detecting anomalous entities via score fusion. ROC-AUC scores indicate an improved ability to capture anomalous entities with the introduction of time-based dimensions. AP scores reveal an increase in false positives with increasing dimension;

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on January 04,2021 at 21:34:21 UTC from IEEE Xplore. Restrictions apply.

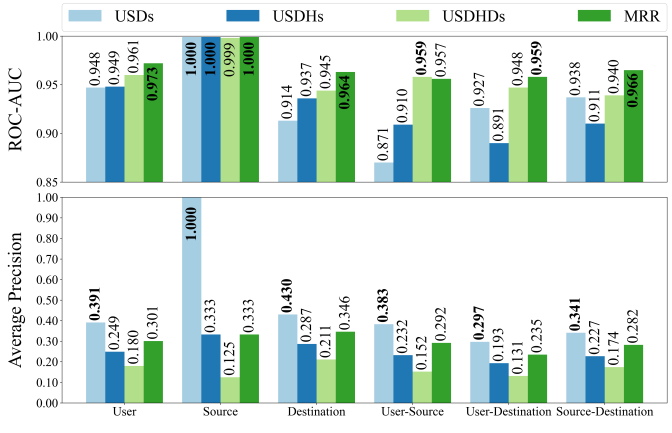

Fig. 4. ROC-AUC and AP scores for anomalous entity detection using tensors with varying dimension. "MRR" shows fusion ranked lists produced by all other tensors.

we suspect this increase is due to added sensitivity to temporal user characteristics. We obtain increased ROC-AUC and AP scores by scoring ranked lists across all tensors, demonstrating that each tensor actually captures complementary anomalies.

The previous benchmark for red team detection on the LANL Unified Host and Network Dataset was established by Sanna Passino et al. with AUC scores of 0.863 and 0.902 when detecting red team events[6] over the matrix with dimensions *User - Source* and *User - Destination*, respectively [6]. Our fusion AUC scores, at 0.957 for *User - Source* and 0.959 for *User - Destination*, demonstrating that jointly learning user behavior patterns over multiple dimensions significantly enhances anomaly detection performance.

VI. Conclusion

We introduce a multidimensional anomaly detection method that is sensitive to anomalous activity over a diverse set of attributes. We show that higher order representations enhance detection of anomalies due to the ability of tensor factorization techniques to extract more predictive activity profiles that describe events simultaneously over multiple dimensions.

Combining information across multiple tensor factorizations demonstrates state-of-the-art results for red team detection. Our framework can be integrated with existing rule-based and statistical intrusion detection systems where post-processing can ease the workload on analysts. For instance, the devices from an alert can be correlated with other weak indicators, such as anomalous process start events [1]. Future work includes augmenting our model to handle "cold starts" by incorporating Sanna Passino's covariate regression model [6].

Acknowledgment

We thank Lissa Moore for helpful suggestions and edits, and Francesco Sanna Passino and Melissa Turcotte for providing valuable feedback and shared attribute mappings [6]. Research presented in this paper was supported by the Information Science and Technology Institute's CyberToaster Research school, and by the Laboratory Directed Research and Development program of Los Alamos National Laboratory (LANL) under project numbers 20190020DR and 20200666DI. LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218CNA000001).

[6]We compare to the PMF model variant that does not have access to covariate information.

References

[1] M. J. Turcotte, J. Moore, N. Heard, and A. McPhall, "Poisson factorization for peer-based anomaly detection," in *IEEE Conference on Intelligence and Security Informatics, ISI 2016, Tucson, AZ, USA, September 28-30, 2016*. IEEE, 2016, pp. 208–210.
[2] "Data breach investigations report 2020," Verizon, Tech. Rep., 2020.
[3] "The cost of cybercrime study," Ponemon Institute, Tech. Rep., 2019.
[4] "Mandiant security effectiveness report," FireEye, Tech. Rep., 2020.
[5] "Cost of a data breach report," IBM, Tech. Rep., 2019.
[6] F. S. Passino, M. J. M. Turcotte, and N. A. Heard, "Graph link prediction in computer networks using poisson matrix factorisation," *CoRR*, vol. abs/2001.09456, 2020.
[7] J. M. Conroy, *Classification of Red Team Authentication Events in an Enterprise Network*, ch. Chapter 9, pp. 179–194.
[8] L. Chapel and C. Friguet, "Anomaly detection with score functions based on the reconstruction error of the kernel PCA," in *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I*, ser. Lecture Notes in Computer Science, vol. 8724. Springer, 2014, pp. 227–241.
[9] E. G. Allan, M. R. Horvath, C. V. Kopek, B. T. Lamb, T. S. Whaples, and M. W. Berry, "Anomaly detection using nonnegative matrix factorization," in *Survey of Text Mining II*. Springer, 2008, pp. 203–217.
[10] M. Price-Williams, M. J. Turcotte, and N. Heard, "Time of day anomaly detection," in *European Intelligence and Security Informatics Conference, EISIC 2018, Karlskrona, Sweden, October 24-25, 2018*. IEEE, 2018, pp. 1–6.
[11] M. Volkovs, G. W. Yu, and T. Poutanen, "Dropoutnet: Addressing cold start in recommender systems," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, 2017, pp. 4957–4966.
[12] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, pp. 455–500, 2009.
[13] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
[14] D. Bruns-Smith, M. M. Baskaran, J. Ezick, T. Henretty, and R. Lethin, "Cyber security through multidimensional data decompositions," in *2016 Cybersecurity Symposium (CYBERSEC)*, 2016, pp. 59–67.
[15] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, vol. 6, pp. 164–189, 1927.
[16] L. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *CoRR*, vol. abs/0903.4530, 2009.
[17] Y. Qi, P. Comon, and L. Lim, "Semialgebraic geometry of nonnegative tensor rank," *SIAM J. Matrix Anal. Appl.*, vol. 37, pp. 1556–1580, 2016.
[18] D. P. Truong, E. Skau, V. I. Valtchinov, and B. S. Alexandrov, "Determination of latent dimensionality in international trade flow," *CoRR*, vol. abs/2003.00129, 2020.
[19] E. C. Chi and T. G. Kolda, "On tensors, sparsity, and nonnegative factorizations," *SIAM J. Matrix Anal. Appl.*, vol. 33, pp. 1272–1299, 2012.
[20] D. Wilson, "The harmonic mean p -value for combining dependent tests," *Proceedings of the National Academy of Sciences*, vol. 116, p. 201814092, 01 2019.
[21] G. V. Cormack, C. L. A. Clarke, and S. Büttcher, "Reciprocal rank fusion outperforms condorcet and individual rank learning methods," in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009*. ACM, 2009, pp. 758–759.
[22] M. J. M. Turcotte, A. D. Kent, and C. Hash, *Unified Host and Network Data Set*. World Scientific, nov 2018, ch. Chapter 1, pp. 1–22.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on January 04,2021 at 21:34:21 UTC from IEEE Xplore. Restrictions apply.