26th Annual INCOSE International Symposium (IS2016)
Edinburgh, Scotland, July 18 – July 21, 2016

On System Dynamics Modeling of Human-Intensive Workflow Improvement – Case Study in Cybersecurity Adaptive Knowledge Encoding

Keith D. Willett, Rick Dove, Robert Cloutier, and Mark Blackburn
Stevens Institute of Technology
Castle Point on Hudson, Hoboken, NJ 07030
kwillett@stevens.edu, rdove@stevens.edu, rcloutier@southalabama.edu,
mblackbu@stevens.edu

Abstract. Modeling and simulation offers one way to *express* a system of systems, to *measure* its performance, and *predict* the effects of prospective modifications. This paper introduces a system dynamics model (SDM) as a systems engineering design decision support tool. The SDM structure, design rationale, and results have broader applicability to many people-oriented workflows that may benefit from 1) working smarter and 2) introducing automation to complement human activities. The SDM structure consists of a sequential time-essential workflow of eleven phases. The simulation establishes a pre-modification performance baseline, identifies the fit of a prospective modification with intent to improve workflow, estimates the modification's effect on performance, and then produces post-modification performance results to compare with the baseline. To provide focus for the SDM, the details herein use cybersecurity operations as one example of a system of systems predominantly operated by people performing insufficiently to address the volume of cyberspace-related anomalies. The SDM is specifically for cybersecurity operations to identify procedural bottlenecks and show the impact of proposed improvements. Cybersecurity decision patterns (CDPs) provide a case study as a prospective improvement. CDPs are one solution under the *integrated adaptive cyberspace defense - joint cognitive system* capability area to help establish and sustain agile cybersecurity operations (Willett 2015a). Lessons learned from the approach, application, and results are more broadly applicable to many systems and system of systems operations.

Introduction

The cybersecurity operations system dynamics model (SDM) approach, structure, method, and lessons learned are applicable to many workflow processes (e.g., manufacturing, materiel movement, and transportation). All of these processes are subject to potential disruptions (threats) and internal weaknesses (vulnerabilities) that include the identification and processing of anomalies in the form of indicators, degraded performance, or failures. The nature and degree of anomaly processing will vary according to the system of interest. The focus herein is on a SDM for cybersecurity operations as a first step toward a more abstract articulation and improvement of anomaly processing in other systems of systems (SoS).

The dynamics of cyberspace operations include the continuous monitoring of assets and dealing with deviations from normal or expected; i.e., monitoring for and addressing *anomalies* in the cyberspace environment. There is a similar approach to address anomalies in any SoS. Anomaly processing consumes resources (e.g., personnel, equipment, time, bandwidth, processing

capacity, and money). An efficient operations design optimizes resource utilization to sustain the ability of the SoS to fulfill its defined mission. This optimization in part requires a continual shift from people being predominantly *in-the-loop* of performing tasks to being predominantly *on-the-loop* to plan and design automation, as well as to validate and modify machine processing. For each operating environment, there is a harmony of people working with machine processing in order to free people from rote tasks in order to focus them on cognitive intense core mission activities.

Cybersecurity operations is a SoS that includes people, machines, and their interaction to ensure organizational efficacy by safeguarding the workflow during its fulfillment of the mission. The many interacting systems provide for monitoring the cyberspace tactical environment (*observe*), understanding what is seen (*orient*), identifying and selecting among viable options (*decide*), and performing tasks to achieve a desired end (*act*). Additionally, there is the need for governance and adjudication to direct strategic and tactical operations (*command*) and a method to deliver commands to all the constituent parts of cybersecurity operations (*control*). To identify and resolve the many considerations, the systems engineer can benefit from *design support* for prospective observe, orient, decide, act (Osinga 2005), command, and control (OODA+C2) solutions within cybersecurity operations. One method of systems engineering (SE) support is modeling and simulation as a "discipline for developing a level of understanding of the interaction of the parts of a system, and of the system as a whole."[1]

Though the model is flexible enough to accommodate many prospective solutions, this paper uses cybersecurity decision patterns (CDPs) and a cybersecurity decision pattern language (CDPL) (Willett, Dove, and Blackburn 2015) as prospective solutions on which to focus the model and simulation. The uniqueness of the approach herein is the use of the SDM to represent anomaly processing in cybersecurity operations. The SDM helps to examine the constituent parts, their interactions, identify areas in need of improvement, apply CDPs as one prospective solution for improvement, and quantify the expected level of improvement for decision support to the systems engineer attempting to design an optimal operating environment among people and machines for anomaly processing.

The model establishes baseline operational performance based on real world data. The method is flexible enough to adapt baseline details to any operational environment. The results of prospective modifications to operational performance are compared to this baseline to determine the nature of the modification, which may be an improvement, degradation, or no change to operational effectiveness or efficiency.

Foundations

A *system of interest* is "the system whose life cycle is under consideration" (INCOSE 2015). A SoS is "a system of interest whose system elements are themselves systems"(INCOSE 2015). A systems engineer provides an interdisciplinary approach and means to enable the realization of successful systems. They define customer needs and desired functionality, document requirements, design the system and its synthesis into the containing whole (e.g., SoS), and provide system validation in context of the holistic problem (INCOSE 2015). Additionally, the systems engineer considers both technical functionality and business/mission priorities for design decisions and for recommending solution investments.

[1] http://www.systems-thinking.org/modsim/modsim.htm, last accessed 14-Sep-2015

Traditional systems engineering security predominantly focuses on fault-tolerance and safety. The increase of cyber-driven systems in an interconnected world recently prompted the International Council on Systems Engineering (INCOSE) to add *systems security engineering* as a core discipline for systems engineers. INCOSE recognizes that cyberspace threats may result in more than just denial of service to a computer or loss of data, but also may include kinetic effects in the real world resulting in physical injury or death (INCOSE 2015). For example, attacks against industrial control systems may cause dam overflows resulting in flooding, furnace overheating resulting in explosion, electrical power plant shutdown, medical device failure (e.g., pacemakers), or remote hijacking of moving automobiles and other vehicles. In addition to the traditional security approach, systems design should consider a threat with intelligence and intent (e.g., state sponsored, organized non-state sponsored (e.g., terrorists, organized crime), and general vandalism (e.g. hacker, script-kiddie)). The details of the cybersecurity operations workflow provide one view for systems engineers on how to identify and address anomalies in a greater system of systems to assure adherence to these security principles.

System Dynamics Modeling Approach

Systems thinking is the process of understanding how things influence one another within a whole including patterns of behavior *within* the whole, behavior patterns *of* the whole, and to *conceptualize local changes* in context of the whole (Gharajedaghi 1999). System dynamics modeling is a computer-aided approach to analysis and design. SDM applies to any system or SoS characterized by interdependence, mutual interaction, information feedback, and circular causality.[2] The SDM tool used to produce the results herein is Stella/iThink[3]. The modeling process entails problem articulation, hypothesis formulation, model formulation, model verification and validation, and policy design and evaluation (Sterman 2000).

The problem at hand is the lack of ability for cybersecurity operations practitioners to *observe*, *orient*, *decide*, *act*, *command*, and *control* within *cyber-relevant time* (Herring and Willett 2014) to maximize utilization of limited practitioner resources. We may be doing the right things, just not doing them fast enough. The case study hypothesis is, if we implement CDPs and the CDPL, we will improve manual processing of cyberspace anomalies. We realize improvements by facilitating people-to-people knowledge sharing where knowledge content minimally includes *context*, *problem*, and *solution*, and is thus more immediately actionable than data or information. Moreover, we will improve machine processing of cyberspace anomalies by creating people-validated repository of knowledge from which to produce machine encoded processing throughout OODA+C2.

Improved machine effectiveness and overall efficiency of cybersecurity anomaly processing will free people from rote tasks and enable them to focus on cognitive-intense tasks that are currently better suited to people. A modeling and simulation approach helps to identify operational structures that need improvement and to determine the degree of prospective improvement to provide decision support on where to apply automation.

---

[2] http://www.systemdynamics.org/what-is-s/, last accessed 9-Sep-2015

[3] Certain commercial software products are identified in this paper. These products were used only for demonstration purposes. This use does not imply approval or endorsement by INCOSE or Stevens Institute of Technology, nor does it imply these products are necessarily the best available for the purpose.

The root of the problem lies within current cybersecurity operations practices and its limitations in available solutions and methods to address anomalies. Therefore, for the purpose of modeling, the historical time horizon is the most recent operations performance statistics; i.e., some data within the previous 12 to 24 months that reflects current practices and solutions for anomaly detection and processing. Additionally, the future time horizon is within the next budget cycle to prioritize solution acquisition and method enhancement to improve anomaly processing.

The model formulation is based on scenarios of how cybersecurity operations practitioners encounter the real world. In brief, they become aware of some phenomenon (observe), they seek to understand it (orient), they identify and select among viable options on how to address it (decide), and then carry out some task to do something about it (act). While OODA is the essence of the workflow process, there needs to be more granularity to focus analysis, identify bottlenecks, and evaluate prospective solutions to introduce effectiveness or enhance efficiency. The SDM structure provides this granularity in eleven workflow phases described in the next section.

Model verification and validation includes subject matter expert (SME) review of the design, rationale behind the design, the method to produce key probability formulas that represent real world workflow, and the results of the model. The specific SDM numbers reflect real world performance as reported in the Ponemon Cost of Data Breach Report 2014 and derived from a major East Coast academic research institution's cybersecurity operations. Any given operating environment will likely vary from these specific numbers. The SME evaluation is to ensure the modeling and simulation method remains sound and flexible enough to accommodate performance variations in any given operating environment. SME's performing the review included a range of mathematicians to validate the mathematics, and modeling and simulation practitioners to review the model design and flow. Cybersecurity operations SMEs reviewed the baseline and the projected results in terms of anomaly quantities processed and time of processing, and variety technical directors and operations managers to review the overall modeling approach and its reflection of real world practices. All input was used to improve the model design and its reflection of cybersecurity operations behavior (Law 2007).

Policy design and evaluation explores what-if scenarios of prospective solutions for cybersecurity operations that provide some aspect of OODA+C2. The model is designed for both machine and manual anomaly processing. For this paper, machine processing is turned off so we may initially focus on the prospective effects of people-to-people knowledge sharing to 1) improve manual processing and 2) to build a foundation of people validated knowledge from which to encode machine processing. The what-if scenarios look at a range of prospective effects of knowledge generation in the form of CDPs as one example of any number of other proposed solutions. The model design can support a variety of prospective effects from a single minor change in one workflow phase to many changes across multiple workflow phases. To reiterate, this case study is a single example of the potential to apply the same SDM approach to any people-oriented workflow process.

SDM Application to Cybersecurity Operations Design

There are two models herein; the first is a *probability distribution function (PDF) determination model* (Figure 1) that produces the probability distributions necessary for an accurate second *system dynamics model* (SDM). The SDM simulates cybersecurity operations workflow to process anomalies using ten workflow phases. These phases are a more granular view of OODA

in context of cybersecurity. This is one example of how to use the OODA concept in a SDM model. Other SoS-related models may use OODA explicitly or adapt OODA to the local workflow. Likewise, the PDF determination model shows one way to discern processing times with basis in reality. The value in both the SDM and PDF determination model for systems engineers are more in the method and approach than in the explicit reuse of the case study details.

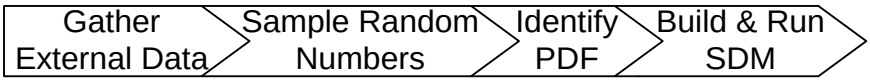

Figure 1: PDF Determination Model Method

System Dynamics Model

Cyberspace threats are increasingly automated and attacks emerge at network speed. Therefore, cyberspace defense must be equally *adaptive*, *dynamic*, and *active*; i.e., there is an increasing need for *agile* cyberspace defense within cyber-relevant time (Herring and Willett 2014). The capability-based engineering of integrated adaptive cyberspace defense (Willett 2015a) provides a fundamental capability framework of OODA+C2 for cybersecurity operations. There are a large number of potential solutions under OODA+C2 umbrella. The judgment as to one being better than another is largely dependent on operational need. To help manage the complexity of SE design choices and investment priorities requires decision support. System dynamics modeling (SDM) is one approach to systems engineering design support and support to identify priority investments.

Security is an enabler for fulfilling core mission by virtue of safeguarding against workflow disruptions. The implication to optimizing anomaly processing is the ability to reclaim resources allocated to security and focus them on core mission fulfillment. For example, people freed from dealing with vehicle manufacturing process anomalies may now engage in the production of the vehicles, or at least in the direct support of vehicle production.

The SDM focus is cybersecurity operations, specifically anomaly processing. The perspective is time to process anomalies and quantity of anomalies processed. The modeling approach is to establish a pre-proposed solution baseline of anomaly processing quantity and time, find the fit for a proposed solution, modify the SDM according to SME determined solution effects or preliminary test results, and produce post-proposed solution results in terms of anomaly processing quantity and time. The approach continues by comparing and analyzing the post-results to the pre-results, and provide conclusions and recommendations according to the findings. The case study herein uses a cybersecurity operations workflow consisting of eleven phases: *monitor*, *detect*, *characterize*, *notify*, *triage*, *escalate*, *isolate*, *restore*, *root cause analysis*, *recover*, and *organizational feedback*. Figure 2 presents these eleven phases with a depiction of the time measurements. Most phases include an OODA loop for that phase where the total time to process is $Observe_t + Orient_t + Decide_t + Act_t$. Even if automated at the Random Access Memory (RAM) level, there is still a measure necessary to determine if responses are within cyber-relevant time.

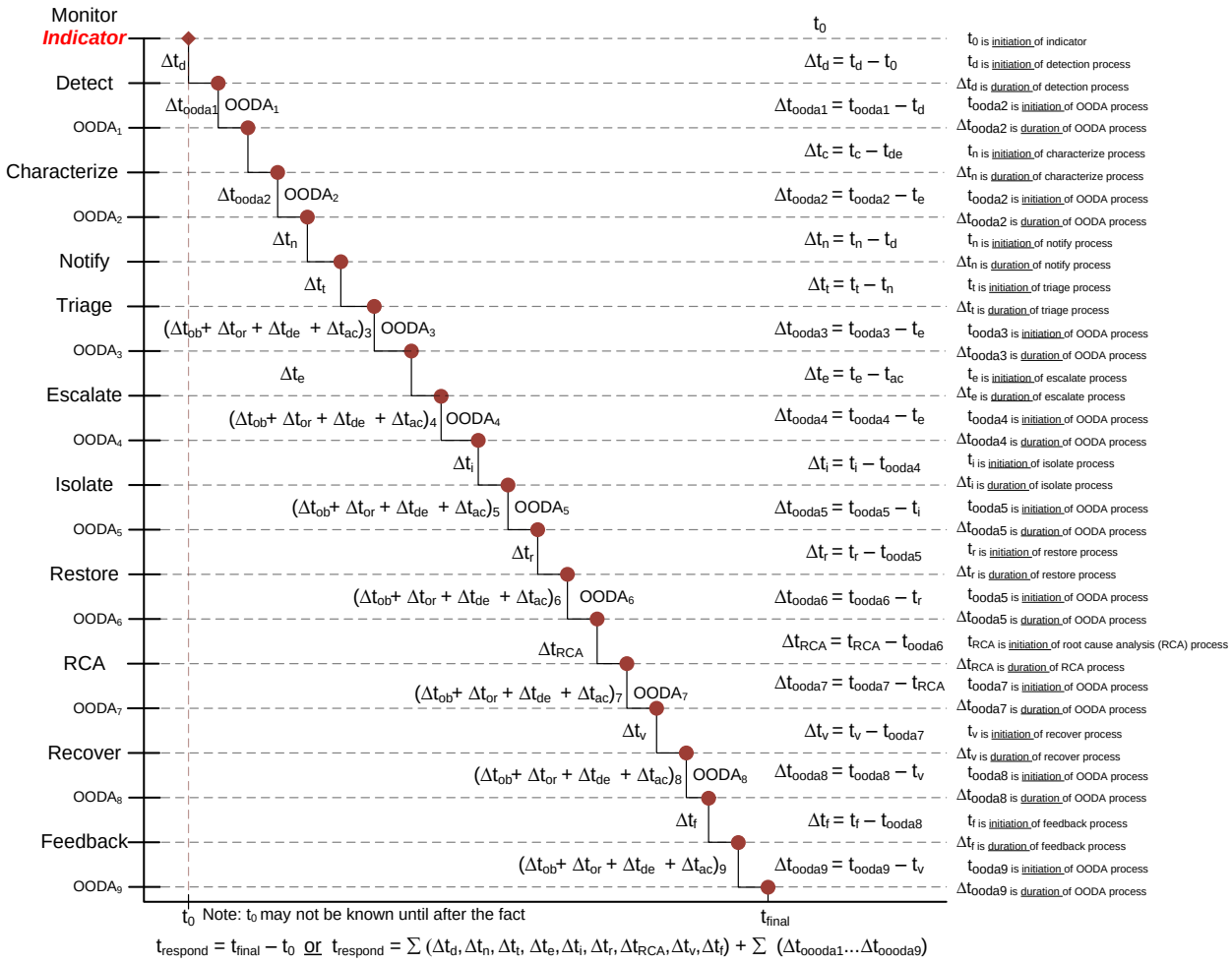

Figure 2. Cybersecurity Operations Time Metrics

Currently, humans are predominantly *in-the-loop* of cybersecurity operations, meaning a large part of workflow requires human monitoring, analysis, identification and selection among viable options, and manually initiating and/or performing tasks. As automation increases, humans become predominantly *on-the-loop* where they increasingly monitor and validate machine processing (Willett 2015b). To establish and sustain effective automation requires capturing and maintaining human knowledge for continual people review and affirmation. People-affirmed knowledge becomes the foundation from which to derive machine-encoded logic that then becomes *continually adapted automation* to reflect the latest best-known practices.

The SDM captures the ten workflow phases, their interactions among each other and among available machine and manual resources. The full SDM is not legible on letter-sized paper (the Annex provides an excerpt). Figure 3 shows a partial causal loop diagram (CLD) that represents the processing of anomalies for the Detect and Characterize phases. Successful manual processing of anomalies produces knowledge in the form of CDPs. CDPs are specific to each workflow phase; e.g., CDPs for Detect and for Notify that help practitioners become better at detection and notification (i.e., they help increase the capacity to detect, notify, etc.), thus accelerating workflow speed. This is a reinforcing loop of successful detection leading to knowledge production, which increases successful detection. CDPs provide for people-to-people knowledge sharing as well as provide a foundation from which to develop automation. The

combination of *people-enhanced cognition* and *machine-enhanced cognition* contribute to improving cybersecurity operations workflow efficiency (i.e., they affect the PDFs in one or more workflow phases).

There is an efficiency limit to manual processing where any person working at maximum capacity may only process a limited number of anomalies in a workday. Resource constraints (e.g., salary, office space) limit the number of people working in parallel. There are similar but different limits to machine processing, where any given machine may process some number of anomalies in parallel. An increase in algorithm efficiency or in processing power (e.g., CPU speed, memory) may increase the capacity of parallel processes per machine. Resource constraints (e.g., data center space, budget, trained administration personnel) limit the number of machines working in parallel. There is a fuzzy upper bound that limits improvements to efficiency. Efficiency improvement may continue at an overall diminishing rate of increase and then level out according to resource constraints that will vary per organization. The balancing loop for available resources and successful detection is one representation of the effects of resource capacity/efficiency limits. The same resource constraints apply to each workflow phase.

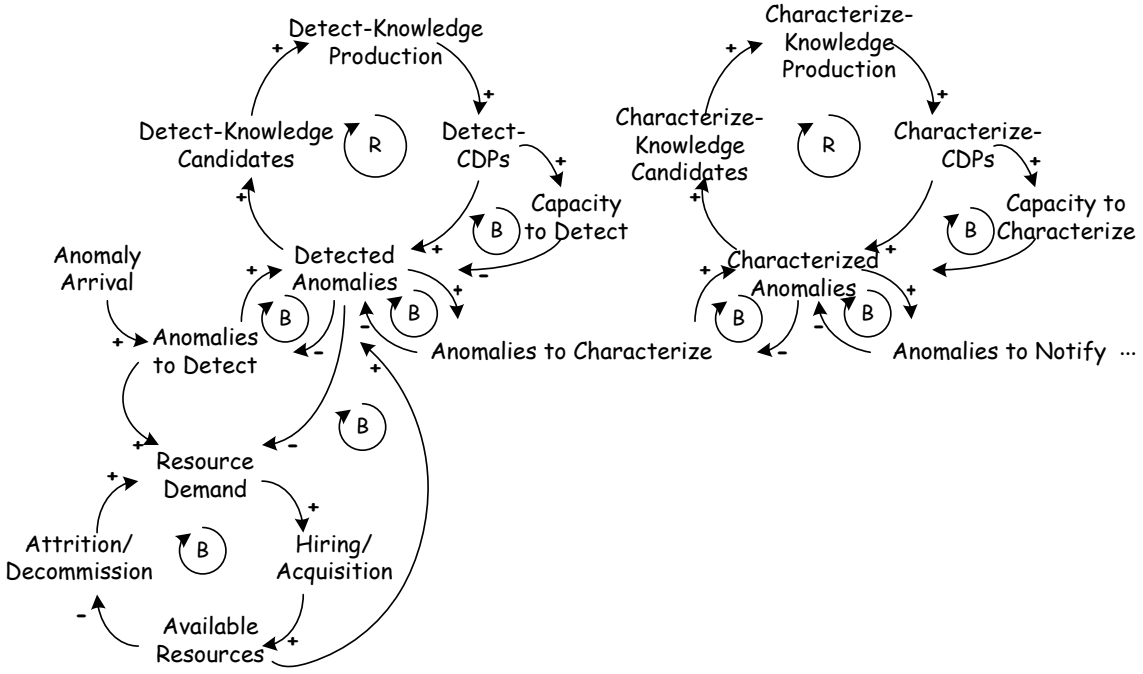

Figure 3. Partial CLD of Cybersecurity Operations

STELLA/iThink represents each workflow phase as a *conveyor* that takes some amount of service time to process an anomaly. Ultimately, the SDM should have dynamic feedbacks that modify service times according to knowledge production and application. This will show effects on the process as a result of CDPs and their influence on manual processing and machine processing over time. The anticipated effects are as people get smarter, they process anomalies they could not before (they become more effective) and process anomalies faster (they become more efficient). The STELLA/iThink software is unable to provide dynamic feedbacks to conveyors, so any CDP effects at this point are derived from a discrete run, modify the model manually, rerun, and the results compared to the baseline to determine overall effects. While still

valid in showing net effect, a continuous adjustment over time is preferable. Further SDM design consideration is necessary to investigate a workaround to this tool constraint.

Probability Distribution Function Determination Model

Systems engineers with experience in modeling understand the challenge to find data representing real world processing. For example, the actual time it takes to detect an anomaly and time to process that anomaly. The approach here uses an industry report representing real world experience from which to generate simulated data that in turn facilitates the generation of probability distribution functions to use in the SDM.

Each SDM workflow phase involves a process that takes some amount of time to complete. The approach to determine anomaly processing times for the SDM is to identify an appropriate scale, find root in reality, create probability distribution function (PDF) tables with basis in reality, generate simulated data according to the PDF tables, and perform goodness of fit tests on the simulated data to calculate the most appropriate PDF for processing times. The resulting PDFs are then used in the SDM to simulate real-world anomaly processing.

Table 1 provides a scale for workflow phase probability distributions that provides groupings for performances that are less than a minute (seconds), less than an hour (minutes), less than a day (hours), less than a week (days), less than a month (weeks), less than a year (months), and up to 2 years (Ponemon_Institute 2014). The numeric representation, unique to this model, is in *minutes*; i.e., any performance in terms of seconds is provided in a fraction of a minute, any performance in terms of hours is provided as a number of minutes greater than or equal to 60, but less than 1440 (the number of minutes in a 24 hour day), and so on. The scale in Table 1 provides ranges of minutes for capturing real world performance of the cybersecurity operations workflow phases.

Table 1: Anomaly Processing Scale

# Minutes	Label	Description
1	seconds	less than a minute (<60 sec)
60	minutes	less than an hour (<60 min)
1440	hours	less than a day (<24 hours)
10080	days	less than a week (<7 days)
40320	weeks	less than a month (<4 weeks)
483840	months	less than a year (<12 months)
1051200	years	up to 2 years (<2 years)

Table 2 provides one example of a PDF table for manual detection, or mean time to detect (MTTD). P(x) represents the individual probabilities, Cum(x) represents the cumulative probabilities (note: Cum(x) must always sum to 1 (100%)), and Time (minutes) is the scale shown in Table 1. The source for the Detection PDF table P(x) is the *Ponemon Cost of Data Breach Report 2014*, Figure 8 (Ponemon_Institute 2014).[4] Similar tables exist for each of the nine additional phases representing *mean time to characterize*, *mean time to notify*, *mean time to*

[4] The Ponemon report data contains an *unknown* category representing 10% of responses. Since the cumulative probabilities must always sum to 1, Table 2 accommodates the unknown responses by distributing the percent unknown equally among the other responses.

triage, *mean time to escalate*, *mean time to isolate*, *mean time to restore*, *mean time to root cause*, *mean time to recover*, and *mean time to feedback*. For the workflow phases with no corresponding real world data, there are P(x) estimates verified and validated by SMEs as likely reflecting real world scenarios. P(x) is easily adjustable to reflect variations in operating environments. The same PDF determination method applies to generate locally relevant PDFs.

Table 2: Example Probability Distribution Function Table for Manual Detection

P(x)	Cum(x)	Time (minutes)
0	0	1
0.122	0.122	60
0.256	0.378	1440
0.322	0.7	10080
0.178	0.878	40320
0.089	0.967	483840
0.033	1	1051200

Table 3 provides an excerpt of probability distribution ranges derived from producing a random number (e.g., r(Det) for Detect) and then performing a lookup using the PDF tables as described in Table 2. For example, comparing the r(Det) of .12539634 to Cum(x) in Table 2 results in a MTTD of 60 because the random number is greater than .122 but less than .378. Similarly, the r(Det) of .83770346 results in a MTTD of 10080 because the random number is greater than .7 but less than .878. The PDF determination model produces 5,000 rows of random numbers for each of the ten workflow phases. This produces 5,000 data points representing *ranges* of MTTD in terms of less than a minute, less than an hour, less than a day, and so on up to less than 2 years.

Table 3: Probability Distribution Ranges Excerpt

r(Det)	Detect (MTTD)	r(Char)	Characterize (MTTC)	r(Not)	Notify (MTTN)	r(Tri)	Triage (MTTT)
0.12539634	60	0.99450602	1440	0.50902739	1	0.82893511	60
0.83770346	10080	0.76424075	60	0.02691813	1	0.38211664	1
0.55457905	1440	0.55108253	1	0.50498849	1	0.85421571	60
0.98757692	483840	0.5296778	1	0.12907582	1	0.48586495	1

Figure 4 provide graphs of the probability distributions as derived from the full set of 5,000 probability distribution ranges in Table 3. In Figure 4, each vertical axis is a quantity of anomalies and each horizontal axis represents the scale in Table 1. While these diagrams are good for a general idea of the distributions, it is not specific enough for the SDM. The next step in the approach produces explicit PDF formulas for each workflow phase in the SDM.

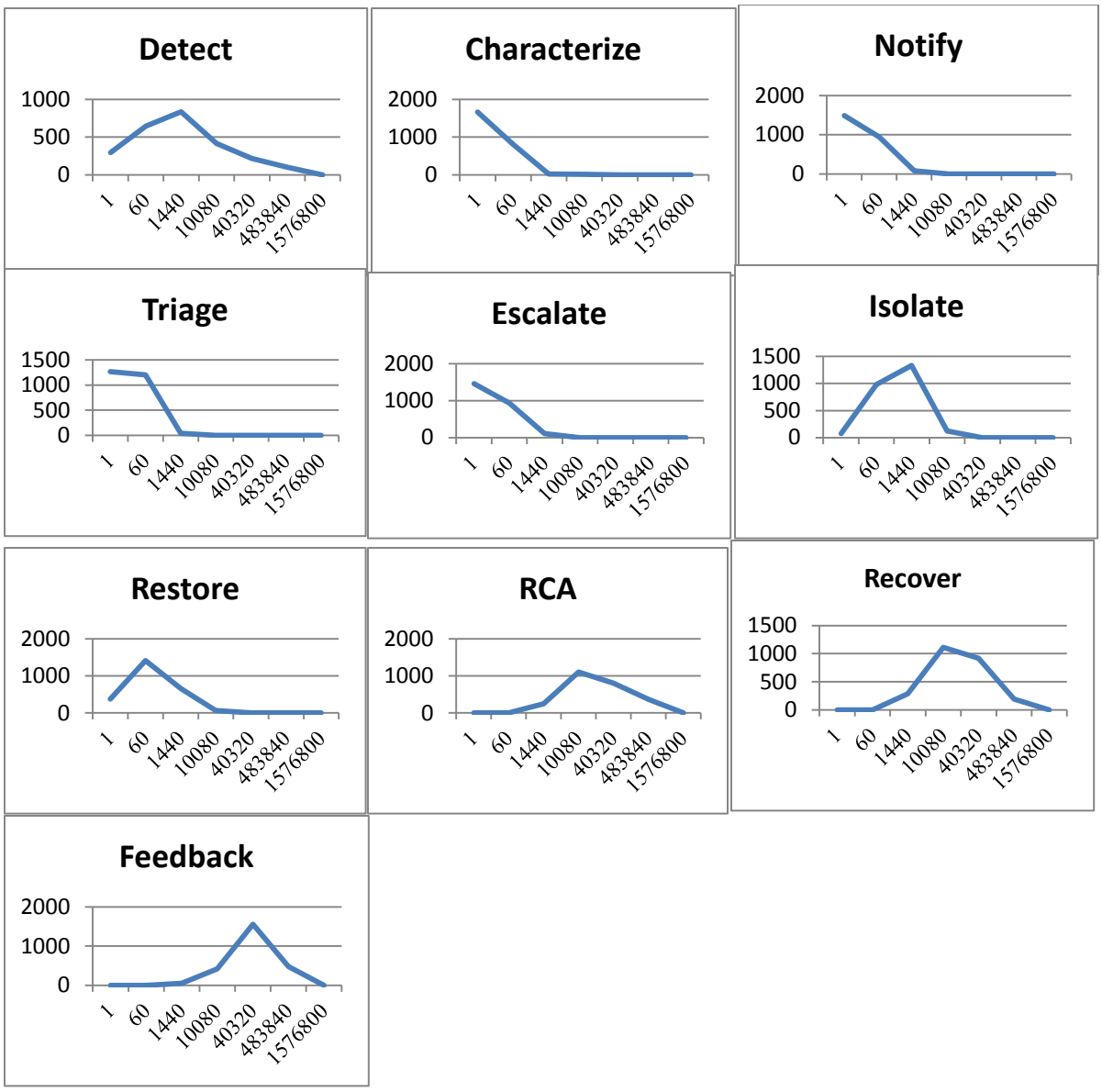

Figure 4: Probability Distribution Range Graphs

The ranges in Table 3 provides an interim step from which to generate simulated raw data necessary for goodness of fit tests that in turn produce the PDFs necessary for the SDM. This raw data is a table (not shown) of uniformly distributed random numbers within the ranges in Table 3. For example, the first MTTD range is 60 representing detection in less than an hour but more than a minute. For each entry of 60, this next step produces a random number between 1 minute and 60 minutes to represent a variety of Detect times that are less than an hour. Similarly for 10080, this next step produces a random number between 1440 minutes and 10080 minutes to represent a variety of Detect times that are less than a week. This represents simulated raw data for workflow processing times with basis in reality. Note: uniform distribution within the ranges is an assumption. The PDF determination approach is flexible enough to accommodate other than uniform distribution given data to support such a variation.

To generate probability formulas accurate enough for the SDM, the PDF determination model uses Oracle Crystal Ball[5], a statistical package add-on to Microsoft Excel[5]. Crystal Ball is configured to run 25,000 trials to produce raw data on which to perform goodness of fit tests to determine the best probability distribution representing processing times for each phase. Figure 5 provides an example for the Detect phase where the best PDF is Weibull with a scale of 1,869 and a shape of .33264. The resulting Stella/iThink SDM formula for manual detection is WEIBULL(.33264, 1869). An important distinction is the SDM has only one arrival time and that is for anomalies. All other PDFs are for processing time (i.e., service time) within the ten workflow phases.

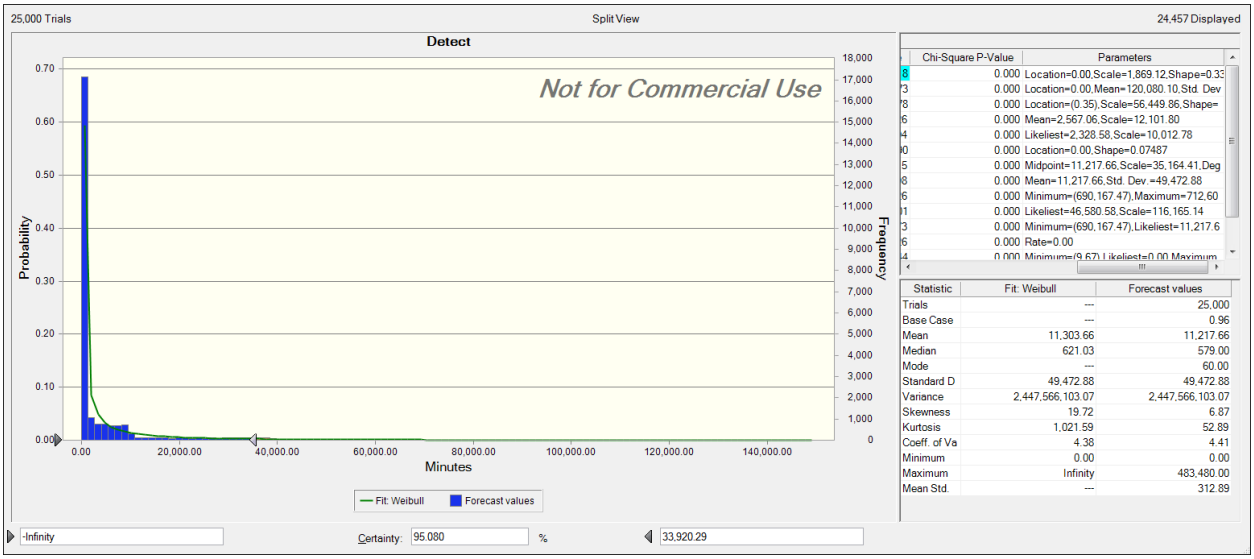

Figure 5: MTTD (Manual) Goodness of Fit Test Results

Table 4 presents the baseline PDFs (pre-proposed solution) for each workflow phase to use in the SDM. The table provides for both manual and machine processing times where machine processing is assumed to be 80 percent faster than manual. Each of the formulas contains a third numeric parameter that acts as a seed for random number generation; i.e., the same stream of random numbers is generated during each run. This facilitates comparing results of multiple runs varying PDFs by eliminating the effect of different random number streams. The current SDM baseline results have machine processing turned off, thus producing results for manual processing only with the assumption that manual processing produces knowledge that eventually translates into future machine processing. Machine processing is shown here as one of many future directions for the SDM.

[5] Certain commercial software products are identified in this paper. These products were used only for demonstration purposes. This use does not imply approval or endorsement by INCOSE or Stevens Institute of Technology, nor does it imply these products are necessarily the best available for the purpose.

Table 4: Baseline PDFs for the SDM

Phase	Pre-<proposed solution> PDFs	
	Manual	Machine
Detect	WEIBULL(.33264,1869,922)	LOGNORMAL(41961.42,58278966.6,923)
Characterize	LOGNORMAL(20,280,924)	LOGNORMAL(1.68,5.14,925)
Notify	LOGNORMAL(40,751,109)	LOGNORMAL(2.33,9.1,110)
Triage	LOGNORMAL(53,920,227)	LOGNORMAL(2.59,10.73,228)
Escalate	LOGNORMAL(53,1252,228)	LOGNORMAL(3.06,14.77,229)
Isolate	WEIBULL(.56032,432,112)	LOGNORMAL(468.59,21153.18,113)
Restore	LOGNORMAL(1117,35405,1962)	LOGNORMAL(108.72,4431.36,1963)
RCA	UNIFORM(41, 483856,1938)	NA
Recover	UNIFORM(41, 483714,1966)	UNIFORM(0, 482793,1967)
Feedback	UNIFORM(43, 483621,1935)	NA

Estimating Prospective Solution Effects

Any proposed solution's prospective effects on processing times are reflected in the probability distribution function table for one or more of the ten workflow phases according to the fit of the proposed solution. For example, a prospective effect of a 5% improvement to manual Detect (Table 2) would shift the range of P(x) by moving 5% from each P(x) to the next lowest row (current assumption). In some cases, the full range of P(x) may not shift and not all the shifts may be equal. After the shifts take place, new PDFs are then calculated using the process above that generates simulated data for a goodness of fit test to recommend the most appropriate PDF for the SDM. The point is to be methodical with rationale as well as flexible to adjust the method according to variances in operating environments.

For proof of concept, Table 5 presents an arbitrary range of potential effects from 1% to 10% for manual processing times specifically for the proposed solution of cybersecurity decision patterns. The SDM may reflect any combination of these speculative effects to determine overall net effect of any proposed solution. Meaning, if the proposed solution only promises some effect in the Detect phase and preliminary lab testing shows a potential improvement of between 5% and 10%, the SDM may be modified to use the Detect PDF for a post-CDP Detect improvement of 5% (LOGNORMAL(154200.65, 104251444.55,922)) and then run to produce the results. Then modify the SDM by substituting the PDF for 10% Detect improvement and then run to produce results. There are now two post-CDP sets of results with which to compare to the baseline to determine a range of overall net effects.

Table 5: Range of Post-CDP PDFs

Phase	Post-CDPs (1% improvement) Manual	Post-CDPs (2.5% improvement) Manual	Post-CDPs (5% improvement) Manual	Post-CDPs (10% improvement) Manual
Detect	LOGNORMAL(143551.89,82229861.89, 922)	LOGNORMAL(138023.37,78478463.84, 922)	LOGNORMAL(154200.65,104251444.55, 922)	LOGNORMAL(148223.09,107706259.16, 922)
Characterize	LOGNORMAL(19.68,264.5,924)	LOGNORMAL(18.27,237.19,924)	LOGNORMAL(18.24,243.6,924)	LOGNORMAL(16.35,204,924)
Notify	LOGNORMAL(39.73,743.74,109)	LOGNORMAL(37.73,699.41,109)	LOGNORMAL(35.17,622.00,109)	LOGNORMAL(32.59,574.06,109)
Triage	LOGNORMAL(50.50,841.15,227)	LOGNORMAL(51.01,864.12,227)	LOGNORMAL(45.39,734.13,227)	LOGNORMAL(42.55,701.07,227)
Escalate	LOGNORMAL(53.02,1249.91,228)	LOGNORMAL(49.89,1144.86,228)	LOGNORMAL(48.45,1104.96,228)	LOGNORMAL(40.34,843.40,228)
Isolate	WEIBULL(.55865,417.92,112)	GAMMA(.42373,1642.36,112)	WEIBULL(.54394,389.6,112)	GAMMA(.38757,1672.44,112)
Restore	LOGNORMAL(1107.69,35920.96,1962)	LOGNORMAL(1131.34,39243.71,1962)	LOGNORMAL(1231.24,50136.44,1962)	LOGNORMAL(1134.13,50499.44,1962)
RCA	UNIFORM(0,483512,1938)	UNIFORM(0,483840,1938)	UNIFORM(0,483700,1938)	UNIFORM(0,483797,1938)
Recover	UNIFORM(0,483858,1966)	UNIFORM(0,483854,1966)	UNIFORM(0,483798,1966)	UNIFORM(0,483802,1966)
Feedback	UNIFORM(0,483815,1935)	UNIFORM(0,483857,1935)	UNIFORM(0,483859,1935)	UNIFORM(0,483787,1935)

SDM Run Time Specifications

The SDM run time specifications are based upon real world encounters by a major East Coast academic research institution's cybersecurity operations. The run time configuration for the SDM is to process per minute for one year (525,600 minutes). Anomaly arrival time and all service times use units of minutes. Anomaly arrivals are a normal distribution between 1 and 2,599 per minute. Perimeter defenses are assumed to block between 67% and 97% of anomalies with those not blocked becoming a stock of *anomalies to detect*. The anomaly detection rate (Table 4 and Table 5) is a flow that controls the entry of anomalies into the workflow phases. The manual resources are assumed to be 55 people working 24x7 (see the *Future* section for prospective modifications to this assumption). Though machine processing is turned off for purposes of this paper, the model is designed to accommodate any number of machine resources that may process a varying number of anomalies in parallel depending on the complexity of the anomaly, the efficiency of the algorithm, and the power of the machine (i.e., CPU, memory, clock speed). This sets up the potential for dynamic feedback over time where machines may become more efficient as we produce better algorithms and technology refresh introduces machines with greater processing power.

Estimating CDP Effects

A supposition for the case study is that the generation of knowledge leads to enhanced performance. The approach is to quantify some amount of knowledge generation and then determine the effects of that knowledge on cybersecurity operations performance. This includes

determining which workflow phases are affected and by how much in terms of shifting the PDF tables and rerunning the PDF determination model to generate new PDFs to reflect the projected effects in the SDM.

The SDM runtime results include knowledge generation. The assumption is that all processed anomalies become knowledge candidates, though not all knowledge candidates are processed into actual knowledge. *Patterns* by definition must meet some level of invariance in real world experience. A single knowledge instance becomes a knowledge candidate and may even be useful to share. However, it takes at least two subsequent encounters to constitute a CDP and CDPs are necessary from which to begin machine encoding to ensure machines will address repeated phenomenon. As the CDP repository grows, subsequent related knowledge candidates may reinforce its status as a pattern, but additional codifying is not necessary by virtue of the CDP already existing. Moreover, knowledge and CDPs decay as they reach the end of their useful life as technology changes and adversaries adapt their tactics, techniques, and procedures.

The initial run of the SDM produces some amount of knowledge that in turn effects people performance by virtue of people-to-people knowledge sharing. That same knowledge provides a foundation from which to develop machine processing. The development and realization of machine processing takes time and emerges over the coming months and years. Machine processing then takes on rote task processing, thus freeing up manual resources for cognitive intensive tasks that in turn lead to additional knowledge. The feedback cycle continues as people work more efficiently at anomaly processing and machines become effective at anomaly processing. Limits to efficiency growth act as an upper bound for improvements where efficiency either grows asymptotically to this limit or oscillates between some upper and lower bound close to a maximum.

SDM Post-CDP Results Analysis

As shown in Table 6, the quantity of anomalies processed increases in most of the workflow phases as improvements increase. This was generally predictable and expected. For 2.5% in the isolate and restore phases, the quantities actually went down. In some cases, the bottlenecks are shifted from one phase to another as local improvements push more anomalies to latter phases that may increase the number of anomalies that take a long time to process, thus decreasing the number of anomalies actually processed in that phase. This shifting of bottlenecks is also expected in general and one benefit of the model is to help identify these types of consequences over time. Note: for the sake of brevity, the results here include only known-knowns (KK), which are anomalies we have seen before and know what to do about them, and manual (Man) processing. Other types of anomalies and machine processing are not included.

Table 6: Results Analysis (Quantity of Anomalies)

Run	KK Characterized Man Total	KK Notified Man Total	KK Triaged Man Total	KK Escalated Man Total	KK Isolated Man Total	KK Restored Man Total	KK RCAed Man Total
Base	1,398.30	1,398.13	1,397.93	1,397.64	172.24	171.90	90.65
1.00%	1,473.80	1,473.74	1,473.66	1,473.58	179.24	178.94	95.01
2.50%	1,444.26	1,444.15	1,443.94	1,443.70	175.90	175.36	96.95
5.00%	1,511.24	1,511.02	1,510.81	1,510.48	185.95	185.62	99.45
10.00%	1,776.47	1,776.35	1,776.18	1,775.88	217.65	217.25	117.79

Run	KK Characterized Man Total	KK Notified Man Total	KK Triaged Man Total	KK Escalated Man Total	KK Isolated Man Total	KK Restored Man Total	KK RCAed Man Total
Base							
1.00%	5.12%	5.13%	5.14%	5.15%	3.91%	3.93%	4.59%
2.50%	3.18%	3.19%	3.19%	3.19%	2.08%	1.97%	6.50%
5.00%	7.47%	7.47%	7.47%	7.47%	7.37%	7.39%	8.85%
10.00%	21.29%	21.29%	21.30%	21.30%	20.86%	20.87%	23.04%

Table 7 shows apparent reductions in efficiency as seen in the increase in mean times to process. The explanation for this is prospective improvements in the lower range (e.g., 1% or 2.5%) only make minor shifts in the PDF table (Table 2). As workflow efficiencies enter into the workflow phases, this begins to accelerate some (not all) processing times, which implies the ability to handle a greater number of anomalies. This in part results in a larger number of anomalies in high time processing phases, thus increasing the overall mean time.

Table 7: Results Analysis (Anomaly Mean Processing Times)

Run	Char KK Man	Noti KK Man	Tria KK Man	Esca KK Man	Isol KK Man	Rest KK Man	RCA KK Man	Reco KK Man	Feed KK Man
Base	20.46	36.59	50.61	47.12	691.09	835.41	167,908.13	169,164.61	130,950.15
1.00%	19.39	35.74	48.75	41.61	659.48	710.03	169,752.36	172,402.66	130,105.57
2.50%	19.80	35.07	47.92	66.93	700.82	836.99	168,163.12	170,674.75	129,303.08
5.00%	18.92	33.36	41.24	48.92	703.79	767.87	174,788.95	171,261.98	130,862.45
10.00%	16.04	31.07	42.20	39.58	655.98	852.81	170,105.83	171,888.20	132,641.70

Run	Char KK Man	Noti KK Man	Tria KK Man	Esca KK Man	Isol KK Man	Rest KK Man	RCA KK Man	Reco KK Man	Feed KK Man
Base									
1.00%	5.22%	2.34%	3.68%	11.69%	4.57%	15.01%	-1.10%	-1.91%	0.64%
2.50%	3.25%	4.16%	5.32%	-42.03%	-1.41%	-0.19%	-0.15%	-0.89%	1.26%
5.00%	7.55%	8.83%	18.52%	-3.81%	-1.84%	8.09%	-4.10%	-1.24%	0.07%
10.00%	21.59%	15.09%	16.62%	15.99%	5.08%	-2.08%	-1.31%	-1.61%	-1.29%

In further elaboration, high processing time phases are bottlenecks in the overall workflow. Bottlenecks restrict not only anomaly processing for that phase, but also limit the number of anomalies able to be passed on to the next phase. As bottlenecks are opened up and allow a greater number of anomalies to be processed, this may exacerbate bottlenecks in downstream phases by now providing a greater number of anomalies potentially subject to high processing time, thus making the immediate effect look negative from a time of processing perspective. However, the overall performance is improving by looking at the number of anomalies processed overall.

Therefore, the process of improving cybersecurity operations workflow may be a multi-step, multi-year process that requires shifting attention according to the improvements realized in one phase and the downstream effects of processing a greater number of anomalies. This is a systems thinking challenge at its finest where continued justification for operational modifications requires a perspective on understanding how things influence one another within a whole including patterns of behavior *within* the whole, behavior patterns *of* the whole, and to *conceptualize local changes* in context of the whole.

Conclusion

The modeling activity provides a first step towards a SDM for cybersecurity operations as an example of a decision support tool for systems engineering design tradeoff analysis. The initial model provides an objective view of cybersecurity operations as one example of a sequential, multi-step, time-essential workflow. The structure, design rationale, and results analysis provide an example of systems engineering decision support. The literal model is adaptable to other cybersecurity operating environments. The model approach and method is representative of how other people-oriented workflows may be represented and analyzed for bottlenecks and process improvement.

The concept of knowledge management provides for people-to-people knowledge sharing and provides a foundation from which to encode machine processing. The specifics elaborate on cybersecurity operations. The predominance of anomaly processing is initially manual. Over time, this may evolve into machine processing. Similarly, anomaly processing is initially reactive. Over time, as the repository of indicators emerges along with quantified associations of cause/effect, the system may predict and preempt efficiency degradation or failure.

The model identifies performance in terms of quantity and time to process anomalies, the ability to analyze workflow, and identify areas in need of improvement. The SDM provides the ability to estimate the effect of a prospective solution to improve anomaly processing and helps the systems engineer view local effects and overall SoS effects of a proposed solution. The results of running the model show a shifting of bottlenecks in anomaly quantity processing as well as both positive and negative effects as part of the path to overall improvement. While "all models are wrong, but some are useful," (Box and Draper 1987) this SDM provides an objective view for systems engineers to evaluate prospective effects of solutions on anomaly processing, thus providing decision support for design tradeoff analysis and establishing investment priorities from the perspective of both quantity of anomalies processed and time of processing. Key findings of this activity for modeling and simulation of cybersecurity operations:

- the SDM is essentially a resource planning tool to:
    - identify bottlenecks in a workflow,
    - discern the effects of local modifications,
    - discern the systemic effects of local modifications;
- the SDM helps anticipate unintended consequences… not that any proposed modification is necessarily *invalid* or *wrong*, it may take multiple steps to realize across the board improvements; i.e., short-term degradation for long-term improvement
- established SDM as one approach to represent a people-oriented workflow, the constituent parts, their interrelationships, and interactions;
- used the SDM as one way to predict the effects of prospective solutions on workflow efficiency; e.g., identify where the prospective solution fits, determine local effects of the prospective solution, determine the systemic effects of the local effects;
- identified and elaborated on a method to represent real world operations in the SDM; i.e., the *PDF determination model* that is adaptable to any particular cybersecurity operations
- applied the results of the PDF determination model in the SDM to produce results based on real world experience;
- the SDM provides decision support for SoS (cybersecurity operations) design including resource acquisition and placement to optimize workflow in terms of anomalies processed;

- discovered that CDPs as a prospective solution in some cases improve local results, in some cases don't improve local results, in some cases shift bottlenecks downstream; however, the macro effect looks like efficiency improvements over time, albeit with some short-term negative effects;
- and, identified areas for SDM improvement as described in the next section.

Future

Future model enhancements are specifically focused on decision support to improve cybersecurity operations. Enhancements include using the results of the SDM as decision support to determine the initial focus for CDPs within the ten workflow phases; i.e. focus CDPs on the greatest need and potential impact areas for anomaly processing. Expand the use of the SDM to run for multiple years that includes the dynamic feedback of CDP production to influence manual and machine anomaly processing; e.g., machine processing turns on after the first year and grows over time according to knowledge production and development delays. Expand people-enhanced cognition production to include influence on machine-enhanced cognition. Add upper bounds to improvements with some oscillation within bounds around a *to be determined* maximum improvement target. Expand the model to include adversary tactics, techniques, and procedures (TTPs) and add people-enhanced cognition and machine-enhanced cognition decay according to adversary change in TTPs; i.e., knowledge is dated and has a useful lifetime. Add improvements to processing times over time; e.g., increase in knowledge implies some modification to processing times. Monetize the model by associating dollars or other relevant currency to processing times (both manual and machine). Modify the manual resources to work 8 hours per day versus 24 hours per day; accommodate a user choice for multiple shifts with X employees per shift.

Annex – SDM Excerpt

Figure 6 shows an excerpt representing manual processing of anomalies, which is complementary to machine processing (not shown). Manual processing ultimately produces knowledge in the form of cybersecurity decision patterns, which is one form of people-enhanced cognition within the joint cognitive system structure that is part of integrated adaptive cyberspace defense. Moreover, Figure 6 only shows the processing of known-knowns. The SDM is extensive when accommodating machine and manual processing of known-knowns, known-unknowns, and unknown-unknowns. The machine processing of unknown-unknowns uses CDPs created from known-knowns and known-unknowns to search for the same or similar patterns in the ever-growing unknown-unknown repository. The intent is to use machine processing to reduce the *perception* of that which we don't know and to focus people on that which we truly don't know.

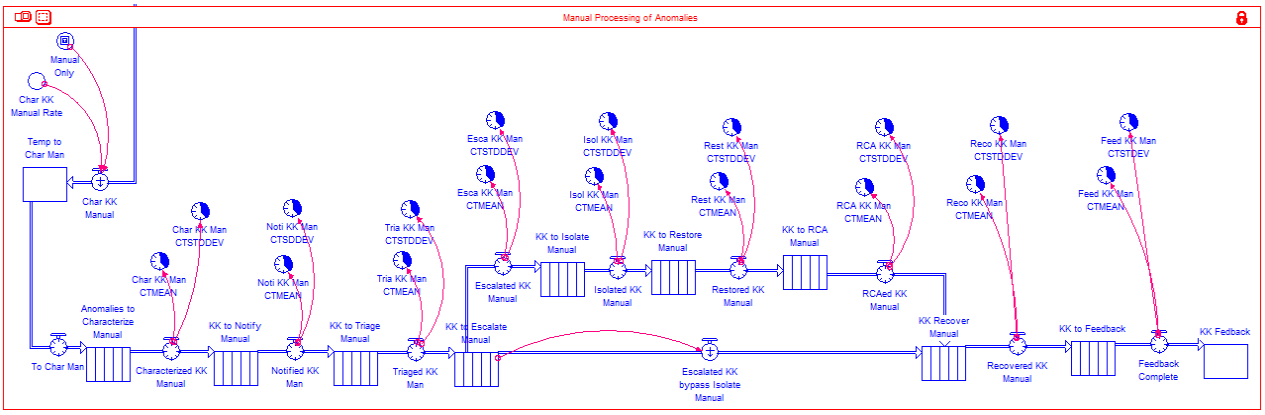

Figure 6. SDM Excerpt of Manual Processing

References

Box, George EP, and Norman Richard Draper. 1987. *Empirical model-building and response surfaces*. Vol. 424: Wiley New York.

Gharajedaghi, Jamshid. 1999. *Systems Thinking - Managing Chaos and Complexity*. Woburn: Butterworth-Heineman.

Herring, Michael J., and Keith D. Willett. 2014. "Active Cyber Defense: A Vision of Real-Time Cyber Defense." *Journal of Information Warfare* 14 (1).

INCOSE. 2015. *INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities v4, INCOSE*: Wiley.

Law, Averill M. 2007. *Simulation Modeling and Analysis, 4th Edition*. New York, New York, USA: McGraw-Hill.

Osinga, Frans. 2005. *Science, Strategy, & War - The Strategic Theory of John Boyd*. The Netherlands: Eburon Academic Publishers.

Ponemon_Institute. 2014. Cyber Security Incident Response: Are we as prepared as we think? : Ponemon Institute, LLC.

Sterman, John D. 2000. *Business dynamics: systems thinking and modeling for a complex world*. Vol. 19: Irwin/McGraw-Hill Boston.

Willett, Keith D. 2015a. "Capability-Based Engineering Approach to Integrated Adaptive Cyberspace Defense." Information Assurance Symposium, Washington D.C., July 1, 2015.

Willett, Keith D. 2015b. "Integrated Adaptive Cyberspace Defense: Secure Orchestration." International Command and Control Research Technology Symposium (ICCRTS), Annapolis, MD., June 2015.

Willett, Keith D., Rick Dove, and Mark Blackburn. 2015. "Adaptive Knowledge Encoding for Agile Cybersecurity Operations." INCOSE International Symposium, Seattle, WA, 13-Jul-2015 to 16-Jul-2015.

Biographies

Mr. Keith D. Willett has a BSc in Computer Science with Mathematics minor from Towson University (1984); an MSc in Business and Information Systems from University of Baltimore (1986); an MSc in Information Assurance from Norwich University (2005); and is a PhD candidate in Systems Engineering Security at Stevens Institute of Technology (est. 2016). Mr. Willett holds $(ISC)^2$ CISSP and ISSAP certifications and has over 30 years of experience in technology and security as an educator, programmer, database administrator, operations manager, systems engineer, enterprise architect, and enterprise security architect. Mr. Willett is the co-author of *How to Achieve 27001 Certification* and *Official $(ISC)^2$ Guide to the ISSMP CBK*; and sole-author of *Information Assurance Architecture* all published by Auerbach Publishing.

Mr. Rick Dove is an INCOSE Fellow and an adjunct professor at Stevens Institute of Technology teaching graduate courses in agile systems and systems engineering. Rick chairs the INCOSE working groups for Systems Security Engineering and for Agile Systems and Systems Engineering. He is CEO/CTO of Paradigm Shift International, specializing in agile systems and agile security R&D and education. As Principal Investigator (PI) he has led agile self-organizing system security R&D on US DHS and OSD funded projects. He was co-PI on the 1991 OSD funded Lehigh study that introduced the concepts of agile systems and enterprises, and led the subsequent DARPA-funded research during the nineties that established basic system fundamentals for agile systems of all kinds. He is author of *Response Ability* (Wiley 2001).

Dr. Robert Cloutier is an Associate Professor of systems engineering in the School of Systems and Enterprises at Stevens Institute of Technology. He has over 20 years of experience in systems engineering & architecting, software engineering, and project management in both commercial and defense industries. Industry roles included lead avionics engineer, chief enterprise architect, lead software engineer, and system architect on a number of efforts and proposals. His research interests include model-based systems engineering and systems architecting using UML/SysML, reference architectures, systems engineering patterns, and architecture management. Rob holds a B.S. from the US Naval Academy, an MBA from Eastern College, and his Ph.D. in Systems Engineering from Stevens Institute of Technology.

Dr. Mark R. Blackburn is an Associate Professor with Stevens Institute of Technology and primarily responsible for research focused on methods, modeling, simulation, visualization, and automated tools for reasoning about computer-based systems. He is the Principal Investigator (PI) on a Systems Engineering Research Center research task sponsored by NAVAIR investigating the most advanced and holistic approaches to model-centric engineering, and co-PI on a related task for Quantitative Risk. He has also been the PI on research tasks for the National Science Foundation, Federal Aviation Administration, and National Institute of Standards and Technology.