



Mathematical Literacy as a Fundamental Component of Training Cyber Warriors: Moving Beyond the Tools and Towards Fishing

Lieutenant Colonel Jonathan Fox and Major Samuel Huddleston

*This paper was initially presented at the Information and Cyberspace Symposium II held
22-24 September 2008 at the Combined Arms Center, Fort Leavenworth, Kansas.*

INTRODUCTION

Several years ago, two officers in the intelligence community were discussing web-based activities. The younger officer with less time in the community said, "Sir, I've been thinking about this new tool. Have we considered ...?" After listening for a few minutes, the officer with more experience in the community replied, "Thinking? Who has time to think? We take the tools they give us and brief what we collect."

Today's Warfighter is busy. On average, a Soldier will spend 1 out of every 2 years "downrange" – an environment filled with complexity and uncertainty. A Soldier involved in cyber-electronic operations returns from downrange missions to face a daily "routine" of complexity and uncertainty while conducting the full spectrum of Computer Network Operations (CNO). These cyber warriors have a variety of tools at their disposal for accomplishing their mission. Unfortunately, as the cyber-electronic environment changes, the cyber warriors have to adapt to new tools and new technologies.

Our proposal is not to develop another tool that will solve one technological problem. Rather, our proposal is to develop within our operators the ability to think critically and creatively and to exploit technology as an enabler.¹ While the buzzwords of "critically and creatively" are used often in the Joint, Interagency, Intergovernmental, and Multinational (JIIM) communities, we offer a unique solution for aiding the development of these skills for our officer, enlisted, and civilian cyber professionals. We recommend the study of the language of mathematics. For the average cyber warrior, learning the language of mathematics allows one to extend their natural thought process into a structured method for solving problems. Mathematics is a way to organize and analyze what is happening around us. Mathematics is a way to quantify both the randomness and trends of events or data collected. Mathematics is a way to see and understand. Without understanding the underlying processes that generate results from technology and software systems, and especially the assumptions made by the mathematical models applied, it is not always possible to interpret the results accurately or apply a tool effectively.

¹ Macgregor, *Transformation Under Fire*.

An extensive review of the literature for mathematical literacy and the cyber-electronic operational environment is beyond the scope and focus of this paper. Previous work by those associated with the United States Army Information Operations (USAIOP) and the United States Army Computer Network Operations-Electronic Warfare Proponent (USACEWP) provides a more in-depth look at some of the research and commentary on the complex operational, cultural, and technical issues involved with regards to the cyber-electronic operational environment. However, the following section establishes a working definition of mathematical literacy and provides a quick review of some of the mathematical components of cyber-electronic operations. Section II also introduces the simulation used to create the intrusion data set presented in the hypothetical applications. Section III gives both a simple and a moderate application of a cyber warrior using mathematical literacy to better understand his assigned sector of the cyber-electronic operational environment. Finally, Section IV offers two proposals for integrating mathematics into current training activities and briefly reviews ten mathematical skills that can serve as our cyber warriors' mathematical foundation.

BACKGROUND

Cyberspace, Cyber Warriors, and Mathematics

So what is mathematical literacy? Several academic organizations have developed definitions that focus on narrow components such as balancing a checkbook to broader statements that encompass visualization skills. For the purpose of this paper, we settled on a definition that is both broad but includes elements of both theory and functionality:

Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments, and to engage in mathematics in ways that meet the needs of the individual's current and future life as a constructive, concerned, and reflective citizen.²

A quick internet search of mathematics and cyber-electronic operations returns a vast array of published research concerning the role of mathematics in both cyberspace and the broader electromagnetic spectrum. Each of the purposed topics for this symposium requires more than an elementary knowledge of mathematics for those cyber warriors who want to truly understand the tools and technology used in each sector of the cyber-electronic environment. From the statistics behind intrusion detection³ to the mathematical models that form the basis for electronic warfare⁴, mathematical literacy provides the cyber warrior the ability to see first, understand first, and act appropriately.

2 (OECD, 1999). The reader is encouraged to examine (Lange, 2007) and (Paulos, 1988) for both a serious and light-hearted view of the consequences of mathematical illiteracy and means for overcoming.

3 (Kazienko & Dorosz, 2004)

4 (Vakin, Shustov, & Dunwell, 2001)

Simulating Cyber Intrusions

The motivation for this research is the application of mathematical literacy to the challenges faced by today’s (and tomorrow’s) cyber warrior. However, the availability and sensitivity of data from current cyber operations presents some challenges for both research and publication. This study proposes the use of a simulated data set of cyber intrusions as a surrogate for the empirical data that might be available to a cyber warrior acting as a cyber defender.

Following previous work on cyber crime⁵, we created a simulation with four distinct agents in order to replicate hypothetical intrusion data. As seen in the table below, each agent (or group) has a distinct preference based on the three given variables: time of intrusion, complexity of site, and value of the information stored on a site.

Table 1. Preferences of Simulated Agents

Agent	Number of Intrusions	Time Period	Complexity	Value
A	900+	0000 – 0800	High	High
B	600+	N/A	Med	Med
C	900+	0800 – 1600	N/A	Low
D	600+	1200 – 2000	Low	N/A

Using a statistical software package, we created over 3,000 simulated intrusions. Simulating the intrusions over time gives us a basis for examining changes to the preferences for complex sites and high value sites. With the assumption that past activity is a good predictor of future activity, we can use a clustering technique to examine the preference structure of actors and identify possible hierarchical structure.

Comments on Analysis

Three theories from the field of environmental criminology suggest that a criminal (or actor) has a set of preferences that are taken into account in deciding where and when to execute a crime (or event). *Intelligent site selection* applies these ideas by using statistical analysis to identify the features that make up the criminal’s preferences in order to cluster group activity or provide input to additional models.⁶ One of the tools cyber defenders can use to accomplish this task is *data mining*. The output of the data mining process is a mathematical model that can be used to predict the nature of future events.⁷

5 (Brown & Gunderson, 2001)

6 (Medby & Glenn, 2003)

7 (McCue, 2007)

SAMPLE APPLICATIONS

Now that we have dusted out the cob-webs of our mathematical attics, let us examine two hypothetical scenarios where a cyber warrior can apply his or her mathematical literacy and knowledge of statistics to better understand operations in the cyber-electronic environment.

Comparing Intrusion Rates

Joe is a cyber warrior for a unit within the larger JIIM community. Recently, his unit has experienced an increase of intrusions across their assigned sector of the grid. His data set provides several time periods of data containing over 3,000 intrusions.

Table 2. Hypothetical Intrusion Data Set Involving Government and Commercial Systems

Time Period	Number of Intrusions	Number of Intrusions (Commercial)	Number of Intrusions (Government)	Proportion of Government Intrusions
1	296	114	127	0.43
2	118	45	63	0.53
3	129	49	71	0.55
4	129	61	62	0.48
5	135	60	69	0.51
6	192	86	89	0.46
7	196	61	127	0.65
8	177	48	117	0.66
9	161	69	86	0.53
10	168	71	81	0.48
11	162	59	96	0.59
12	165	64	82	0.50
13	183	88	95	0.52
14	334	119	213	0.64
15	317	164	151	0.48
16	309	145	162	0.52
17	360	155	200	0.56

While some time periods appear to differ significantly in the proportion of government targeted intrusions, Joe’s Commander wants to know if the unit is identifying an increase in the number of attacks on government systems. Joe goes home and gathers up his high school mathematics books. He recalls learning of means, null hypothesis, and random samples.⁸ He decides that what he needs to focus on is the following:

⁸ A complete review of basic statistical analysis is beyond the scope of this essay. But the interested reader is encouraged to examine (Wackerly, Mendenhall, & Scheaffer, 2002), (Walpole & Myers, 1972), and (Ross, 2005).

$$H_o : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 < 0$$

where p_2 is the proportion of intrusions involving government systems in time period 4 and p_1 is the proportion of intrusions involving government systems in time period 16.

Joe wants to test the hypothesis that the proportion of intrusions involving government systems has not changed over the test period. A failure to reject the null hypothesis would support the case for neither an increase nor a decrease in intrusions involving government systems (steady state). A rejection of the null hypothesis would support further investigation into research on the complexity of terrorist incidents throughout the study period. According to some researchers an increase in the complexity of intrusions or incidents might indicate a strengthening of a core group or a concentrated effort against a particular sector.⁹ On the other hand, an increase in the number of “normal” intrusions might indicate a growing field of participants. Joe decides to use the following test statistic:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

The rejection region is given by $Z < Z_{0.95} = -1.645$. Computing to find the observed value of the test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = \frac{.52 - .48}{\sqrt{\frac{(.52)(.48)}{309} + \frac{(.48)(.52)}{129}}} = 0.76$$

Since the value does not fall in the rejection region, Joe fails to reject the null hypothesis and concludes that sufficient statistical evidence does not support the claim that the cyber-electronic environment is experiencing an increase in the number of government related intrusions. While this failure to reject the null hypothesis does not provide a definitive statement on the increase of cyber players or attacks in general, it does support Joe continuing his investigation by examining the patterns of intrusions over the study period.

Mining Known Intrusions for Hierarchical Clusters

As Joe moves into his next phase of analysis, he returns home and gathers up his and college probability and statistics books. He recalls learning of event probability, temporal analysis, clustering, and maximum likelihood.¹⁰ Though he can’t recall all the details, some quick review highlights a connection between his classes on spatial choice theory and his classes on statistics. If he accepts the theory that actors have a set of preferences that are taken into account when deciding the location and time to commit an act, Joe should be

⁹ (Enders & Sandler, 2000)

¹⁰ Again, a complete review of statistical analysis is beyond the scope of this essay. But the adventurous reader is challenged to examine (Everitt, 1993), (Moye, 2008), and (Gelman and Hill, 2007).

able to determine the cyber actor’s *intelligent site selection* process using *clustering*. Clustering is a method of using algorithms to group objects based on perceived or identified similarities. Joe can use data mining of the historical incident data to discover each actor’s or group of actors’ preferences. The payoff for identifying preferences over traditional methods of analyst heuristics is to provide a statistical foundation to the field experience and serve as input for additional analysis and modeling.¹¹

Joe first examines the visual evidence offered by the data.¹² Looking at the temporal density, Joe can distinguish what appears to be two peak times for activity:

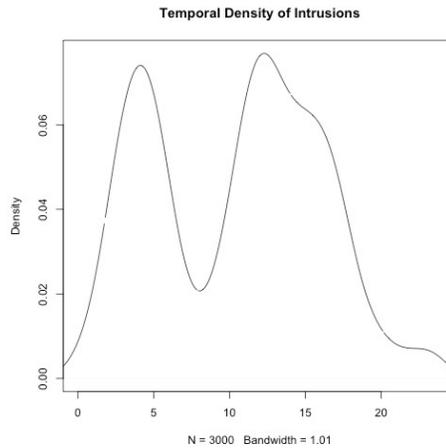


Figure 1. Density plot of events per time period

However, when Joe attempts to examine the scatterplot of all intrusions, he faces the challenge of examining multivariate data in a 2-dimensional plane.

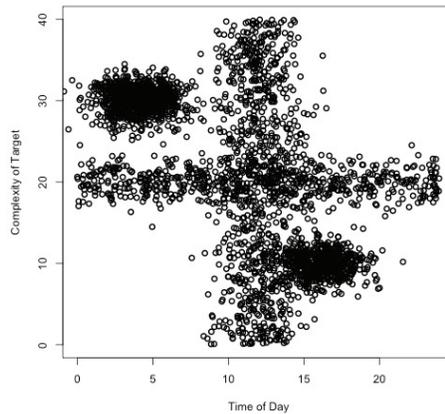


Figure 2. Scatterplot of simulated intrusions

¹¹ (Brown & Gunderson, 2001) and (Riese, 2001)

¹² (Tufte, 2001)

So how can Joe extract the features or preferences from a collection of event data? Joe decides to use clustering. Clustering is a method for organizing data or grouping similar objects together. While some might call this profiling, others prefer to think of this as applied multivariate statistics or data mining. Joe wisely decides to leave the ethical appropriateness questions to other agencies. For this application, Joe specifically decided to use a method known as Cluster Specific Saliency Discovery (CSSD)¹³. CSSD is a nine step process that enables the cyber warrior to group all intrusions that share like features across a space or sub-space. The result of this grouping identifies sets of preferences that may indicate a connection by a particular agent or actor.

Joe realizes the details of the CSSD processes are important but might not be interesting to his Commander. Using the following graphic, Joe explains CSSD as a form of projection pursuit.

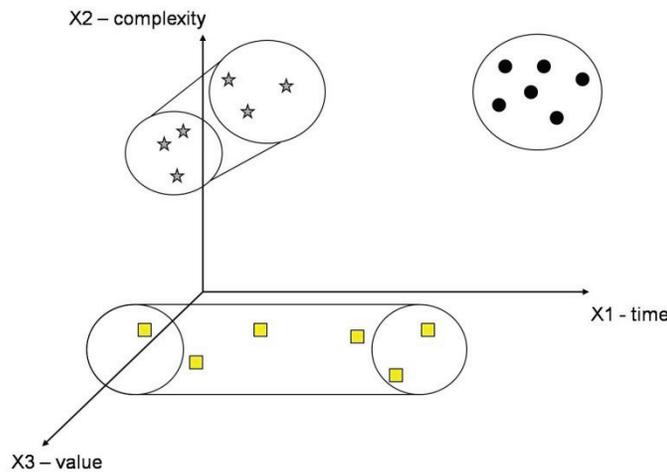


Figure 3. Graphical view of clustering

If Joe projects some sample intrusions points against a multivariate set of predictors, he can form clusters based on the minimum distance between observations across the space. Starting with all the data, Joe clusters across all dimensions and removes the grouping with the smallest variance. In this sample problem, the smallest variance would be the events late in the temporal blocks with high complexity and high values of the targets (circles). Joe continues the CSSD algorithm by removing the circle cluster and clustering again in only two dimensions. Now he removes the star cluster. Finally, Joe clusters against the remaining dimensions and removes the last cluster.¹⁴

Applying the CSSD against his intrusion dataset, Joe can now create the following clustered projection.

¹³ (Brown & Gunderson, 2001)

¹⁴ (Brown & Gunderson, 2001)

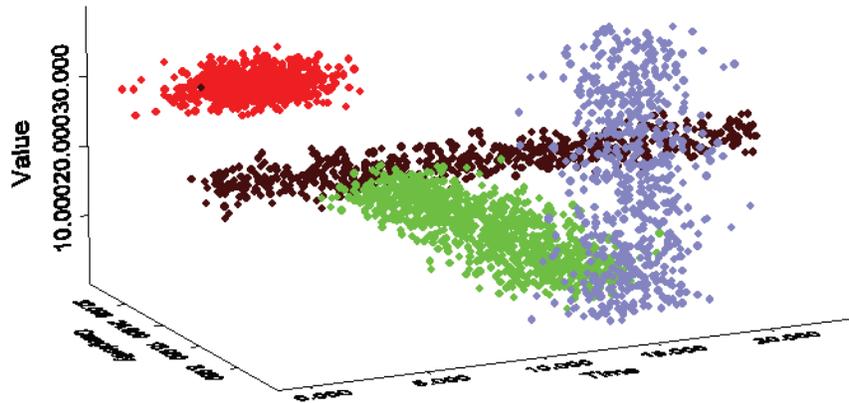


Figure 4. Clustered intrusions from simulated dataset

Joe can now clearly see four distinct clusters with varied preferences across each cluster. Clearly one group operates across all time periods while the group focused on complex, high value targets appears to operate exclusively in the morning hours. Joe can also determine that of the four clusters, three appear to have distinct preferences for activity time. The following table provides summary statistics for the identified clusters.

Cluster	Number of Intrusions	Time Period	Mean Complexity Score	Mean Value Score
1	899	2300 – 0800	30.1	30.0
2	601	0000-2400	20	20
3	900	0730 – 1730	20.1	10
4	600	1145 – 2130	9.9	19.5

Table 3. Summary statistics for intrusion clusters¹⁵

Joe reminds his Commander that clustering does not mean that only four groups are operating but that rather the intrusions can be associated based on similar patterns of activity. The results of clustering are built upon an assumption that past activity is a good predictor of future activity. Additional investigative work would allow Joe to make a stronger claim on hierarchical or group associations and forecast specific intrusions. But before Joe does that he decides to visit the library for some more statistics books.

¹⁵ As expected, the results from the clustering match the criteria used to establish the simulation. The reader is cautioned not to see this as perfection in the method or to discount the results based on the simplicity of the simulation.

RECOMMENDATIONS AND CONCLUSION

Teaching Them to Fish

Finally, we reach the second part of the title. As the old proverb says: “Give a man a fish and he will eat for a day; teach a man to fish and he will eat for a lifetime.” But what does teaching people to fish have to do with operations in cyberspace? If you have been following along you know the answer: Mathematical literacy is a fundamental skill that will enable our warriors to be successful in the cyber-electronic operational environment regardless of the technology available. Teaching our cyber warriors the basic mathematical constructs of the cyber-electronic environment offers the JIIM community the best opportunity for identifying, defending, and exploiting the gaps in technology today and tomorrow.

A recent industry white paper was titled *Making Analysis Relevant: It's More Than Connecting the Dots*. While we appreciate the efforts of our academic and industry partners, from our perspective the JIIM community currently focuses on providing our cyber warriors a lot of fish.¹⁶ These fish could be today's intrusion detection software or tomorrow's next generation cyber-electronic decision support tools. Each of these fish is necessary in the cyber operational environment but having the fish alone is not sufficient. Teaching our cyber warriors the basic mathematical constructs of the cyber-electronic environment offers the JIIM community the best opportunity for identifying, defending, or exploiting the gaps in technology today and tomorrow.

One method for incorporating mathematical literacy training for current officers would be to include a requirement for probability and statistics as part of the Expanded Graduate School program. Adding two courses of graduate study would not greatly increase the burden on the officer student. And the impact of understanding statistics and probability would reach across the Army's focused disciplines of operational skills, diplomacy, security, and governance.

A second method for incorporating mathematical literacy training for future cyber warriors would be to include a requirement for mathematical modeling and statistics as part of any Advanced Civil Schooling programs that support the cyber force. Additionally, as we continue to identify and select our future cyber warriors, we should consider past performance and education in the field of mathematics and or engineering.

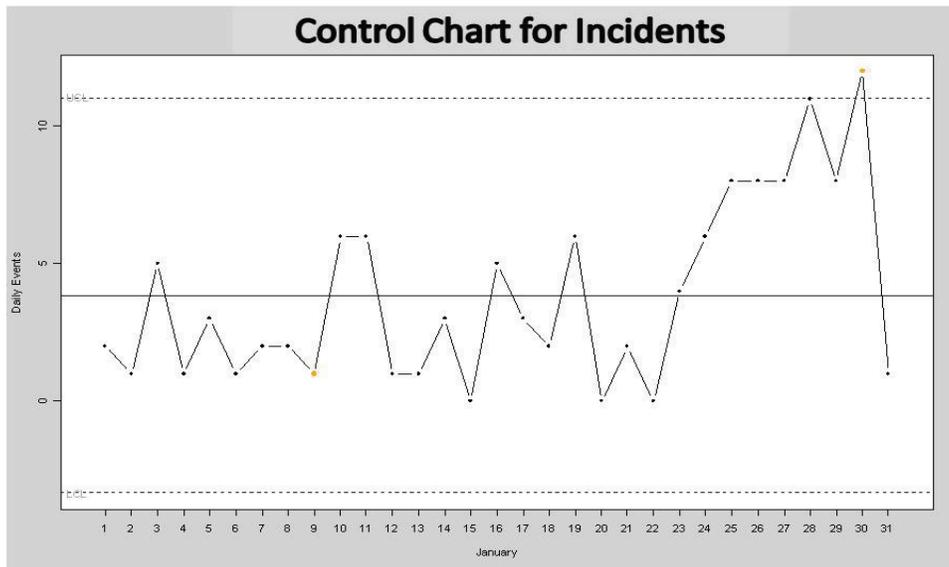
Fishing Skills

But what do we teach our future cyber warriors? The following ten skills provide a foundation in theoretical and practical applications of mathematics for the cyber warrior. The details provided for each skill are brief but the corresponding graphics and references for each skill offer the interested reader additional information regarding applicability to the cyber environment today and tomorrow:

¹⁶ AFCEA Intelligence Committee, April 2005. The authors recognize the focus given to “human systems effectiveness” in the white paper, but we still feel that those efforts miss the importance of actually making the analyst “smarter.”

1. Probability and Statistics. While not a single skill, it is appropriate for future cyber warriors to understand the elementary components of probability and statistics that form the foundation for the remaining skills.

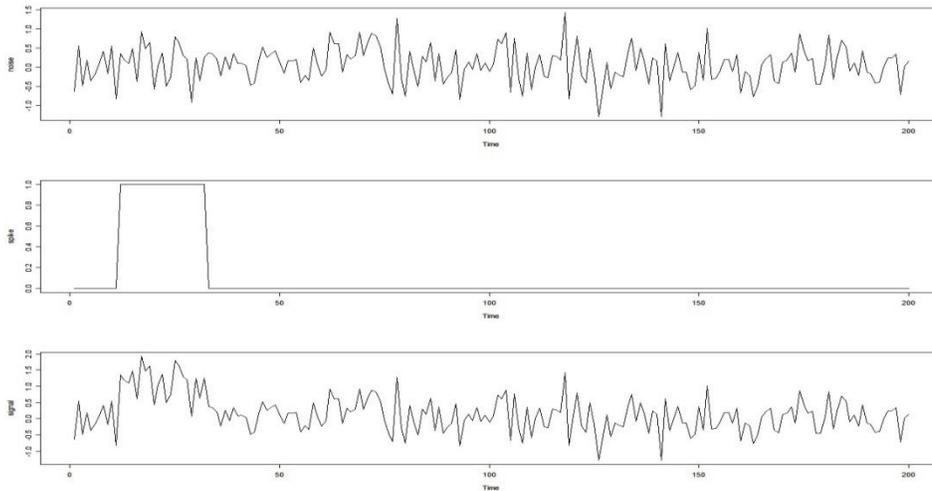
2. Change Detection. A process for using statistics to detect changes in the behavior of a system. Visual observation of simulated incidents during a period with a “pulse event” can be enhanced using a Shewhart control chart to highlight the numerical outlier of 12 events on 30 January.



While the increase in events leading up to and on the pulse event is not unexpected, additional analytical work could focus on determining possible causation given similar patterns in real world data. And similar to the work done with environmental data, the cyber warrior might incorporate multivariate control charts as a means of examining the daily network activity in order to more rapidly identify “out of control” situations.¹⁷

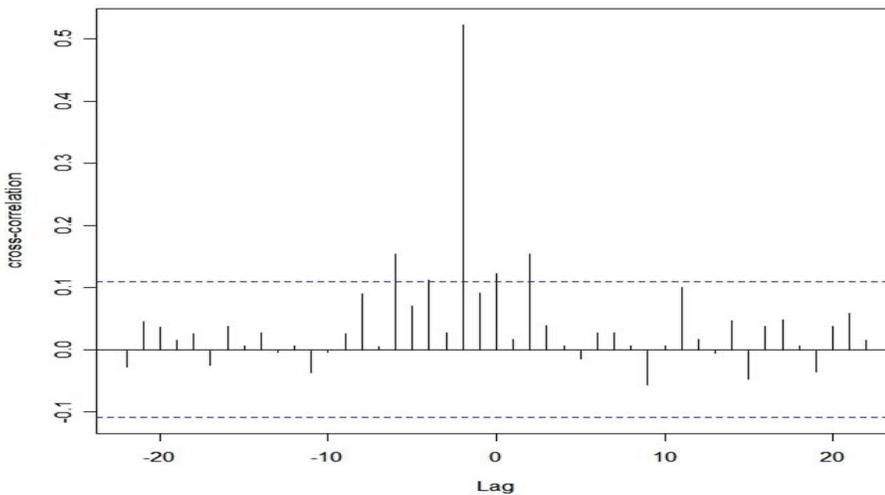
3. Pattern Analysis. A method of examining data against know patterns or from extracted information.

¹⁷ Davis, Ginger, and K B Ensor. (2006).



Superposition of a random noise (top) with the signal amplitude (middle) that results in the composite signal (bottom) that might be observed across the Electromagnetic Spectrum. A cyber warrior that understands the probability distributions of a given noise source has a better chance of successfully jamming a target signal or “clearing the clutter” from a desired transmission.¹⁸

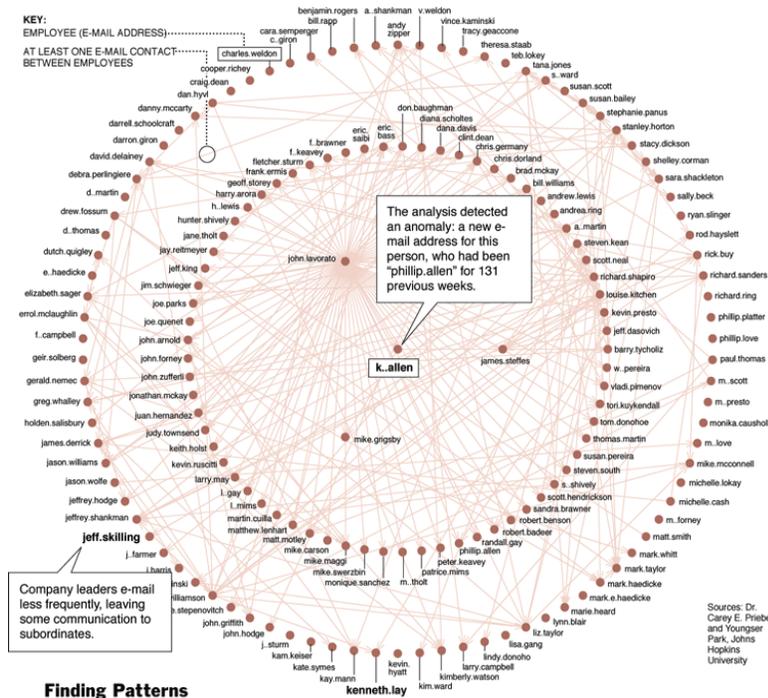
4. Temporal Analysis. Examining data as a time series supports signal processing and change detection.



Temporal analysis gives the cyber warrior a method for determining correlation between two data sets. The spike in the graph above indicates a positive correlation between two simulated datasets over time.¹⁹

¹⁸ Schlesinger, R. J., et al. (1979).
¹⁹ Enders, W., & Sandler, T. (2000).

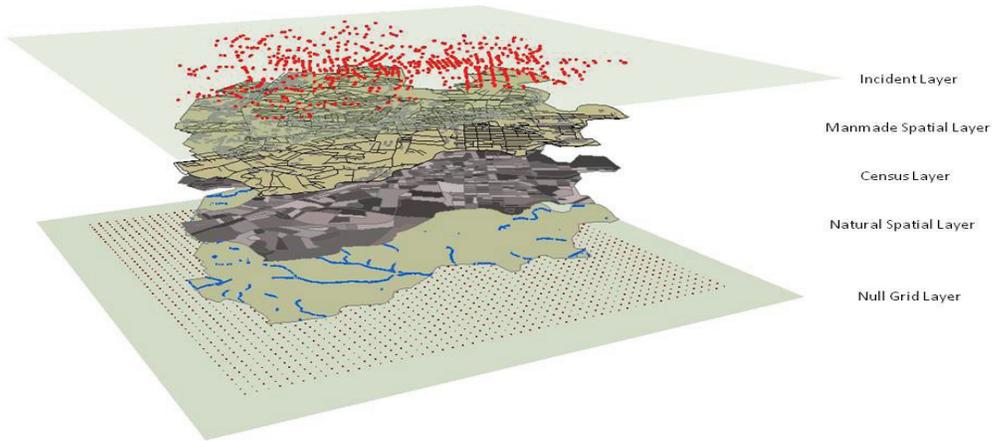
5. Link Analysis. Examining nodes based on known or discovered relationships.



The graphic above depicts the link analysis resulting from a treatment of emails as graphs. A cyber warrior can use similar techniques to monitor network activity of groups and individuals.²⁰

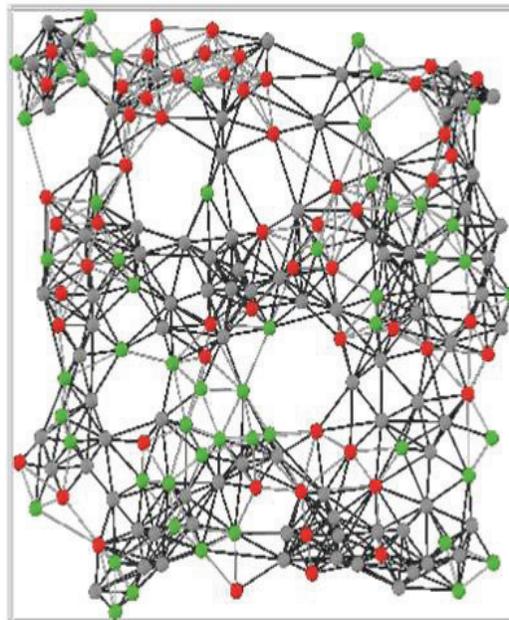
6. Spatial Analysis. Analysis based on the geographic location and relationships of the observed data.

20 Priebe, et al (2005). Graphic from the New York Times (available on-line at http://graphics8.nytimes.com/images/2005/05/21/weekinreview/nwr_KOLATA1_CHART.gif).



Spatial analysis allows the cyber warrior to understand patterns across space for specific activities. The graphic above depicts the Feature-Space methodology for identifying spatial choice patterns related to an actor's site selection decision. Feature-space uses past activity of an actor or group to determine the probability of future events in a given space.²¹

7. Modeling and Simulation. Using mathematical models to gain understanding of asset interactions.²²



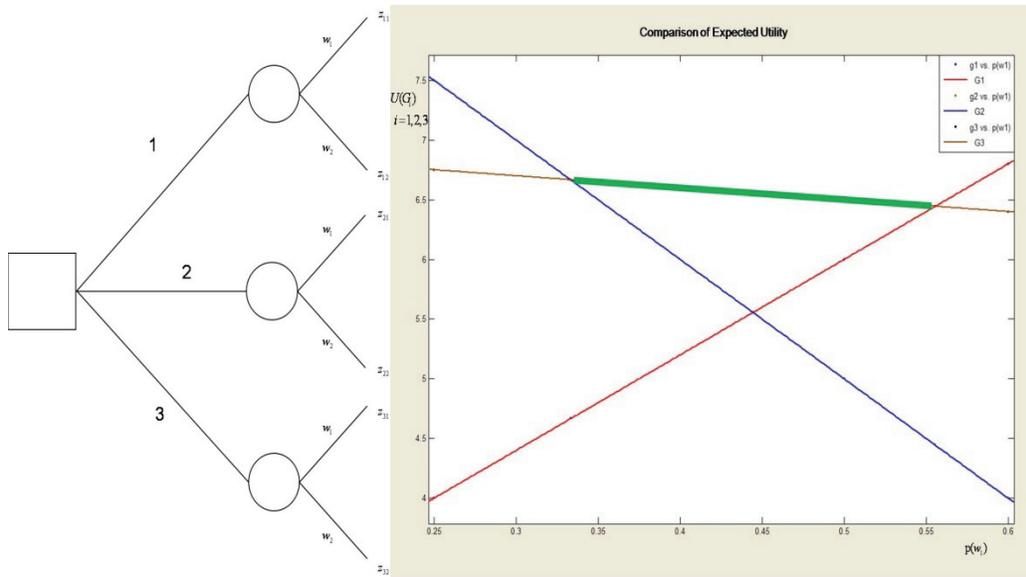
²¹ Brown, D.E, Dalton, J., an

²² Adamy, D. (2001).

Modeling and simulation provides a method for cyber warriors to examine virus spread on a network with different topologies and information assurance policies.²³

8. Clustering. A generic term used to describe numerical methods for identifying groups within an observed data set.²⁴

9. Decision-Making. Determining the rational decision given a choice of options, results, and associated probabilities.²⁵



The decision tree on the left and the resulting expected utility comparison offers a cyber warrior a formal method for evaluating courses of actions using varied probabilities assigned to outcomes.²⁶

10. Data mining. Again, data mining is not a single skill, but a process for using mathematical methods for finding information and knowledge within data.

23 Stonedahl, F. and Wilensky, U. (2008).

24 Everitt, B. S. (1993).

25 Riese, S. R. (2001).

26 Lindley, D.V., (1985).

Learning for a lifetime

We live in an increasingly complex world. The changes in the cyber-electronic environment are rapid and continuous. While our cyber warriors must have the best tools and technologies we can provide, our success in cyber-spaces depends on our ability to have the best command of known facts and the best predictions of the unknown facts. Mathematical literacy is the core knowledge and skill that enable this understanding and these forecasts. Educating and training our cyber warriors in mathematics does more than provide a tool; it teaches them to fish.

Lieutenant Colonel Jon Fox is an active duty Army officer currently assigned to Fort Gordon, Georgia. He is also pursuing a PhD in Systems and Information Engineering from the University of Virginia. Commissioned as an infantry officer, he has deployed in support of operations in both Iraq and Afghanistan.

Major Sam Huddleston is an active duty Army officer assigned to the United States Military Academy. He recently obtained a Masters Degree in Systems and Information Engineering from the University of Virginia. Commissioned as an armor officer, his most recent operational assignment was as a planner for one of the Stryker brigades while deployed to Iraq.

REFERENCES

- Adamy, D. (2001). *EW 101: A First Course in Electronic Warfare*. Boston. Artech House.
- Brown, D.E, Dalton, J., and Holye, H., (2004). "Spatial Forecast Methods for Terrorist Events in Urban Environments" in *Proceedings of the Second NSF/NIJ Symposium on Intelligence and Security Informatics, Lecture Notes in Computer Science*, Springer-Verlag Heidelberg.
- Brown, D. E., & Gunderson, L. F. (2001). Using Clustering to Discover the Preferences of Computer Criminals. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* , 311-318.
- Davis, Ginger, and K B Ensor. (2006). "Outlier detection in environmental monitoring network data: an application to ambient ozone measurements for Houston, Texas." *Journal of Statistical Computation and Simulation*, 407-422.
- Enders, W., & Sandler, T. (2000). Is Transnational Terrorism Becoming More Threatening? A Time Series Investigation. *The Journal of Conflict Resolution* .
- Everitt, B. S. (1993). *Cluster Analysis*. Edward Arnold.
- Kazienko, P., & Dorosz, P. (2004, Jul 23). *Intrusion Detection Systems (IDS) Part 2*. Retrieved sep 1, 2008, from www.windowsecurity.com.
- Lange, J. D. (2007). *Mathematics for Literacy*. Mathematical Association of America.
- Lindley, D.V., (1985) *Making Decisions*, John Wiley & Sons.
- McCue, C. (2007). *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. New York: Butterworth-Heinemann.
- Medby, J. J., & Glenn, R. W. (2003). *Street Smart: Intelligence Preparation of the Battlefield for Urban Operations*. Rand.
- Moye, L. A. (2008). *Elementary Bayesian Biostatistics*. Boca Raton : Chapman & Hall/CRC.
- OECD. (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: Organization for Economic Cooperation and Development.
- Paulos, J. A. (1988). *Innumeracy*. New York: Hill and Wang.
- Priebe, et al (2005). *Scan Statistics on Enron Graphs*. available at HYPERLINK "http://research.cs.queensu.ca/~skill/proceedings/priebe.pdf" <http://research.cs.queensu.ca/~skill/proceedings/priebe.pdf> .

- Riese, S. R. (2001). *Estimating the Probability of Landmine Contamination in a Non-Combat Environment*. Charlottesville: University of Virginia.
- Ross, S. (2005). *First Course in Probability*. Prentice Hall.
- Schlesinger, R. J., et al. (1979). *Principles of Electronic Warfare*. Los Altos. Peninsula Publishing.
- Stonedahl, F. and Wilensky, U. (2008). NetLogo Virus on a Network model. HYPERLINK “<http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork>” <http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press.
- Vakin, S. A., Shustov, L. N., & Dunwell, R. H. (2001). *Fundamentals of Electronic Warfare*. Artech House Publishers.
- Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2002). *Mathematical Statistics with Applications*. Duxbury.
- Walpole, R. E., & Myers, R. H. (1972). *Probability and Statistics for Engineers and Scientists*. Collier-Mac.

**UNITED STATES ARMY COMBINED ARMS CENTER
FORT LEAVENWORTH, KANSAS**

The NEXUS is published by the USACEWP, a directorate of the Combined Arms Center (CAC) at Fort Leavenworth, Kansas. Permission is granted to print single copies of this article for personal, non-commercial use. Requests for reprint of articles in other publications, lesson plans, or similar multiple copy venues must be submitted via e-mail to: USACEWPStratcomm@conus.army.mil; or in writing to the following address: USACEWP, 950 Bluntville Ave, Ft Leavenworth, KS 66027. Also, it should be noted that The NEXUS does not fact check statements or reputed sources in documents published. The responsibility for the accuracy of information asserted, sources, and (where appropriate) security clearance for release is deemed the personal and professional responsibility of the author.