Pattern discovery in intrusion chains and adversarial movement

Nima Asadi
Computer and Information Sciences
Temple University
Philadelphia, USA
nima.asadi@temple.edu

Aunshul Rege
Department of Criminal Justice
Temple University
Philadelphia, USA
rege@temple.edu

Zoran Obradovic
Computer and Information Sciences
Temple University
Philadelphia, USA
zoran.obradovic@temple.edu

***Abstract*—Capturing the patterns in adversarial movement can present crucial insight into team dynamics and organization of cybercrimes. This information can be used for additional assessment and comparison of decision making approaches during cyberattacks. In this study, we propose a data-driven analysis based on time series analysis and social networks to identify patterns and alterations in time allocated to intrusion stages and adversarial movements. The results of this analysis on two case studies of collegiate cybersecurity exercises is provided as well as an analytical comparison of their behavioral trends and characteristics. This paper presents preliminary insight into complexities of individual and group level adversarial movement and decision-making as cyberattacks unfold.**

***Index Terms*—adversarial movement, social networks, time series analysis, cyber security, intrusion chains**

I. Introduction

Organizations worldwide are facing an increasingly sophisticated threat landscape with highly persistent cyberadversaries. Therefore, defenders can only be successful by understanding how adversaries coordinate, make decisions, and adapt to various situations.

Cyberadversaries execute their attacks in distinguishable steps, known as the intrusion chain stages. Multiple intrusion chain models have been proposed in the open literature with varying levels of detail and structure [1]–[3]. While intrusion chain models originally aimed to provide responders with a framework for understanding intrusions, they can be employed as a groundwork to delve deeper into how cyberadversaries advance through a cyberattack. However, in order to carry out such analysis, it is necessary to create a quantitative framework which is suitable for information-based methodologies. This paper aims to create such quantitative framework through data science approaches, namely, time series analysis and social networks. Data from two cybersecurity exercises are used to not only analyze the patterns in time allocation and execution of the intrusion stages, but also how cyberadversaries move through intrusion chains, and whether there are certain traits in the adversarial movements of the group members.

This paper is organized as follows. In the next section, We first explain the case study data collected during two real-time cybersecurity exercises and then discuss the employed data science techniques including time series analysis, clustering, adversarial movement networks, and structural comparisons measurements. We then discuss the empirical results of the analysis on these exercises in section 3. Also, conclusions and further discussion based on the analysis are provided in section 4.

II. Methodology

A. Case Studies

Data were collected at a regional (October 2017) and a national (November 2017) Collegiate Penetration Testing Competition (CPTC). These competitions help students build and refine their skills to discover and mitigate vulnerabilities found in computer systems via a simulated environment that mimics real world networks [4]. At each six-hour competition, one student team was observed and interviewed before, during, and after the exercise. In this paper, we denote these two teams as team 1 (from the regional CPTC) and team 2 (from the national CPTC). Team 1 included 7 members while 6 members formed team 2.

B. Time Series Analysis

The goal of temporal analysis is to discover the patterns and trends within a process. In this paper, temporal analysis is aimed to find the actionable patterns in the activities of the adversarial teams. In order for this analysis to be carried out, it is necessary to use the qualitative observations from the training sessions to create a framework for quantitative analysis. To develop such a framework, we converted the collected qualitative observational data from the case studies to time series using the time stamps and duration of focus of the team on each intrusion stage. We employ an intrusion chain model that is comprehensive and cyclical [5]. The aggregated number of minutes that the entire observed team focused on an intrusion stage within each 10-minute period was calculated to generate the value of the time series at each time point. As

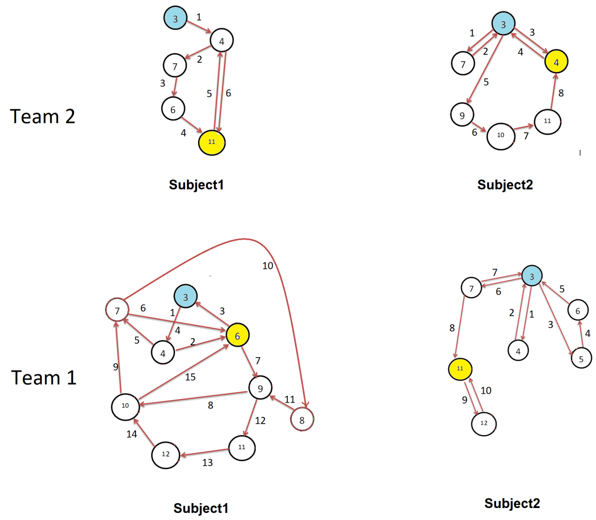

Fig. 1. The adversarial movement topological graph of two subjects from each team. The values in circles show the intrusion stage the subjects performed, and the edge numbers show the order of their movement. Blue nodes are the start/roots nodes and yellow nodes are the end/sink nodes. The intrusion stages based on the employed intrusion chain model include :1-define targets, 2-find and organize accomplices, 3-build or acquire tools, 4-research target infrastructure/employees, 5-test for detection, 6-deployment, 7-initial intrusion, 8-outbound connection initiated, 9-expand access and obtain credentials, 10-strengthen foothold, 11-exfiltrate data, 12-cover tracts and remain undetected [5].

a result, one time series representation was generated for each intrusion stage, where each time point represented a 10-minute period of the exercise, and the value of the time series at each time point was the total time that the entire team spent on the intrusion stage within that time period [5].

After generating the time series data for each intrusion stage and each team, we were able to perform temporal analysis of the teams activities through data mining techniques developed for time series analysis. The correlation level between the time series representation of the team's focus on intrusion stages can indicate the level of co-occurrence of intrusion stages throughout the adversarial movement, thus gaining further capability for capturing the decision making patterns of adversarial teams. In this study, we performed clustering of generated time series as a tool for achieving a measurement of similarity of the patterns in the team's focus on intrusion stages. Clustering places the time series into groups based on their Euclidean distance [6]. Therefore, high consistency between the temporal representation of two intrusion stages indicates that the focus level of the team for those stages showed similar peak/valley patterns throughout the exercise. This analysis results in deeper insight into the decision making process of the adversarial team, which is difficult to achieve through mere observation.

Agglomerative hierarchical clustering, which is a bottom up data grouping technique, was used for this analysis, where pairs or clusters of time series are merged to create the similarity hierarchy [7].

C. *Adversarial Movement Network*

Analysis of patterns in adversarial movement throughout the intrusion can provide crucial information about the decision making process of the teams during cyberattacks. In order to capture such patterns, we can use a network representation of the adversarial activities based on the topological ordering of their movement from one intrusion stage to another. Therefore, in this representation, each node of the graph is the intrusion stage performed by the team member, and the directed edges show the direction of movement from one stage to another [8], [9]. Note that in this representation, we create one graph for each team member as opposed to the time series representation where we created one time series representation for each intrusion stage. An example of such graphs is displayed in Figure 1 which includes the adversarial movements of two subjects from each observed team. As mentioned previously, in that figure each node is an intrusion stage denoted by the node number. Also, each edge number shows the order of the adversarial movement. As an example, the first three movements by subject 1 of team 1 include starting by stage 3 (build and acquire tools), and then moving to stage 4 (research target infrastructure), and then to stage 7 (initial intrusion) based on the intrusion chain model [5]. After creating the adversarial movement graphs, we can analyze them using social network techniques. In order to perform such analysis, we extract and compare their structural characteristics which present important insight into the patterns within the networks [10]. One of the measures used in this analysis is the longest path in the networks which is defined as the maximum number of edges between two nodes in the graph [11]. Therefore, the length of the path between two intrusion stages in our representation is measured by the number of edges that it takes a team member to traverse between them. This measure was selected due to its ability to display the linearity level of the movement of the subjects during the adversarial movement (linear movement here means that it does not contain loop backs to a previously focused stage). By this definition, a completely linear movement is where the team member does not return to a previously taken stage during the entire adversarial movement.

The other analytical measure employed for this study is the number of edges to the number of nodes ratio. In this analysis, the edge to node ratio reveals the frequency of shifts between various intrusion stages by the team member. This is due to the fact that if a subject moves between stages frequently, it creates more edges between intrusion stages, therefore, increasing the edge to node ratio of the network. The results of the adversarial movement network analysis on both case studies are discussed in the next section.

III. Results

A. *Time Series Analysis*

The time series representations were created for each intrusion stage for both teams as explained in the methodology. Example time series created for intrusion stage 4 (research target infrastructure) for both teams is presented in Figure 2.

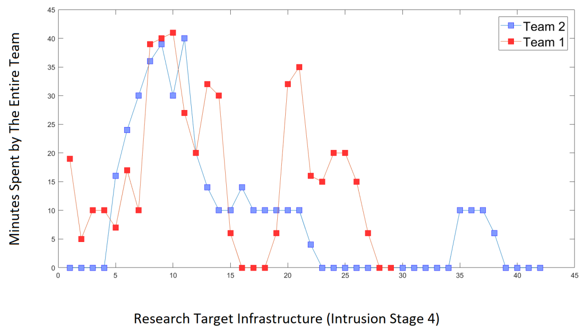

Fig. 2. Time series generated based on the focus (time allocation) of the two teams on intrusion stage 3 (Research target infrastructure).

Subject	Max Path Length	Edge to node ratio
S1	5	1.55
S2	4	1.42
S3	2	1.85
S4	3	1.2
S5	2	2
S6	4	1.58
S7	4	1.76

Subject	Max Path Length	Edge to node ratio
S1	4	1.2
S2	4	1.33
S3	3	1.166
S4	5	1.83
S5	4	1.2
S6	5	1.6

TABLE I

LEFT: THE STRUCTURAL CHARACTERISTICS OF THE ADVERSARIAL MOVEMENT GRAPH FOR TEAM 1; RIGHT: THE STRUCTURAL CHARACTERISTICS OF THE ADVERSARIAL MOVEMENT GRAPH FOR TEAM 2

Also, the results of the agglomerative clustering on the intrusion stage time series of both teams are provided as dendrograms in Figure 3. In that figure, the vertical axis represents the Euclidean distance between two time series. Also, the horizontal red line represents the clustering threshold, which was assigned as the middle of the maximum Euclidean distance. As can be seen in that figure, in both teams, the time allocation time series of intrusion stages 3 (build and acquire tools) and 4 (research target infrastructure) showed higher similarity (lower Euclidean distance). This shows the overall similarity in peaks and valleys of time allocation to those two intrusion stages. For team 1, stages 9, 10, and 11 also showed a higher similarity, but for team 2, the time allocation of intrusion stage 9 did not show similar patterns to intrusion stages 10 and 11. Another noticeable similar pattern in the time allocation between the two teams is in stages 6 and 7, which are placed in the same cluster. However, for team 2, they cannot be considered as one cluster due to the fact that their distance is above the clustering threshold. This analysis shows the similarity and differences between time allocation behavior of the two adversarial teams, which provides further insight into their decision making process.

Furthermore, the correlations between the time series of the two teams are provided in Figure 4. We can observe that in general, the similarities between the time series corresponding to similar intrusion steps are higher than their similarities with time series of other intrusion steps (maximum correlation is 0.5182, and mean correlation is 0.1531).

B. Adversarial movement Network

Adversarial movement networks were also created for each team member in order to capture the traits in alterations between intrusion stages. The structural features of networks were then extracted, which are provided in tables 1 and 2 for teams 1 and 2, respectively. We can observe in those tables that the maximum path length for both teams is 5, and the minimum path lengths is 2 for team 1, and 3 for team 2. Overall, the path lengths are non-homogeneous among the team members, despite the fact that extreme outliers are not observed. The maximum edge to node ratios for team 1 and team 2 are 2 and 1.83, respectively, and the minimum ratio for both teams is 1.2. The average path length of team 1 is 3.42, and the average path length of team 2 is 4.16. Also, the average edge to node ratios for team 1 and team 2 are 1.65 and 1.47 respectively. These results show a quite consistent path length and edge to node ratio among the members of the two teams. In the next section, we provide the conclusions that we drew from the analytical results.

IV. CONCLUSION

1) There is higher similarity between the time allocation patterns of some of the adversarial stages

Observations based on the clustering results of the time series shows that the similarity is higher between intrusion stages 3 (build and acquire tools) and 4 (research target infrastructure/employees), as well as intrusion stages 10 (strengthen foothold) and 11 (exfiltrate data). However, this similarity is not observed among other intrusion stages. Also, one of the noticeable differences between the clustering results on two case studies includes the higher similarity among intrusion stage 9, 10 and 11 for team 2, which is not observed in the time allocation behavior of team 1. This is an indicator of the difference in time allocation pattern between the two teams.

2) Adversarial movements are not linear

Social network analysis of the adversarial movements of the team members involved in the two case studies indicates that most of the generated movement graphs contain loop-back edges between intrusion stages. This shows that the team members shifted their focus to the intrusion chain stages they performed previously, therefore creating a movement that was not sequential. The reasons for returning to previously performed stages can include the failure in progress through the intrusion chain, differences of the objectives among the team members, or the possibility of the subjects being involved in multiple stages. While the intrusion chain model offers a basic set of sequential stages, it fails to capture the non-linearity of movement between different stages [5].

3) Adversarial movements are not homogeneous

By comparing the information provided in Tables 1 and 2 as well as Figure 1, we can observe that although some similarities can be found in the analytical measures of adversarial movement, they are not consistent among all team members. As a result, a conclusion based on this analysis is that the

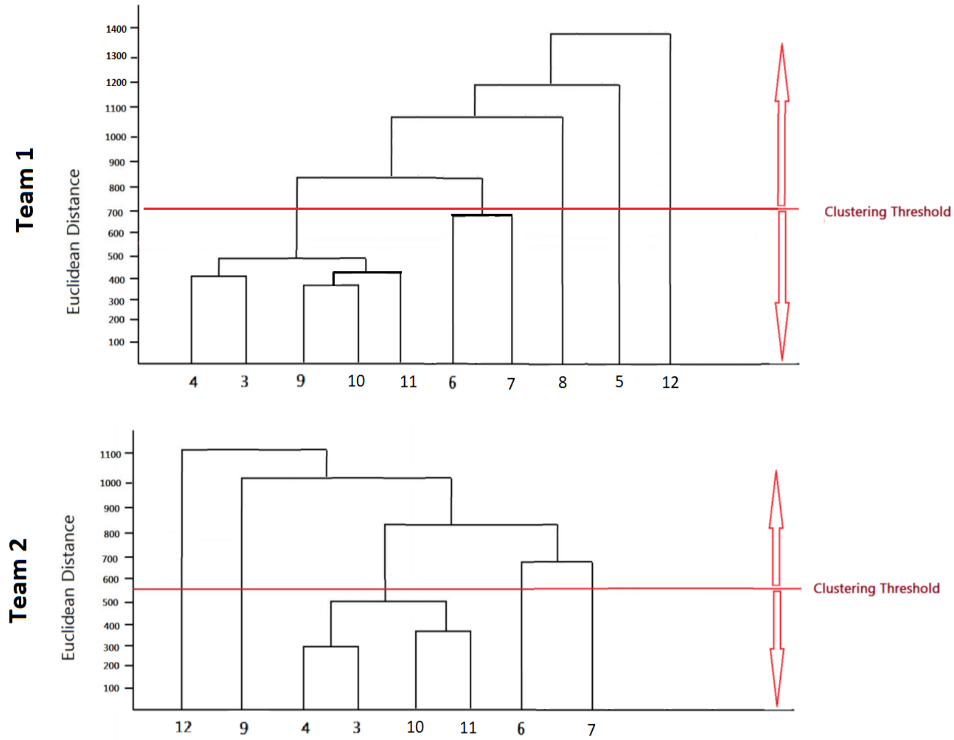

Fig. 3. Clustering results of the intrusion stage time allocation time series.

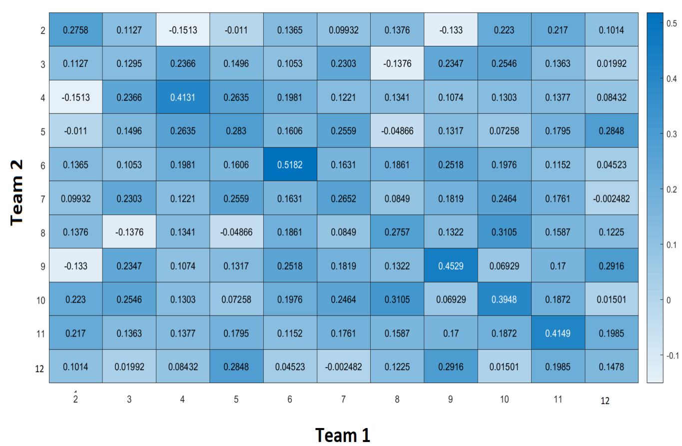

Fig. 4. Correlation heatmap of the time series of the two teams

overall decision making of the adversarial team throughout the exercise is rather individual than based on a unified process.

Despite offering the investigation of only two case studies, this work provides a verifiable analytical framework that sheds light on the complexity and caveats of adversarial movement. We hope this work lays the groundwork for a discussion on intrusion stage models and the different movement patterns and directions of individual adversaries as well as their group dynamics. As future plan, further quantitative experiments could be conducted to further explore adversarial dynamic and movement.

V. Acknowledgement

This material is supported by the National Science Foundation Award No. XXXXXXX.

References

[1] D. 2012, "Lifecycle of an advanced persistent threat," http://www.redteamusa.com/PDF/Lifecycle of an Advanced Persistent Threat.pdf, accessed: 2016-12-20.

[2] D. 2015, "Advanced persistent threats: Understand the threat." http://www.secureworks.com/cyber-threat-intelligence/advanced-persistent-threat/understand-the-threat/, accessed: 2015-06-10.

[3] S. Barnum, "Standardizing cyber threat intelligence information with the structured threat information expression," http://www.mitre.org/sites/publications/stix.pdf.

[4] "Cptc (collegiate penetration testing competition)," http://www.nationalcptc.org/, accessed: 2017.

[5] N. Asadi, A. Rege, and Z. Obradovic, "Analysis of adversarial movement through characteristics of graph topological ordering," in *2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*. IEEE, 2018, pp. 1–6.

[6] O. Maimon and L. Rokach, *Data mining and knowledge discovery handbook*. Springer, 2005, vol. 2.

[7] S. Guha and N. Mishra, "Clustering data streams," in *Data Stream Management*. Springer, 2016, pp. 169–187.

[8] A. D. Kalvin and Y. L. Varol, "On the generation of all topological sortings," *Journal of Algorithms*, vol. 4, no. 2, pp. 150–162, 1983.

[9] D. E. Knuth, *The art of computer programming*. Pearson Education, 1997, vol. 3.

[10] J. D. Guzman, R. F. Deckro, M. J. Robbins, J. F. Morris, and N. A. Ballester, "An analytical comparison of social network measures," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 1, pp. 35–45, 2014.

[11] D. Knoke and S. Yang, *Social network analysis*. Sage, 2008, vol. 154.