Optimizing Alert Data Management Processes at a Cyber Security Operations Center

Rajesh Ganesan[1(✉)], Ankit Shah[1], Sushil Jajodia[1], and Hasan Cam[2]

[1] Center for Secure Information Systems, George Mason University, Mail Stop 5B5, Fairfax, VA 22030-4422, USA
{rganesan,ashah20,jajodia}@gmu.edu
[2] Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD 20783-1138, USA
hasan.cam.civ@mail.mil

Abstract. Alert data management is one of the top functions performed by a Cyber Security Operation Centers (CSOC). This chapter is focused on the development of an integrated framework of several tasks for alert data management. The tasks and their execution are sequenced as follows: (1) determining the regular analyst staffing of different expertise level for a given alert arrival/service rate, and scheduling of analysts to minimize risk, (2) sensor clustering and dynamic reallocation of analysts-to-sensors, and (3) measuring, monitoring, and controlling the level of operational effectiveness (LOE) with the capability to bring additional analysts as needed. The chapter presents several metrics for measuring the performance of the CSOC, which in turn drives the development of various optimization strategies that optimize the execution of the above tasks for alert analysis. It is shown that the tasks are highly inter-dependent, and must be integrated and sequenced in a framework for alert data management. For each task, results from simulation studies validate the optimization model and show the effectiveness of the modeling and algorithmic strategy for efficient alert data management, which in turn contributes to optimal overall management of the CSOCs.

1 Introduction

The desiderata of a CSOC enterprise can broadly be structured into the following major elements: (1) all alerts must be investigated in a timely manner, (2) resources (analysts) must be optimally managed, and (3) desired performance must be achieved. Alert data management of a CSOC consists of several tasks that influence the above elements, and it is imperative that the tasks are optimized to achieve the best CSOC performance. Among the different tasks at a CSOC, this chapter presents three most important tasks, and shows how they are inter-dependent, integrated, sequenced, and optimized. These tasks include (1) determining the regular analyst staffing of different expertise level for a given

S. Jajodia et al. (Eds.): Adaptive Cyber Defense, LNCS 11830, pp. 206–231, 2019.
https://doi.org/10.1007/978-3-030-30719-6_9

alert arrival/service rate, and scheduling of analysts to minimize risk, (2) sensor clustering and dynamic reallocation of analysts-to-sensors, and (3) measuring, monitoring, and controlling the level of operational effectiveness (LOE) with the capability to bring additional analysts as needed.

The framework for alert data management system is shown in Fig. 1. The framework consists of the CSOC alert analysis block, which is central to the system. In this block the alerts are analyzed as per Fig. 2, which is described in Sect. 2. The inputs to the central block are analyst-to-sensor allocation and the alerts per sensor generated by the IDS and any backlog from the previous shift of operation. The outputs include the determination of innocuous and significant alerts, which are further processed as shown in Fig. 2. Additionally, the backlog of unanalyzed alerts is also measured and monitored, which indicates the LOE status of the CSOC. There are three optimization models that are integrated in Fig. 1 and each of them perform an important task. The tasks briefly presented next while the details are presented later in this chapter. The tasks are sequenced in the following order which serves as the road map of the integrated model framework shown in Fig. 1 that is described in this chapter.

Task 1 determines (1) the staffing levels for regular analysts who are categorized into junior, intermediate and senior analysts based on their level of expertise, and (2) the analyst schedule over a two-week (14-day) work cycle that provides adequate analyst expertise mix to handle the alerts in every shift of operation. Under normal operating condition, the above staffing and scheduling of analysts would maintain the level of operational effectiveness of a CSOC and the risk (% of unanalyzed alerts per shift of operation) is minimized and kept at the desired level of performance. Regular analyst levels can be determined based on the alert arrival rate for the given number of sensors, and the alert service rate of each analyst of a particular expertise level using queueing theory [2], which maintains a baseline queue of alerts at any given time (the number of unanalyzed alerts which constitutes an acceptable or desired baseline risk). Regular analyst levels also determine one portion of the analyst budget for alert analysis at a CSOC. Ideally, a zero baseline risk would be desired but as per queueing theory, a zero queue length would need many analysts that could be impractical from a budgetary standpoint. Hence, the queue length for a given arrival rate of alerts and service rate by hired analysts (number of hires within the budget) is deemed to be acceptable, against which the CSOC's LOE performance is measured. An optimization model as shown at the bottom in Fig. 1 achieves the scheduling of regular analysts, which is described in [3]. Under continuous CSOC operation, Task 1 also receives input from Task 3 for any additional analysts that are needed when the queue length exceeds the desirable level (risk increases and the LOE degrades). The additional analysts are also scheduled by the optimization algorithm for scheduling as shown in Fig. 1. The additional analysts constitute another portion of the analyst budget for alert analysis at a CSOC. In summary, the schedule optimization block considers the regular and additional analyst staff and produces a shift schedule for the analysts who must report to work for the immediate following day (24 h or 2 shifts).

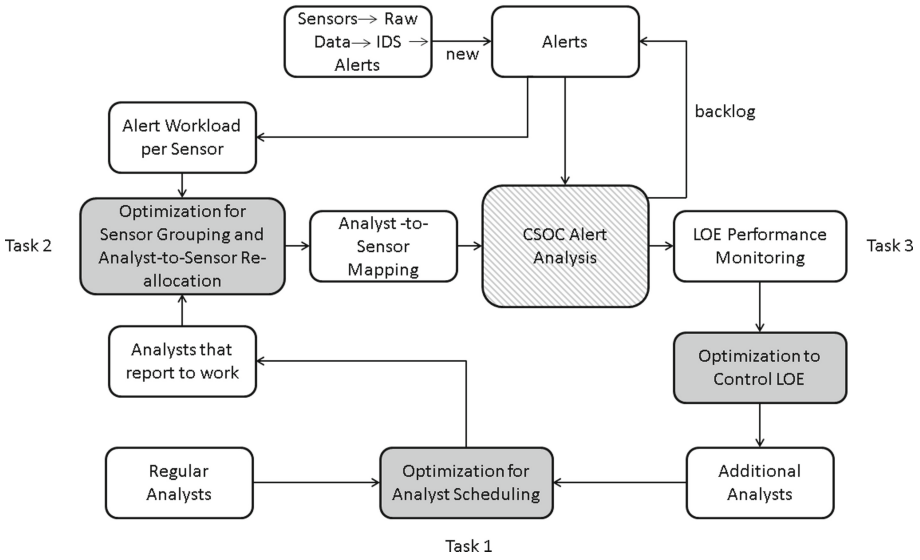

Fig. 1. Framework for alert data management

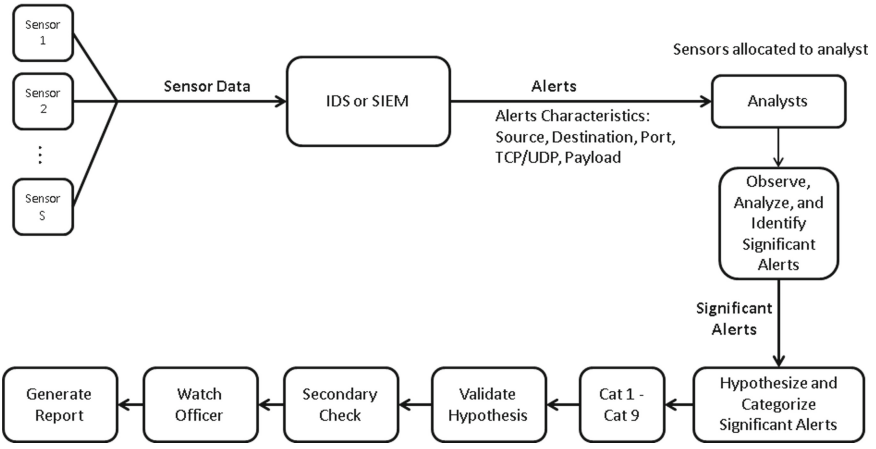

Fig. 2. Alert analysis process [1].

Task 2 performs sensor clustering and dynamic reallocation of analysts-to-sensors as shown in Fig. 1. The sensor grouping and dynamic allocation block is another optimization model that considers the alert workload expected per sensor for the next shift of operation (new alerts and backlog alerts per sensor) and the available number of analysts of various expertise levels that report to work (output of Task 1) in order to generate groups of sensors and the analyst-to-sensor allocation. An optimization model determines the analyst to sensor

mapping such that the tooling and credential expertise to analyze alerts from a sensor are met along with capability to investigate the rate of alert generation by the sensor. The output of Task 2 and the alerts per sensor generated by the IDS and the backlog of alerts are passed into the central analysis block where alert investigation happens based on the process laid out in Fig. 2. The LOE performance is monitored as the output of the CSOC alert analysis process, which leads to Task 3.

Task 3 uses the LOE status as the input or trigger to the LOE optimization, which outputs the number of additional analysts (on-call) required to handle the backlog that is above the normal or baseline backlog queue. The optimization algorithm is a reinforcement learning model which operates in a dynamic mode to determine the additional number of analysts per shift of operation. The output of Task 3 is one of the inputs to Task 1 for scheduling the analysts as shown in Fig. 1. The three tasks are integrated and loop over in order to achieve effective alert data management of a CSOC.

The chapter is organized as follows. Section 2 describes the major elements that provide context for the alert data management of a CSOC, which includes the alert analysis process along with three major characteristics: alert, performance (LOE), and resource characteristics. In Sect. 3, the description of an integrated framework of three optimization models one for each of the above tasks, and their respective roles in effective alert data management are presented. Section 4 presents the related literature. Section 5 concludes the chapter with the major contributions.

2 Alert Analysis Process of a CSOC

Alerts are generated and analyzed by cyber security analysts as shown in Fig. 2. In the current system, the number of analysts that report to work remains fixed, and sensors are pre-assigned to analysts. A 12 h shift cycle is used, and analysts work six days on 12 h shift and one day on 8 h shift, thus working a total of 80 h during a two-week period. There is a very small overlap between shifts to handover any notes and the work terminal or workstation to the analyst from the following shift. The type and the number of sensors allocated to an analyst depend upon the experience level of the analysts. The experience level of an analyst further determines the amount of workload that they can handle in an operating shift. The workload for an analyst is captured in terms of the number of alerts/hr that can be analyzed based on the average time taken to analyze an alert. In this chapter, three types of analysts are considered (senior L3, intermediate L2, and junior L1 level analysts), and their workload value is proportional to their level of expertise.

A cybersecurity analyst must do the following: (1) observe all alerts from the IDS such as SNORT or a Security Information and Event Management (SIEM) tool such as ArcSight [4], (2) thoroughly analyze the alerts that are identified as significant alerts that are pertinent to their pre-assigned sensors, and (3) hypothesize the severity of threat posed by a significant alert and categorize the

significant alert under Category 1–9. The description of the categories are given in Table 1 [5]. If an alert is hypothesized as a very severe threat and categorized under Cat 1, 2, 4, or 7 (incidents) then the watch officer for the shift is alerted and a report is generated (see Fig. 2). The Level of Operation Effectiveness (LOE) of a CSOC is measured at the end of every day of operation.

Table 1. Alert categories [5]

Category	Description
1	Root Level Intrusion (Incident): Unauthorized privileged access (administrative or root access) to a DoD system
2	User Level Intrusion (Incident): Unauthorized non-privileged access (user-level permissions) to a DoD system. Automated tools, targeted exploits, or self-propagating malicious logic may also attain these privileges
3	Unsuccessful Activity Attempted (Event): Attempt to gain unauthorized access to the system, which is defeated by normal defensive mechanisms. Attempt fails to gain access to the system (i.e., attacker attempts valid or potentially valid username and password combinations) and the activity cannot be characterized as exploratory scanning. Can include reporting of quarantined malicious code
4	Denial of Service (DOS) (Incident): Activity that impairs, impedes, or halts normal functionality of a system or network
5	Non-Compliance Activity (Event): This category is used for activity that, due to DoD actions (either configuration or usage) makes DoD systems potentially vulnerable (e.g., missing security patches, connections across security domains, installation of vulnerable applications, etc.). In all cases, this category is not used if an actual compromise has occurred. Information that fits this category is the result of non-compliant or improper configuration changes or handling by authorized users
6	Reconnaissance (Event): An activity (scan/probe) that seeks to identify a computer, an open port, an open service, or any combination for later exploit. This activity does not directly result in a compromise
7	Malicious Logic (Incident): Installation of malicious software (e.g., trojan, backdoor, virus, or worm)
8	Investigating (Event): Events that are potentially malicious or anomalous activity deemed suspicious and warrants, or is undergoing, further review. No event will be closed out as a Category 8. Category 8 will be re-categorized to appropriate Category 1–7 or 9 prior to closure
9	Explained Anomaly (Event): Events that are initially suspected as being malicious but after investigation are determined not to fit the criteria for any of the other categories (e.g., system malfunction or false positive)

2.1 Alert Characteristics

Alert Generation. The network data collected by the sensors is analyzed by an IDS or a SIEM, which automatically analyses the data and generates alerts. Most of the alerts are deemed insignificant by the IDS or SIEM, and about 1% of the alerts generated are classified as significant alerts.[1] The significant alerts are those with a different pattern in comparison to previously known alerts. The significant alerts must be further investigated by cybersecurity analysts and categorized.

Based on the past alert generation rate per day, a historical daily average alert generation rate can be derived, which is used as a baseline for determining a static workforce size, their expertise levels, and their daily work schedule. In reality, the number of alerts generated per sensor per hour varies throughout the day. On days when the number of alerts generated exceeds the above historical daily average alert generation rate, the static workforce size cannot cope with the additional workload, which will result in many alerts that will not be thoroughly investigated. Consequently, the backlog also increases (LOE is reduced). Hence, dynamic scheduling of cybersecurity analysts is a critical part of cybersecurity defense, which includes both the static workforce and a dynamic (on-call) workforce to meet the everyday varying demands on the workforce for alert investigation. In this chapter, the alert generation is modeled as a Poisson distribution, whereas the variation in alert generation per sensor is modeled as a Poisson distribution. The sum of the above distributions taken together will generate the historical daily-average alert generation per day (referred as the baseline alert generation rate). The parameters of the above distributions can be altered as needed based on historical patterns in alert generation, and the dynamic programming model presented in this chapter will adapt and converge to find the optimal dynamic schedules for the analysts that minimizes the backlog, which is the metric to measure the LOE.

Alert Prediction. The uncertainty in the alert generation rate is the primary driver for modeling a dynamic (on-call) workforce in addition to the static workforce that report to work daily. In order to determine the size and expertise composition of the static workforce, the historical daily-average for alert generation is used. However, to determine the size of the dynamic (on-call) workforce on a daily basis, one of the key inputs to the stochastic dynamic programming model is the number of additional alerts (over and above historical daily-average) estimated per sensor for the next day. It should be noted that the dynamic scheduling of analysts is required not only due to the dynamic increase in alert traffic generation rate of the sensors but also the detection of very important attacks/exploits/vulnerabilities such as the first-time detection of zero-day attacks and vulnerabilities (e.g., heartbleed vulnerability and exploit), which

[1] We arrived at the 1% figure based on our literature search and numerous conversations with cybersecurity analysts and Cybersecurity Operations Center (SOC) managers. Our model treats this value as a parameter that can be changed as needed.

could trigger an increase in alert generation rates for the shifts and days following the attack or requires additional monitoring as explained below. When a new zero-day attack is detected or reported in the news, additional dynamic (on-call) analysts are required to determine (i) whether such (zero-day) attacks have already exploited any vulnerability in the network, (ii) what defensive mechanisms such as new signatures (or attack detection rules) must be developed and used to detect (zero-day) attacks, and (iii) what and how attack detection should be reported to upper level management and other agencies. Hence, workload of cybersecurity analysts is increased significantly when zero-day attacks are detected or reported in the industry, even if the traffic rate of sensors during this period may not have necessarily increased. This type of significant event is expected to increase the workload between shifts and the team work of analysts includes not only thorough inspection of events but also preparing and sharing reports, and developing new attack detection rules if needed. In this research, a one-day (one-shift) look-ahead on-call analyst selection model will be run every day (shift) at an appropriate time such that there is sufficient time for the dynamic force to report to work prior to the starting of their shift.

The chapter assumes a Poisson distribution for the baseline average hourly rate of alert generation and a Poisson distribution to introduce variability and spikes in the hourly rate of alert generation. A prediction model for alert estimation using real-world data collected by the CSOC can replace the Poisson distribution in practice. To use the dynamic programming model in practice, the cyber-defense organization could develop statistical models to analyze their data patterns, and replace the distributions that are used in this chapter for making hourly alert predictions for each day of operation. The chapter assumes that the organization has developed a statistical model for alert prediction using historical actual alert generation data, and has determined that the alert generation rate comprises of two distributions. Since, real alert data was not available, the chapter assumes another stream of data to mimic the actual alert generation rate that draws a single random number using only a Poisson distribution whose average is the sum of average of the Poisson distributions that was used to generate the predicted stream of data. In summary, in the real-world, the actual alert rate will come from the intrusion detection system itself and the predicted alert rate will come from the statistical alert prediction model developed by the organization. The avgTTA/hr (LOE status) is estimated using the above rate of alert generation a explained next.

2.2 Performance Characteristics

Performance Metric: LOE. The readiness level of a CSOC is paramount to achieving the above mission successfully. The readiness level must be quantified and measured so that it provides a manager with full understanding of the impact of the interdependencies between various factors that affect the dynamics of the CSOC operations, and take corrective actions as needed. Some of these factors include (1) backlog of alerts that depends on the alert generation and processing rates, (2) the false positive and negative rates of analysts, (3) the

optimal allocation of analysts to sensors, (4) optimal scheduling of the analysts with the right expertise mix in a shift, (5) grouping of sensors, (6) triaging of alerts, (7) the availability of tooling and credentials of analysts in a shift, and (8) effective team formation with highest collaborative scores among the analysts. In this chapter the readiness of the CSOC is defined as the level of operational effectiveness (LOE) of a CSOC, which is a color-coded scheme that indicates the timely manner in which an alert was investigated at the CSOC [6]. The LOE is continuously monitored for every hour of the work shift. Among the factors given above that affect the LOE of a CSOC, this chapter investigates two factors, namely, (1) the dynamic optimal scheduling of CSOC analysts to respond to the uncertainty in the day-to-day demand for alert analysis, and (2) the dynamic optimal allocation of CSOC analyst resources to the sensors that are being monitored. Thus, the objective of this research is to maintain the LOE of a CSOC at the desired level through the dynamic optimal scheduling and allocation of CSOC analyst resources.

In this chapter, the LOE of a CSOC is monitored as follows. The chapter identifies a common metric that is influenced by the disruptive factors that affect the normal operating condition of a CSOC, and this metric is the total time for alert investigation (TTA) for an alert after its arrival in the CSOC database. Any delay in data transmission between the IDS and the CSOC is ignored, and is not part of the TTA metric. In this chapter, it is assumed that an alert will be immediately queued after it arrives in the CSOC database. The TTA of an alert consists of the sum of two parts as shown in Fig. 3: (1) waiting time in queue, and (2) time to investigate an alert, after it has been drawn for investigation by the analyst. Clearly, when the rate of alert generation increases or a new alert pattern decreases the throughput of the system or when the CSOC capacity is reduced by analyst absenteeism the immediate impact is felt in terms of the delays experienced by the alerts waiting in the queue for investigation. Since all the alerts must be investigated, the queue length could become very long. The above means that the alerts stay much longer in the system and the average TTA calculated for each hour (avgTTA/hr) of operation of the CSOC increases.

The avgTTA/hr is calculated at the end of each hour of CSOC operation by using the individual values of TTA for all the alerts that completed investigation during that hour. A baseline value for avgTTA/hr is established for normal operating condition of the CSOC as shown in Fig. 4. It is a requirement of the CSOC that the avgTTA/hr remain within a certain upper-bound (four hours, for example), which is referred as the threshold value for avgTTA/hr. If the avgTTA/hr is maintained below the threshold during any given hour of CSOC operation then the LOE is said to be *optimal*, however, if the avgTTA is maintained at the baseline value then the LOE is said to be *ideal*. Different tolerance bands are created both below and above the threshold value of avgTTA to indicate a color-coded representation of LOE status (see Fig. 4).

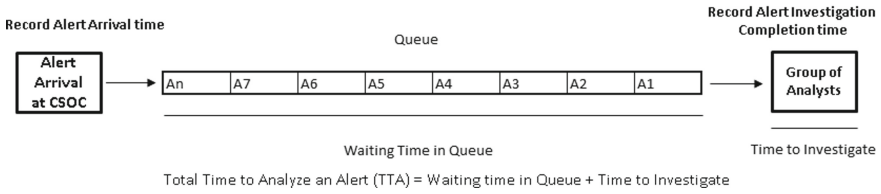

Fig. 3. Total time for alert investigation (TTA) [6]

Performance Metric: Notion of Risk per Sensor. A formal definition for the risk per sensor used in this chapter is presented below. Recently, an optimal scheduling for the cybersecurity analysts was published to minimize risk, where risk was measured as a single overall metric for all the sensors by computing the % of significant alerts that remained unanalyzed by the analysts at the end of a shift [1,3]. However, the context of the problem in this chapter is different. In this chapter, the sensors generate a certain number of alerts per shift, however, in some instances the % of significant alerts among all the alerts generated could vary between the sensors. In another instance, the % of significant alerts between two sensors could be the same but the significant alerts of one sensor takes more time than the other due to a new alert. The consequence of the above is that the number of alerts per sensor that remain unanalyzed (constitutes risk per sensor) is uneven among all the sensors at the end of the shift. This condition results in an imbalance among the risk values obtained from each sensor. In order to balance the number of unanalyzed alerts among all sensors, this research uses a modified notion of risk r_s per sensor s at any time of observation as follows:

$$r_s = V_s - c_s \quad \forall s \tag{1}$$

where r_s is the number of unanalyzed alerts per sensor, which is also the unanalyzed alert queue length for a sensor, V_s is the number of alerts generated from the start of the shift till the time of observation, and c_s (alert coverage) is defined

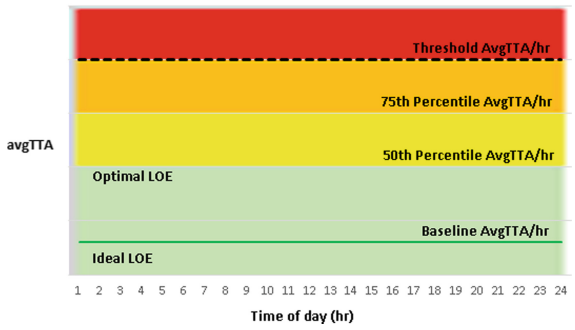

Fig. 4. Color-coded representation of (LOE) [6]

as the number of alerts thoroughly analyzed by the analysts from the start of the shift till the time of observation.

It should be clearly noted that at any observation time during the progress of a shift, the risk per sensor is measured from the number of unanalyzed alerts that have queued up for that sensor (observable to a shift manager), and the queue might contain some significant alerts that are detected upon alert investigation. In other words, one of the metrics to initiate sensor grouping and reallocation decision at any time during a shift is based upon the magnitude of the unevenness in the number of alerts that remain unanalyzed per sensor, which is captured by the alert queue statistics of each sensor. The other metric is analyst utilization. Furthermore, it should also be noted that the adaptive reallocation model is both reactive and proactive in the analyst to sensor decision making process. The model is reactive to the alert queue statistics that is observed for each sensor. The model assumes that the causes for an imbalance in alert queue length (risk per sensor) between sensors would persist after reallocation. Hence, the reallocation model is proactive because it uses the above assumption to compute both the expected alert and significant alert rates for the remainder time in the shift. Using the above reactive and proactive computations, the model determines a new analyst to sensor reallocation decision that will balance the risk per sensor among all sensors as the shift progresses.

In general, a shift manager would observe the length of the unanalyzed alert queue that builds up for each sensor, which is defined as r_s. An imbalance among r_s for all sensors is used as a metric to perform reallocation.

It is true that unless an alert is thoroughly analyzed, its category or severity is unknown. Also, the time taken to analyze an alert depends on its category or severity, whether or not it is a known or a new pattern of alert, and the expertise level of the analyst. Therefore, at the time of drawing an alert from the queue for investigation, since its category or severity is unknown, the time to analyze an alert in this chapter is based upon an average time from a probability distribution, which can be obtained from historical real world data. The total time needed to thoroughly analyze all the alerts and significant alerts can be compared to the total time available, which is based on the current capacity of the organization (number and expertise mix of analysts), their sensor-to-analyst allocation rules, and shift-schedules, in order to determine the % of significant alerts that would remain unanalyzed (risk). Such a risk metric could be used to initiate actions to build analyst capacity for the organization with optimal number of analysts, expertise mix in a work-shift, sensor-to-analyst allocation, and optimal shift schedules. Hence, the scope of the chapter is focused on capacity building for a cyber-defense organization through the optimal allocation and scheduling of its analysts, regardless of the type of alert (category or severity), using the notion that some alerts will need more time than the others. Several parameters are considered in this chapter to calculate the alert investigating capacity of the organization, which includes number of sensors, an average alert generation rate for the sensors, number of analysts, their expertise level, sensor-to-analyst allocation, analyst time to investigate an alert, and their

work-shift schedule. The chapter assumes that all the alerts that were thoroughly investigated were also accurately categorized. It should be noted that as a second metric (quality), once a significant alert has been detected by thorough alert analysis, a different definition of risk can be used to measure the quality of work performed by capturing the true positive and false negative rates. Furthermore, the severity of the threat that an alert poses to the organization, and actions to mitigate the threat can be taken. However, such a definition of risk and the actions to mitigate are beyond the scope of this chapter.

2.3 Resource Characteristics

The analysts have certain characteristics as well. They differ from each other in terms of their expertise levels such as junior, intermediate, and senior analysts. They also have different tooling knowledge and individual credentials (security clearance levels such as confidential, secret, top secret, and so on) to investigate certain types of alerts. From our conversations with CSOC managers, it was learnt that tooling knowledge was correlated to the level of expertise. For example, junior analysts would have access to basic tools while senior analysts would have access to the entire tool-set. Also, their alert service rate, and false positives and negatives rate are not the same among them. Typically, higher expertise is associated with lower false positives and negatives rate. Similarly, optimal matching of analyst's tooling knowledge and credentials with sensor requirements could reduce the number of unanalyzed alerts (backlog). The following are the characteristics of analysts (resources) who investigate alerts.

1. L3 - senior analyst. L3 analysts are assigned 4–5 sensors and they and can handle on average 12 alerts per hour (5 min/alert).
2. L2 - intermediate analyst. L2 analysts are assigned 2–3 sensors and they can handle on average 7–8 alerts per hour (8 min/alert).
3. L1 - junior analyst. L1 analysts are assigned 1–2 sensors and they and can handle on average 5 alerts per hour (12 min/alert).
4. Analysts work in two 12-h shifts, 7 PM–7 AM and 7 AM7 PM. However, the optimization model can be adapted to 8 h shifts as well.
5. Each analyst on regular (static) schedule works for 80 h in 2 weeks (6 days in 12-h shift and 1 day in 8-h shift)
6. When a group of analysts are allocated to a group of sensors by the optimization algorithm, the alerts generated by that group of sensors are arranged in a single queue based on their arrival time-stamp, and the next available analyst within that group will draw the alerts from the queue based on a first-in-first-out rule.
7. Based on experience, an analyst spends, on average, about the same amount of time to investigate alerts from the different sensors that are allocated, which can be kept fixed or drawn from a probability distribution such as Poisson or Uniform.
8. Analysts of different experience levels can be paired to work on a sensor.

9. Writing reports of incidents and events during shifts is considered as part of alert examining work, and the average time to examine the alert excludes the time to write the report. Analysts spend 80% of their time on alert analysis and the remaining time on training and writing reports.
10. L1 analysts are not scheduled on-call because the purpose of on-call workforce is to schedule the most efficient workforce to handle the additional alerts above the historical daily-average that are generated.
11. Analysts of different experience levels can be paired to work on a sensor.

3 Alert Data Management

The framework for alert data management system is shown in Fig. 1. This section describes the requirements and modeling assumptions, which is followed by the detailed description of the tasks.

3.1 Effective Alert Analysis at a CSOC- Requirements

The requirements of the cybersecurity system can be broadly described as follows. The cybersecurity analyst scheduling system,

1. shall ensure that LOE is maintained at the baseline that is established for normal operating conditions,
2. shall ensure that an optimal number of staff is available and are optimally allocated to sensors to meet the demand to analyze alerts,
3. shall ensure that a right mix of analysts are staffed at any given point in time, and
4. shall ensure that weekday, weekend, and holiday schedules are drawn such that it conforms to the working hours policy of the organization.

3.2 Effective Alert Analysis at a CSOC - Model Assumptions

The assumptions of the optimization model are as follows.

1. At the end of the shift any unanalyzed alert is carried forward into the next shift. The backlog indicates the avgTTA/hr, which in turn indicates the LOE status of the CSOC.
2. All alerts that were thoroughly investigated were also accurately categorized. Hence, false positives and false negatives are not modeled in this chapter.
3. The optimization model is run for 24-h to determine the sensor-to-analyst allocation for that day. Simulation statistics on risk and analyst utilization are calculated at the end of the 24-h day.

In the following, the three tasks are described in detail along with their optimization models and results. Tasks 1, 2, and 3 loop over as in Fig. 1, which achieves the effective alert data management in a CSOC.

3.3 Task 1: Scheduling of Analysts to Minimize Risk

Task Description. The objective of Task 1 is to formulate and test an adaptive and dynamic analyst scheduling strategy for effective cyber-defense that is capable of using an estimate of the varying future alert generation rates and scheduling an optimal number of cybersecurity analysts at different expertise levels that minimizes the risk and maintains risk under a pre-determined upper bound for a set of system defined parameters and constraints.

Simulation and Optimization Model. The scheduling optimization model takes inputs from (1) a static mixed-integer programming model for obtaining the minimum number of analysts (static or regular workforce) for a historical daily-average alert generation rate calculated over the past two-week period, and (2) a dynamic LOE model (Task 3) based on stochastic dynamic programming to obtain the minimum number of additional workforce and their expertise level that is needed (dynamic or on-call workforce) based on the estimated additional alerts per sensor for the next day. The mathematical details of the models, algorithms, and implementation guidelines are available in [1,3].

Scheduler Module: The input to the 14-day static scheduling module is the number of personnel needed per level per day, which is derived from the integer programming optimization module. An optimal schedule for the static workforce can be derived based on the following constraints.

1. Each analyst gets at least 2 days off in a week and every other weekend off.
2. An analyst works no more than 5 consecutive days.
3. An analyst works 80 h per two weeks counted over 14 consecutive days between a Sunday and a Saturday. Both 12 h and 8 h shift patterns are allowed.

The objective of the static workforce scheduling algorithm is to find the best days-off schedule and days-on schedule for both 12 h and 8 h shifts for all analysts in the organization subject to the above scheduling constraints. A mixed integer programming scheduling model is used to obtain the 14-day static schedule. During the 14-day schedule, the dynamic programming algorithm would assign on-call status to those analysts who have the day-off. The number of on-call analysts that actually report to work in a day is drawn from those who have been designated with the on-call status.

Results of a Heuristic for Static and Dynamic Workforce Scheduling. The days-off scheduling heuristic is given in [7]. The minimum number of employees needed W as per the scheduling constraints is given as follows.

$$W_1 \geq \lceil \frac{k_2 max(n_1, n_7)}{k_2 - k_1} \rceil \tag{2}$$

$$W_2 \geq \lceil \frac{1}{5} \sum_{j=1}^{7} n_j \rceil \tag{3}$$

$$W_3 \geq max(n_1, \ldots, n_7) \tag{4}$$

$$W = max(W_1, W_2, W_3) \tag{5}$$

where k_1 weekends are off in k_2 weekends, and $n_1, \ldots, n_7$ is the number of employees needed on $Sunday, \ldots, Saturday$ respectively. For a sample scenario of 10 sensors and 6 L1, 6 L2, and 8 L3 analysts required per day (split equally in two 12 h shifts), $k_1 = 1$, and $k_2 = 2$, and $n_1, \ldots, n_7 = 20$. The value of W is 40 (12 L1, 12 L2, and 16 L3), which is the number of employees that the organization must hire (be on payroll) to meet the days-off constraints given above. It should be noted that in the above situation, there are no part-time analysts and all full-time analysts work 12 h shifts ($12 * 7 = 84$ h in every 14-day cycle).

Table 2 shows the combined output of the scheduling heuristic for scheduling static and a fixed dynamic workforce in which X represents days-off for analysts, and c indicates the days on which on-call analysts are scheduled at each level of expertise. The issue with fixing the number of people that are on-call per day at the beginning of the 14-day period is that the cyber defense system is no longer adaptable to higher alert generation rates that exceed the alert rates covered by the fixed on-call workforce. In contrast to the above, the dynamic programming algorithm will select the actual number of on-call workforce required for the next day from the available on-call workforce for that day, which provides greater scheduling flexibility and adaptability to varying alert generation rates. L1 (junior) analysts are not scheduled for on-call workforce.

3.4 Task 2: Sensor Clustering and Dynamic Allocation of Analysts-to-Sensors

Task Description. Understanding the importance and relationship between sensors, analysts, and shift characteristics leads to the two essential properties that must be met in performing the grouping of sensors into clusters, and the allocation of analysts to clusters. The following properties serve as objectives for Task 2. Property 1: meeting the cluster's requirement for specific analyst expertise mix of junior, intermediate, and senior analysts, complete tools coverage that allows the analysts to handle the type of alerts generated by the sensors in the cluster, and analyst credentials such as security clearances needed for the cluster. Property 1 ensures that a high quality of work performance is maintained. In other words, a cluster with a sub-optimal mix of expertise, or with analysts lacking credentials or tooling knowledge is said to perform inefficiently. We define quality of work performed in terms of metrics such as minimizing false positive and negative rates. However, it must be noted that the chapter does not measure the quality metrics directly, instead, it expects the quality of alert analysis to

be high if the above property is met. Property 2: minimizing and balancing the number of unanalyzed alerts in each cluster at the end of the daily work-shift because a large number or an imbalance of unanalyzed alerts among clusters, due to factors such as lack of analyst credentials or tooling expertise in a cluster, would pose a security risk to the organization whose network is monitored by these sensors. Property 2 deals with quantity of unanalyzed alerts and ensures that the overall number of unanalyzed alerts are minimized. It also ensures that no cluster has been unduly disadvantaged in the grouping and allocation process that has resulted in some clusters having a higher number of unanalyzed alerts over other clusters. The quantity of unanalyzed alerts per cluster is measured in this chapter and the motivation behind this property is explained below.

Simulation and Optimization Model. Figure 5 shows the framework of the adaptive model presented in [8], which consists of an optimization and a simulation model. The optimization model used for grouping of sensors to form clusters, and analyst allocation to the clusters is modeled and solved using mixed integer programming. Analyst and sensor characteristics are provided as inputs to the optimization model. A minimum mix of analyst expertise levels, a complete toolset coverage, and analyst credential requirements on the clusters are provided as constraints to the model. The outputs of the optimization model are clusters (groups of sensors) and the analyst to cluster allocation, which are then provided as inputs to the simulation model. The simulation model is used for verification and validation in which a CSOC work-shift is simulated with the analyst and sensor characteristics used in the optimization model. The number of unanalyzed alerts that remain per cluster at the end of the shift is measured by replicating the shift several times in order to achieve a 95% confidence interval.

Results. A case with an increase in alert generation rate is considered. For other results refer to [8]. Due to an increase in the number of alerts generated from some sensors in the last few (one or more) shifts, the estimated average alert generation rate on the respective sensors is increased, though the number of scheduled analysts from various expertise levels, at the start of the shift, remain the same as in the nominal case. In this *case study*, the average alert generation rate per day on twenty sensors were increased to a range between 100 to 150 significant alerts (*i.e.* 1% of 10000 to 15000 total alerts).

The outputs for the case with an increase in alert generation rate is shown in Table 3. It is to be noted that the workload (number of alerts generated) has increased while the resource capacity (number of alerts that could be analyzed by analysts) has remained the same as compared to the nominal case. As a result, there are more number of alerts that remained unanalyzed at the end of the shift compared to the nominal case. In order to minimize the maximum number of unanalyzed alerts at the end of the shift among all the clusters, more number of clusters (11) were created compared to the nominal case (8).

The analyst credential requirement $C5$, for sensors S17, S19, S38, and S40 is met by allocation of analyst A12 to the cluster Q1, while analyst credential

Table 2. Scheduling of L1, L2, and L3 level analysts for both static and a fixed dynamic workforce using days-off scheduling heuristics, X- days-off, and c- on-call [3]

Day →			1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Level ↓	Analyst ID ↓	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun
L3	1	x	c	x	x						c	x			x	x	x
	2	x	x	c	x	x						c	x			x	x
	3	x	x			c	x	x					x	c		x	x
	4	x	x				c	x			x			x	c	x	x
	5	x	c	x	x						c	x			x	x	x
	6	x	x	x	c	x						x	c			x	x
	7	x	x			x	x	c					x	x		c	x
	8	x	x				x	c			x			x	x	c	x
	9			c	x				x	x	x	c			x		
	10			x	c	x			x	x		x	c				
	11					c	x	x	x	x			x	c			
	12						c	x	x	x	x			x	c		
	13			x	x				c	c	x	x			x		
	14			x	x	x			c	c		x	x				
	15					x	x	x	c	c			x	x			
	16						x	x	c	c	x			x	x		
Day →			1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Level ↓	Analyst ID ↓	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun
L2	1	x	c	x	x						c	x			x	x	x
	2	x	x	c	x	x						c	x			x	x
	3	x	x			c	x	x					x	c		x	x
	4	x	x				c	x			x			x	c	x	x
	5	x	c	x	x						c	x			x	x	x
	6	x	x	x	c	x						x	c			x	x
	7	x	x			x	x	c					x	x		c	x
	8	x	x				x	c			x			x	x	c	x
	9			c	x				x	x	x	c			x		
	10			x	c	x			x	x		x	c				
	11					c	x	x	x	x			x	c			
	12						c	x	x	x	x			x	c		
Day →			1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Level ↓	Analyst ID ↓	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun
L1	1	x	x	x	x						x	x			x	x	x
	2	x	x	x	x	x						x	x			x	x
	3	x	x			x	x	x					x	x		x	x
	4	x	x				x	x			x			x	x	x	x
	5	x	x	x	x						x	x			x	x	x
	6	x	x	x	x	x						x	x			x	x
	7	x	x			x	x	x					x	x		x	x
	8	x	x				x	x			x			x	x	x	x
	9			x	x				x	x	x	x			x		
	10			x	x	x			x	x		x	x				
	11					x	x	x	x	x			x	x			
	12						x	x	x	x	x			x	x		

requirement $C4$, for sensors S16, S18, S20, S36, S37, and S39 is met by allocation of analysts A10, A11, and A12 to the clusters Q5, Q6, Q10, and Q11. There is a complete tool-set coverage, and the minimum mix of analyst expertise levels

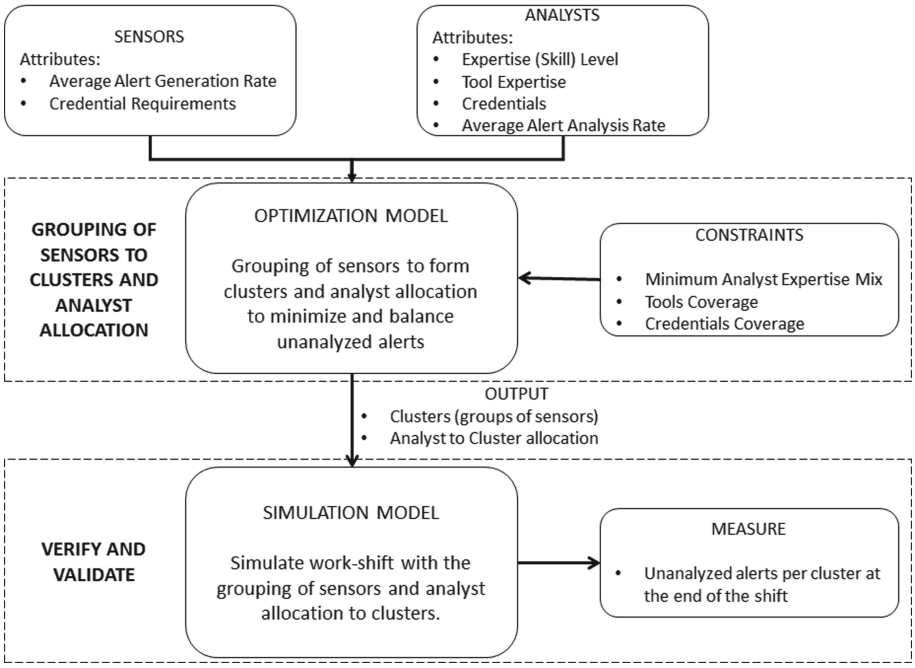

Fig. 5. Adaptive grouping of sensors into clusters and allocation of analysts to clusters model [8].

is maintained on each of the clusters. The average number of unanalyzed alerts per cluster at the end of the shift is balanced as shown in Table 3 with 50 as the maximum average number of unanalyzed alerts that remained on either of the clusters.

The following meta-principles are derived for sensor grouping and allocation to analyst optimization model.

1. It was observed from the clusters that the number of alerts per cluster could vary among them, however, the more important metric is to minimize and balance the number of unanalyzed alerts among the clusters at the end of the shift.
2. It was observed that an analyst allocated to only one cluster may not be a good strategy because if the alert generation rate among the group of sensors in the respective cluster decreases during the shift, the analyst will be idling.
3. The integrated grouping and allocation methodology was able to generate *optimal* sensor to analyst allocation for a given alert generation rate by sensors and known alert service rate by analysts that met the expertise, tooling, and credentials requirements of the cluster.
4. It was observed that the maximum number of clusters that could have been generated (upper-bound, R) is the minimum of the product between the num-

Table 3. Outputs for increase in alert generation rate [8]

Clusters	Groups of sensors	Analysts allocated	Avg. alerts gen.	Avg. alerts unan.
Q1	S17, S19, S38, S40	A2, A4, A6, A12	213	50
Q2	S5, S7	A4, A7, A11	78	50
Q3	S14, S22, S34	A4, A6, A10	122	50
Q4	S1, S13, S21	A1, A6, A11	78	48
Q5	S6, S8, S16, S32	A3, A9, A11, A12	176	50
Q6	S3, S27, S30, S36, S37, S39	A3, A8, A9, A10	297	50
Q7	S2, S11, S23, S31, S35	A1, A9, A12	204	50
Q8	S9, S28	A5, A7, A12	70	50
Q9	S24, S33	A3, A7, A11	93	50
Q10	S4, S10, S20, S25, S29	A5, A6, A12	209	50
Q11	S12, S18, S15, S26	A1, A2, A7, A11	100	50

ber of analysts at each level, and the maximum clusters that could be allocated to an analyst from the respective level.

5. The number of unanalyzed alerts per cluster at the end of the shift can be brought to zero if and only if there are sufficient numbers of analysts hired with various levels of expertise, tooling knowledge, and credentials.
6. With limited analyst resource, the goal is to ensure that the number of unanalyzed alerts is minimized and balanced among the clusters by meeting the expertise, tooling, and credentials requirements of the cluster.

3.5 Task 3: Measuring, Monitoring, and Controlling LOE

Task Description. The objective of Task 3, is to develop an intelligent and adaptive decision support tool for the CSOC manager to take optimal actions (when and how much) to allocate the additional resources in order to maintain an optimal LOE status throughout the 14-day cycle of the CSOC. Due to the dynamic and sequential decision making framework of the 14-day CSOC operation, the chapter presents a reinforcement learning (RL)-based model for representing the manager's decision making process under uncertainty. The RL model takes the continuously monitored LOE status of the CSOC operation as one of its inputs, and takes corrective actions depending on the extent of deviation of the current avgTTA/hr value from the baseline avgTTA/hr value for the CSOC system. The decisions made by the RL model is compared with greedy and rule-based uniformly distributed resource actions to demonstrate the superior decision making ability of the RL model in the face of uncertainties due to disruptive factors. While the rule-based actions are limited in its use of future resources in advance (not adaptive), the greedy actions are myopic in nature that responds to surges in alert backlog by allocating additional resources without any consideration of the future resource needs. For further details refer to [9].

Simulation and Optimization Model. A framework for the dynamic LOE optimization model is provided in Fig. 6. The dynamic optimization model consists of two main blocks - the alert analysis process simulation block, and the RL-based optimization block, which is executed one after another. As explained later, the time over the 14-day work cycle is indexed in 1 h time steps. As shown in Fig. 6, at each time step, the state of the CSOC system is observed, and a decision is made by the RL agent. The decision is then ratified by the CSOC manager, and implemented for the next hour of CSOC operation. The details of the above blocks are presented next.

Simulation Model of Alert Analysis Process: The simulation model of the alert analysis process consists of four main blocks as shown in Fig. 6. They include the CSOC system inputs (system parameters and alert generation by IDS), the uncertain events (both internal and external) that affect the CSOC system inputs, the alert analysis process block in which the work shift is simulated, and the performance metrics block that captures the LOE status of the CSOC at each point in time using the avgTTA/hr metric.

Work-day Simulation: A work-day at the CSOC is simulated and each simulation run corresponds to one operation day of 24 h. Alerts are generated using a Markovian distribution. Analysts are considered as resources, and they investigate alerts from a single queue of alerts populated by the IDSs in a first-come-first-served (FCFS) manner. The time taken to investigate an alert by an analyst, T, is the average time taken based on historical statistics observed in the organization. In this chapter, T is maintained constant except when a new alert pattern causes an increase in T, although it could also be drawn from a probabilistic distribution for each alert. It is assumed that all analysts spend 80% of their effort in a shift toward alert analysis, and the rest of the time is spent on report writing, training, and on generating signatures. Hence, an analyst could increase their effort on alert analysis up to 20% when the need arises, which will increase the service rate of alerts investigated in a day. The alert analysis process of a CSOC is considered to be in steady state under normal operating conditions, which means that the average alert arrival rate per hour, average queue length, average waiting time in queue, and average alert investigation time are all normal. Hence, a baseline avgTTA/hr value for the CSOC system can be established using queueing theory, and the LOE status is ideal as shown in Fig. 4. A threshold value for avgTTA/hr is also established. The scheduled analyst staffing levels are adequate to maintain a pre-determined acceptable avgTTA/hr (and LOE status) of the CSOC.

RL-based Optimization Model for Decision Making: The past hour performance of the CSOC from the real-world alert analysis process (simulation block in this chapter), presents the avgTTA/hr and LOE status to the CSOC manager. Disruptive events, if any, from the past hour are known. It is imperative that the CSOC manager considers the current avgTTA/hr metric and LOE state of the system in order to make a decision to add additional resources or do nothing for the next hour of CSOC operation. The decision is non-trivial because of the uncertainties in the future disruptive events and the limited additional resource

that is available to the CSOC manager. Accurate prediction of future disruptive events such as a zero day attack is very hard. However, one can observe from the history of past events, such as resources that were needed to mitigate a past disruptive event and the frequency of occurrences of each type of disruptive events, and build a probability distribution that can simulate the arrival process of real-world uncertain events. By interacting with the unknown environment via simulation and by learning from the past decisions, the goal of the RL-based decision support system is to optimally plan the allocation of additional resources such that in the long run (over several 14-day cycles), the CSOC system with an adaptive RL-based decision performs far better (in terms of its LOE) than making ad hoc or greedy or a rule-based decision.

When a disruptive event occurs, a CSOC manager, in the order of preference as determined through our discussions with CSOC managers at the Army Research Lab, would utilize the remainder 20% of analyst time on alert analysis, spend some of their own time to assist the analysts in clearing the alert backlog, and bring on-call analysts to supplement the regular analyst workforce. The RL model manages the additional resource allocation that follows the CSOC manager's order of preference, and decides the quantity of additional resource and the timing of when to allocate the additional resources. For further details refer to [9].

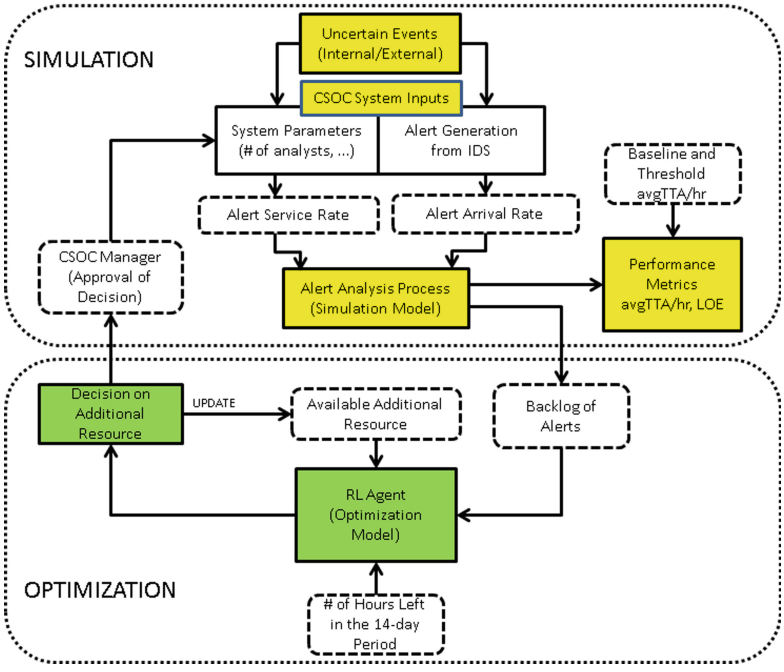

Fig. 6. Dynamic LOE optimization model framework [9].

Results. The following section presents the results from an experiment in which five uncertain disruptive events occurred over the 14-day work cycle. It is reemphasized that in the case of an occurrence of an uncertain event, the additional 20% effort for alert analysis, which is available during every hour of the shift from resources such as analysts and watch officers is utilized first. The on-call analysts are called upon only if needed as a second source of additional resource. Figure 7(a–c) shows only the depletion of on-call additional resource. Where ever the plot is horizontal, it means that the on-call resource was not utilized because (1) either the backlog was 0 (avg. queue length is 1,175 alerts as in the baseline case), or (2) that the backlog was low and was cleared by adding the additional 20% analyst effort for alert analysis.

In the greedy approach as shown in Fig. 7(a), the additional resources were utilized in a myopic manner. An additional resource is called upon only when the workload that is worth the resource's time is accumulated. Such a strategy is commonly employed in various organizations where on-call resources are limited and expensive. For example, an analyst is called upon as soon as four hours of workload is accumulated. As shown in Fig. 7(g), since there is a waiting time for the four hours of workload to accumulate, the LOE (and the alert backlog as shown in Fig. 7(d)) is observed to climb into the yellow zone. As soon as the additional resources are assigned, the LOE is restored into the green zone. However, since this strategy is also myopic (greedy), it can be observed from Fig. 7(g) that there were no additional resources left after ten days, and the LOE climbed into the red zone on the eleventh day.

In the rule-based uniformly distributed approach, the following rules are followed. The CSOC's additional resources are evenly distributed at the start of the 14-day period. As the days progress, any unused resource is rolled over into the following day. Resources allocated to future days cannot be used in advance, which sharply differs from the greedy approach that can exhaust as much resource as needed. The decision to allocate additional resources depends on the available resource on that day, and it is a reaction to the magnitude of the uncertain event that occurs. Figure 7(b) shows that there are unused resources at the end of the 14-day period because resources were evenly distributed and no major event occurred toward the end of the 14 days that consumed all of the remaining resources. Due to the rule that future resources could not be used in advance because they are reserved for future uncertainties, the LOE was found to have higher variance (see Fig. 7(h)) than the greedy and RL approaches as shown in Figs. 7(g) and (i), respectively. Despite being better than greedy in reacting to the uncertainties, the LOE eventually crosses the red band (4-h avgTTA threshold) with the onset of the 4th uncertain event.

There are two critical decisions that are learned with reinforcement learning, (i) when to call the additional resources with respect to the time available in the 14-day work cycle, and (ii) how many additional resources to call upon such that an optimal LOE is maintained over the 14-day work cycle. It can be seen from Fig. 7(c) that very few resources were utilized until the fourth event on the tenth day as compared to Figs. 7(a) and (b) in which the additional on-call resources

were exhausted. As a result, there were fluctuations in the avgTTA values but the LOE was maintained in the green zone. With the event on the tenth day ($t = 240$), the majority of the additional on-call resources were called upon to keep the LOE in the green zone. With resources still remaining at the end of the 14-day work cycle as shown in Fig. 7(c), it can be observed from Fig. 7(i) that the LOE was maintained in the green zone throughout the 14-day work cycle.

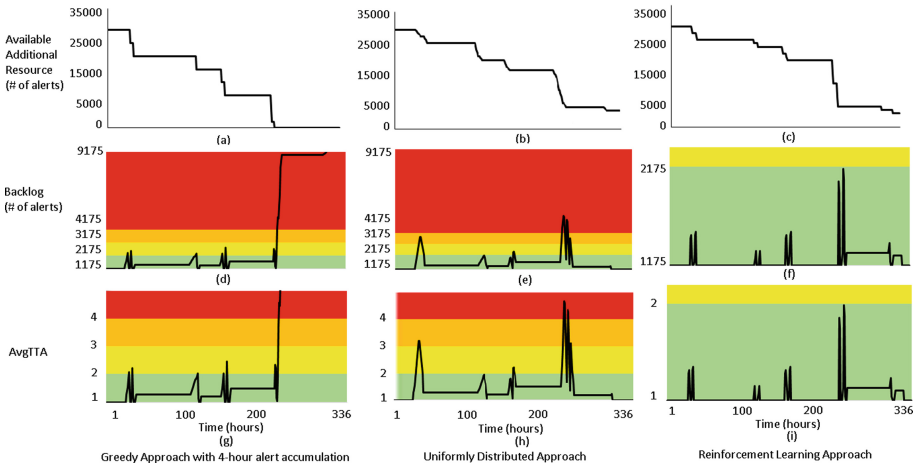

Fig. 7. 5 uncertain events: (a–c) available additional resource, (d–f) backlog, and (g–i) AvgTTA (LOE) [9]

4 Related Literature

D'Amico and Whitley [10] identified six analysis roles of cybersecurity analysts: triage analysis, escalation analysis, correlation analysis, threat analysis, incident response, and forensic analysis. The amount of time an analyst spends in triage analysis is a function of the alert generation rate and is bounded by no more than 80% of the analyst's time (effort) [11]. The rest of the time is spent on writing reports, updating signatures in the IDS from new alert patterns, and training. Triage analysis is the fundamental function of a CSOC. In triage analysis, the large amount of data (alerts) that are generated from IDS using pattern matching techniques [12,13], and automated techniques for malicious behavior [14,15], are investigated to identify suspicious activities (significant alerts). The thorough analysis of a significant alert requires adequate analyst time, which varies between analysts depending on their level of expertise and the category of the alert from the sensor. The alert data received at the CSOC often contains false positive alerts which lead to wastage of time of the analysts. Similarly, there could be false negatives (missing alerts) in the alert data due to unknown vulnerabilities. The experience of an analyst (expertise level) can help in reducing

the number of false positives and false negatives in a shift [16]. It is the desiderata of a CSOC to adequately staff analysts such that all the generated alerts are investigated in a timely manner, and to maintain a proper mix of expertise levels among the analysts allocated to groups of sensors such that the number of false negatives and false positives are minimized.

Managing a CSOC requires critical decision making on scheduling the optimal number of cybersecurity analysts of various expertise levels and an optimal allocation to the sensors in a manner that minimizes the risk to the organization while meeting the resource, work schedule, and organizational constraints. Recent work in literature has focused on optimally scheduling the cybersecurity analysts [17] and their allocations to sensors such that the total number of unanalyzed alerts that remain at the end of the shift is minimized [1,3,18,19], and on improving the efficiency of cybersecurity analysts [20–22]. In practice, at several CSOCs, groups of sensors are clustered together and allocated to analysts for investigation. Sensors have attributes such as historical average alert generation rates, and analyst credential requirements for monitoring and investigating issued alerts. By creating clusters of sensors and allocating them to analysts help in providing context during alert investigation. Analysts allocated to the same cluster of sensors are able to investigate alerts efficiently from the respective sensor during an alert campaign by an adversary. The clusters are adjusted once every few months when a sensor (or a site) is added or removed at the CSOC. This results in an uneven number of unanalyzed alerts that remain at the end of the shift on each cluster. The need to cluster sensors has been recognized, but it has not been implemented often enough, and in a unified manner that takes human factors (analyst attributes) into consideration. To the best of the authors' knowledge, a unified model that creates clusters (groups of sensors), and allocates cybersecurity analysts to the clusters by taking into account the unique attributes of both, sensors and analysts, to minimize and balance the work (unanalyzed alerts) that remain at the end of the shift has not been studied or researched in published literature.

A CSOC is a unique amalgamation of people, processes, and technology. A CSOC performs many roles in terms of variety of services offered, which are broadly categorized into reactive, proactive, and security quality management services [23]. Alert management, incident handling, and vulnerability handling are categorized under the reactive services, while intrusion detection services [24] and development of security tools [22] are categorized under the proactive services. D'Amico and Whitley [10] conducted a cognitive task analysis to study the analytical process that transforms data into security situation awareness, which is categorized as a security quality management service provided by a CSOC. A complete list of services offered by a CSOC is given in Killcrece et al. [23].

Real-time work schedule adjustments or reactive-scheduling has been studied for over three decades. Early research work on reactive-scheduling used rolling horizon technology for the job-shop scheduling [25] while real-time schedule adjustment in a call center environment had been studied to provide a high level of customer service [26]. A sequential mixed-integer programming with loose

bounds has been shown to achieve higher profit improvement than experienced managers in real-time work schedule adjustment decisions at a quick service restaurant study [27]. In the hospitality and tourism industry, research efforts have focused on using mixed-integer programming to solve tour scheduling problems to minimize labor cost and meet service standards [28,29]. In manufacturing systems, predictive-reactive scheduling has been studied where schedules are revised in response to real-time events such as machine breakdowns and random job arrivals [30]. The relationship between situational constraints (schedule or allocation disruption) and the effect on worker performance has been a topic of interest in organization studies [31] which relates excessive schedule disruptions to worker morale and to worker turnovers.

5 Conclusions

The chapter presented an innovative integrated framework for efficient alert data management, and highlighted three very important tasks and the strategies to optimize them. The desiderata of a CSOC enterprise can be achieved by the integrated model in which all alerts could be investigated in a timely manner, the CSOC resources (analysts) could be optimally managed, and the LOE desired performance could be achieved. The tasks considered in this chapter are (1) determining the regular analyst staffing of different expertise level for a given alert arrival/service rate, and scheduling of analysts to minimize risk, (2) sensor clustering and dynamic reallocation of analysts-to-sensors, and (3) measuring, monitoring, and controlling the level of operational effectiveness (LOE) with the capability to bring additional analysts as needed. The chapter demonstrated the framework under which the inter-dependent tasks can be integrated, sequenced, and optimized, which is very useful for CSOC managers to make shift-to-shift decisions on scheduling analysts, allocation them to sensors, and maintaining the LOE of the CSOC. The algorithmic details of each optimization model are available in the cited references under each task, and the best strategy is to optimize individual tasks such that the outputs of one task are the inputs to another as described in the introduction to the chapter. As on-going and future research, trade-off analysis of competing factors would be studied, along with other tasks such as alert prioritization and team formation in order to increase the fidelity of the integrated model.

Acknowledgment. The authors would like to thank Dr. Cliff Wang of the Army Research Office for the many discussions which served as the inspiration for this research. This work is partially supported by the Army Research Office under grant W911NF-13-1-0421.

References

1. Ganesan, R., Jajodia, S., Cam, H.: Optimal scheduling of cybersecurity analyst for minimizing risk. ACM Trans. Intell. Syst. Technol. **8**(4), 52:1–52:32 (2017)

2. Gross, D., Shortle, J., Thompson, J., Harris, C.: Fundamentals of Queuing Theory. Wiley, New York (2008)
3. Ganesan, R., Jajodia, S., Shah, A., Cam, H.: Dynamic scheduling of cybersecurity analysts for minimizing risk using reinforcement learning. ACM Trans. Intell. Syst. Technol. **8**(1), 1–21 (2016)
4. Bhatt, S., Manadhata, P.K., Zomlot, L.: The operational role of security information and event management systems. IEEE Secur. Privacy **12**(5), 35–41 (2014)
5. CIO: DON cyber crime handbook. Department of Navy, Washington, DC (2008)
6. Shah, A., Ganesan, R., Jajodia, S., Cam, H.: A methodology to measure and monitor level of operational effectiveness of a CSOC. Int. J. Inf. Secur. **17**(2), 121–134 (2018)
7. Pinedo, M.: Planning and Scheduling in Manufacturing and Services. Springer, New York (2009). https://doi.org/10.1007/978-1-4419-0910-7
8. Shah, A., Ganesan, R., Jajodia, S., Cam, H.: Optimal assignment of sensors to analysts in a cybersecurity operations center. IEEE Syst. J. **13**, 1060–1071 (2018)
9. Shah, A., Ganesan, R., Jajodia, S., Cam, H.: Dynamic optimization of the level of operational effectiveness of a CSOC under adverse conditions. ACM Trans. Intell. Syst. Technol. **9**(5), 51:1–51:20 (2018)
10. D'Amico, A., Whitley, K.: The Real Work of Computer Network Defense Analysts. In: Goodall, J.R., Conti, G., Ma, K.L. (eds.) VizSEC 2007. MATHVISUAL. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78243-8_2
11. West-Brown, M.J., Stikvoort, D., Kossakowski, K.P., Killcrece, G., Ruefle, R.: Handbook for computer security incident response teams (CSIRTs). DTIC Document CMU/SEI-2003-HB-002 (2003)
12. Bejtlich, R.: The Tao of Network Security Monitoring: Beyond Intrusion Detection. Pearson Education Inc., Boston (2005)
13. Crothers, T.: Implementing Intrusion Detection Systems. Wiley, New York (2002)
14. Di Pietro, R., Mancini, L.V. (eds.): Intrusion Detection Systems. Advances in Information Security, vol. 38. Springer, New York (2008)
15. Northcutt, S., Novak, J.: Network Intrusion Detection, 3rd edn. New Riders Publishing, Thousand Oaks (2002)
16. Kott, A., Wang, C., Erbacher, R.F.: Cyber Defense and Situational Awareness. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11391-3
17. Altner, D.S., Rojas, A.C., Servi, L.D.: A two-stage stochastic program for multi-shift, multi-analyst, workforce optimization with multiple on-call options. J. Sched. **21**, 517–531 (2017)
18. Ganesan, R., Shah, A.: A strategy for effective alert analysis at a cyber security operations center. In: Samarati, P., Ray, I., Ray, I. (eds.) From Database to Cyber Security. LNCS, vol. 11170, pp. 206–226. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04834-1_11
19. Ganesan, R., Shah, A., Jajodia, S., Cam, H.: A novel metric for measuring operational effectiveness of a cybersecurity operations center. In: Wang, L., Jajodia, S., Singhal, A. (eds.) Network Security Metrics, pp. 177–207. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66505-4_8D
20. Erbacher, R.F., Hutchinson, S.E.: Extending case-based reasoning to network alert reporting. In: 2012 ASE International Conference on Cyber Security, pp. 187–194 (2012)
21. Sundaramurthy, S.C., et al.: A human capital model for mitigating security analyst burnout. In: Eleventh Symposium on Usable Privacy and Security (SOUPS 2015), pp. 347–359. USENIX Association (2015)

22. Sundaramurthy, S.C., McHugh, J., Ou, X., Wesch, M., Bardas, A.G., Rajagopalan, S.R.: Turning contradictions into innovations or: how we learned to stop whining and improve security operations. In: Twelfth Symposium on Usable Privacy and Security (SOUPS 2016), pp. 237–250. USENIX Association (2016)
23. Killcrece, G., Kossakowski, K.P., Ruefle, R., Zajicek, M.: State of the practice of computer security incident response teams (CSIRTs). Technical report CMU/SEI-2003-TR-001, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, table 9, p. 66 (2003)
24. Scarfone, K., Mell, P.: Guide to intrusion detection and prevention systems (IDPS). Special Publication 800-94, NIST (2007)
25. Nelson, R.T., Holloway, C.A., Mei-Lun Wong, R.: Centralized scheduling and priority implementation heuristics for a dynamic job shop model. AIIE Trans. **9**(1), 95–102 (1977)
26. Cleveland, B., Mayben, J.: Call Center Management on Fast Forward: Succeeding in Today's Dynamic Inbound Environment. Call Center Press, Annapolis (1997)
27. Hur, D., Mabert, V.A., Bretthauer, K.M.: Real-time work schedule adjustment decisions: an investigation and evaluation. Prod. Oper. Manag. **13**(4), 322–339 (2004)
28. Love, R.R., Hoey, J.M.: Management science improves fast-food operations. Interfaces **20**(2), 21–29 (1990)
29. Loucks, J.S., Jacobs, F.R.: Tour scheduling and task assignment of a heterogeneous work force: a heuristic approach. Decis. Sci. **22**(4), 719–738 (1991)
30. Vieira, G.E., Herrmann, J.W., Lin, E.: Rescheduling manufacturing systems: a framework of strategies, policies, and methods. J. Sched. **6**(1), 39–62 (2003)
31. O'Connor, E.J., Peters, L.H., Rudolf, C.J., Pooyan, A.: Situational constraints and employee affective reactions: a partial field replication. Group Organ. Stud. **7**(4), 418–428 (1982)