National Cybersecurity Institute Journal

Volume 3, No.1

ISSN 2375-592X

National Cybersecurity Institute | 2000 M Street, Suite 500 | Washington, D.C. 20036
Excelsior College | 7 Columbia Circle | Albany, NY 12203-5159

NATIONAL CYBERSECURITY INSTITUTE JOURNAL

Volume 3, No. 1

PEER REVIEWERS

The *National Cybersecurity Institute Journal* gratefully acknowledges the reviewers who have provided valuable service to the work of the journal:

NATIONAL CYBERSECURITY INSTITUTE JOURNAL

The National Cybersecurity Institute at Excelsior College is a research center based in Washington, DC, dedicated to increasing knowledge of the cybersecurity discipline and its workforce demands. Published three times a year, the peer-reviewed National Cybersecurity Institute Journal covers topics that appeal to a broad readership within the cybersecurity discipline, with a particular focus on education, training, and workforce development.

The manuscripts submitted to the journal are reviewed for their contribution to the advancement of applied research in the area of cybersecurity.

Submission guidelines for authors can be found at http://ncij.excelsior.edu/.

FROM THE EDITOR

Greetings and welcome to the first issue in Volume 3 of the National Cybersecurity Institute Journal. As the cybersecurity community is aware, our mission at NCI is to increase knowledge of the cybersecurity discipline, and assist the government, industry, military, and academic sectors to better understand and meet the challenges in cybersecurity policy, technology, education, and training. Much attention has been given lately to the role of cybersecurity in our national defense and the roles the next generation of cyber professionals must be prepared for. This edition of the journal provides informative articles that relate to those and other timely cybersecurity issues to better educate and inform our readers. NCI is proud to publish relevant and noteworthy articles that will serve to enlighten those with a vested interest in the cybersecurity field.

In this edition, Charles Parker gives us a view of online higher education from the important perspective of a professor. Audie Hittle discusses the capabilities of innovative data-driven Intelligent Data Storage (IDS), which is followed by work from Ezhil Kalaimannan and Caroline S. John who provide us with information concerning a security development life cycle framework for web-based applications. James Angle reminds us that, increasingly, medical devices have become a major point of attack in the healthcare industry and offers his work about managing the risk of those devices.

We have become increasingly aware of the important role that all sectors play in the defense of our cybersecurity. To that end, Wm. Michael Volk discusses the innovations community colleges are making to actively develop and refine their cybersecurity programs to increase the flow of skilled entry-level employees into the cybersecurity pipeline. In their offering, Ruth Agada and Jie Yan provide details on how they "…developed an animated agent that serves as a virtual commentator in small-scale cybersecurity competitions to educate and engage novice spectators…," which is a very interesting read.

Finally, we conclude with the work of R. David Parker and Michael D. Regier, who note the importance of cybersecurity to our national defense and offer their article "Vulnerability Risk Modeling: Combining Cybersecurity & Epidemiology to Enhance National Defense," which is a must-read for everyone with a vested interest in the cybersecurity of our nation and our critical infrastructure.

This edition of the National Cybersecurity Journal is, once again, timely in bringing you the latest information relating to cybersecurity and the important issues in that arena. The articles you find on these pages will provide you with insight that you will bring to the workplace, and instill in those you speak with a desire for further thought on the myriad of cybersecurity issues that swirl around us.

The National Cybersecurity Journal Journal is never the work of one individual but always a collaboration of dedicated individuals here at NCI whose hard work results in the informative product you have before you. My sincere thanks go to all the notable authors, the administration, and our hard-working staff for their outstanding efforts in bringing the National Cybersecurity Institute Journal to you once again. I hope that everyone in the cyber community will find this journal informative as you work within your respective cyber areas. As always, I look forward to your comments, suggestions, and future submissions.

Jane LeClair

Jane A. LeClair, EdD
Editor in Chief

Cybersecurity Matriculation: Analyzing Online Higher Education from the Professor's View

Charles Parker II

ABSTRACT

The industry of higher education has been in a state of flux for years. Originally the education was delivered via scheduled classes in brick-and-mortar buildings. At a point in time, the addition of distance learning was entered into the field. As technology advanced, online learning was utilized to further reach the masses yearning to achieve university degrees. A subset within the university degrees involves information technology (IT) and information security. The research on the topic included a survey with university professors seeking their opinion of the online learning process and its beneficial or detrimental aspects. The results indicated this mode of delivery is viable and beneficial to the students overall.

INTRODUCTION

The prior model of IT education involved the traditional brick-and-mortar buildings, with ivy grown well over the brick façade. There would be the professor standing in front of the class talking about the theory and application of this to the subject matter. This standard model had been in application for over a century and is well known.

The higher learning environment, as with most subjects, is not static. This model began to change as technology advanced. A symptom of this had been the computer systems evolving from rooms of wires and tape drives to a laptop or desktop performing the same tasks, which continues to be engineered to operate faster in a smaller physical size. The speed at which technology has advanced is an application of Moore's Law (Moore's Law, 2015). Some stated that the in-person classrooms were superior to the online mode, as the physical classroom students would achieve better grades than the online students (Harrell, 2008). The early thoughts also included researchers postulating the learning for the individual student would decrease (Kirtman, 2009). These theories did not come to fruition.

The increased use of the Internet, the better connectivity, and better quality of the equipment provided the framework for the paradigm shift of higher education. This has manifested itself with the advent and acceptance of online learning, also known as Web-based learning (Brown, 2012). The student sets aside time for class which fits within their schedule. The student may be at their home, the library, or at lunch during the workday and logs into the university's portal for the online learning platform. The university may use a number of various applications for the online courses, including but not limited to Moodle or Blackboard. The student chooses which class(es) to continue to work with. The student may take one or more courses during a semester. This coursework may take the form of research for a paper, downloading homework to complete, adding to discussions on a topic, or turning in homework or their research paper. This also allows the student to contact other students in the class to share ideas, work as partners on research papers as approved by the instructor, and other uses beneficial to the student's studies. This also allows the student to initiate and continue a social contact with students also in the university, which may be of assistance in later classes.

This newer avenue of learning has allowed many more students the option to matriculate and obtain bachelor's degrees and master's degrees who normally may not have this opportunity. The online form of education is only going to increase in usage (Anthony, 2012) and relevance for the universities due to its usefulness and flexibility. One academic discipline that has noticeably been present in this mode of education is IT. Other disciplines within the university would not be such a good fit due to their associated laboratory assignments, such as chemistry or physics. Within these two, as an example, it may be difficult to complete the laboratory assignments due to the equipment needed. A subset of the IT discipline, which is growing in importance, would be the Information Assurance and Security curriculum.

This new format methodology for the coursework, while providing a clear benefit for the students, has provided challenges for the professors. With decades of physical classroom experience, there were very few surprises in the traditional mode of teaching at this level. The professor arrived at the class, taught the class in front of the students, fielded questions, and graded the assignments and tests. The administration and professors did not have to think about the course framework significantly. The professor would know well ahead of time the courses to be taught, when the class would meet, the room in which the class would meet, the syllabus would be written and copied, and on the first day of class the materials would be handed to the students.

The new mode of delivery has provided opportunities and challenges for the professors and students alike. There are multiple platforms to teach from. Two of the many are Moodle and Blackboard. Each of these is materially different and requires time to become accustomed to its processes.

PROBLEM STATEMENT

The online format for students to matriculate has clearly become a viable option for the students to successfully enroll, complete courses in various concentrations, and graduate with their respective degree. This has also increased the revenue from the online tuition for the schools. This increase is from the number of students taking these classes and also moving away from other universities which do not have the online option.

Although there are positive attributes overall, the research to date on this topic had been focused on the student's responses and opinion. The student's responses may not be entirely without bias. Their grade may skew the response to the respective study. As an example, they may believe a higher grade was earned than what was received. The student's responses may also be skewed by the amount of time the student has available to study in comparison to the recommended amount of time to study. This may, in certain instances, have added stress to the student's life, which may adjust their responses from the normal responses. The research specifically directed to the information security courses and their benefits from the professor's view has been relatively sparse over the last few years.

The professor's view on the current state of online learning in the cybersecurity field provides an insight into the inner processes and workings. This may also provide a less altered assessment of the impact on the student as the professor is acting as a third person reviewing the data. The benefits and negative aspects with online cybersecurity university learning environment from the professor's view have not been adequately researched.

The research problem is centered on this. The project researched whether online learning for information security courses was useful for the students and what attributes would a successful student show in this case.

LITERATURE REVIEW

COMPUTER SCIENCE

The online higher education model has been applied to most academic principles. This is common with general business, accounting, computer programming, and others. The field has shown itself to be fluid and dynamic. Earlier, delivery was completed via the student logging in, gathering an assignment, and later emailing this into the professor. The technology has improved exponentially so that the delivery includes an active chat room between the students from around the planet, online quizzes, and remarks or comments for papers being provided on the original document.

As this was the environment, the research articles from 2012 forward were included. Any older research articles may not fully represent the findings and research at that time on the subject.

Ali and Smith (2014) researched computer literacy courses taken online in comparison to courses taken in person with the professor present. The measure included in the researcher's study was the student's performance. The research problem was whether, in the computer literacy course, the student performed better with the online courses or in person with the face-to-face interaction. The limitation on this research study was the data were only from one semester and the performance was measured using only the final grade in the class. There were other mitigating factors involved with this research study—for example, the student's background and prior online course usage—were not taken into account. The research indicated, between the in-person and online courses, that the students in the online courses did not perform or achieve a statistically significantly better score than the students in the face-to-face courses.

This differential was also researched in an online e-commerce programming course that took place in 2010 with a sample of 13 students during the spring and fall semesters (Arslanyilmaz & Sullins, 2013). The independent variable for this research study was in the spring semester class when the instructor interacted with the students at the end of the course's time period designated for discussion. The fall semester had the instructor interacting with the students in the minimum once in the middle of the course and again once prior to the completion of the period designated for discussion. The dependent variable was the students' level of message posting. An additional measure was the students' grades on their respective assignments. The research indicated there was not a statistically significant difference between the two groups or the students' interactions with the timing and number of interactions with the instructor. There was not an effect on the level of the student's participation or grades.

CYBERSECURITY

The computer science and computer information systems have a wide variety of subtopics in place. One of these and the subject of the research project is cybersecurity. Here also there has been a significant amount of research on this topic, as will be explored. This historically has been researched from the student's view and noted effect. The view of the professor of the process and students has not been researched.

One method of teaching cybersecurity involved utilizing workflow technology with case studies (He, Kshirsagar, Nwala, & Li, 2014). The case studies have been used to actively engage the students in the learning process. This methodology had proven to be beneficial to the students, in that they had the opportunity to apply the lessons from the case study to their workplace environment. This has been shown to provide a greater level of student involvement. The researchers studied this via introducing the Kepler workflow technology into the course framework. The workflow technology uses a graphical user interface (GUI) to bring the students together. A portion of the respondents favored the workflow technique as it divided the case study into steps, versus attempting to work on the whole at once. Overall the students had a positive experience with the new framework as it related to cybersecurity case studies.

VIRTUAL LABS

Teaching cybersecurity with a lab setting has been researched in varying capacities. One aspect reviewed this from the online laboratory implementation. Willems and Meinel (2012) researched this with a virtual lab they developed. This platform was name Tele-Lab, which provided a virtual environment for learning and training. There were also training exercises with the application. As the student completed the tasks, they were able to self-assess their advancement. The researchers noted that this had been implemented with positive results.

Additionally, Son, Irrechukwu, and Fitzgibbons (2012) researched virtual labs as a way to support online education. This provided an added layer of functionality as a vast number of students are able to learn directly from exercises in the lab. The researchers noted the importance of the traditional learning process, and emphasized the virtual lab as a supplemental learning tool to be integrated into the online process. The virtual lab also has shown itself to be very scalable. For instance, up to 300 users may be utilizing this tool at any particular time. With this many users, there were no connectivity issues. Although this has shown itself to be beneficial, there may not be an adequate level of support for certain levels of users.

Salah (2014) researched teaching cybersecurity and training in the cloud. The researcher, for this study, used the Amazon Web Server for the cloud provider. This cloud-based service bypassed the typical issues of the traditional mode, such as configuring and installation. This, coupled with the labor of a person teaching, be it a technician or instructor, only made this a lesser cost in comparison to the traditional mode. The researcher noted several exercises and applications to be used with this form of training. The research indicated cybersecurity is a topic well-suited for cloud learning as nearly all of the exercises for cybersecurity may be completed here. The students involved with the research indicated this was a positive and beneficial course of action. The students found this to be a motivator for them and a convenience. The instructor also found this as a positive experience for himself and a method to conserve time.

Cybersecurity is a growing field due to the global expansion and evolution of criminal activity and cyber espionage. To meet the challenge, cybersecurity professionals have had to adjust to the new environment. These adjustments include modifying the delivery of the cybersecurity courses, practical experiences, and training. This assists those students who are not geographically close to a campus and with limitations on time availability for the course or curriculum to attend (LeClair, Abraham, & Shih, 2013). The availability of online courses is beneficial for working students and allows these students to learn and grow as cybersecurity professionals.

METHODOLOGY

The subject of this research paper is the online cybersecurity matriculation at the university level for the students, as gauged from the professor's view. This study was qualitative in nature.

PARTICIPANTS

The participants for this research study were the professors teaching courses in an information technology doctoral program. The concentration in this program was information assurance and security. Ten professors asked to comment on this situation in the form of an email. Of these, six responded to the email. The sample is clearly much smaller than what would normally be utilized for a research project. This was due to the mitigation factor of the limited time frame. There was only one pass for the questions for the professors. A follow-up was not completed.

MATERIALS AND PROCEDURES

This was a very simplistic study in that there were two questions asked referencing the online format: (a) what were the instructor's impression of the online learning format and (b) what were the benefits of this format. The questions were emailed to the professors for their voluntary response.

RESULTS

The results for the research were manifested by the professor's responses. The responses were direct and elucidated the issue. The professors stated with online higher education at the university level in this academic area, there needs to be a greater focus on information acquisition on the part of the students. The students have to be motivated to gather the information in comparison to other learning methods where a person may provide the information directly. In the online mode of delivery, the student has to log into the university's website and find the location of the information on the website.

This mode of education delivery is more flexible than the traditional method. With these programs the student is able to log in at any time—day or night—to review assignments, enter and post discussion questions and answers, and check for the instructor's messages. The student can also log in from anywhere there is a viable Internet connection.

Because the model provides, in general, for a less subjective form of knowledge transmission—for example, readings, chat sessions, and email—in comparison to face-to-face gatherings, the professor should be aware of the need to have more individualized, personal contact with the students. The messages sent and received are presented in the course virtual room in a letter format, not generally in a verbal format. When these messages are posted in the course room in the verbal format, they are posted as listen-only files.

It is noteworthy that the professors remarked on the students' internal motivation. With the level of the program and the online option utilized, the student needs to put much more work into the course in comparison to the traditional brick-and-mortar school. The students, due to the format limitations and parameters, have to do more self-learning as there are not students proximate to them to directly and immediately ask questions. There is the opportunity for the student to attempt contact with the other students via email. This format tends to have the best results for students reading a book on a topic they would enjoy, which is a form of information gathering.

The online mode of delivery is best suited and engineered for the graduate students versus undergraduate students. In this aspect, the graduate students have been through the undergraduate courses and overall curriculum and appreciate how the education system works when applied to them. In comparison, the undergraduate students may not have the background and be accustomed to the university level work flow or amount of work that is required for an acceptable grade.

DISCUSSION

The research study responses were indicative of the present state of online learning at the university level in the cybersecurity curriculum. The learning environment is significantly different than that of the traditional method and has been upgraded as the technology would allow. The new model is much more flexible (Cain & Phillip, 2013) than the traditional brick-and-mortar mode. The students are allowed to attend classes at their leisure from anywhere there is an internet connection without connectivity issues. This convenience (Buckley & Narang, 2014) has brought many people to the online format (Bender, 2003). The online higher learning methodology is a benefit for the student as they are able to log into the course at any time. This allows those working unique hours or with other special requirements to matriculate and successfully earn the degree.

One respondent noted the online option is best suited for the graduate students. The graduate students have had the opportunity to learn how the undergraduate educational works and is applied to them. The graduate students generally would have a greater grasp of multitasking and time management, as they have already achieved one degree and are moving towards a graduate degree. However, this scenario may also be well-suited for the transfer student, who has an appreciation for how the system works from their specific experiences.

A common theme through the responses was that the students need to be self-motivated to take the courses and be successful. Because there is more work to complete with these courses alongside the lack of direct peer support, the students need to be motivated to complete the work. The lack of the face-to-face interactions with the professor and other students furthers the need for the student interested in self-learning. This lack of interaction also does not allow the student to have the benefit of social cues (Brown, 2012). Without this drive, the student may have issues with completing the course or achieving a grade indicating success.

FURTHER RESEARCH

As noted, this was not an exhaustive research study of the topic. The sample size was small, yet held opinions and data that are focused and are representative of the underlying framework of this newer mode of online learning. However, this is certainly a topic deserving of further research and analysis. The online learning environment is a completely viable mode of learning. This has grown from the original model of distance learning to fully integrating technology to bring the student into the virtual classroom and improving the user experience (UX). The future and present students continue to be exceptionally busy with work and their personal lives and other uses of their time. This trend is not going to cease. The need for quality online education will only continue to grow. This topic is certainly worthy of a much larger study, which would generate a much larger sample.

Further research on this topic is warranted as online learning grows to be a larger part of the academic community. As time passes, more attentiveness will be applied to this topic. Online learning is not a simple fad that will fade away as time passes. Further topics for future research may include expanding the sample and research time frame. A larger set of responses would certainly be more beneficial and prudent for a robust research project. This would also provide for a wider array of responses to analyze. Additional professors from the various institutions may also note the different aspects from their perspectives and add a deeper understanding of this new learning model.

The students' withdrawal rates within this program may add new insight into the paradigm. As noted, the online higher education format provides a layer of flexibility. This is a positive attribute for the students, as it accommodates their needs. A negative bi-product of this flexibility may be the ability of the student to withdraw or drop all of their classes via a simple log-in and selection to drop courses the student previously signed up for. The online format may provide less of a personal attachment to the university as there is a lack of face-to-face contact with the professor and other students. This ratio may also be compared to the withdrawal rate from the traditional brick-and-mortar universities. Similarly, the effect of social interactions with other students in the same university should be researched as this may have a distinct effect on the students and their willingness to continue with the class.

In the short-term, the students are focused on completing the course work and the degree so they may enter the workforce or improve their position in the marketplace. A longitudinal study researching the success rates of persons graduating with the IT degree and comparing the online graduates to the traditional students would be of interest. One challenge would be defining "success." This could be measured extrinsically with the person's title at the time of the study, based on an increase in pay from one year post-graduation to the current year, or intrinsically with their job satisfaction. The time frame would be from 5 to 10 years post-graduation.

CONCLUSION

Historically, online higher education university usage has increased. This demand has been driven by the student's lack of discretionary time when the traditional classes would meet along with other constraints. From a survey of professors currently teaching in the IT doctoral program at an online university in the Information Assurance and Security curriculum, this mode of learning is beneficial. This allows the students to matriculate when otherwise they would not be able to. This is an especially positive choice for those students who are driven by intrinsic factors for success, such as self-motivation and a thirst for knowledge.

REFERENCES CITED

Ali, A., & Smith, D. (2014). Comparing student's performance in online versus face-to-face courses in computer literacy courses. *Competition Forum,* 12(2), 118–123.

Anthony, K.V. (2012). Analyzing the influences of course design and gender on online participation. *Online Journal of Distance Learning Administration,* 15(3).

Arslanyilmaz, A., & Sullins, J. (2013). The extent of instructor participation in an online computer science course: How much is enough? *Quarterly Review of Distance Education,* 14(2), 63–74, 121.

Bender, T. (2003). Discussion-based online teaching to enhance student learning: Theory, practice, and assessment. New York, NY: Stylus Publishing, LLC.

Brown, J.M. (2012). Online learning: A comparison of web-based and land-based courses. *Quarterly Review of Distance Education,* 13(1), 39–42.

Buckley, I.A., & Narang, H. (2014). A study: Exploring the feasibility of developing a computer science online degree program at Tuskegee University. *Higher Education Studies,* 4(3), 48–57.

Cain, M., & Phillip, S. (2013). An exploration of students' experiences in and online primary teaching education program. *Journal of Online Learning and Teaching,* 69(3), 304–315.

Harrell, I.L. (2008). Increasing the success of online students. *Inquiry,* 13(1), 36–44.

He, W., Kshirsagar, A., Nwala, A., & Li, Y. (2014). Teaching information security with workflow technology-A case study approach. *Journal of Information Systems Education,* 25(3), 201–210.

Kirtman, L. (2009). Online versus in-class courses: An examination of differences in learning outcomes. *Issues in Teacher Education,* 18(2), 103–116.

LeClair, J., Abraham, S., & Shih, L. (2013). An inter-disciplinary approach to educating an effective cyber security workforce. *InfoSec CD '13 Proceedings of the 2013 InfoSec CD '13: Information Security Curriculum Development Conference,* 71. doi:10.1145/2528908.2528923

Moore's Law. (n.d.). Moore's law, or how overall processing power for computers will double every two years. Retrieved from www.mooreslaw.com.

Salah, K. (2014). Harnessing the cloud for teaching cybersecurity. *Proceedings of the 45th ACM Technical Symposium on Computer Science Education,* 529–534. doi:10.1145/2538862.2538880

Son, J., Irrechukwu, C., & Fitzgibbons, P. (2012). Virtual lab for online cyber security education. *Communications of the IIMA,* 12(4), 81–96.

Willems, C., & Meinel, c. (2012, April). Online assessment for hands-on cyber security training in a virtual lab. In *Global Engineering Education Conference (EDUCON), 2012* IEEE, 1–10.

AUTHORS

Charles Parker II (charlesparkerii@gmail.com) is currently the application security architect for an automaker. His research interests include encryption, next generation antivirus, and biometric security. Parker has earned a Master of Business Administration, Master of Science in Administration, Doctor of Law, and Master of Laws (LLM). He is currently completing his doctorate in information assurance and security (ABD).

Data-Driven Cybersecurity Leveraging Intelligent Data Storage (IDS) Capabilities

Audie Hittle

ABSTRACT

This paper addresses research and innovative data-driven Intelligent Data Storage (IDS) capabilities that contribute to various cybersecurity resiliency functional areas. The intent and focus of this paper is to enhance awareness, discussion, and interaction to stimulate innovation for the purpose of accelerating the transition and creative technology application of IDS capabilities. There is a growing recognition of deficiencies associated with traditional perimeter security or intrusion detection and the importance of creating systems that are more proactive and resilient to cyber attacks. Market research shows that 86% of respondents believe big data analytics would significantly improve their organization's cybersecurity. The same survey further shows that 61% of information technology (IT) managers could more effectively detect an ongoing security breach by leveraging big data analytics. This type of proactive cybersecurity can be enhanced through the application of IDS capabilities, also referred to as Software-Defined Storage (SDS) capabilities. These SDS capabilities, available as commercial off-the-shelf (COTS) solutions, offer the potential of creating a proactive, data-driven cybersecurity environment. With the global cybersecurity market anticipated to grow to $170 billion by 2020, the demand is high for innovative solutions. Government agency officials and industry leaders have acknowledged the importance and mission-critical nature of cybersecurity. To address this, SDS capitalizes upon changes occurring at the bit-level, and supports analysis in real-time, near-real-time, and Hadoop big data analytics. It also supports the growing trend of cyber resiliency, as advocated by leading research and development centers. SDS capabilities exist today with rapid technology application advances underway that can deliver simple, efficient, and resilient solutions.

INTRODUCTION

Cybersecurity is taking on a more prominent and dominant place in the Information Technology (IT) and business communities. Inadequate cybersecurity has the potential of bringing operations to a halt and cyber investments consume increasing portions of available IT and business operating budgets. With the global cybersecurity market anticipated to grow to $170 billion by 2020 (Markets And Markets, 2015), and the U.S. Federal market being valued at $65 billion during this 2015–2020 period (Market Research Media, 2015), the operational demand for innovative and efficient solutions is growing rapidly and tends to rank as a top priority.

Recent market research, concepts, and technologies associated with the innovation of data-driven Intelligent Data Storage (IDS), and specific capabilities that contribute to various cybersecurity functional and mission areas, are introduced in this paper. The concept of cybersecurity resiliency is introduced, as is its relevance to IDS. Industry standard guidelines and requirements are also discussed relative to an IDS architecture and implementation framework. In addition, the visualization of a "data lake," a new data storage paradigm, is introduced that helps operationalize the IDS concept and technology. The importance of these proactive, data-driven concepts and capabilities were validated by leading researchers at MeriTalk, the Government IT Network, who surveyed over 300 federal, state and local cybersecurity professionals in March of 2015 (MeriTalk, 2015). Their research shows that 86% of respondents believe big data analytics would significantly improve their organization's cybersecurity. The same survey further indicated that 61% of IT managers could better detect an ongoing security breach by leveraging big data analytics. It is exactly this type of cybersecurity big data analytics scenario which can be dramatically improved through the application of IDS capabilities, or as it is increasingly referred to in the industry, as

Software-Defined Storage (SDS) which recognizes the flexibility, automation, and efficiency of storage where the hardware can be separated from and controlled by the software (Webopedia, 2015).

Using this SDS acronym also avoids the potential confusion with traditional and well-established cybersecurity intrusion detection system (IDS) terminology. These SDS capabilities, available as existing commercial off-the-shelf (COTS) solutions, offer the potential of creating a proactive, data-driven cybersecurity environment. These capabilities basically work by ensuring that any and all attempts or changes to the data are instantly logged with alerts sent to appropriate Security Information Event Management (SIEM) systems or big data analytic solutions such as Splunk, RSA Security Analytics, Hortonworks, Cloudera, etc., and any accidental or malicious attempts to alter the data can be instantly correlated and proactively dealt with by automated policy-based management systems, or escalated to human-in-the-loop awareness and intervention.

CYBER RESILIENCY AND THE OUTLAW CYBER SEA

Top U.S. federal agency officials and commercial industry leaders have acknowledged the importance and "mission-critical" nature of cybersecurity, perhaps best summarized by Dean of the Tufts University Fletcher School and former U.S. Navy Admiral James G. Stavridis, North Atlantic Treaty Organization (NATO) Supreme Allied Commander for Europe, and commander of U.S. European Command, when he testified before the U.S. Senate Armed Services Committee stating:

"Among the greatest concerns that impacts both military and civilian realms is cybersecurity. Today, we have a billion devices that are accessing the Internet. Our economies are entangled in this Internet Sea, and it is an outlaw sea. Nothing exists in the norms of behavior. There is a military aspect to it, but it's all of society. At some point, there needs to be a very global conversation on this challenge." (Daniel, 2010)

To address the market needs, growth potential, and the Internet "outlaw sea" that Admiral/Dean Stavridis refers to, SDS supports numerous proactive big data cybersecurity objectives, as well as cyber forensics, which capitalize upon changes occurring in data storage at the bit-level and support analysis in real-time, near-real-time and Hadoop big data analytics. It also supports one of the latest trends and techniques—cyber resiliency—as advocated by Federal Laboratories and leading edge Federally Funded Research and Development Centers (FFRDCs) and government think tanks (MITRE, 2015). With some creative thinking about the cybersecurity needs of customers, and the leveraging of some innovative COTS solutions to address the type of internet cyber sea operational requirements identified above, SDS is positioned to deliver simple, efficient, and resilient solutions across the governments and commercial sectors today and to future-proof investments for tomorrow.

Today's most critical operational requirements and most valuable capabilities increasingly address this concept of cyber resiliency. According to a recent report by Government Sales Specialists (GSS) on security priorities and trends across federal agencies, cybersecurity was slated to be the top security priority of most agencies in 2015 (GSS, 2014). This same report cited a Defense News Leadership poll indicating that 67% of federal agencies were unprepared to fend off hackers, as perceived by federal cybersecurity specialists—hence the importance of making systems more resilient to hacking and cyber attacks. The concept and capability are succinctly defined in the original 2012 and recently updated FFRDC MITRE Corporation report, "Cyber Resiliency Engineering Aid—The Updated Cyber Resiliency Engineering Framework and Guidance on Applying Cyber Resiliency Techniques" (MITRE, 2015). That report defined cyber resiliency as a nation's or organization's ability to effectively develop a strategy for mission assurance and the capabilities to fight through and recover from cyber attacks. The MITRE report is an excellent source of reference material for a deeper dive into resiliency, as it compares security with a conventional focus, as covered by Federal Information Processing Standard (FIPS 199, 2004) and other foundational documents such as NIST SP 800-53R4 and CNSS 1253, with considerations for more sophisticated cyber threats and cyber resiliency, as outlined under the Joint Task Force Transformation Initiative (JTFTI, 2014) publications such as the NIST SP 800-39, NIST SP 800-53R4 and NIST SP 800-30R1 (MITRE, 2015).

Increasingly, cyber resiliency is recognized as one element of an overall cyber defense strategy consistent with operational mission assurance objectives. However, it is differentiated from more traditional cybersecurity strategies, such as firewalls, intrusion detection, or various types of perimeter security, which are primarily focused on keeping adversaries out. In addition, cyber resiliency leverages and supports cyber forensics to help ensure critical operational capabilities continue in spite of successful attacks on individual components. Cyber resiliency addresses the reluctant acknowledgement that we cannot, it seems, keep the adversaries out; therefore we must design solutions that enable us to fight through modern cyber attacks.

Intelligent data storage, or SDS, has automated the processes and provides solid enterprise-level data protection, security, management, and performance management capabilities that can contribute significantly to cyber analytics. This includes the ability to support streaming real-time analytics correlated with existing data through long-term forensic analysis, as well as the associated business continuity and recovery operations.

Complementing resiliency is the concept of cyber forensics and how super-efficient SDS supports forensics. In an article published for the Economic Crime Institute (ECI) at Utica College, New York, Joseph Giordano, technical advisor at the Air Force Research Laboratory (AFRL) Information Warfare Branch, and Chester Maciag, program manager of the AFRL Digital Forensics Program, outlined some of the data storage and analysis needs of cyber forensics. While the concept of forensic computer analysis has been around since the early days of computer intrusion detection research done in the 1980s, the term cyber forensics is relatively new to the vocabulary and was not introduced until about 2002.

Likewise, while the benefits of protecting national defense information infrastructure clearly require significant real-time assessment and analysis, the IT system itself is the likely target. The Giordano and Maciag article notes how the primary sources of corroborative evidence, eventually used to create an attack timeline, are the connective elements and the information system itself. Much effort must go into the collection, protection, recovery, and analysis of such broadly distributed digital information in order to develop an understanding of a cyber attacker's intent or objectives. With regard to the forensic process, Giordano and Maciag state that this does in fact drive the military's cyber attack and recovery functions (Giordano & Maciag, 2002). Therefore, when assessing cyber forensics considerations, cyber forensics could be summarized as follows:

Cyber forensics explores and applies scientifically proven methods to collect, store, and analyze digital evidence in order to:

- Keep a detailed log of all network events and activities and cyber attacks
- Associate, understand, and anticipate potential damaging/adversarial events and operational impact
- Deliver data and digital assets necessary and sufficient to support the investigative process

SDS leverages these same forensic foundations in the high-performance computing (HPC), scientific research, and regulatory/audit compliance realms currently in operation. SDS has automated processes and enterprise-level solutions that already contribute significantly to traditional HPC and compliance realms and supporting business continuity operations, and are directly relevant for these new cyber forensic missions.

CYBER RESILIENCY SUPPORTED BY INTELLIGENT DATA STORAGE (SDS)

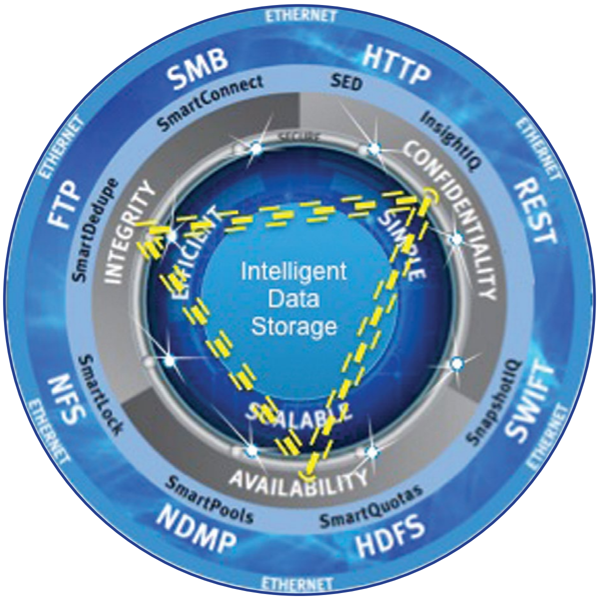
Cyber resiliency can certainly be achieved in a multitude of ways. However, there are certain foundational building blocks and guidelines that can enhance software architecture's ability to be compliant with such guidelines and requirements, or increase the probability of designing a more resilient solution. There are numerous advanced features and functions that are inherent in SDS that directly support key cyber resiliency operations and objectives. Some of these are described in detail in the following paragraphs. However, for quick reference, Table 1 provides an alphabetized list of many of the SDS capabilities correlated with their respective FIPS 199 Confidentiality, Integrity, and Availability (CIA) objectives (FIPS, 2004).

TABLE 1: INTELLIGENT DATA STORAGE WITH FIPS 199 CIA CORRELATION

FIPS 199 CIA CATEGORY	INTELLIGENT DATA STORAGE CAPABILITY
C	Access Zones, controls access to a segment of the file system
C	ACL policies and Cross-protocol permissions
I	Antivirus, Interoperability with ICAP for PCI DSS compliance
I	Auditing of SMB, NFS and HDFS events (current and evolving)
C	Authentication
I	Automated cluster replication and failover
C	Compliance Mode- locks data from accidental or malicious changes
I	Distributed architecture for efficient Recovery Time Objectives (RTOs)
A	Efficient and Reliable, Forward Error Correcting with 80+% efficiency
C	Encryption of Data At Rest (DAR)
C	Fine-grained access control to the file system
A	Flexible, efficient data protection (erasure encoding, non-RAID based)
A	Fully distributed single file system
I	Integrity Scan, examines file system, generates alerts and auto-repairs
C	Identity Management
C	Integration with 3rd-party tools to monitor events and to encrypt data in transit
A	No single point of failure
A	Proactive failure detection and preemptive drive rebuilds
I	Protocol Checksums, blocks, metadata and RBM protocol management
A	Redundancy, operational flexibility for policy-based, data-driven levels
C	Role-Based Access Control (RBAC) for System Administration
I	Snapshots, provides Recovery Point Objective (RPO) support
A	Tolerance for multi-failure scenarios
C	User and ID mapping to associate one user with one ID
C	Write-Once Read Many (WORM) storage

This is not intended so much as a tutorial on FIPS 199 as it is to highlight where and how an SDS capability can address specific cyber resiliency requirements. While the FIPS 199 was originally established to facilitate federal agencies' abilities to meet the requirements of the Federal Information Security Management Act (FISMA), it has also become a standard for industry and government regarding security categorization. FIPS 199 addresses the level of criticality and sensitivity, referred to as security categorization, based upon the potential impact on operations, assets, or people given a security breach due to the loss of these CIA objectives. For instance, loss of confidentiality, for example, unauthorized disclosure of information; integrity, for example, unauthorized modification of information; or availability, for example, denial of service (FIPS, 2004). These CIA objectives also provide the foundational guidelines that form the core capabilities and architecture associated with data-driven Intelligent Data Storage, as shown in Figure 1.

FIGURE 1: INTELLIGENT DATA STORAGE ARCHITECTURE CORRELATION WITH THE CONFIDENTIALITY, INTEGRITY, AND AVAILABILITY (CIA) INFORMATION FRAMEWORK



This CIA core for Intelligent Data Storage is also sometimes referred to as the Information Security Triangle. Subsequent rings around the core architecture add intelligent or "smart" features, for functions like smart connections, quota management, snapshots, de-duplication, etc. Additional rings add multi-protocol support for enhanced collaboration across an enterprise and various levels of Ethernet physical connectivity for operational flexibility. The elements of Figure 1's Intelligent Data Storage architectural core are briefly described as follows, consistent with FIPS 199, and key components are outlined in subsequent sections:

- Confidentiality—Preserving authorized restrictions on access and disclosure, including the means for protecting personal privacy and proprietary information
- Integrity—Guarding against improper information modification or destruction and ensuring information nonrepudiation and authenticity
- Availability—Ensuring timely and reliable access to and use of information

INTELLIGENT DATA STORAGE AND CONFIDENTIALITY

Confidentiality, highlighted in Figure 1, is defined as the preservation of authorized restrictions on access and disclosure as a core element of the SDS architecture. Confidentiality can be achieved by various means. Several of the features and functions that contribute directly to the attainment of a confidential environment are described as follows (EMC, 2014):

- **Role-Based Access Controls (RBAC) for Administration:**

  SDS needs this capability to enable management by role; for example, separate roles for security, auditing, storage, and backup—all of which can be further tailored for execution privileges.

- **Compliance Mode, Write Once Read Many (WORM) and Root Account Controls:**

  - Compliance Mode is a selectable feature that protects critical data from accidental, premature, or malicious alteration or deletion to help address stringent federal compliance regulations such as the Security Exchange Commission (SEC) 17a-4 and other federal compliance regulations requiring a non-erasable, non-rewritable format. To comply with federal regulations, this feature must include a tamper-proof compliance clock and disablement of the root account that cannot be restored. Note that for organizations planning Hadoop operations, elimination of the root could complicate management and administration, and should be carefully weighed with all operational considerations.

  - Solutions that can smartly secure or "SmartLock" the data should also provide operational flexibility to lock down specific directories with WORM whether or not the data storage environment or cluster is in Compliance Mode.

- **Directory Services and Identity Management Systems:**

  - Securely support industry standard protocols like Network File System (NFS), Server Message Block (SMB), and Hadoop Distributed File System (HDFS). Intelligent Data Storage would typically connect to clients via these protocols via some Identity Management Systems, such as Active Directory, Network Information Service (NIS), and Lightweight Directory Access Protocol (LDAP). This authenticates users and groups that verify identities and trigger creation of an access token that contains information about a user's identity. Subsequently, this controls access to directories and files at the level of the files system via a comparison of information in the access token with permissions associated with a directory file to deny or allow access at a specified level.

- Meet compliance regulations like the Federal Information Security Management Act (FISMA), Health Insurance Portability and Accountability Act (HIPAA), Sarbanes-Oxley, SEC 17a-4, and the Payment Card Industry Data Security Standard (PCI DSS).
- Identify and authenticate users and groups by directory services support account management
- Provide rules to map identifiers from multiple external directory services to a single, unique user identity (ID).
- Work with directory services to authenticate users and control access to files, again helping to satisfy compliance regulations for a unique ID for each user as well as for authentication and access control as an Identity Management System.
- Leverage Kerberos to support authentication by using Active Directory or a stand-alone MIT Kerberos 5 key distribution capability.
- Combine ID Mapping and User Mapping from different identity management systems into a single access token with a unique identity and identifier as is necessary to help meet regulatory requirements for one user with one ID.
- Authorize users and groups and control access across different protocols by using POSIX mode bits NTFS Access Control Lists (ACLs), or some optimal merging of them.
- Implement a consistent, predictable permissions model across all file-sharing protocols to preserve the intended security settings for files, directories, and other objects in the file systems.
- Include ACL policies that are created to ensure compliance. These policies preserve Access Control Entries (ACEs) that are set to deny access to individual groups or users. These are set by policy and are designed to facilitate optimal performance tuning to meet access control objectives.
- Provide Access Zones as a virtual security context for authentication and authorization functions to enable the SDS to connect to directory services, authenticate users, and control access to a segment of the file system.
- Offer enhanced data security via Self-Encrypting Drives (SEDs). SDS can leverage Data at Rest (DAR) technology to protect data and drives against theft, enable failed drives to be returned to vendor, and enable automatic key management operations to simplify security operations.
- Protect Data in Transit with partner solutions. SDS can leverage various encryption agents to address end-to-end data protection, depending on operational considerations.

INTELLIGENT DATA STORAGE AND INTEGRITY

Figure 1 also shows how Integrity, the second leg of the security triad as integrated into an effective and efficient SDS architecture, involves guarding against improper information modification or destruction and ensuring information nonrepudiation and authenticity. This is accomplished by implementing the following types of features and functions:

- Protect and stripe the data with Forward Error Correction (FEC) algorithms. Intelligent Data Storage inherently includes advanced features like these for efficient storage and protection and to support basic data CIA functionality.
- Interoperate with antivirus solutions and be compatible with Internet Content Adaptation Protocol (ICAP) servers to help fulfill PCI DSS Requirement 5.1 regarding the deployment of anti-virus software.
- Examine the file system for inconsistencies conducting a type of integrity scan. Read every block, verifying checksum, detecting any mismatch, generating an alert, and automatically attempting repairs as necessary to ensure file system accuracy and integrity.

- Protect the file system structures against corruption and ensure file integrity with 32-bit CRC checksums for all SDS blocks, generate alerts, logs, repairs, and return repaired blocks, to maintain integrity.
- Verify Remote Block Management (RBM) protocol data, a unicast RPC-based solution for high performance networks, in addition to checksum verification for blocks and metadata.
- Automate Cluster Replication and Failover and Failback operations to simplify and expedite operational flexibility.
- Provide flexible data storage Snapshot features to meet Recovery Point Objectives (RPO).
- Facilitate flexible backup and restore to tape and other devices via flexible Network Data Management Protocol (NDMP) capabilities.
- Leverage a distributed architecture to accelerate efficient reconstruction of data from failed drives in a parallel process to ensure Recovery Time Objectives (RTO) are met.

INTELLIGENT DATA STORAGE AND AVAILABILITY

As the final leg of the CIA Information Security Triangle, Figure 1 highlights the focus on Availability, as a means to ensure timely and reliable access to and use of information. In addition, a system that is designed to provision a high level of system availability and uptime will inherently be a platform that is highly resilient—an end goal of our next generation cybersecurity system. The following items summarize some of the characteristics of such an Intelligent Data Storage solution (EMC, 2015):

- An architecture that supports these availability objectives:
  - Fast, flexible, and efficient data protection
  - Tolerance for multi-failure scenarios
  - Proactive detection of failures and preemptive drive rebuilds
  - No single point of failure
  - Fully distributed single file system
  - Fully journaled file system
  - Fast drive rebuild
- Operational efficiency and reliability
- The ability to stripe data to guard it with parity blocks at the file level instead of parity disks
- The ability to protect data with Forward Error Correction (FEC) — a highly efficient method of reliably protecting data. FEC encodes a file's data in a distributed set of symbols, adding space-efficient redundancy

An SDS architecture that is inherently designed for higher availability would be a scale-out data storage system different from traditional hardware-based, Redundant Array of Independent Disk (RAID), scale-up systems. The architecture of such a system would contain no single master for the data and no concept of a high-availability (HA) pair. Instead, a SDS Scale-Out would be a fully distributed system that consists of nodes, implemented in a modular HW arrangement, such as a cluster. Such an arrangement would combine the memory, Input/Output (I/O), Central Processing Units (CPUs), and enterprise grade storage disks or SSDs, of all the nodes into a cohesive storage solution to present a global namespace as a single file system.

When compared with traditional scale-up systems, a scale-out architecture provides a more resilient foundation for data protection and availability. As noted in its report entitled “Critical Capabilities for Scale-Out File System Storage,” Gartner rated the Intelligent Data Storage the highest among all data storage evaluated for resiliency—defined as the platform's capabilities for provisioning a high level of system availability and uptime (Gartner Group, 2013). For innovative vendors, this resiliency and availability supports efficiencies that deliver as much as 80% overall disk utilization efficiency, and, according IDC, 40% to 50% better Capital Expense (CAPEX) and Operating Expense (OPEX) efficiency (IDC, 2011).

A DATA-DRIVEN PROACTIVE CYBERSECURITY SOLUTION

With the ability to detect changes or attempted changes to the data at the bit level, coupled with the ability to monitor and log this activity, combined with log data for RBAC and Access Zones, an Intelligent Data Storage solution is a data-driven, proactive cybersecurity solution. Alerts and notices generated as events at the byte- and bit-level are logged to reflect the node and disk that clients connect to. This log data is stored and is simultaneously accessible to event monitoring and auditing capabilities which integrate with formal auditing software, such as Varonis DatAdvantage or Symantec Data Insight, and can be streamed to Security Information Event Monitoring (SIEM) tools or the Representational State Transfer (REST) API for further leveraging of the data or data storage management system. Logs typically roll over to a new file once they reach a certain size—for example, 1 gigabyte—and there is usually a default protection scheme for the audit log files that should be confirmed to be consistent with Governance Risk Compliance (GRC) policies and cybersecurity objectives.

The chronological flow of a data-driven cybersecurity solution might be something like this: Following an event logging, a forwarding service sends the event to an alert detection point or common event enabler with an HTTP PUT operation, which would then forward the event to a decision support point, such as an auditing software solution or a SIEM. Depending on the policies established at that point, automatic actions could attempt correlations of the suspect ID or IP address with other unusual or anomalous activity across the network. Depending on a variety of factors, such as the location of the occurrence, frequency of the event, and so forth, a human decision maker could be notified to take appropriate action or a policy-based automatic network action could be taken based upon the urgency and certain trigger thresholds being exceeded. In either case, the data storage itself would have been the initial early warning or trigger event, prior to any subsequent forensic cyber analysis determination of need to act, thereby creating a proactive data-driven cybersecurity solution.

FROM THE INTERNET OUTLAW SEA TO SECURITY IN THE DATA LAKE

The Internet outlaw sea that Admiral/Dean Stavridis described earlier presents numerous cybersecurity and data management challenges that need to be addressed. The challenges need to be addressed conceptually in a manner in which people can appreciate the problems and potential solutions, and technically in a simple, secure and efficient way. In July 2011, a visionary technology analyst and writer named Dan Woods speculated on what the big data architecture of the future might look like (Forbes, 2011). He noted that based on his research, it all needed to start with the big data repository, and he had discovered that within the prior few months, an innovative chief technology officer (CTO) named James Dixon was apparently the first to use and advocate the use of the concept called a “Data Lake” (Forbes, 2011). Mr. Dixon made a critical distinction between a data lake and a data warehouse in that a data warehouse is typically pre-categorized, defines the structure of data as it is ingested, and thereby places constraints on how it can be analyzed. He noted that this contrasts and conflicts with the world of big data, where we frequently do not know the structure of the data, what questions to explore, or the potential value of the data. Therefore, as Mr. Dixon proposed, “storing data in some ‘optimal’ form for later analysis doesn’t make any sense” (Forbes, 2011). As an alternative, what Mr. Dixon suggested, was a cheap, easily accessible, massive, data repository that would leverage the technology of the day, rather than bet on the technology of the future. His data lake concept supported the analytics on a huge scale by allowing that the data could be organized and searched at the time future questions arose, rather than bear the cost upfront and attempt to structure the data to address the uncertainty of unknown questions of the future.

While the data lake concept has been around for a few years now, leading industry analysts, such as IDC, still consider the concept new and have recently published their insights in a report on these Enterprise Data Lake Platforms (EDLPs) (IDC 2014). The IDC analysts state that it may still be too early to establish a separate taxonomy for these data lakes, but that it is clearly better and easier to do big data analytics in-place in such repositories. Reasons cited include the use of open standards instead of proprietary technologies and formats, and the ability to use multiple analytic tools, such as Hadoop, simultaneously, on the data streaming into the data lake. IDC further observed that given such macro trends like the Internet of Things (IoT) and evolution to the Third Platform solutions (social, mobile, analytics, cloud), enterprises are likely to continue the move to the data lake for their large repositories (IDC, 2014).

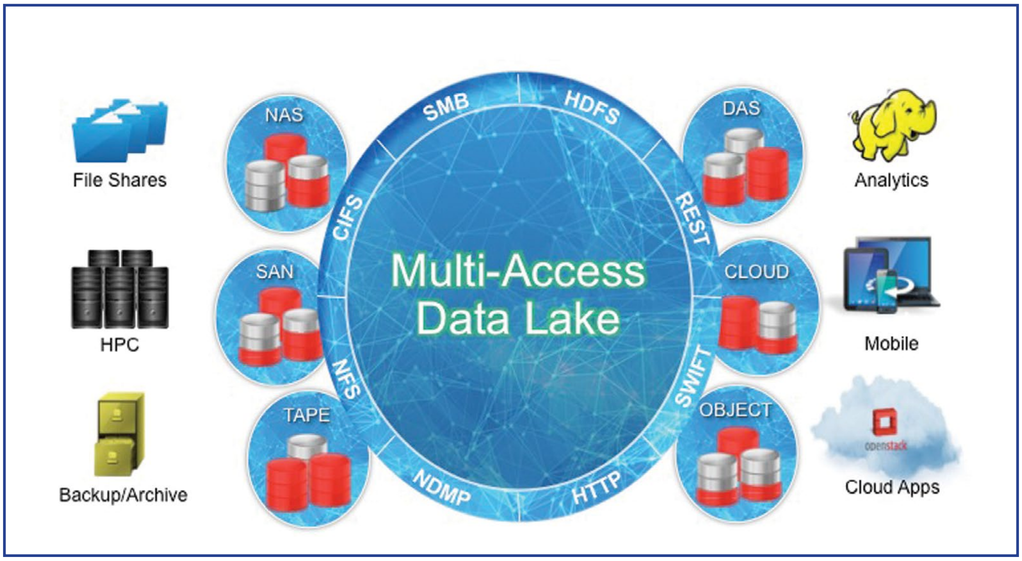
Translating this concept means that the data lake will increasingly become the one place that all “streams” of data flow into. It will become the preferred place to store, protect, secure, and manage all the unstructured enterprise data, irrespective of traditional or next-generation workload. The data lake is not placid, but rather, is a secure, active, and dynamic capability that supports a multi-protocol (HDFS, NFS, SMB, CIFS, FTP, HTTP, etc.), multi-access (REST, S3, open standard API, etc.) collaboration environment (EMC, 2015), as shown in Figure 2. As the IDC concept outlines, and this figure shows, the data lake is an environment that enables any single file — for instance, from an HPC application creating a file and storing it over a CIFS or NFS interface — to be shared in the data lake and simultaneously leveraged by all other authorized users and applications on traditional or next-generation workloads, for example, by an analytics program accessing the data via HDFS. It is a true collaboration environment that enables cross-correlation and fusion between multiple sources of data in a Hadoop, Splunk, RSA Security Analytics, or other big data analytics solution. It is Third Platform ready, and can provide secure, synchronized mobile access with smart tablet or smartphone technology for access on the go, or access in a mobile or tactical work environment.

Perhaps most importantly, the data lake concept and capability provides the business benefits of consolidating all of the unstructured data in the lake, thereby eliminating the "silos of storage," for example, data created by an analytics program, stored via HDFS on Direct Attached Storage (DAS), HPC in a SAN, etc., which have precluded effective secure, enterprise-wide big data sharing, use, and analysis in many cases. When these data lakes are built upon the innovative scale-out SDS technologies and intelligent data storage solutions, they can simply and efficiently scale-out, on a massive level, to address the demands of any enterprise. They can ingest data from all of the sources shown in Figure 2, including File Shares, HPC, Backup/Archive, Analytics, Mobile, Cloud, etc., over any industry standard protocol, such as CIFS, NFS, SMB, HDFS, etc., or next generation access method, such as REST, SWIFT, and so forth. Data lakes can also instantly make this data available to other authorized users across the enterprise for other purposes, such as cybersecurity situational awareness or big data analytics. There are instances of this already implemented in COTS architectures which deliver the type of simplicity and ease of use the cybersecurity industry demands, for example, scaling to over 50 petabytes in a single volume global namespace, along with the performance necessary to support cybersecurity missions, for example, over 200 gigabytes per second throughput or nearly 4 million input/output operations per second (IOPS). This level of performance, coupled with the distinctive resiliency characteristics of intelligent data storage, provides a robust and flexible platform to support rapidly evolving cybersecurity operations.

A SIMPLE, SECURE, AND EFFICIENT SOLUTION

Similar to the concept of Software Defined Networking (SDN), Software Defined Storage (SDS) or Intelligent Data Storage is a concept whose time has come. Intelligent Data Storage is available today in select COTS offerings that address storage options ranging from file-based unstructured data, block-based-data, object-based data, and hybrid cloud capabilities. Most impressive are

FIGURE 2: MULTI-ACCESS DATA LAKE ELIMINATING DATA SILOS



how some of these options are able to scale-out from a few terabytes to over 50 petabytes in a single file system global name space. These options have been recognized as industry leading by organizations like Gartner Group and IDC for their cost effectiveness, operational efficiency, and resiliency (Gartner, 2013), (IDC, 2011).

Equally impressive to overcoming the traditional scaling challenges is efficient disk utilization. Current SDS solutions typically start their disk utilization where other systems frequently peak, and grow from there to 80 or 85 percent disk utilization efficiency (EMC, 2015). They also include significant amounts of automation for things like initial system configuration, performance tiering, load balancing, fail-over and fail-back, etc., which tend to add to the overall system resiliency.

CONCLUSION

Cybersecurity is capturing a more prominent role in the Information Technology (IT) and business communities that are themselves converging. With the global cybersecurity market anticipated to grow to $170 billion by 2020, and the U.S. Federal market being valued at $65 billion during this 2015 – 2020 period, opportunities abound for innovative and efficient solutions. Top government agency officials and industry leaders have acknowledged the importance and "mission-critical" nature of cybersecurity, which has the potential to bring operations to a grinding halt, threaten national security, and impact future commercial profits through downtime or loss of intellectual property. Recent market research, concepts, and technologies associated with the innovation of Intelligent Data Storage (IDS), or as it is increasingly referred to in industry as Software Defined Storage (SDS), and specific paradigms such as the data lake, that contribute to various cybersecurity functional and mission areas, were introduced in this paper. Research shows that 86% of respondents believe big data analytics would significantly improve their organization's cybersecurity and 61% of IT managers could better detect an ongoing security breach by leveraging big data analytics. This type of cybersecurity big data analytics can be dramatically improved through the application of SDS capabilities. These SDS capabilities are available as existing commercial off-the-shelf (COTS) solutions and offer the potential of creating a proactive, data-driven cybersecurity environment. These capabilities work by ensuring that any and all attempts or changes to the data are instantly logged with alerts sent to appropriate Security Information Event Management (SIEM) systems or big data analytic solutions, and any accidental or malicious attempts to alter the data can be automatically correlated and proactively dealt with by policy, or escalated to a human-in-the-loop for awareness and intervention.

ACKNOWLEDGMENT

The author would like to recognize his mentors, too numerous to mention by name from his career in the United States Air Force, MITRE, MIT Lincoln Laboratory, various startups, and small businesses he was a part of, as well as the Fortune 500 companies that have provided him with an opportunity to learn, apply his knowledge, and expand his personal and professional leadership potential. He would also like to acknowledge support and assistance from Team EMC and MeriTalk in the preparation of this manuscript. And last, but not least, he recognizes the ongoing love and support from his wife, Karina McMillan Hittle, who has sacrificed personal time and supported the editorial review of this journal article.

REFERENCES CITED

Daniel, L. (2010, March 9). *Global threats demand broad response.* Retrieved from http://archive.defense.gov/news/newsarticle.aspx?id=58248

EMC (2014, August). *Security and compliance for scale-out Hadoop data lakes.* Retrieved from http://www.emc.com/collateral/white-paper/h13354-wp-security-compliance-scale-out-hadoop-data-lakes.pdf

EMC (2015). *The federation business data lake.* Retrieved from http://www.emcfederation.com/solutions/business-data-lakes.htm

FIPS 199 (2004, February). *The federal information processing standards (FIPS) publication for security categorization of federal information and information systems.* Retrieved from http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf

FIPS 200 (2006, March). *The Federal Information Processing Standards (FIPS) publication.* Retrieved from http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf

Forbes (2011, July 21). *Big data requires a big, new architecture, rethinking repositories.* Retrieved from http://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/

Gartner (2014, January). *Critical capabilities for scale-out file storage.* Retrieved from https://www.gartner.com/doc/2314015/critical-capabilities-scaleout-file-storage

Giordano, J. & Maciag, C. (2002). Cyber forensics: a military operations perspective. *International Journal of Digital Evidence,* Volume 1, Issue 2, Summer.

GSS (2014, November). *Report on the government technology trends to watch in 2015.* Retrieved from http://www.gssfedsales.com/wp-content/uploads/2014/11/2015-Government-Technology-Trends.pdf

IDC. (2011, November). *Quantifying the business benefits of scale-out network attached storage solutions.* Retrieved from https://www.emc.com/collateral/analyst-reports/idc-ar-business-benefits-of-scaleout.pdf

IDC. (2014, January). *Report #245940 defines data lakes as large repository of unstructured data.* Retrieved from https://idc-community.com/groups/it_agenda/storageanddatamanagement/are_data_lakes_real_idc_reviews_enterprise_data_lake_platforms.

JTFTI (2014, July 17). *The Joint task force transformation initiative interagency working group unified information security framework for the federal government.* Retrieved from http://www.fismapedia.org/index.php?title=Joint_Task_Force_Transformation_Initiative

Market and Markets. (2015). *Cybersecurity market forecast 2015–2020.* April 2015. Retrieved from http://www.marketsandmarkets.com/PressReleases/cyber-security.asp

Market Research Media. (2015). *U.S. federal cybersecurity market forecast 2015–2020.* April 2015. Retrieved from http://www.marketresearchmedia.com/?p=206

MeriTalk. (2015, April 28). *Go big security survey.* MeriTalk and Splunk survey of 302 Federal, State, and Local IT leaders. Retrieved from http://www.meritalk.com/go-big-security

MITRE. (2015, May and 2012, December). *Cyber resiliency engineering aid—the updated cyber resiliency engineering framework and guidance on applying cyber resiliency techniques.* Original MITRE Corporation Study, December 2012. Retrieved from http://www.mitre.org/work/tech_papers/2012/11_4436/11_4436.pdf

Webopedia. (2015). *Software defined storage definition, infrastructure managed and automated by intelligent software, as opposed to the storage hardware itself.* Retrieved from http://www.webopedia.com/TERM/S/software-defined_storage_sds.html

Woods, D. (2015). Analyst, writer, and explainer, who finds technology for early adopters. Retrieved from https://www.linkedin.com.

AUTHOR

Audie Hittle (audie@alum.mit.edu) is the chief technology officer (CTO) federal and cybersecurity for the Emerging Technologies Division (ETD) of EMC. His leadership spans more than 15 years of corporate experience plus a 22-year USAF IT-focused career with nationally recognized government-industry collaboration across 16 federal, civilian, defense, and intelligence agencies, and technical staff experience at MIT Lincoln Laboratory. At EMC, Hittle translates operational requirements and technology capabilities to ensure a mutual appreciation of the business challenges and technological solutions for all involved. He envisions his CTO position as a “Chief Translation Officer” role, helping to create a common understanding of how to accelerate solution adoption. He is a member of the ETD Global Office of the CTO, their first CISSP, and the first EMC Corporation member of the ASIS International Chief Security Officer (CSO) Roundtable. He has also served on several boards, including the Federal Laboratory Consortium for Technology Transfer (where he was awarded the U.S. National Technology Utilization Foundation Lifetime Achievement Award), MIT Sloan Fellows, and the AUVSI New England Chapter. His diverse career includes leadership roles in several start-up companies, small businesses, and Fortune 500 Corporations. Hittle’s education includes three degrees, including a joint master’s degree in the Management of Technology (MOT) from the MIT School of Engineering and Sloan School of Management, Master of Science in engineering management from Western New England University, and a Bachelor of Science in electrical engineering from the University of New Mexico.

Security Development Life Cycle Framework for Web-based Applications

Ezhil Kalaimannan and Caroline S. John

ABSTRACT

Information security is an essential component for any organization that offers web-based services for electronic commerce. While this is true, companies often fail to understand that computer security is a constant process and failure to continually review current computer security controls could lead to events that may possibly damage the organization's reputation, cause a loss in productivity, or cause financial loss at any particular time. In view of these potential risks that could occur unexpectedly from the failure to continually review computer security controls, this paper suggests a security development life cycle (SecDLC) framework for an organization to maintain and update security for its web-based applications. This framework will demonstrate the necessary steps to maintain the effectiveness of security controls in place, a practical know-how to implement efficient protective and preventative controls, the necessary steps to detect an attack against the company's web-based assets, and the guidelines to develop an appropriate response/recovery plan.

Keywords: information security, security development life cycle, assessment, protection, detection, response, web services.

INTRODUCTION

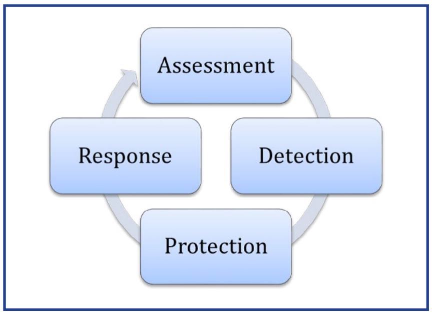
Maintenance and governance of information security is a complex process for organizations that offer web services for electronic commerce. It needs to be continually monitored, reviewed to be able to protect against recent attacks and vulnerabilities that originate within the technologies for web services (Maconachy, Schou, Ragsdale & Welch, 2001). Vulnerability can be defined as the "weakness or fault in a system or protection mechanism that opens it to attack or damage" (Whitman & Mattord, 2006). In order to continuously monitor, review, and update information security for websites, implementation of a Security Development Life Cycle (SecDLC) would be helpful and hence can alleviate the risks of operating an e-commerce website.

SECURITY DEVELOPMENT LIFE CYCLE

The goal of the SecDLC is to maintain, preserve, monitor, and improve information security. The four stages of SecDLC are shown in Figure 1 and are described briefly as follows:

- **Assessment:** This phase is considered to be the rudimentary block of SecDLC, since this is where a user can both assess their own system for well-known vulnerabilities and simultaneously gather details about the tradeoff between the predicted vulnerability and its associated risks which could be imposed on the system.
- **Detection:** The detection phase involves the process of discovery of malicious activity through monitoring tools and external services.
- **Protection:** This phase targets active monitoring of network traffic through tools or software, providing visibility and intelligence to detect and respond to targeted attacks and threats.

FIGURE 1: PHASES OF A SECURITY DEVELOPMENT LIFE CYCLE



- **Response:** The detection and response phases of SecDLC often work in tandem to provide necessary protection to an information asset.

In the following sections of the paper, we will illustrate each stage of the SecDLC with relevant tools and show how they facilitate the process of monitoring, managing, and securing web-based applications.

ASSESSMENT

"The Internet and intranets are in a state of constant change—new protocols, new applications, and new technologies and a company's security practices must be able to adapt to these changes. To adapt, the security process should be viewed as forming a circle. The first step is to assess the current state of security within one's intranet and along the perimeter" (Conorich, 2011).

Assessments should be completed for all systems that support the websites and included networks. Website assessments focus on the code of the website, the databases that support them, and the systems that support the website. Automated web scanners, such as Netsparker, Burp Suite, Hailstorm, or WebInspect, can greatly reduce the cost and training required to perform assessments. Choosing the right one for an organization can be assisted by reviewing the cost and capability reviews from reputable websites or articles from computer security publications.

TABLE 1:

A LIST OF WELL-KNOWN VULNERABILITY SCANNERS

PRODUCT	SOURCE
NESSUS	http://www.tenable.com
MICROSOFT SECURITY BASELINE ANALYZER	http://technet.microsoft.com
SAINT	http://www.saintcorporation.com
GFI LANGAURD	http://www.gfistore.com
RETINA	http://www.beyondtrust.com
CORE IMPACT	http://www.coresecurity.com
QUALYS GUARD	http://www.qualys.com
MCAFEE VULNERABILITY MANAGER	http://www.mcafee.com

Network assessment is another level of assessment, which should be completed to ensure a thorough end-to-end procedure of maintaining an organization's computer security program. This involves performing assessments on an organization's network devices from the perimeter firewall to the internal network switches and routers. Fortunately, most of the vulnerability scanners discussed earlier are able to perform automated assessments on the network devices as well. Table 1 provides a detailed list of well-known vulnerability scanners available in today's market.

Another level of assessment that is often over looked is configuration assessment. Configuration assessments differ from normal vulnerability assessments in that they often look for best practices and not for specific vulnerabilities or missing patches. Again some of the automated system scanners can also perform automated configuration assessments; however, not all systems or applications have the capability to perform automated assessments. Some must have manual assessments performed in order to assess the security configurations (National Institute of Standards and Technology [NIST], 2011).

Maintaining an internal assessment team can prove to be too expensive for small companies or in some cases it may require third parties to perform assessment on web-based applications due to regulations such as the Health Insurance Portability and Accountability Act (HIPAA), the National Institute of Standards and Technology (NIST), or the Payment Card Industry Data Security Standard (PCI-DSS). Despite of the nature of the case, there are many large and small independent testing companies that offer a wide range of services for website assessments that can be tailored for any size company. Some of the vendors today are PWC, Sudo Secure, Aleta Technologies, and Booz Allen Hamilton. It is always commendable to research the provider to ensure that it meets all necessary requirements and legal needs of the organization.

PROTECTION

Protecting the web-based application must include a defense-in-depth strategy. In this approach, an organization will perform the necessary steps to implement information security in all aspects affecting the security of information systems.

Information security policies lay the foundation for complete organizational security architecture. Policies must state how the organization's servers will be secured and monitored, who should be allowed to perform security functions, and how incidents will be handled. The three vital policies that govern the protection of web-based applications are (Whitman & Mattord, 2012):

- User polices, which determines what is acceptable, what the user's responsibilities are and what the consequences for violations of these security policies are.
- Configuration policies provide the framework for enabling repeatable steps to prevent or quickly recover from attacks.
- Change management policies provide a framework for recording and authorizing changes to information security systems.

Defense starts at the perimeter of an organization's network. An organization may opt to use hardware based network firewalls, which protect all of its systems at the network level to block unwanted traffic and protect systems and websites from attacks such as Denial of Service (DOS) to ensure that web applications and resources are always available. In addition, depending on additional needs, there may also be a need to implement host-based firewalls. Host based firewalls only offer protection at the system level. There are also application-based firewalls that can be used for protection for just the website. A review of the configuration, rules to allow or block traffic, and upgrades for firewall systems must be continuously monitored and updated.

One of the first lines of defense in protecting web-based applications is to set up a patch management plan. A patch management plan includes patch testing, implementation strategies, risk mitigation, and deployment strategies. The organization should be aware of the latest vulnerabilities and patch updates using vendor websites such as Microsoft, Apple, Oracle, Linux, Cisco, and Adobe. There are websites that provide detailed information about the latest vulnerabilities such as Open Web Application Security Project (OWASP), SysAdmin Audit Network Security (SANS), United States Computer Emergency Readiness Team (US-CERT), and Common Vulnerabilities and Exposures (CVE). While these websites provide a foundation for understanding the latest patches, the organization may consider using automated patch management tools such as RH-satellite Server, Windows Server update services, LanRev, Bigfix, Patchlink, Luminsion, and Kace to be able to provide an overall protection boundary against malicious attacks (National Institute of Science and Technology [NIST], 2005).

Configuration management is another key element in protecting web-based applications. The configuration management system should automatically assist in tracking key changes to the systems' security configurations baselines. It is recommended to utilize the experiences from other similar organizations when setting security baselines. There are several available from the Internet today such as United States Government Configuration Baseline (USGCB), the Center for Internet Security (CIS), and Security Technical Implementation Guides (STIGs) from the Defense Information Systems Agency (DISA). There are also automated tools that can track or prevent unauthorized configuration changes such as Tripwire or Bit9. Several system scanners have this feature built in or supplied as an add-on. Antivirus scanners are an additional class of utility software which can track and identify unauthorized configuration changes as well as attempts to subvert the security of the programs running on the systems. Some of the most popular vendors today are McAfee, Symantec, Trend Micro, Kaspersky, and Sophos (National Institute of Science and Technology [NIST], 2011).

Training is also an important tool used in protecting websites and if used properly can provide a great return on investment. Formal education through college degrees, technical training offered by companies like Microsoft, and employee awareness form a strong triangle to help protect computer systems. There may also be compulsory requirements such as those from the federal government (Department of Defense [DoD], 2012).

As a well-known statement, not all the security postures discussed above will render an organization's web applications free from attacks. The protection strategy must include elements for continued operations and physical security such as having recommended backups to include offsite storage, disaster recovery, and access controls. Additionally, both physical and logical fixes should include encryption, role separation, logging and auditing, and redundant systems for high availability (Erickson, 2008). This is by no means an exhaustive list, but certainly a good beginning.

DETECTION

Once an organization has begun to understand the threats that its website faces and institutes some security controls, it is paramount that the organization should be able to detect the occurrence of an attack. Even though there are protections in place that are intended to stop attacks from occurring, technology and hackers are evolving at an alarming pace and are always finding new vulnerabilities to exploit. For this reason, it is important to implement some sort of an Intrusion Detection System (IDS). SANS compares an intrusion detection system to an automobile's burglar alarm. "For example, the lock system in a car protects the car from theft. But if somebody breaks the lock system and tries to steal the car, it's the burglar alarm that detects that the lock has been broken and alerts the owner by raising an alarm" (SANS Institute InfoSec Reading Room, 2001). The same concept applies when considering the security of a website. It is vital to have a system in place that will alert the administrators of a possible attack that is about to occur. "If internal and external users are aware that an organization has an IDS in place, they are less likely to probe or attempt to compromise it, just as criminals are much less likely to break into a house that has been clearly marked as having a burglar alarm installed in it" (Whitman & Mattord, 2006). The main reasons for employing IDS to protect an organization's web resources are (Whitman & Mattord, 2006):

- To prevent problem behaviors by increasing the perceived risk of discovery and punishment for those who would attack or otherwise abuse the system.
- To detect attacks and other security violations that are not prevented by other security measures.
- To detect and deal with the preambles to attacks (commonly experienced as network probes and other doorknob-rattling activities).
- To document the existing threat to an organization.
- To act as quality control for security design and administration, especially of large and complex enterprises.
- To provide useful information about intrusions that do not take place, allowing improved diagnosis, recovery, and correction of causative factors.

There are two commonly known implementations of IDSs in a network infrastructure: host- based and network-based. Network-based IDSs are connected to a segment of a network that monitors the traffic on that specific network segment (Whitman & Mattord, 2012). This can allow for earlier intrusion detection before it actually reaches the host. Host-based IDSs reside on a specific host and only monitor activity on that specific system (Whitman & Mattord, 2012). This monitoring is not as broad based as network-based IDSs. It is also important to understand that IDSs can be passive or active. A passive IDS system will simply raise an alarm when it notices an attack. An active IDS will not only raise an alarm when it detects an attack, but will attempt to quarantine or remove the attack. This can be helpful in mitigating the effects an attack more quickly; however it can also lead to interrupting legitimate network traffic due to false positives.

It is important that the website developers understand and consider various attacks that could take place on the website to ensure that there is monitoring in place to recognize these attacks. One such attack that has caused a great deal of damage to companies and individuals is SQL injection attack. Many websites are tied to databases that are constantly queried for information, which is returned to the requester. These databases contain all types of sensitive information such as credit card information, customer addresses, and so forth, which would be useful to hackers. By fooling the database into thinking it is receiving a valid request for information, a hacker is able to steal valuable sensitive information. There are several software applications on the market that can help detect these vulnerabilities and attacks such as Acunetix Web Vulnerability Scanner and Netsparker. Such applications provide the organization with complete vulnerability assessments and serve as a tool to help monitor against new attacks as the technology evolves.

Another threat that websites face is being compromised with the introduction of malware to the website. Whitman and Mattord (2012) define malware as "software designed to damage, destroy, or deny service to the target system." Cyber criminals can use malware as a helping agent to achieve large financial gains. IBM outlines various ways through which cyber criminals use malware as exploitation agents to infect web-based applications as described in Table 2 (IBM Developer Works, 2013). It is very important that the website is continuously monitored for these types of infections to ensure that they are detected quickly before they are disseminated to the public.

TABLE 2:

MULTIPLE WAYS TO USE MALWARE EXPLOITS TO INFECT WEB-BASED APPLICATIONS

CYBER CRIMINAL INTENT WITH MALWARE	
DISPLAYING AND CLICKING ADS	Stealing confidential data
HIJACKING CONFIDENTIAL DATA	Compromising user login credentials
STEALING FINANCIAL INFORMATION	Making fraudulent purchases
CREATING SPAM	Launching denial-of-service attacks

There are various solutions to help identify and protect an organization's web resources against these attacks. One such offering is Malware Scanning from SiteLock. SiteLock will scan the web-based applications in a timely manner and detect any malware that has been introduced and generate notifications based on its findings. There are a variety of similar products available and many should be considered to find the best solution based on the organization's needs and budget. It is also very important to understand that the threats we have discussed for detection represent an extremely small sample size of the types of threats that any web-based application faces. The important baseline from this stage of the SecDLC is that the web-based applications must be constantly monitored for threats that get past the security measures that were put in place in the Protection phase. With the rate at which information technology is changing and the large financial gains that hackers stand to make, an organization can be certain that a cyber-criminal will eventually find a vulnerability in its web-based resources.

RESPONSE

The last phase of the SecDLC is the Response phase. In this phase, the organization needs to be able to respond to an attack that it has detected on its website. Without a proper response, the attack will be successful in its intent and can cause greater harm to the organization and its customers. The organization needs to ensure that it is able to respond to an attack quickly and efficiently so that business can continue to be conducted as usual.

To be able to properly respond to an attack, or an incident that is taking place, an organization must have well-drafted policies in place and a defined incident response plan. "An incident response (IR) plan is a detailed set of processes and procedures that anticipate, detect, and mitigate the effects of an unexpected event that might compromise information resources and assets" (Whitman & Mattord, 2006). It is vital that the business has this well-defined IR plan in place because once an attack occurs, it is way too late to start thinking about what can be done to contain or mitigate the attack. For a response plan to be effective the organization must have clear steps and responsibilities defined before any attack or incident ever takes place. Whitman and Mattord (2006) provide a brief explanation of how an organization with a proper IR plan would react to a potential virus infection with the following steps:

- Verify the presence of the virus by examining antivirus software, system logs and other monitoring systems.
- Determine the extent of the exposure.
- Attempt to quarantine the infestation by first disconnecting infected systems from the network and then looking for evidence of continued spread.
- If not contained, take appropriate measures such as isolating network segments, terminating server sessions, disconnecting the internet connection, or shutting down network servers to contain infection.
- Once infection is contained, continue to look for "flare-ups."
- Disinfect systems by running antivirus software and searching for spyware.

We can also consider the examples of an SQL injection attack and website infection with malware to see how an organization could respond to such attacks. Cleaning up a website after an SQL injection attack is a clear example of the importance of implementing proper protections on the front end. There are so many different ways that an attacker could use an SQL injection, which makes it very difficult to respond and clean them up. As an example of the complexity of this type of clean up, Cherry (2011) describes an SQL injection attack that included an HTML iframe tag into each row of a table. Once this gains a spot to reside inside a customer's computer, viruses and spyware can be easily installed. To remove this iframe, it "means going through every record of every table looking for the attack code that is pushing the iframe to the customer's web browser" (Cherry, 2011).

As we can assume, this process can take a great deal of time and some clever coding to find and replace all of this information to make the database safe for use on the web again. This brings up the importance of regular back-ups of the database, so that the database can possibly be restored quickly from the most recent backups. By having proper detection methods in place, the website administrator will receive a notification when such a threat is encountered.

The preliminary steps to take when alerted to this type of attack is to apply short term solutions to help mitigate the intrusion:

- Temporarily shut down the compromised system.
- Disconnect the compromised system (or network) from the local network (or Internet).
- Disable access to compromised file systems that are shared with other computers.
- Disable system services, if possible.
- Change passwords or disable accounts.
- Monitor systems and network activities.
- Verify that redundant systems and data have not been compromised.

After the intrusion is contained and does not pose a threat to the outside world, the next step is to eliminate all sources of intruder access, determine if the website can be restored from back-ups or if the website will need to be cleaned up, and finally to return the website to normal operations with necessary fixes.

CONCLUSION

It is imperative that an organization can effectively manage its web-based applications from various threats by following the various phases in a SecDLC. Further, these phases can be advantageous in creating theoretical or conceptual based assignments pertaining to computer security concepts and laboratory exercises that can facilitate the learning of various phases in a SecDLC along with its implications. The learner would also gain knowledge about the various tools that can be used in conjunction with each cycle of SecDLC as described in this paper.

The probability that any web-based application will face threats and encounter attacks is practically close to 100%. This makes it absolutely necessary that the organization take proactive measures to protect and defend against these attacks. To be able to perform these tasks, the organization will need to devise a Security Development Lifecycle (SecDLC) framework for all web-based applications. This should be monitored and implemented as a continuous process, which must be reviewed and tested as needed. By adopting and devising the SecDLC framework, an organization can mitigate the potential risk faced by its web-based applications and the organization can show evidence of due diligence for an attack on their resources being successful and cause damage to its investors or customers.

REFERENCES CITED

Cherry, D. (2011). *Securing SQL server: protecting your database from attackers.* Burlington, MA: Elsevier, Inc.

Conorich, D. (2007). *Internet security: securing the perimeter.* (H. Tipton & M. Krause Eds.). Information Security Management Handbook Part 3 (6th ed., pp. 2051–2060). Boca Raton: Auerbach.

Department of Defense. (2012). *Information assurance training, certification, and workforce management.* (DoD Publication No. 8570.01-M). Washington D.C.: U.S. Government Printing Office.

Erickson, J. (2008). *Hacking: the art of exploitation* (2nd ed., pp. 398–405). San Francisco, CA: No Starch Press, Inc.

IBM Developer Works. (2013). *Loaded pages: How your website can infect visitors with malware: A developer's introduction to malicious websites.* Retrieved from: http://www.ibm.com/developerworks/web/library/wa-loadedpages/index.html?ca=drs-. [Accessed: June 05, 2015].

Maconachy, W.V., Schou, C.D., Ragsdale, D., & Welch, D. (2001). A model for information assurance: an integrated approach. *Proceedings of IEEE workshop on information assurance and security, United States Military Academy* (pp. 5–6). West Point, NY.

National Institute of Standards and Technology. (2005). *Creating a patch and vulnerability management program.* (NIST Special Publication No. 800-40 rev2). Gaithersburg, MD: U.S. Government Printing Office.

National Institute of Standards and Technology (2011). *Guide for security-focused configuration management of information system.* (NIST Special Publication No. 800-128). Gaithersburg, MD: U.S. Government Printing Office.

SANS Institute InfoSec Reading Room. (2001). *Intrusion detection systems: definition, need and challenges.* Retrieved from: http://www.sans.org/reading-room/whitepapers/detection/intrusion-detection-systems-definition-challenges-343.

Whitman, M. E., & Mattord, H. J. (2006). *Principles of incident response and disaster recovery.* Boston, MA: Course Technology.

Whitman, M.E., & Mattord, H.J. (2012). *Principles of information security.* Boston, MA: Course Technology.

AUTHOR

Ezhil Kalaimannan, PhD, (ekalaimannan@uwf.edu) is an assistant professor in computer science at the University of West Florida, Pensacola, Florida, working in the areas of cybersecurity and digital/information forensics. He received his master's and doctoral degrees in computer engineering with specialization in information assurance and security from the University of Alabama in Huntsville. His research has been largely in the areas of digital forensic investigation, network security, and cybersecurity education.

Caroline Sangeetha John (csj0005@uah.edu) is a doctoral candidate in electrical engineering at the University of Alabama in Huntsville with a minor in information assurance and engineering. She works as a graduate research assistant and also serves as an adjunct instructor in the College of Business Administration. Her wide areas of interest include cybersecurity, in which she holds a post-bachelor's certificate.

Medical Devices: Managing the Risk

James Angle

ABSTRACT

Medical devices are becoming a major point of attack in the healthcare industry. Most medical devices have the capability to access and transmit data on the hospital network both wired and wirelessly. These devices collect and transmit real-time Electronic Protected Health Information (ePHI) and they often rely on out-of-date software that can be susceptible to malware. The software used in medical devices is the same as in other computing devices and subject to the same vulnerabilities. Add to this the fact that healthcare is the target of choice for criminal hackers and there is a recipe for disaster. While it is possible for hackers to gain access to medical devices for nefarious purposes, a more likely scenario is that they will use this access to identify, access, and exfiltrate medical records. Hackers look for a vulnerable system, gain access to the system, and use that as a pivot point to find systems containing the information they are looking for. This paper will identify vulnerabilities and recommend a risk management framework for mitigating the risks.

INTRODUCTION

Medical devices increasingly rely on complex software to manage critical functions (McCaffery, Burton, Richardson, 2009). Medical devices collect, process, and store ePHI as well as manage life-critical functions (Hanna, Rolles, Molina-Markham, Poosankam, Fu, and Song, 2011). Many of these devices run on Commercial Off The Shelf (COTS) software such as Linux, Windows, and Oracle (Integrating the Healthcare Enterprise, 2009). Running COTS software makes the device susceptible to the same vulnerabilities as any other computer. Compounding the problem associated with COTS software, device manufacturers continue to use old technologies such as Windows CE 5 and earlier in their devices because they know regulators will approve the software (Healey, Pollard, and Woods, 2015). This means device manufacturers sell devices when the software has already passed the main support window. In addition to the vulnerabilities in the software, the manufacturer may install backdoors in the systems in the form of hard coded administrator passwords (O'Brien, 2014).

According to the SANS Health Care Cybersecurity Report (Filkins, 2014), 7% of malicious traffic found in the healthcare environment originated from radiology imaging software. In a three-year period, the Veterans Administration (VA) experienced 142 instances of malware affecting a broad range of medical devices (Kramer, Baker, Ransford, Molina-Markham, Stewart, Fu, and Reynolds, 2012). Another issue is the use of medical devices as a pivot point for hacking medical records. In a recent study, TrapX Security identified a Picture Archive and Communications System (PACS) infected with hacking software that was moving laterally through the network looking for an appropriate target and exfiltrating data to China (TrapX Security, 2015). A number of conditions such as misconfiguration (Filkins, 2014), unpatched software (Healey, et al., 2015), hard coded passwords (O'Brien, 2014), and open and unprotected ports (Paul, 2015) can cause these issues.

This paper breaks down the risk management functions into three areas. The assumption made is the organization has identified the threats. The first area of concern deals with legacy medical devices. There are some distinct actions required to identify and mitigate vulnerabilities in these devices. The second area identifies requirements for purchasing new devices to ensure the identification and mitigation of vulnerabilities. The third area deals with continuous monitoring of the devices to ensure the mitigating control effectiveness.

LEGACY MEDICAL DEVICES

Since the management of medical devices primarily falls under clinical engineering and not the IT department, there may not be an accurate inventory of devices connected to the network. The first step is to conduct a complete scan of the network. A scan using a vulnerability scanner can accomplish multiple things. The scan will identify all of the devices connected to the network including the operating system on the device and it will identify vulnerabilities on the devices. The vulnerabilities identification includes the Common Vulnerability Enumeration (CVE) number and the vulnerability rating as high, medium, or low. After identifying the vulnerabilities, the organization must prioritize the mitigation effort based on the severity of the vulnerability (Joint Security and Privacy Committee, 2007). After identifying the devices, the organization should create a naming convention that clearly identifies each device as a medical device. Using the same naming convention used for workstations adds confusion to the processes of vulnerability and patch management.

The next step is to remotely run a virus scan on the medical devices identified by the vulnerability scanner. Do not automatically clean malware from the devices. The antivirus software can incorrectly identify critical parts of the software as malicious (Thompson, 2014) and removing or quarantining it can render the device inoperable. Identify any suspected malware and manually review the list prior to cleaning the devices. This is a time consuming but necessary step.

Finally, the organization should develop an architecture that isolates the medical devices from the rest of the network. All medical devices should be isolated from the rest of the network regardless of how current the software. The use of Virtual Private Networks (VLANs) protect the segments from penetration and can contain virus outbreaks if they should occur (Integrating the Healthcare Enterprise, 2011). It is easy to think that running current software versions such as Windows 7 to update the device is possible whenever a patch is available; however, the vendor must approve the patch. This will slow down the patching process, so regardless of how current the software, segment the device.

Here are some things to consider before making the decision to exempt newer devices from isolation. The lifecycle of the device is most likely longer than the lifecycle of the software (Fu, 2011) and will eventually require isolation. Additionally, medical device patches require a greater level of testing by the vendor ensuring the patch does not affect the intended operation of the device (Joint Security and Privacy Committee, 2004). Prior to approving a patch, the vendor must ensure the patch has no effect on any of the components of the software installed on the device. This will likely slow down the patching process, requiring the device to be isolated from the rest of the network.

The isolation architecture should include an air gap to ensure the device does not allow access to the Internet. The Internet is a common cause of malware (Kramer et al., 2012). Remove this access since medical devices generally do not require Internet access. In one case, a PACS system compromise occurred when a user visited a malicious website (TrapX, 2015). If the device requires Internet connectivity, for patching, implement a VPN connection and restrict the device's Internet browsing capability.

The next step is to implement a secure configuration for all devices—especially devices using COTS software as the operating system. The National Institute of Standards and Technology (NIST) maintains a secure configuration check lists for some of the COTS products at https://web.nvd.nist.gov/view/ncp/repository. Remember these devices have the same vulnerabilities as any other IT device (Integrating the Healthcare Enterprise, 2011) using COTS software. Medical devices using COTS software may have ports open that are not required for the functionality of the devices. The Industrial Controls Systems Cyber Emergency Response Team (ICS-CERT) recently issued a warning about a drug infusion pump that has port 20/FTP and port 23/TELNET open. These ports were not required for the operation of the device. The ICS-CERT recommends disconnecting the device from the wireless network until the unused ports are closed (ICS-CERT, 2015). Ensure all ports that are not required for the operation of the device are disabled.

As part of the configuration, turn off the ability to check email on computers that manage medical devices. The use of email is prevalent on computers managing medical devices and susceptible to phishing attacks. According to the Verizon 2015 Data Breach Investigation Report, phishing evolved in recent years to incorporate installation of malware with the intent to gain an initial foothold into a network (Verizon, 2015). The phishing attack and subsequent user interaction remains almost entirely about the attackers establishing persistence on user devices and continuing stealth activity inside the network. Once inside the network, attackers can find and exfiltrate data.

Anytime an attacker has access to a system, the attacker can infect it with malware. Removable media such as USB devices can be used to accomplish this. The difficulty of controlling mobile systems increases the ability to infect the systems (Joint Security and Privacy Committee, 2007). Service personnel using USB devices can unknowingly infect a medical device when updating or patching the medical device. To mitigate this threat, scan all removable media before using the media.

The final step for legacy medical devices is to ensure they are at the current patch level. While on the surface, this seems like a simple task, patching medical devices is a complex process. Healthcare organizations seem to have the misconception that the Food and Drug Administration (FDA) prohibits patching devices or they require the device go back through the approval process (Thompson, 2014). This is not the case. The FDA does say to "restrict software or firmware updates to authenticated code" (Center for Devices and Radiological Health, 2014 p. 5). This means the manufacturer should test the patch and supply it to the using organization through a trusted connection. Things the manufacturer test for include ensuring the device works as intended, ensuring the patch does not compromise the safety and effectiveness of the device, and ensuring the patch is free of malware (Joint Security and Privacy Committee, 2004).

Once the organization has the patch, who is responsible for applying the patch? If the organization pushes a patch to a device connected to a patient and it causes the device to reboot, how does that affect the patient? Who applies the patch depends on the device's proximity to the patient. When identifying and segmenting the devices, the organization should determine responsibility for patching each type of device. Some devices require the vendor to patch; some require patches by Clinical Engineering (CE), and in some cases, the IT department will apply the patch. One way to assign responsibility is to assign degrees of separation from the patient. The following table is an example:

TABLE 1: DEFINING DEGREES OF SEPARATION

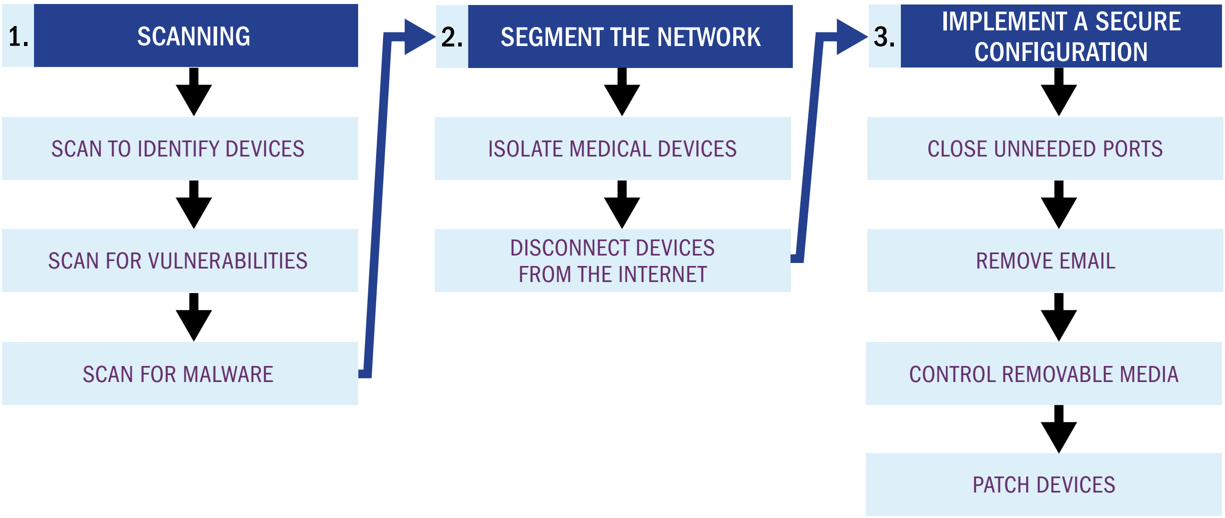
DEGREES OF SEPARATION	DEFINITION FOR DEGREES OF SEPARATION	SUPPORT RESPONSIBILITY
0 DEGREES	Means the device touches the patient	Vendor or Clinical Engineering
1 DEGREE	Means it does not touch the patient but it is doing measurements with patient vital signs, fluids, or data	Vendor or Clinical Engineering
2 DEGREES	Means it does not touch the patient, but it may be doing something still vital to proper patient diagnosis	Vendor or Clinical Engineering
3 DEGREES	Means it is removed from the patient, and is an operational tool more than a diagnostic or clinical device	Vendor or IT

Both CE/IT should identify life support and other related critical devices. Special protocols for priority service of these devices require clear communication to the user.

Items identified with zero to two degrees of separation require either the vendor or Clinical Engineering patch the device. Those identified as having three degrees of separation are either the vendor or the IT department's responsibility to patch. When the vendor is required to patch a device, it is a function of the contractual obligation on the part of the vendor. Regardless of who applies the patches, patching is still necessary, time consuming, and costly.

The process of securing legacy medical devices will require a significant expenditure in time and resources; however, if we do not secure these devices, we leave our systems, data, and patients at risk. Figure 1 shows the steps required for securing legacy medical devices.

FIGURE 1: STEPS FOR SECURING LEGACY MEDICAL DEVICES



NEW MEDICAL DEVICES

Many medical device purchasing decisions are made in isolation, which is not the best way considering that the knowledge required in order to make sound, risk-based decisions is distributed among different people in different departments (Hinrichs, Dickerson, and Clarkson, 2013). Purchasing new medical devices should be a multifunctional effort requiring representation from clinicians, clinical engineering, IT, and information security (Graves, 2011). Addressing concerns prior to purchasing the device ensures the interests of all concerned. The information security representative ensures the manufacturer followed the FDA guidelines for cybersecurity and assesses the risk involved with implementing a new device. The information security representative can assess the requirements to run, support, and secure the device (O'Brien, 2014). The FDA defines cybersecurity as "the process of preventing unauthorized access, modification, misuse, or denial of use, or the unauthorized use of information that is stored, accessed, or transferred from a medical device to an external recipient." (Center for Devices and Radiological Health, 2014 p. 3). To this end, the FDA recommends the manufacturer follow the cybersecurity framework established by the National Institute of Standards and Technology (NIST). The cybersecurity framework requires manufacturers to address the following core functions: Identify, Protect, Detect, Respond, and Recover (National Institute of Standards and Technology, 2014).

The security officer is required to perform a risk assessment on the device. The FDA recommends the following documentation on the cybersecurity of devices:

1. Hazard analysis, mitigations, and design considerations pertaining to the cybersecurity risk associated with their device
2. A traceability matrix linking the cybersecurity controls to the risk considered
3. A plan for providing validated updates and patches throughout the device lifecycle
4. A summary of security controls in place ensuring the integrity of the software
5. Instructions including specifications related to the cybersecurity controls

Keep in mind that these are only recommendations and are not required by the FDA (Center for Devices and Radiological Health, 2014).

The Manufacturer Disclosure Statement for Medical Device Security (MDS2) is another possible source of information. The form is the result of collaboration between the Health Information Management System Society (HIMSS) and the National Electronic Manufacturers Association (NEMA). The MDS2 contains the device description, information on the management of private data, and information on the security capabilities of the device (NEMA, 2013). While no regulatory requirement exists for the manufacturer to fill out the MDS2 form, the recommendation is for the organization to require the form.

Either of these documents will give the organization information necessary to make a risk-based decision on the purchase of the device. Having the information to make a risk-based decision prevents a situation where the device is purchased and the organization later discovers the device has little or no built-in security controls. If a healthcare organization purchases a device without reviewing the security controls, the decision to use the device becomes an economic decision and not a risk-based decision. One example of this is an organization that purchased glucometers that connect to the network wirelessly and stores PHI. The device requires no authentication to use and stores the SID and password in clear text.

After making the decision and prior to purchase of the device, some items must be included in the contract. Primarily the contract should require the manufacturer to supply a life-cycle management plan for the device. Require the manufacturer to incorporate software into the life-cycle management plan. In addition, include the responsible party for installing any updates or patches. When the responsibility for installing the patches lies with the manufacturer, include a time requirement for completion. When the healthcare organization is responsible for installing the patches, include a time requirement for the manufacturer to supply the patches. Table 2 gives an example of time requirements for supplying patches based on the criticality of the vulnerability.

TABLE 2:
TIME REQUIREMENTS FOR SUPPLYING PATCHES

CRITICALITY	TIME REQUIREMENT
HIGH	2 weeks
MODERATE	3 weeks
LOW	4 weeks

The determination of the time requirement depends on organizational risk tolerance and the manufacturer's ability to test and deliver the patches. In addition to the above requirements, healthcare organizations must insist that manufacturers do not include hard coded passwords in their devices.

After purchasing new devices, the healthcare organization has some steps to follow prior to allowing the device to be operational. First, configure the device to meet the organization's policies, include removing all unneeded services and closing unused ports. Next, scan the device for vulnerabilities. The scan will also identify vulnerabilities associated with Operating System (OS) misconfiguration. Purchasing new devices does not ensure there are no vulnerabilities associated with the device. Use the scan results to document the OS patch level. If the device requires updates, complete them before proceeding. The device manufacturer does not know what the organization's requirements are concerning OS configuration so work with them to configure the device.

After completing updating and configuration, put the device on the medical device VLAN with the legacy medical devices. Placing the devices on a segmented network allows the organization to manage all of the devices without having to find where they are located. Additionally, this allows the organization to identify devices where IT can push patches or where CE needs to patch. Referring to table 1, only push patches to devices with three degrees of separation. Pushing patches to any other device will affect patient care. Devices with less than three degrees of separation require individual application of patches.

CONTINUOUS MONITORING

Continuous monitoring is part of the organization-wide risk management strategy. Understanding what assets require protection and their security posture are key pieces of information for situational awareness (Dempsey, Chawla, Johnson, Johnston, Jones, Orebaugh, Scholl, & Stine, 2011). Additionally, an understanding of the threat is necessary for understanding security requirements. This information is critical when making risk-based decisions regarding medical device security and essential in developing a continuous monitoring plan.

An important part of the plan is determining both the type and frequency of monitoring requirements. Does the organization require logging on the device? And if so, what type of logging? When making these decisions, the organization acquires this information from the risk assessment. At a minimum, scanning the devices for vulnerabilities and viruses is required. The frequency varies from monthly to quarterly based on the threat. In addition to scheduled scans, events can trigger the need for immediate scans. One such example is the discovery of an unauthorized access to the device. These scans will identify vulnerabilities necessitating actions to mitigate the risks.

When patches are available, the vendor will test the patch and supply the organization with the approved patch. While it is the vendor's responsibility to provide the patch in the required time, it incumbent on the organization to ensure the vendor complies. Once the patch is received, the identified organizational element schedules and patches the device.

An analysis of the information gained from monitoring is conveyed to management. This analysis allows management to update their risk-based decisions as needed. The NIST cybersecurity framework provides an excellent tool for conveying this information to the organization's management. The FDA recommendations that premarket submissions for cybersecurity management follow this framework. The framework provides the organization with the means to identify their current security posture and to set goals for improving security (NIST, 2014).

The Cybersecurity Framework consists of five core functions: Identify, protect, detect, respond, and recover (NIST 2014). These core functions provide the guidelines for the organization's cybersecurity activities. The organization compares their current security posture with the industry as a whole. The organization then establishes a guide for reaching the desired security posture.

CONCLUSION

While manufacturers may try to improve the security of their devices, for various reasons there will always be issues. Devices will continue to have outdated operating system software installed. No matter how new the device, the operating system software may become outdated. This is due in part to the lifecycle of the device being longer than the lifecycle of the software. This will require manufacturers to manage the patch process proactively by providing tested and verified patches in a timely manner and updating their software to the current version as soon as possible.

For their part, healthcare organizations must take steps to segment their networks. The organization must isolate devices from other systems and from the Internet. Medical devices require scanning for vulnerabilities and malware on a regular basis. Patching requires planning and coordination in order to avoid affecting patient care. The organization must diligently execute their continuous monitoring program by identifying and mitigating new threats and vulnerabilities.

Currently, information security looks at a single host in isolation, setting a security baseline and controls for each system and centrally managing policy and configuration (Crane, 2013). This model is less and less effective as the size and complexity of networks grow. Adding a segment for medical devices adds even more complexity to the network. This will require organizations to look at new technologies that proactively address threats as they develop. Some of the new technologies monitor networks for behavioral changes identifying and reacting to threats before they cause major damage. Recent research explored using emergent behaviors for network defense to increase the defensive security posture when many endpoints are uncontrolled or unmanageable from a central command and control capability. This type of technology seems ideal since most medical devices are uncontrolled or unmanageable from a central command and control capability (Crane, 2013). It is incumbent on the healthcare organization to manage the security of their medical devices in a manner that mitigates vulnerabilities to the greatest extent possible.

REFERENCES CITED

Crane, E. N. (2013). *Emergent network defense* (Doctoral dissertation). Retrieved from http://media.proquest.com/media/pq/classic/doc/2865848201/fmt/ai/rep/NPDF?_s=8qldv%2F%2Frsx0FBkqtwgfdfvhLtGo%3D

Dempsey, K., Chawla, N. S., Johnson, A., Johnston R., Jones, A. C., Orebaugh, A., Scholl, M., & Stine, K. (2011). *Information security continuous monitoring for federal information systems and organizations* (Special Publication No. 800-137). Gaithersburg, MD: National Institute of Standards and Technology. Retrieved from http://csrc.nist.gov/

Filkins, B. (2014). *Health care cyberthreat report: Widespread compromises detected, compliance nightmare on horizon* (SANS Analyst Whitepaper). Retrieved from SANS website: https://www.sans.org/reading-room/whitepapers/analyst/health-care-cyberthreat-report-widespread-compromises-detected-compliance-nightmare-horizon-34735

Fu, K. (2011, April) Trustworthy medical device software. *Institute of Medicine.* Workshop on public health effectiveness of the FDA 510(k) clearance process. Retrieved from https://spqr.eecs.umich.edu/papers/fu-trustworthy-medical-device-software-IOM11.pdf

Graves, K. (2011). Global best practices in medical device procurement—A road map to system success. *Journal of Medical Marketing*, 11(2), 101–108. doi: 10.1057/jmm.2011.1

Hanna, S., Rolles, R., Molina-Markham, A., Poosankam, P., Fu, K., & Song, D. (2011) *Take two software updates and see me in the morning: The case for software security evaluations of medical devices.* Paper presented at the 2nd USENIX workshop on health security and privacy, San Francisco, CA. Abstract retrieved from https://spqr.eecs.umich.edu/papers/hanna-aed-healthsec11.pdf

Healey, J., Pollard, N., & Woods, B. (2015). *The healthcare Internet of things: Rewards and risks*. Retrieved from the Atlantic Council website: http://www.atlanticcouncil.org/images/publications/ACUS_Intel_MedicalDevices.pdf

Hinrichs, S., Dickerson, T., & Clarkson, J. (2013). Stakeholder challenges in purchasing medical devices for patient safety. *Journal of Patient Safety* 9(1), 36–43. doi: 10.1097/PTS.0b013e3182773306.

Integrating the Healthcare Enterprise, Patient Care Device Technical Committee. (2009). Technical Framework White Paper Retrieved from http://wiki.ihe.net/index.php?title=Special%3ASearch&search=medical+equipment+management&go=Go

Integrating the Healthcare Enterprise, Patient Care Device Technical Committee. (2011). (Cybersecurity White Paper). Retrieved from http://wiki.ihe.net/index.php?title=Special%3ASearch&search=medical+equipment+management&go=Go

Industrial Controls Systems Cyber Emergency Response Team. (2015). *Hospira LifeCare PCA Infusion System Vulnerabilities* (Advisory ICSA-15-125-01B). Retrieved from https://ics-cert.us-cert.gov/advisories/ICSA-15-125-01B

Joint Security and Privacy Committee, National Electrical Manufacturers Association. (2004, October). *Patching off-the-shelf software used in medical information systems*. Rosslyn, VA. Retrieved from www.nema.org/medical/spc

Joint Security and Privacy Committee, National Electrical Manufacturers Association. (2007, October). *Information security risk management for healthcare systems*. Rosslyn, VA. Retrieved from www.nema.org/medical/spc

Kramer, D. B., Baker, M., Ransford, B., Molina-Markham, A., Stewart, Q., Fu, K., & Reynolds, M. R. (2012). Security and Privacy Qualities of Medical Devices: An Analysis of FDA Postmarket Surveillance. *PLoS ONE*, 7(7), e40200. Doi:10.1371/journal.pone.0040200

McCaffery, F., Burton, J., & Richardson, I. (2010). Risk management capability model for the development of medical device software. *Software Quality Journal 18*, 81–107. doi: 10.1007/s11219-009-9086-7

National Electrical Manufacturers Association. (2013). HIMSS/NEMA Standard HN 1-2013, *Medical disclosure statement for medical device security*. Retrieved from www.himss.org/resourcelibrary/MDS2

National Institute of Standards and Technology. (2014). *Frameworks for improving critical infrastructure cybersecurity*. Retrieved from http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214.pdf

O'Brien, G. (2014). *Wireless medical infusion pumps* (Use case). Retrieved from The National Institute of Standards and Technology website: https://nccoe.nist.gov/sites/default/files/nccoe/NCCOE_HIT-Medical-Device-Use-Case.pdf

Paul. (2015). Researcher: drug pump the least secure IP device I've ever seen. *The Security Ledger*. Retrieved from https://securityledger.com/2015/05/researcher-drug-pump-the-least-secure-ip-device-ive-ever-seen/

Thompson, J.J. (2014). Securing medical devices while maintaining FDA compliance. *ISSA Journal* 12(5), 24–30. Retrieved from https://www.issa.org/global_engine/download.asp?fileid=CDA41084-FF25-46FA-BDD5-6B0A48D18E5F&ext=pdf

TrapX Security. (2015, June). *Healthcare hospital PACS MEDJACK*, San Mateo, CA. Retrieved from www.trapx.com

U.S. Department of Health and Human Services, Food and Drug Administration Center for Device and Radiological Health. (2014). *Content of premarket submissions for management of cybersecurity in medical devices*. Retrieve from http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089543.htm

Verizon. (2015). 2015 Data Breach Investigations Report. Retrieved from www.verizonenterprise.com/DBIR/2015/

AUTHOR

James Angle (anglejl@trinity-health.org) has a doctorate in business administration with a specialization in computer and information security. He has over 20 years of experience in the field of information technology with 15 years of information security. Angle is currently an information security officer for a major healthcare organization.

Community Colleges as Innovators in Cybersecurity Workforce Development

Wm. Michael Volk

ABSTRACT

It is well documented that the supply of cybersecurity workers has not kept pace with employer demand. Community colleges and universities are actively developing and refining cybersecurity academic programs designed to produce a well-trained entry-level cybersecurity workforce pipeline. The next step to help accelerate the development of a trained cybersecurity workforce is to increase access to relevant cybersecurity training designed to help incumbent entry-level and mid-level cybersecurity workers advance in career paths. It is also important for community colleges to help organizations of all shapes and sizes reduce cyber risk by offering meaningful cybersecurity awareness training.

The purpose of this paper is to provide recommendations about how community colleges can be innovators in cybersecurity workforce development and training by providing noncredit options aligned with industry and employer needs. The recommendations discussed in this paper are based primarily on the findings of three sources: H4CKER5 Wanted—An Examination of the Cybersecurity Labor Market (published by the RAND Corporation in 2014), The Job Market Intelligence: Cybersecurity Jobs, 2015 Report (published by Burning Glass Technologies), and the 2015 Data Breach Investigations Report (published by Verizon). There is room for additional research into the cybersecurity labor market, but these sources provide a logical starting point to guide community college innovation in noncredit cybersecurity training.

The recommendations based on the literature review include how community colleges can accelerate career advancement in cybersecurity by increasing access to noncredit skills-based cybersecurity training; a conceptual framework for the implementation of cybersecurity training pathways to enhance curriculum development and increase collaboration with employers; and a strategy to help organizations reduce risk in cyberspace through meaningful cybersecurity awareness training. The recommendations and conclusions presented in this paper are based on pilot initiatives being undertaken by Anne Arundel Community College, Cyber and Technology Training group.

COMMUNITY COLLEGES AS INNOVATORS IN CYBERSECURITY WORKFORCE DEVELOPMENT

Businesses and government agencies across all industries struggle to find qualified candidates for in-demand cybersecurity positions. The well documented cybersecurity workforce shortage is clearly articulated in a recent cybersecurity labor market study published by Burning Glass where they report “238,158 postings for cybersecurity-related jobs nationally” and that “cybersecurity postings have grown 91% from 2010–2014” (Burning Glass Technologies, 2015). Community colleges have historically proven to be nimble and adaptable to meeting workforce demands in high-growth areas and have been quick to answer the call for job-relevant programs that develop a trained pipeline of cybersecurity workers. However, much of the work at the community college level focuses on traditional academic programs designed for entry-level workers. To meet the immediate workforce needs in cybersecurity, community colleges must also innovate in noncredit cybersecurity training designed to help the incumbent cybersecurity workforce advance in their careers.

The approach at the core of many community college-led initiatives is by design built to support entry-level cybersecurity careers where job roles are well defined and are relatively similar across industry and government organizations. The result of these efforts, for good reason, is structured, methodical programs where each component is necessary to achieve the stated result of building a well-trained robust cybersecurity workforce pipeline. According to a recent study funded by the RAND Corporation called “Hackers Wanted, An Examination of the Cybersecurity Labor Market” by Martin C. Libicki, David Senty and Julia Pollak, “the rising difficulty of finding and retaining qualified individuals at what are considered reasonable wages—is predominantly at the high end of the capability scale: roughly the top 1–5 percent of the overall workforce” (Libicki, Pollak, & Senty, 2014). Community college programs targeting new entrants into the workforce are well established and often times very effective. A growing concern further down the workforce pipeline is the large gap between entry-level careers in cybersecurity and those at the middle and senior levels within an organization. The next opportunity to innovate lies outside of the traditional approach taken by community colleges and focuses on training initiatives for the incumbent middle-tier cybersecurity workforce in medium and large organizations as well as emerging cybersecurity employers.

LITERATURE REVIEW

To better articulate the need for new approaches in cybersecurity education and training at the community college level, it is important to examine the current workforce landscape. The literature review provided is not exhaustive; rather it is intended to establish a baseline for the current cybersecurity labor market issues impacting cybersecurity recruitment and workforce development. To achieve this, the literature review focuses on three main sources: the study by Libicki, Pollak, and Senty (published by the RAND Corporation in 2014), the Job Market Intelligence: Cybersecurity Jobs, 2015 Report (published by Burning Glass Technologies), and the 2015 Data Breach Investigations Report (published by Verizon).

Libicki, Pollak, and Senty provide a comprehensive examination of the current cybersecurity labor market and offer meaningful insight and context to help explain the rising number of unfilled cybersecurity jobs. Where many reports focus too narrowly on the growing shortage of skilled cybersecurity workers, Libicki, Pollak, and Senty take a much broader and realistic approach that examines the current issues surrounding the cybersecurity labor market in the context of traditional economic theory. For example, they argue the labor market trends seen in cybersecurity are not uncommon and state that, "whenever rapid demand increases hit a profession with nontrivial skill and/or education requirements, economic theory suggests that rapidly rising compensation packages and strong competition for workers can be expected" (Libicki, Pollak, & Senty, 2014). They also highlight organizations that have crafted successful cybersecurity workforce development strategies and provide recommendations about what can be done to improve conditions going forward. The primary conclusion of the report is that, "the difficulty in finding qualified cybersecurity candidates is likely to solve itself, as the supply of cyber professionals currently in the educational pipeline increases, and the market reaches a stable, long-run equilibrium" (Libicki, Pollak, & Senty, 2014). The suggestion is not that cybersecurity education and training should be discontinued; rather they caution against "overdoing existing programs."

The most relevant finding for the purpose of this paper presented by Libicki, Pollak, and Senty is that, "the larger organizations—both private and public—have found ways of coping with tightening labor markets, in large part through internal promotion and education" (Libicki, Pollak, & Senty, 2014). Large public and private organizations also take different approaches based on the unique environments in which they operate. For example, the NSA is constrained by pay bands and is not able to compete with private industry in the cybersecurity labor market on salary alone. In response, instead of investing at the top-tier of the cybersecurity labor market where they are unable to compete, they employ long term strategies of hiring entry-level employees and growing their cybersecurity workforce from within. This strategy targets students early in their cybersecurity academic careers and continues into employment where employees participate in "a very intensive internal schooling system, lasting as long as three years for some" (Libicki, Pollak, & Senty, 2014). While hard to replicate in other industries, it is a successful strategy that aligns very well with cybersecurity programs at colleges and universities.

Private organizations, particularly those in the defense contracting space have also developed effective cybersecurity workforce development strategies. Organizations that fall in this category typically start with a relatively technically skilled workforce, have more flexibility in offering competitive pay to the top-tier cybersecurity talent, and have flexibility to fund internal training and development initiatives. These conditions make it possible to grow a cybersecurity workforce from the inside rather than relying solely on external recruitment efforts where there is a limited supply of qualified candidates from which to choose. One large defense contractor interviewed by Libicki, Pollak, and Senty reports that they require thousands of employees across the organization to go through a two-week cybersecurity course. This approach allows the organization to maintain a trained cybersecurity workforce and from there, "the talented ones, as defined by their performance and behavior, are sent through further education, culminating in six to nine months of focused training" (Libicki, Pollak, & Senty, 2014). This approach helps the organization exert some level of control over cybersecurity recruitment and insulate itself from the external issues that others face when building their cybersecurity workforce. Because they have flexibility to offer competitive salaries, there is less of a risk of training top-tier talent only to have them leave to a competitor.

A final important takeaway from Libicki, Pollak, and Senty is that organizations across the board are finding it difficult to attract and retain top-tier cybersecurity professionals. Education and training initiatives will help this issue in the long term, but the impact that community colleges, universities, and training providers have in the short term is minimal. Top-tier cybersecurity professionals possess skills that are hard to develop and difficult for employers to measure using traditional screening techniques. Contrary to the popular notion that the hard to find cybersecurity experts are young hackers, the RAND research finds that the "the top-tier is not necessarily comprised of young geniuses so much as those who possess the right combination of technical talents and organizational experience (notably administrative, managerial, bureaucratic, and/or marketing smarts). Typically, they tend to be in their 30s (or older), not in their 20s" (Libicki, Pollak, & Senty, 2014). Over time, the supply of top-tier cybersecurity professionals will increase as individuals currently employed in cybersecurity roles advance in their career paths. Education and training initiatives designed to quickly increase the number of top-tier cybersecurity professional are more difficult to develop and successfully deliver, because experience is such a critical component of the natural process.

Another recent report that provides insight into the evolving cybersecurity job market is the 2015 report published by Burning Glass Technologies. Libicki, Pollak, and Senty focus their analysis primarily on large government agencies and defense contractors, because they have historically shown the largest demand for cybersecurity workers. However, there appears to be a new trend in the cybersecurity job market that is worth noting. According to the job posting data from 2010–2014, "sectors managing increasing volumes of consumer data such as finance, healthcare, and retail trade have the fastest increases in demand for cybersecurity workers" (Burning Glass Technologies, 2015). The defense sector continues to show significant job posting growth with a 57% increase from 2010 – 2014. In the same time period job postings in the finance sector grew by 131%, postings in the healthcare sector grew by 118% and postings in the retail trade sector grew by 120% (Burning Glass Technologies, 2015). The numbers indicate an increase in demand for cybersecurity workers in these industries, but they do not provide insight into hiring and retention rates or if these emerging cybersecurity employers have sophisticated workforce development strategies for cultivating the unique cybersecurity workforce that they demand.

The Burning Glass report also corroborates the findings of Libicki, Pollak, and Senty in that the significant experience requirements for cybersecurity positions make the skills gaps hard to close with short-term solutions. According to cybersecurity job posting data, 83% of cybersecurity positions require at least three years of experience (Burning Glass Technologies, 2015). Hands-on, job-relevant training that prepares individuals for the workforce can help make the transition between the classroom and the work environment easier, but training alone cannot replace work experience. This point is emphasized in the findings of the Burning Glass report that states "employers and training providers must work together to cultivate a talent pipeline for these critical roles" (Burning Glass Technologies, 2015). This is especially true for industries that are showing the fastest growth in their demand for cybersecurity workers.

According to the findings in the Burning Glass report, the need for enhanced collaboration between training providers and employers is especially important for emerging cybersecurity employers in the finance, healthcare and retail trade sectors. Employers in these sectors are struggling to fill "hybrid jobs" that require industry specific knowledge combined with IT and cybersecurity skills. This new development is widening the skills gap, because hybrid jobs require "skillsets that are not traditionally trained together. This often results in skills gaps where employers struggle to find employees that meet these skill needs" (Burning Glass Technologies, 2015). The emergence of hybrid cybersecurity jobs in industries that have not historically been leading employers of cybersecurity professionals adds a new dynamic to the already ambiguous cybersecurity workforce development landscape.

A final issue worth noting is a new development that shows a link between sectors experiencing significant growth in demand for cybersecurity professionals and the increasing role that end users play in the cybersecurity posture of organizations. According to the 2015 Data Breach Investigations Report published by Verizon, there were more incidents of insider misuse in 2015 than ever before. The primary motivating factor prompting insider misuse is personal gain but, "coming in a not-so-distant second is the motive of convenience (basically using an unapproved workaround to speed things up or make it easier for the end user), and while this is not something that is intended to harm the organization, it certainly often has the same result" (Verizon, 2015). And according to the Verizon report, the most affected sectors are Public, Healthcare and Financial Services.

This issue is not traditionally seen as something that impacts the cybersecurity skills gap nor is knowledge of cybersecurity awareness a skill employers are seeking in candidates to fill in demand cybersecurity positions. While this is true, improved security across the organization can impact the demand for a cybersecurity workforce. According to Libicki, Pollak, and Senty, when trying to address cybersecurity labor market issues, "one route that gets little attention is reducing the demand for cybersecurity professionals by finding other ways to reduce cybersecurity issues" (Libicki, Pollak, & Senty, 2014). The growth in insider misuse in healthcare and financial services coinciding with an increase in demand for cybersecurity workers in the same sectors is worth noting as an area where training can potentially have a positive impact.

RECOMMENDATIONS

The cybersecurity workforce development landscape and labor market is evolving as new employers are emerging with cybersecurity workforce needs. Cybersecurity programs being developed at the college and university level are preparing a trained workforce pipeline ready for entry-level cybersecurity careers. Where they once struggled, large organizations that have historically shown the greatest demand for a cybersecurity workforce have demonstrated effective strategies to deal with the tight cybersecurity labor market. While successes in these areas are progressing, employers in sectors that have not historically shown a strong demand for a cybersecurity workforce are experiencing a new demand for a hybrid cybersecurity worker. It is also clear that top-tier cybersecurity professionals are difficult to develop and continue to be a sought after commodity by all employers.

Community colleges already play a significant role in the development of the cybersecurity workforce pipeline by offering comprehensive academic programs. But it is clear that additional innovation at the community college level is needed to accelerate career advancement and the role of noncredit continuing education should be more clearly defined. This next section provides recommendations to help guide noncredit cybersecurity training initiatives at community colleges that address the current, most pressing needs organizations face in the area of cybersecurity recruitment.

- **Accelerated Noncredit Cybersecurity Skills-Based Training** — Community colleges can accelerate career advancement in cybersecurity and enhance workforce development efforts by making accelerated skills-based cybersecurity training more readily available. Based on the information presented in the literature review above, there are existing training options for entry-level workers, and top-tier cybersecurity professionals build their skills over time in the workplace. Outside of large employers who offer comprehensive internal training programs to current employees, there are limited in-person hands-on training options for individuals currently working in entry-level and mid-level IT and cybersecurity professions.

  Community colleges can help fill this void by offering hands-on training that isolates the skills existing entry- and mid-level employees need to advance in cybersecurity career paths. The instructor-led option provides alternatives for individuals who prefer live instruction to online formats. Incumbent IT and entry-level cybersecurity workers already have a technical foundation and are likely to already have some level of formal education. Therefore noncredit training does not need to replicate or mirror comprehensive courses that might be found in an academic program. This allows continuing education cybersecurity training options to focus on critical skills required for the job thereby reducing time participants need to spend in the classroom.

  Short skills-based training consisting of 8–20 contact hours delivered over several meetings on nights and weekends, provides accessible options for individuals balancing work and life commitments. Accelerated skills-based training developed in collaboration with employers allows individuals to stack sequences of training together to build competencies in broader technical areas over time. Individuals also have the flexibility to choose the skills that are most important to their chosen cybersecurity career path without making semester-long commitments. This approach also enhances the prospect of collaboration with industry by making it possible to pair technical skills-based courses with industry specific non-technical training to help cultivate a hybrid cybersecurity worker.

One characteristic of continuing education that makes community colleges especially well positioned to respond to this need is the noncredit training development that is typically accelerated compared to traditional academic programs. Further, most community colleges offering cybersecurity academic programs have facilities built that can be leveraged for this type of training. The combination of responsiveness and brick-and-mortar training locations allows development and delivery to be accelerated to meet employer needs.

- One example of how this strategy can be implemented is exemplified in a pilot project being launched by Anne Arundel Community College (AACC) Cyber and Technology Training group. In the winter of 2016 AACC launched a new 10 contact hour Introduction to Digital Forensics: Evidence Handling and Incident Response course. AACC currently offers an Information Assurance and Cybersecurity, Cyber Forensics Option Associate of Applied Science degree and a Cyber Forensics academic certificate. This instructor-led accelerated noncredit digital forensics course is a new approach in addition to the existing programs. The development and offering of the new noncredit Introduction to Digital Forensics course addresses a gap in AACC's Cyber and Technology Training catalog and provides an innovative skills-based training option designed specifically for incumbent workers that need a low barrier entryway to accelerated training options in this area. The development of this course was led by one of AACC's Assistant Professors in the Science and Technology department and aligns with the NICE Workforce Framework Incident Response specialty area.

- **Cybersecurity Training Pathways** — Community colleges can accelerate career advancement in cybersecurity by implementing approaches to cybersecurity training that encourage collaboration with employers and enhance or support their existing workforce development strategies. According to Libicki, Pollak, and Senty, organizations that have been able to cope with the competitive cybersecurity labor market leverage internal promotion and education strategies. This minimizes fierce competition for a limited number of qualified external candidates. While effective, this strategy requires significant financial and staffing resources as well as time to develop and implement complex initiatives. Organizations that cannot compete on salary for top-tier cybersecurity workers risk investing in internal development initiatives only to have trained workers leave for organizations that offer higher salaries. These constraints make it difficult to replicate the strategies that Libicki, Pollak, and Senty found in large government agencies and defense contractors.

Community colleges can help small, medium, and emerging cybersecurity employers in the community by providing the necessary resources to help employers build internal cybersecurity training initiatives by leveraging the resources of the college to offer employees training pathways provided by the community college. A promising approach that enables this type of collaboration between community colleges and employers is the use of cybersecurity training pathways. This provides a platform for community colleges to group existing training around in-demand cybersecurity roles and to guide the development of employer-driven cybersecurity skills-based training. This approach capitalizes on the responsiveness of the community college as the training pathway tool is designed to be dynamic and based on employer, strategic partner, and jobseeker feedback. This input enables the creation and development of new skills-based courses and training pathways in response to new or unique job roles identified by employers.

The use of cybersecurity training pathways is a common approach across education and training that works very well for noncredit cybersecurity training initiatives. The use of the tool enables continuing education areas of the community college to leverage best practices in cybersecurity education, such as the DHS National Initiative for Cybersecurity Education (NICE) Workforce Framework, while reducing the duplication of offerings already provided by an academic program. Secondly, grouping training around in-demand cybersecurity job roles with the ability to customize training pathways based on specific needs makes the tool especially useful for emerging cybersecurity employers that have unique requirements for hybrid cybersecurity workers.

Industry driven cybersecurity training pathways also guides noncredit cybersecurity curriculum development efforts. The use of training pathways allows accelerated skills-based courses to be developed aiming towards the target of an in-demand cybersecurity job role. This allows participants to take short courses in a sequence along a defined path where they can apply what they learn in the classroom immediately while stacking training to build a more comprehensive skillset over time. Where this type of effort would be a considerable undertaking for an organization that is not in the business of workforce development and training, creating cybersecurity training pathways is manageable for community colleges. Provided the community colleges have an existing catalog of cybersecurity courses and are familiar with the NICE Workforce Framework, cybersecurity training pathways can be launched quickly.

- An example of how this approach can be implemented is this first iteration of AACC's Cyber and Technology Training Pathways that was launched in the fall of 2015. The first iteration of AACC's Cyber and Technology Training Pathways was built based on four of the 31 NICE Workforce Framework specialty areas. After an extensive review of the knowledge, skills, and abilities (KSAs) in the Information Systems Security Operations, Information Assurance & Compliance, Customer Service & Technical Support, and Incident Response specialty areas, AACC's Cyber and Technology Training group mapped its existing courses to as many KSAs in these specialty areas as possible. The results of the mapping are four training pathways that help guide individuals towards training options related to these specialty areas.

  One early success that is an outcome of the implementation of the training pathways approach is the identification of two gaps in AACC's noncredit cybersecurity training inventory. These are in the areas of Digital Forensics and Malware Analysis related to the Incident Response specialty area. As a result, AACC is launching a new 10 contact hour Introduction to Digital Forensics course discussed above, and is now offering an Introduction to Malicious Code Analysis course in collaboration with industry partner OmegaCor Technologies.

- **Meaningful Cybersecurity Awareness Training** — Reducing cybersecurity risk is a strategy used to slow the demand for a cybersecurity workforce. As identified in the Verizon report, insider misuse is a growing concern particularly for organizations required to handle large amounts of sensitive data. Cybersecurity technology can only promise a certain level of protection for an organization. Security tools—even the most advanced ones—still operate based on the parameters of how technology is expected to work in the known threat landscape. Unfortunately it is not always possible to predict with absolute certainty the behavior of the human end user. This uncertainty includes typical issues such as leaving an unencrypted device in a public place or accidently clicking on malicious link in a phishing email. It also includes more deliberate activities such as bypassing security controls when cybersecurity policies do not align with the business processes of the organization.

  With the appropriate policies for technology use in place, a commitment to create a cyber-aware culture, and with the appropriate training, the end-user can be transformed into an asset for an organization. Well-trained end-users that are included in the development of acceptable technology use and cybersecurity policies can become active sensors in a network and complement technology-based incident detection capabilities. This ideal end result begins with creating a cyber-aware culture and is a top down as well as a bottom up process. Leaders need to know some details about the scope of the current threat landscape and the typical threat actors that may target their industry or type of business. Technology users from the bottom up need to be aware of typical threats they may encounter and strategies to use to avoid them. It is also important to help all members of the organization understand the importance of cybersecurity policies. This way, when employees feel they are being asked to sacrifice convenience for security they understand the importance. Conversely, individuals also need a way to communicate when policies interfere with workflow. Without this, users will often work around the policies and thereby expose the organization to increased risk.

  While this strategy is not something that can be deployed overnight, community colleges in the continuing education areas can push cybersecurity awareness training to new levels. The first and most obvious strategy is to foster a cyber-aware culture in an organization by offering meaningful interactive cybersecurity training in formats that are convenient for employees. This includes short sessions focusing on specific aspects of cybersecurity offered in-person throughout the year. Interactive online training is also possible, and new training technologies and platforms make online cybersecurity awareness training more viable than ever before.

Another strategy that community colleges can push internally with minimal effort from employers is to integrate cybersecurity awareness into existing technical and non-technical continuing education courses. For example, when developing and delivering courses that help individuals learn the fundamental technologies and applications to be successful in a modern business environment, general cybersecurity awareness should be incorporated into this training. Many community colleges also offer corporate training options to help organizations develop critical skills across the workforce. These include critical skills such as leadership, management, communication, customer service, and others. Integrating cybersecurity topics into these courses and making new options available to businesses can help improve the baseline cybersecurity awareness with minimal effort on the part of the organization. Both of these strategies can be implemented at the community college level and stand to take cybersecurity awareness training to new levels.

The implementation of this strategy will differ from one community college to the next and will also be driven by the needs of the community. The approach that AACC's Cyber and Technology Training took to move this recommendation forward was to begin developing curriculum and offering open enrollment instructor-led cybersecurity awareness training. The two cybersecurity awareness courses currently being offered by AACC are nine contact hours each and are titled "Simplifying Security in the Cyber Age" and "Securing your Cyber World." The courses are designed for all audiences and increase in complexity as participants advance through the training. The next step that follows the recommendation above is to infuse cybersecurity awareness topics into popular noncredit cyber and technology courses. This strategy leverages AACC's existing cybersecurity awareness curriculum and helps reach a broader audience by integrating relevant cybersecurity awareness concepts into courses that do not currently address cybersecurity awareness issues. Development of the pilot for second phase of AACC's cybersecurity awareness strategy in process as of spring 2016.

CONCLUSIONS

The demand for a trained cybersecurity workforce is growing, but the situation is not as dire as we are often led to believe. Over time as economic theory suggests, the supply of cybersecurity workers will begin to even out with the demand. New technologies that reduce risk and a workforce that is more aware will contribute to a decreasing demand for top-tier cybersecurity workers. The efforts led by community colleges and universities to offer cybersecurity academic programs provide a workforce pipeline that will over time help the cybersecurity labor market stabilize. Access to academic programs is increasing as evidenced by the growth of the National Centers of Academic Excellence in Information Assurance Research, and a two-year education (CAE/IAE, CAE-R and CAE2Y) initiative led by the Department of Homeland Security (DHS) and National Security Agency (NSA). There are currently 181 colleges across the U.S. and the Commonwealth of Puerto Rico with a CAE designation (National Initiative for Cybersecurity Careers and Studies, 2015). Once employed, entry-level cybersecurity workers can then work to develop top-tier skills through continuing education and on-the-job experience.

After a closer examination it appears that the most pressing issue regarding the cybersecurity labor market is finding innovative ways to help entry-level workers advance in cybersecurity careers. At the same time, all organizations may benefit from a new emphasis on meaningful cybersecurity awareness training. Formal academic programs offered at the community college and university level must continue to develop, but more innovation is needed in the areas of continuing education and noncredit training so new entrants into the cybersecurity workforce have access to the continuing education they need to advance in a cybersecurity career.

Community colleges innovating in the areas of noncredit cybersecurity training should therefore caution against replicating programs that are already offered as part of an academic program. On the other end of the training spectrum, community colleges should proceed cautiously when developing training designed for the top-tier cybersecurity workforce. Because experience is such a critical factor for the top-tier cybersecurity workforce, high-end training initiatives must have clearly defined goals developed in collaboration with employers and cost of development, delivery, and maintenance should be carefully considered.

The outlook for cybersecurity education and training is positive. The strategies outlined above offer insight into areas where community colleges can innovate in cybersecurity education and serve as a community resource for organizations with cybersecurity workforce needs. As the U.S. Labor Secretary Tom Perez accurately stated, "community colleges are incubators of innovation and opportunity" (Perez, 2014). Leading advancements in cybersecurity continuing education training is an excellent example of how community colleges can live up to this statement.

REFERENCES CITED

Burning Glass Technologies. (2015). *Job Market Intelligence: Cybersecurity Jobs, 2015*. Boston: Burning Glass Technologies.

Libicki, M. C., Pollak, J., & Senty, D. (2014). *H4CKER5 Wanted — An Examination of the Cybersecurity Labor Market*. Washington, DC: RAND Corporation.

National Initiative for Cybersecurity Careers and Studies. (2015, 09 04). *National Centers of Academic Excellence (CAE)*. Retrieved from https://niccs.us-cert.gov/education/national-centers-academic-excellence-cae

Perez, S. T. (2014, 29 2014). *Community Colleges: The Secret Sauce.* Retrieved from U.S. Department of Labor Blog: http://blog.dol.gov/2014/09/29/community-colleges-the-secret-sauce-2/

Verizon. (2015). *2015 Data Breach Investigations Report*. Verizon Enterprise Solutions.

AUTHOR

Wm. Michael Volk (wmvolk@aacc.edu) is an instructional specialist for the Cyber and Technology Training department at Anne Arundel Community College (AACC). In addition to his role at AACC, he serves as the chair of the Chesapeake Regional Tech Council Workforce Development Forum. Michael has experience working with individuals and organizations to identify and address their cybersecurity and workforce training needs. Prior to joining AACC he worked for the Mayor's Office of Employment Development as the Cybersecurity Navigator and Project coordinator. Michael holds a bachelor's degree in political science from McDaniel College, a master's degree in public administration from the University of Baltimore, and is CompTIA Network+ certified.

Leveraging Automated Animated Agent Commentary to Improve Sense-Making for Novice Users at Cybersecurity Competitions

Ruth Agada and Jie Yan

ABSTRACT

For this paper, we developed an animated agent that serves as a virtual commentator in small-scale cybersecurity competitions to educate and engage novice spectators. This virtual commentator is capable of expressing life-like behavior, much like a sensitive and effective human commentator. It provides a remarkable communication instrument to the human computer interface. Given the different cybersecurity competition exercises, the agent will act in much the same way a human commentator would act. Our goal in designing and implementing the virtual agent is to make the experience accessible to a wider community by providing an environment that would educate and prepare those without an extensive prior expertise. By observing the interactional behaviors of different professional commentators working with spectators, we identified specific verbal and nonverbal signals and cues exchanged during those interactions. We investigated two issues in this experiment: 1) the comprehension level of the spectator and 2) the perception of their learning experiences after interacting with the system. Experimental results showed that with the virtual commentator imbued with the same capabilities as its human counterparts, it has the same effect of educating novice level spectators to the cybersecurity dangers they may face in their daily lives. In addition, there is the added benefit of this system raising awareness in cybersecurity field.

Keywords: cybersecurity competitions, intelligent commentary system, animated agent.

INTRODUCTION

In recent years, public and private sector authorities in the United States have acknowledged that there is a shortage in the number of competent information assurance and cybersecurity professionals in the U.S. workforce and that there is a critical need to address this shortfall (Cheung, Cohen, Lo, Elia, & Carrillo-Marquez, 2012; O'Leary, 2012; Turner et al., 2015). In an effort to provide assistance in addressing this critical need, the academic community has responded by developing new programs in information assurance and devising creative ways to attract and train the next generation of cybersecurity professionals. Capstone courses with a significant hands-on element form a fundamental component of the cybersecurity curriculum. As such, these courses offer small-scale cyber defense exercises between groups of students which play an increasingly important role in providing hands-on knowledge (Werther, Zhivich, Leek, & Zeldovich, 2011). Also providing spectators, with little to no experience, with a vicarious experience.

As with any sporting event, in an effort to educate novice spectators about the game, there are always people to explain the highlights—what key players are doing, team stratagem, player/team statistics, history, and so forth. As evidenced by the millions of people who watch the sport—all at differing levels of experience—one can see why it is necessary to have experienced people, such as sports commentators, educate the spectators on the highlights of the game. This rule applies to any kind of sporting event from contact sport to non-contact sport. While there is some commentary in cybersecurity competition, they lack the same play-by-play commentary as seen in other sporting events. To meet this challenge, we propose an automated animated agent that serves as a virtual commentator in small-scale cybersecurity competitions, but an agent that is flexible enough to be employed in the Collegiate Cyber Defense Competitions (CCDCs) environment, an intelligent commentary system (ICS).

Studies (Bloom, 1984; VanLehn, 2011; Yu & Zhiping, 2008) have shown that effective human tutoring is the most powerful mode of teaching. In traditional learning environments, the teacher monitors the learner's affective state to promote not only positive learning affects, but by linking these states to more profound cognitive learning states, the learner achieves deeper states of learning. In trying to translate the aforementioned relationship to computers, there are several tools that are designed for the educational arena (Alexander, Sarrafzadeh, & Hill, 2006; Graesser et al., 2004; Malekzadeh, Mustafa, & Lahsasna, 2015). Recent literature (Malekzadeh et al., 2015) indicates that there is a similar symbiotic relationship in computerized learning environment. As opposed to the classic teacher—student interaction, for the competition environment it becomes necessary for the computerized learning environment to incorporate another skill set.

In a classroom setting, there are several options open to the instructor to help their students learn. First, they could leave the students to take their own learning into their own hands—an option generally left for the more independent student with a genuine interest in learning. Secondly, students could work one-on-one with tutors for additional instructions, which is the most effective, but is not practical (Bloom, 1984; VanLehn, 2011; Yu & Zhiping, 2008). The alternative to both previously mentioned strategies is to create a computer-aided learning system that follows the desired curricula commonly referred to as an intelligent tutoring system. Nevertheless, how does one define an intelligent tutoring system? As defined by several researchers (Cooper, Nam, & Si, 2010; Kollu, 2011; Malekzadeh et al., 2015; Merrill, Reiser, Ranney, & Trafton, 1992; Xochihua, 2012), intelligent tutoring systems (ITS) are advanced computer aided tutoring systems that provide customized instruction to the student, based on the domain and tutoring knowledge of the experts (Xochihua, 2012).

In this paper, we discuss a virtual commentator system for cybersecurity competitions that offer a positive learning experience for spectators — especially at the novice level. We present an intelligent commentary system with an embedded virtual human and in concert with other systems in the competition environment, the agent engages spectators by presenting information pertinent to understanding the real-time details of the game in a usable and spirited manner. In much the same way that the ITS serves as a supplementary tool in the classic teaching framework, the intelligent commentary system must adopt traits of an intelligent tutoring system and combine it with its programmed behavior as a sportscaster.

The main goal of this project is to develop a system to help the spectator make sense of all the information disseminated during the competition. The interdisciplinary term "sense-making" (Carvalho et al., 2012; Goodall & Sowul, 2009; Jajodia, Noel, Kalapa, Albanese, & Williams, 2011; Natarajan & Huang, 2010; Paul & Morris, 2009; Petre, 2013; Zhuo & Nadjin, 2012) is defined broadly as finding meaning in a situation (Paul & Morris, 2009). This concept is quite important in the field of human computer interaction (HCI), as it refers to the cognitive act of understanding information (Fisher, Counts, & Kittur, 2012; Paul & Morris, 2009). As an integral aspect in learning, any tool designed for teaching or otherwise will serve to enhance learning and understanding in network security principles and practice, especially in the form of a virtual sportscaster.

RELATED WORK

An overview of research into developing believable animated agents, capable of natural face-to-face conversations with people, are discussed in a variety of works (Bellegarda, 2010; Cole et al., 2003; Gratch, Marsella, & Rey, 2001; Johnson, Rickel, & Lester, 2000). Currently, researchers have generated powerful systems for modeling and controlling behavioral facial expressions of animated agents to make them believable, personable, and emotional. These behaviors and expressions range from facial emotions, gestures, gazes, and emotions conveyed in speech. Considering the novelty of the intelligent commentary system and the fact that such a system is built on the ITS, we need to review literature concerning ITSs. Table 1 lists some of the more popular intelligent tutoring systems. Malekzadeh, Mustafa and Lahsasna (2015) provide comprehensive work on the research on ITS's and listed below is a condensed version of the table they created.

TABLE 1: SUMMARY OF SEVERAL ITS AND THEIR IMPLEMENTATIONS

APPLICATION	IMPLEMENTATION
WEB-BASED LEARNING SYSTEM (STRAIN & D'MELLO, 2011)	Case—based system used to teach the U.S. Constitution and Bill of Rights.
VIRTUAL TUTOR FOR DATA STRUCTURE (CHAFFAR, DERBALI, & FRASSON, 2009)	Problem—based system to teach data structures in web courses.
WAYANG INTELLIGENT TUTOR (WOOLF ET AL., 2009)	Heuristic driven tool to teach mathematics.
AUTOTUTOR (GRAESSER ET AL., 2004)	Conversational agent to teach Newtonian physics, computer literacy, and critical thinking.
ITS FOR DATABASE DESIGN (ZAKHAROV, MITROVIC, & JOHNSTON, 2008)	Feedback—based system that teaches database design.
INTELLIGENT E-LEARNING SYSTEM (MAO & LI, 2009)	Affect detecting system to teach concepts in affect computing.
CHINESE TEXT-BASED E-LEARNING SYSTEM (TIAN ET AL., 2014)	Interactive text based system to recognize negative affect.
ITS FOR SCATTERPLOTS (RODRIGO ET AL., 2012)	Conversational agent to help interpret scatterplot data.

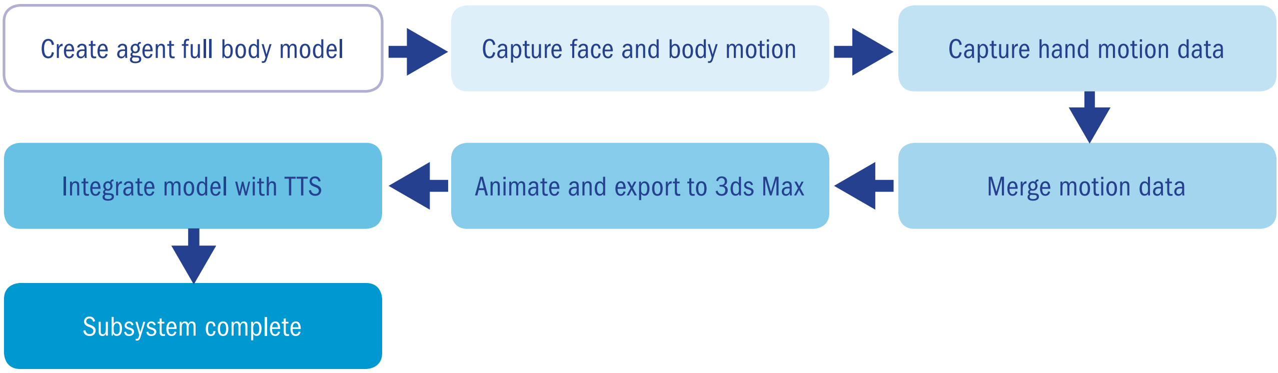
With the above-mentioned applications and the levels of success for the area they were applied to, we see no instance of an intelligent tutoring system acting as a commentator for competitive events. The competition environment can be viewed as a different kind of classroom where the instructor is not only boisterous but has real-time updates about the state of all the "learners in the classroom." This is where the novelty of this application appears.

INTELLIGENT COMMENTARY SYSTEM FRAMEWORK

Separate from the other system that make up the competition environment, the ICS goes through several steps before it is included in the entire environment. Figure 1 illustrates the design workflow of the ICS subsystem.

FIGURE 1: AGENT DESIGN WORKFLOW

This figure illustrates the design and development of the ICS.


Create agent full body model
Capture face and body motion
Capture hand motion data
Merge motion data
Animate and export to 3ds Max
Integrate model with TTS
Subsystem complete

DATA COLLECTION

In order for the agent to simulate human commentator behavior, we initially used video footage of contact sports commentators. In Figure 2, we see example footage of a commentator from NFL roundtable discussions, Super Bowl commentators, and Apollo Robbins' TED talks (Robbins, 2013; Sports, 2014; Walkowicz, 2016). Each video file is roughly 9 minutes in length. These contain commentator motion data that can be applied to the animated agent. In October 2014 and again in October 2015, we attended the Maryland Cyber Challenge in which we interviewed (and recorded) several sponsors of the competition as well as team coaches, competitors, and spectators. In 2014, the interview with the sponsors yielded 30 minutes of footage as seen in Figure 3. In 2015, interviews of several participants of the competition and some sponsors yielded roughly 45 minutes of raw footage. We were able to create frame-by-frame shots to aid in the analysis of facial and body gestures.

FIGURE 2: SAMPLE VIDEO CLIPS FROM YOUTUBE.COM

Clips of sports hosts and speakers making varied gestures.

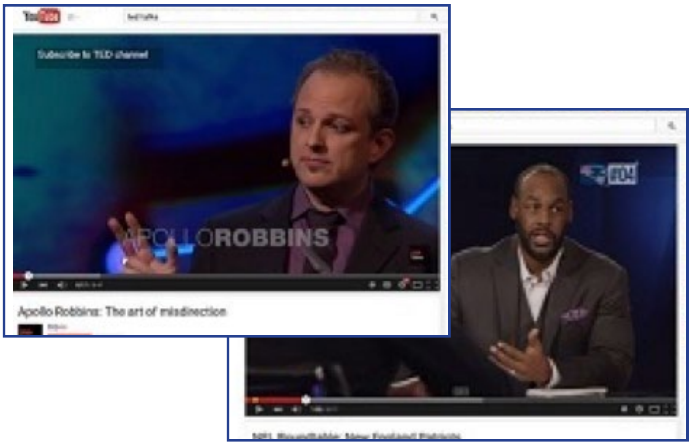
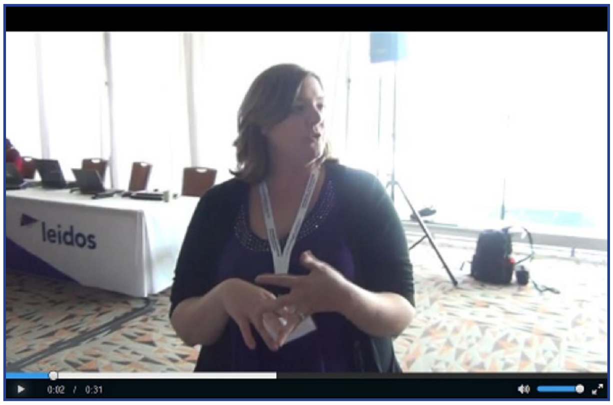

Subscribe to TED channel
APOLLO ROBBINS
Apollo Robbins: The art of misdirection
NFL Roundtable: New England Patriots

FIGURE 3: CYBERSECURITY COMPETITION FOOTAGE

Data collected from an October 2014 cybersecurity competition in Maryland.

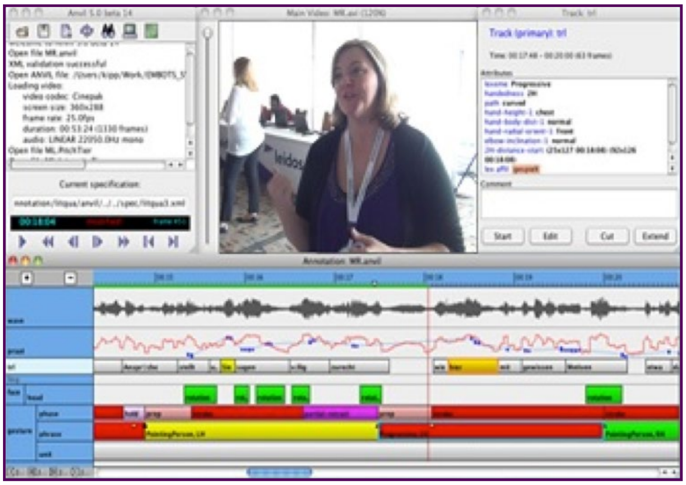


DATA ANNOTATION

We observed several motions/gestures across all the collected data. To annotate the footage, we used a tool called ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) as seen in Figure 4. With ELAN a user can add an unlimited number of annotations to audio and/or video streams. An annotation can be a sentence, word or gloss, a comment, translation, or a description of any feature observed in the media. Annotations can be created on multiple layers, called tiers. Tiers can be hierarchically interconnected. An annotation can either be time-aligned to the media or it can refer to other existing annotations. The textual content of annotations is always in Unicode and the transcription is stored in an XML format.

FIGURE 4: ELAN USER INTERFACE

Annotated video clip using ELAN.



Open file MR.anvil
XML validation successful
Open ANVIL file: /Users/kipp/Work/EMBOTS_S
Loading video:
video codec: Cinepak
screen size: 360x288
frame rate: 25.0fps
duration: 00:53:24 (1330 frames)
audio: LINEAR 22050.0Hz mono
Open file ML.PitchTier

Current specification:
nnotation/lingua/anvil/../../spec/lingua3.xml

Track (primary): trl

lexeme: Progressive
handedness: 2H
path: curved
hand-height-1: chest
hand-body-dist-1: normal
hand-radial-orient-1: front
elbow-inclination-1: normal

With ELAN, we set up multiple tiers to investigate specific behaviors. Since the data is time-aligned, this allows us to mark the start and end time for each behavior. Listed below are some of the more frequently used gestures for each tier.

- Hand gestures
  - Counting with fingers
  - Pointing
  - Palm facing out
  - Clasped hand
  - Ushering motion
  - 2-finger pointing
- Body motion (generally kept to a minimum)
  - Gentle body sway
  - Lean forward/back
  - Pitch forward/back
- Facial expression / head motion
  - Raised eyebrows
  - Subtle smile
  - Surprise
  - Head nod up-down/left-right
  - Head tilt

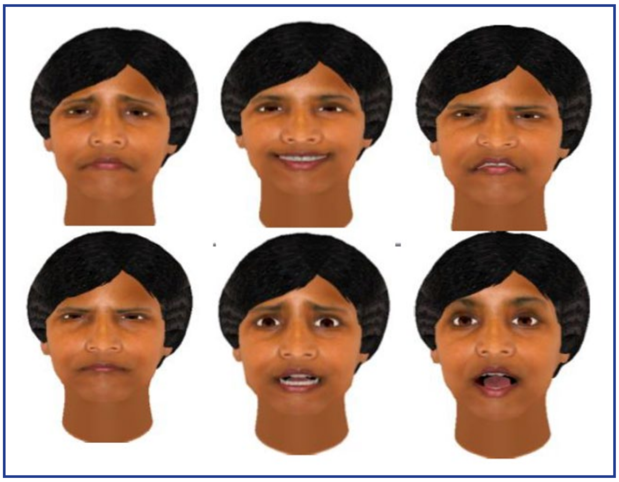
We observe that in all these interactions, there is a lot of gesticulating with the hands to emphasize certain points or to draw attention to certain areas of the competition environment—or in the case of the data we collected—areas of the screen displaying competition status. We also observe that all facial expressions range from subtle to extreme positive expressions. These are challenges that we have to take into account when modeling the expressions to the embedded virtual human in the ICS.

AGENT DEVELOPMENT

In order to build an affect-capable ICS, we implement a method for modeling different affective states exhibited by most engaging sportscasters. The system-generated images in the 3-D scene are generated in two steps. First, in the modeling step, the system produces a precise description of the agent, in terms of graphics primitives. Then it acquires the vertex data required for the primitives from 3-D modeling software that can generate vertex data. The data provided for the development of the animated agent contain over one thousand points in 3-D space. Secondly, in the rendering step, the vertex data also serve as a means to draw the model to the screen. Connecting the vertex data using a series of polygons generates a wireframe image. This image applies textures to the frame to create the face of the agent. The agent is able to animate starting from a neutral position and then transitioning into one of the six basic facial expressions (fear, happiness, surprise, anger, sadness, and disgust) as shown in Figure 5.

FIGURE 5: SAMPLE AGENT MODEL

The six basic emotions the agent can express. The emotion set from top-left to bottom-right are sadness, happiness, disgust, anger, fear, and surprise.



For the system to simulate a smooth transition in the agent's facial expression as well as different full body gestures, a method would need to generate a sequence of vertices given that the only known vertices are the start and end vertices. By studying and annotating the video clips we collected, we were able to model realistic animation patterns. With the help of third-party applications, the animation is natural-looking and includes head and face movements combined with facial movements (eyes, eyebrows, nose, cheeks, and so forth) and full body movements (arms, hands, fingers, torso, legs, and so forth).

We modeled anthropomorphic characteristics and behaviors of expert commentators in the 3-D animated agent by macro and micro rules and patterns governing hundreds of different nuances characteristic of different modes of communication, interaction, and feedback. At the macro level, semantic intent, dialog act, and past context of the agent and interlocutor both express the overall mood and emotion of the agent. At the micro level, additional nuances relating to discourse, expression, and backchannel communication are involved. For example, when a character intends to talk, he/she usually begins by smiling and then slightly tilting the head up and down in order to get the user's attention. After speaking for a while, he/she may blink, change the head position a bit, and so forth.

To generate a marked up animation sequence or add additional markup to a sequence, a series of animation tags are inserted at the phoneme level into a previously generated animation sequence that was inputted manually or obtained from a speech recognition alignment or speech synthesizer.

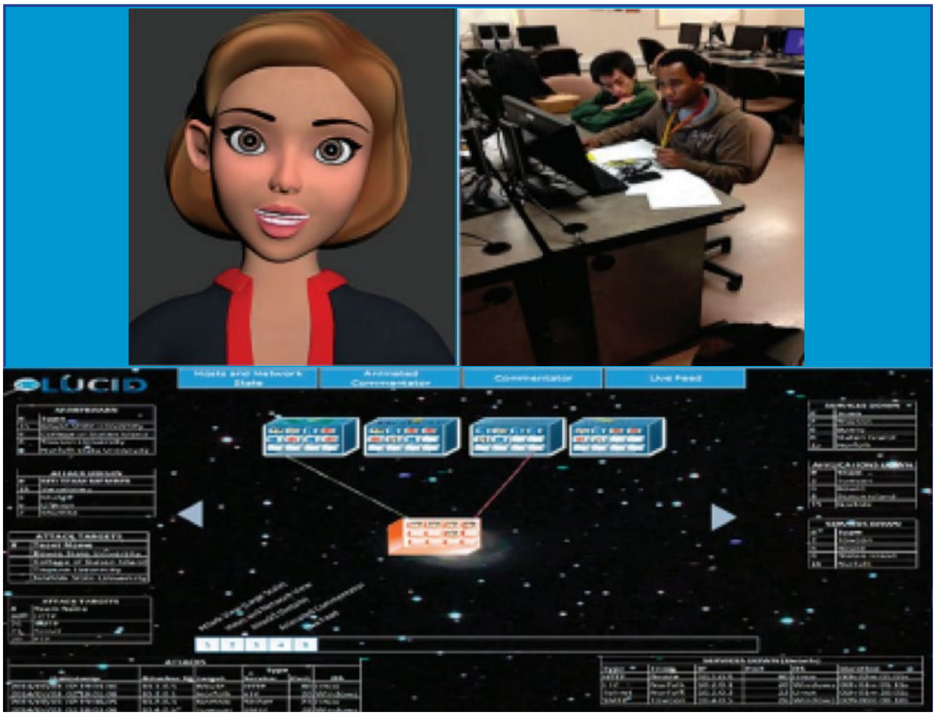
DISCUSSION

We conducted initial experiments on the virtual human alone to determine the impact of an affect capable intelligent commentary system on its viewers. In an experiment performed by Agada (Acquaah, Agada, & Yan, 2013; Agada & Yan, 2012), participants were presented with three different versions of the same virtual agent. In presenting randomly selected participants with one of the three clones—all at different affect levels—they can make certain correlations between the learners' comprehension and agent perception to the affect level of the virtual agent. In those experiments, we see that the affect capable virtual agent outperformed all other clones in comprehension tests and agent perception.

Even though the results from the previously mentioned results were favorable, the system was designed as a teacher and expressed behaviors exhibited by teachers. Challenges arise if the same effects are observed when the agent is used as an intelligent commentary system. As Baylor (2011) noted, the impact a virtual human has on the learner's motivation learning gains yields the same results (Alexander et al., 2006; Samuel Thomas Vaughan Alexander, 2007). On its own, the intelligent commentary system is a sufficient education tool but in concert with other aspects of the entire competition environment, it has a potential to be a very powerful tool. The ICS would receive data about the network and competitors from network, visualizations, and scoring subsystems and would relay that information to the novice spectator. In Figure 6, we see a mockup of the overall system integrated with the ICS with embedded agent.

FIGURE 6: SYSTEM MOCKUP

The interface shows one of the models act as the virtual commentator.



CONCLUSION

This article discusses the implementation of an embedded virtual human in an intelligent commentary system targeted to improving spectators' ability to understand and make sense of cybersecurity competitions. The system aims to engage the spectator by presenting information pertinent to understanding the real-time events of the game as well as relevant fundamental information. Our goal in designing and implementing the virtual agent is to make the experience accessible to a wider community by providing an environment that would educate and prepare those without an extensive prior expertise. By observing the interactional behaviors of different professional commentators working with spectators, we identified specific verbal and nonverbal signals and cues exchanged during those interactions. We investigated two issues in this experiment: 1) the comprehension level of the spectator and 2) the perception of their learning experiences after interacting with the system. Experimental results showed that with the virtual commentator imbued with the same capabilities as its human counterparts, it has the same effect of educating novice level spectators to the cybersecurity dangers they may face in their daily lives.

ACKNOWLEDGMENT

This research was conducted in the computer perception and animation lab (CPAL) which is headed by Dr. Jie Yan of the Computer Science Department at Bowie State University. This work is a NSF supported grant: NSF-DUE 1303424.

REFERENCES CITED

Acquaah, K., Agada, R., & Yan, J. (2013). A real-time emotion mirror system. In *Annual International Conference on Computer Games* (p. 131). http://doi.org/10.1037/e551172013-008

Agada, R., & Yan, J. (2012). Research to improve communication by animated pedagogical agents. *Journal of Next Generation Information Technology*, 3(1), 58. Retrieved from http://connection.ebscohost.com/c/articles/76342306/research-improve-communication-by-animated-pedagogical-agents

Alexander, S., Sarrafzadeh, A., & Hill, S. (2006). Easy with Eve: A functional affective tutoring system. In *Workshop on motivational and affective issues in ITS-8th international conference on intelligent tutoring systems* (pp. 38–45). Taiwan.

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, 59(2), 291–300. http://doi.org/10.1007/s11423-011-9196-3

Bellegarda, J. R. (2010). A Data-Driven Affective Analysis Framework Toward Naturally Expressive Speech Synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5), 1113–1122. http://doi.org/10.1109/TASL.2010.2078808

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. http://doi.org/10.3102/0013189X013006004

Carvalho, M., Bradshaw, J. M., Bunch, L., Eskridge, T., Feltovich, P. J., Hoffman, R. R., & Kidwell, D. (2012). Command and Control Requirements for Moving-Target Defense. *IEEE Intelligent Systems*, 27(3), 79–85. http://doi.org/10.1109/MIS.2012.45

Chaffar, S., Derbali, L., & Frasson, C. (2009). Inducing positive emotional state in intelligent tutoring systems. *Frontiers in Artificial Intelligence and Applications*, 200(1), 716–718. http://doi.org/10.3233/978-1-60750-028-5-716

Cheung, R. S., Cohen, J. P., Lo, H. Z., Elia, F., & Carrillo-Marquez, V. (2012). Effectiveness of Cybersecurity Competitions. In *Proceedings of the International Conference on Security and Management (SAM)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). Retrieved from http://world-comp.org/p2012/SAM6108.pdf

Cole, R., Van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., … Yan, J. (2003). Perceptive animated interfaces: First steps toward a new paradigm for human-computer interaction. *Proceedings of the IEEE, 91(9)*, 1391–1404. http://doi.org/10.1109/JPROC.2003.817143

Cooper, S., Nam, Y. J., & Si, L. (2010). Initial results of using an intelligent tutoring system with Alice. In *17th ACM annual Conference on Innovation and Technology in Computer Science Education*. Retrieved from http://dl.acm.org/citation.cfm?id=2325332

Fisher, K., Counts, S., & Kittur, A. (2012). Distributed sensemaking. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems—CHI '12* (p. 247). New York, New York, USA: ACM Press. http://doi.org/10.1145/2207676.2207711

Goodall, J. R., & Sowul, M. (2009). VIAssist: Visual analytics for cyber defense. In *2009 IEEE Conference on Technologies for Homeland Security, HST 2009* (pp. 143–150). http://doi.org/10.1109/THS.2009.5168026

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers, 36(2)*, 180–192. http://doi.org/10.3758/BF03195563

Gratch, J., Marsella, S., & Rey, M. (2001). Tears and Fears : Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the fifth international conference on Autonomous agents* (pp. 278–285). New York: ACM. http://doi.org/10.1145/375735.376309

Jajodia, S., Noel, S., Kalapa, P., Albanese, M., & Williams, J. (2011). Cauldron mission-centric cyber situational awareness with defense in depth. In *2011—MILCOM 2011 Military Communications Conference* (pp. 1339–1344). IEEE. http://doi.org/10.1109/MILCOM.2011.6127490

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated Pedagogical Agents : Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.

Kollu, K. (2011). *Prototype of an intelligent tutoring system using the java expert system shell*. Temple University.

Malekzadeh, M., Mustafa, M. B., & Lahsasna, A. (2015). A review of emotion regulation in intelligent tutoring systems. *Educational Technology and Society*, 18(4), 435–445.

Mao, X., & Li, Z. (2009). Implementing emotion-based user-aware e-learning. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems—CHI EA '09* (p. 3787). New York, New York, USA: ACM Press. http://doi.org/10.1145/1520340.1520572

Merrill, D. c., Reiser, B. j, Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences*, 2(3), 277–205. Retrieved from http://www.jstor.org/stable/1466610

Natarajan, S., & Huang, X. (2010). An interactive visualization framework for next generation networks. In *Proceedings of the ACM CoNEXT Student* ... (p. 1). http://doi.org/10.1145/1921206.1921222

O'Leary, M. (2012). Small-Scale Cyber Security Competitions. In *Proceeding of the 16th colloquium for information systems education*. Lake Buena Vista: Towson University.

Paul, S. A., & Morris, M. R. (2009). CoSense. In *Proceedings of the 27th international conference on Human factors in computing systems—CHI 09* (p. 1771). New York, New York, USA: ACM Press. http://doi.org/10.1145/1518701.1518974

Petre, M. (2013). MOOCs schmoocs. *ACM Inroads*, 4(4), 22–23. http://doi.org/10.1145/2537753.2537762

Robbins, A. (2013). *TED talks—The art of misdirection*. TED talks. Retrieved from https://www.youtube.com/watch?v=GZGY0wPAnus

Rodrigo, M. M. T., Baker, R. S. J. D., Agapito, J., Nabos, J., Repalam, M. C., Reyes, S. S., & San Pedro, M. O. C. Z. (2012). The Effects of an Interactive Software Agent on Student Affective Dynamics while Using ;an Intelligent Tutoring System. *IEEE Transactions on Affective Computing*, 3(2), 224–236. http://doi.org/10.1109/T-AFFC.2011.41

Samuel Thomas Vaughan Alexander. (2007). *An affect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor*. Massey University.

Sports, F. (2014). *NFL Roundtable: San Francisco 49ers*. USA: Fox sports live. Retrieved from https://www.youtube.com/watch?v=V8KYDaBfiUA

Strain, A. C., & D'Mello, S. K. (2011). Emotion Regulation during Learning. In *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 566–568). http://doi.org/10.1007/978-3-642-21869-9_103

Tian, F., Gao, P., Li, L., Zhang, W., Liang, H., Qian, Y., & Zhao, R. (2014). Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems. *Knowledge-Based Systems*, 55, 148–164. http://doi.org/10.1016/j.knosys.2013.10.019

Turner, C., Yan, J., Richards, D., Brien, P. O., Odubiyi, J., & Brown, Q. (2015). LUCID: A visualization and broadcast system for cyber defense competitions. *ACM Inroads*, 6(2), 70–76. http://doi.org/10.1145/2746408

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. http://doi.org/10.1080/00461520.2011.611369

Walkowicz, L. (2016). *TED talks — Let's Not Use Mars as a Backup Planet*. TED talks. Retrieved from https://www.youtube.com/watch?v=h2KQoHMCwlw

Werther, J., Zhivich, M., Leek, T., & Zeldovich, N. (2011). Experiences In Cyber Security Education: The MIT Lincoln Laboratory Capture-the-Flag Exercise. In *Proceedings of the 4th Workshop on Cyber Security Experimentation and Test*.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. http://doi.org/10.3758/BRM.41.3.591

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognising and responding to student affect. *International Journal of Learning Technology*, 4(3/4), 129–164. http://doi.org/10.1504/IJLT.2009.028804

Xochihua, O. A. (2012). *A mixed-response intelligent tutoring system based on learning from demonstration*. Texas A&M University. Retrieved from http://repository.tamu.edu/handle/1969.1/ETD-TAMU-2012-05-10793?show=full

Yu, S., & Zhiping, L. (2008). Intelligent pedagogical agents for intelligent tutoring systems. *2008 International Conference on Computer Science and Software Engineering*, 516–519. http://doi.org/10.1109/CSSE.2008.414

Zakharov, K., Mitrovic, A., & Johnston, L. (2008). Towards Emotionally-Intelligent Pedagogical Agents. In *Intelligent Tutoring Systems* (pp. 19–28). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-69132-7_7

Zhuo, W., & Nadjin, Y. (2012). MalwareVis: entity-based visualization of malware network traces. ... *of the Ninth International Symposium on Visualization* ..., 41–47. http://doi.org/10.1145/2379690.2379696

AUTHORS

Ruth Agada (agadar0208@students.bowiestate.edu) is a doctoral candidate of the Department of Computer Science at Bowie State University, advised by Dr. Jie Yan. Her research interests include computer vision and machine learning, the areas of facial expression recognition, multi-view face detection, tracking and recognition, 3D character animation, animated agents for tutoring applications, hand gesture recognition, and animation synthesis. Currently, she has been investigating holistic features in facial expression data for valence classification. She earned her bachelor's and master's degrees in computer science from Bowie State University in 2009 and is currently working on her doctorate in science.

Dr. Jie Yan (jyan@bowiestate.edu) is an associate professor in the Department of Computer Science at Bowie State University. She received her doctorate from the Harbin Institute of Technology in China in 1999. After graduation, she worked as an associate researcher for Microsoft Research Asia. Prior to joining Bowie State University, she was a research associate in the Center for Spoken Language Research at the University of Colorado at Boulder. Her research interests include computer graphics, animation, computer vision, and pattern recognition. She has published numerous research articles in scholarly journals and conferences. She also holds three patents in her research area.

Vulnerability Risk Modeling: Combining Cybersecurity and Epidemiology to Enhance National Defense

R. David Parker and Michael D. Regier

ABSTRACT

The FY2016 U.S. defense budget includes $14 billion in funding across the entire government for cybersecurity-related programs. In the U.S. and abroad, one type of program which is not apparent is a system to quantify vulnerability estimates. Identifying security gaps provides opportunities to direct resources before an attack and may focus on offensive or defensive methods. Traditional approaches may lack an individual level focus, especially one which considers aspects which are most susceptible to social engineering. For instance, a system can be negatively impacted by the number of users, percent of active antivirus use, password strength, and adherence to security protocols. Recent international events demonstrate that the individual is an often overlooked but serious potential security flaw.

Validated methods which combine outcomes from people and can include other measurable, social factors are available. For instance, infectious diseases modeling, which combines biological, environmental, and social data could add value and save resources if adapted specifically for cybersecurity deployment. Extant data may be used to determine areas of security concern via easily interpretable metrics of evaluation. This proven methodology requires only adaptation, testing, and adoption for both defensive and offensive operations.

Keywords: privacy, security and protection, modeling methodology

INTRODUCTION

The documented frequency and occurrence of cyber attacks against the United States government, businesses, and citizens is increasing with an estimated one-third of the population impacted in 2014 alone (DiMase, Collier, Heffner, & Linkov, 2015). While the defining concepts of cyber warfare and cyber terrorism are nebulous, there is an increasing use of cyber tactics (Schmitt, 2015). As the use of third parties for attacks grows around the world, offensive measures may prove less effective against unknown opponents than would more stringent defensive measures. The use of these attacks have been to disable infrastructure, increase chaos in areas of civil unrest, as well as to exact retribution against businesses, groups, and governments who have been perceived to perpetrate negative actions against others (Schmitt & Vihuil, 2014).

EPIDEMIOLOGIC & PUBLIC HEALTH MODELS

Two concepts which could significantly benefit national defense through adaptation in cybersecurity are public health surveillance systems (PHSS) and estimated risk prediction. These methods are currently used to measure the effect and risk of infectious diseases across the world (Broz et al., 2014; Lanier, Johnson, Rolfs, Friedrichs, & Grey, 2012; Meynard et al., 2008). Given our premise of quantifying macro level risk emanating from micro level causes, infectious diseases modeling would be similar given how many such diseases are spread through social networks. Using the general framework of a PHSS, data could be collected across multiple domains, groups, and units of interest. Units of interest may be defined as individuals, businesses, governmental organizations, countries, or other geographic areas for which data are collected and for which statistical models are developed to identify system or protective gaps. Prevention is a key aspect in public health as it is easier and more cost effective to prevent illness or to stop cascading effects once a vulnerability has been exploited. These are very similar concepts in security and defense systems.

PHSSs are dynamic systems importing and linking data from multiple sources to build a comprehensive overview, allowing effective identification of potential intervention points (WHO, 2000, 2011). Data collection is standardized allowing consistent evaluation and analyses which increase understanding of the issues at present and predict the direction of spread throughout a population. Due to their communicable nature, infectious diseases

(ID) and their modeling require a different approach than that of chronic diseases. One of the primary distinctions is the social aspects and their ability to enhance the spread of disease by increasing the potentially susceptible population.

Recently, the Ebola outbreak in Western African nations dominated global—especially Western—media and was quickly declared a major security and defense concern (Gostin, Waxman, & Foege, 2015). Similarly, the United Nations declared AIDS an international defense issue in 2000 *(United Nations Security Council Resolution 1308 (2000) on the Responsibility of the Security Council in the Maintenance of International Peace and Security: HIV/AIDS and International Peace-keeping Operations,* 2000). The connection between international defense and health security is increasingly established (Aldis, 2008; Parker, 2011). As the connection between the two is better understood, the ability to adapt established methods from one area into another is a natural extension.

Cyber defense, as part of national defense, possesses many components that mirror the spread of diseases. For instance, the terms virus and bacteria in information technology were borrowed from the field of epidemiology as the actions of written code impacts IT systems from a similar perspective as viruses attack biological systems. In epidemiology, identification and understanding a pathogen in an outbreak requires a combination of multiple data sources based on the location or unit of interest. The primary identification point is that of susceptible and infected units of interest; in ID the unit is persons. Susceptibility is measured as the probability that a person is at risk of exposure and acquisition of the condition of interest. Infected persons are exposed, susceptible persons who contracted the condition and may therefore infect other susceptible persons. Not all susceptible persons will be exposed or become infectious. Not all infectious persons will transmit the condition if they are not exposed to a susceptible person.

This is where we join the two worlds of epidemiology and defense, through the exportation of disease modeling using identifiable, susceptible units (or vulnerabilities) ahead of time on an aggregate level, such as a government. Table 1 lists risk domains with examples of potential measurements within each area. Identification of vulnerable areas using consistent measurement allows a pre-emptive allocation of resources. Better understanding of risky conditions at the individual level allows a more precise determination of aggregated risk. For effectiveness and replicability, a statistic or model that is easily interpretable, and can be validated and retested in an expeditious manner would deliver the most information.

TABLE 1: CONCEPT DOMAINS FOR RISK ASSESSMENT

RISK DOMAINS	CONCEPT DOMAIN MEASUREMENT
INFORMATION TECHNOLOGY SATURATION	Cumulative numbers derived from the measurement of factors such as mobile phone use, cellular, and mobile data trends, internet use, among others in the estimated population at midpoint of year.
FOREIGN POLICY	Key factors for threats for conventional war or disagreements. Cumulative measurement derived from total offensive and defensive measures against others (internal and external).
ALLIES	The impact of foreign governments with which positive formal or informal alliances exist.
ASSETS	Resources other nations may want to possess.
LEVEL OF SECURITY	The ability or perceived ability of a state to defend itself against a cyber attack or a traditional military operation.
POPULATION CHARACTERISTICS	Elements which may influence the impact of a cyber attack on the citizenry (general or subpopulation). Age Education level /literacy Internet use Income Geography (rural v. urban) General health Special health considerations linked to critical care

ESTIMATING RISK

Risk estimation is utilized in many fields, including business, defense, and public health (S. L. Cutter, Boruff, & Shirley, 2003; S. L. F. Cutter, Christina, 2008; Schmidtlein, Deutsch, Piegorsch, & Cutter, 2008; US, 2009). Table 1 provides an overview or risk domains. However, often times, the attempts to predict risk actually use models which are designed not to predict, but rather use associative modeling to describe the relationship between factors. Associative and predictive models can be fundamentally similar from a mathematical perspective, but they differ in data use as well model construction, assessment, and validation (Kuhn & Johnson, 2013). Additionally, previous research has demonstrated that when there is a dearth of data, the known data can be utilized to construct an empirical approximation (estimate) of the underlying population from which additional, similar data may be simulated. This approach increases the overall ability to predict outcomes in the short term (O'Neill, Balding, Becker, Eerola, & Mollison, 2000; Yang, Longini Jr, Halloran, & Obenchain, 2012).

STATISTICAL MODELS

Multivariable models are used to estimate risk across a diverse array of fields, ranging from epidemiology and statistics to risk assessment for insurance, banking and fraud detection. The methods are utilized more and more in other fields for prediction, such as the US Geological Service predicting wild fire debris fields and in the prediction of mobile malware attacks and defense in IT (Dunham, 2008; Rupert, Cannon, Gartner, Michael, & Helsel, 2008). A commonality across all the diverse areas of risk modelling is the increased use of machine learning (statistical learning) models and algorithms. Two well-known machine learning statistical models are linear regression (linear discriminant analysis) and logistic regression. We will use these two models as illustrative archetypes.

A fundamental objective when using a regression model (e.g. linear or logistic) is to relate a set of variables (e.g. predictors, explanatory variables, regressors) to an observed response (e.g. outcome, regressand). (Freedman, 2009) In disease modeling, regressions are selected because of the ability to include numerous variables for consideration. These models not only articulate the relationships between the regressors and the outcomes modeled, but also have the potential to shed light on other types of relationships inherent in the collected data (e.g. conditional R2, additive interaction, multiplicative interaction, collinearity, and curvature). Regression techniques are widely used, but frequently they come with unsatisfactory limitations.

Two such limitations are linearity in the regressors and linearity of the model coefficients. The first limitation—linearity in the regressors—can be addressed by transforming the data. Variable transformations may range from the simple, such as adding a power term (e.g. quadratic term), to the more complex, such as spherical smoothing. In addition, there are regression techniques designed to incorporate such transformations (e.g. polynomial regression and fractional polynomials) or by approximating a complex relationship using piecewise linear models (e.g. splines), allowing the models to better characterize the underlying structure of the data (Du, 2012; Gillies, 2010; Regier & Parker, 2015; Royston, Ambler G Fau—Sauerbrei, & Sauerbrei). The use of nonlinear models addresses the constraint of linearity of the model coefficients by permitting nonlinear functions of the parameters. The exponential model and fractional polynomials are regression examples, while random forests and neural networks are examples from machine learning (Amato et al., 2013; Breiman, 2001; Regier & Parker, 2015; Rencher & Schaalje, 2008; Royston & Altman, 1994).

While there is much activity surrounding the development and application of methods to overcome the two linearity constraints, the literature is sparse concerning the inferential objectives that underscore statistical modeling. Whether a decision is implicitly or explicitly made, statistical models are constructed, in general, for either associative (e.g. descriptive, causal) or predictive relationships (e.g. prediction, forecasting). Associative models allow an assessment of the impact on the outcome for the unit at risk. These techniques can accommodate variables from different types and sources of data providing an interpretable measurement of each variable's contribution to explaining the outcome conditional on the variables included in the model. Therefore, the contribution of information within one variable may change as different variables are included or removed from the associative model and the associated impact of the change can be quantified (e.g. conditional R2).

For a predictive model, the target of inference is not necessarily the contribution of a variable, coefficient magnitude or effect size to the explanation of the observed outcome, as indicated in the previous paragraph. The target of inference, for predictive models, is how accurately we can project forward that an event will (or will not) occur. The projection forward may be in time or to a new observation from a similar or related population. This stands in contrast to the underlying objective of an associative model; the prediction model paradigm does not feature parameter interpretation as a required product of the modeling exercise.

With a different target of inference for predictive modeling, a different approach to the use of the collected data is required. For associative models, the primary objective is to describe the variable-outcome relationship. The data is collected, prepared for analysis, and then analyzed using an appropriate associative model. The entire data set is used to construct and assess the model; it is a single use model construction process. In contrast, a predictive model is constructed to say something about a new or potentially new observation; the inference is to data not yet collected. A single data set must be used to reflect the inferential goal of constructing a predictive model, assessing the quality of the predictions, and then having a utility for predicting an outcome using a new, independent observation or data set. To appropriately assess the quality and utility of a predictive model, it must be constructed in a manner that reflects the intended use. To reflect the intended use of a predictive model, the collected data is re-used in order to construct (train) the model and then to provide an accurate assessment of the model's utility.

The simplest data re-use approach is to split the data into two parts: a training and a test data set. The first part is used to construct the predictive model and the second part assesses the model (e.g. accuracy of prediction, sensitivity, and specificity). This is the simplest way to obtain a training data set used to construct the predictive model and an independent test data set to assess the model's utility. Given the high costs of data, this may be an inefficient use of a valuable resource, thus data re-use methods have been developed that can provide unbiased measures of model quality (e.g. bootstrap and cross-validation) (Hastie, Tibshirani, & Friedman, 2009; Kuhn & Johnson, 2013).

The inherent challenge with these preferred methods is the diversity of models obtained from the data re-use methods and the lack of methodologies for integrating the results for a coherent exposition of predictive importance and strength of association. A further complication is that common methods used to assess the quality of an associative model are often not the preferred methods for assessing predictive models. However, this does not prevent individuals from using associative models as predictive models, though the inferential objectives are not the same, nor are the metrics of model quality and utility.

There is a resulting paucity of literature about the extraction of a predictor-outcome relationship measure for predictive models, examples of which are shown in Table 2. Additionally, there is limited research for the identification of important predictors, those that are critically important for the development of a highly useful predictive model. A notable exception has been with random forests where effort has been dedicated to

TABLE 2: EXAMPLE OF IMPORTANT PREDICTOR VARIABLES IDENTIFIED BY A PREDICTIVE MODEL

RISK DOMAINS	CONCEPT DOMAIN MEASUREMENT	(e.g. RANDOM FORESTS)
INFORMATION TECHNOLOGY SATURATION	Number of computers in use	✓
	Number of Internet users	✓
	Number of active mobile phone numbers	✓
	Number of users of social media (Facebook, Twitter)	
POPULATION CHARACTERISTICS	Factors which may influence the impact of a cyber-attack on the general or special population.	
	Age	✓
	Education level/literacy	✓
	Internet use	✓
	Income	
	Geography (rural v. urban)	
	General health	✓
	Special health considerations linked to critical care	

developing method for identifying a subset of variables that are critically important (Table 2) (Breiman, 2001; Zani, 2006). Table 2 is an example identifying, with check marks, the subset of predictors that are deemed most important over the entire forest of classification trees explored by the random forests methodology.

Although random forests provide information about which predictors are important, no interpretable quantification of the strength of the relationship between the predictor variable and outcome is provided (Table 3). Table 3 is an example of the strength of relationship, using the domain specific odds ratio that would provide insight about the domain-outcome associations. The challenge is to link predictive methodologies and results with associative model interpretations to provide high utility predictive models that yield deeper insight into domain-specific relationships. This link will provide richer information for developing interventions and will require an evolution in the concept of statistical practice (Hoerl & Snee, 2011; Meng, 2009). An initial strategy is to use the identification of important predictors as a model selection technique (Table 2) and then use the identified predictors for the development of associative statistical models (Table 3). The use of correct modeling, with existing data, and an interpretable outcome can move our cyber defense strategies forward beyond our current capability.

TABLE 3: EXAMPLE OF INTERPRETABLE DOMAIN OUTCOME

FIELD	ODDS OF SUCCESSFUL ATTACK	SIGNIFICANCE LEVEL	95% CI
OVERALL SYSTEM	1.40	0.01	1.14, 1.71
SECTOR (PUBLIC)	2.85	0.01	1.30, 6.21
CREDIT UNIONS 1	2.86	0.01	1.46, 5.61
STOCK MARKETS 1	7.59	0.01	2.21, 26.12
LOCAL LEVEL GOVT 2	27.56	0.01	6.17, 122.99
STATE LEVEL GOVT 2	11.83	0.01	2.73, 51.30
1 compared to banking institution			
2 compared to national government			

TRANSLATION TO PRACTICE

Next steps include system design and development. The major need is for data in areas that are known as high risk, areas of perceived low risk, and areas of known low risk. Such a system would enable the United States to determine a cumulative probability of successful cyber attacks against multiple systems (i.e. financial, governmental, communications, and critical infrastructure—including civilian and defense). Methods based on validated epidemiologic statistical modeling for infectious diseases using real world, extant data, creates a flexible, scalable and extendible risk prediction framework. Within each domain, risk probabilities can be calculated for smaller discrete components, such as counties within a state, banks within the financial sector or departments within an institution. The macro and micro level probability estimates allow the most ‘at risk’ areas to be targeted for resource allocation, thereby reducing duplicative efforts (i.e., securing SCADA systems within the electrical grid).

The novelty of this approach fills a distinct void in cybersecurity by using established modeling principles for cybersecurity risk analysis as a prevention tool. The primary limitations are access to and the quality of imported data. In a proof of concept project, we plan to test our methods predicting cybersecurity risk on a small scale domain. The extension of the pilot project will target state and national-level risk prediction as well as real time modeling. For example, within a governmental organization where IT security procedures are well developed, documented, and distributed, a reliance on the mere existence of these protocols could develop into a potential risk if the organization does not test the use and efficacy of its protocols. Organizations must adopt policies that balance the need for security with ease of use by creating states that do not allow access to information to which a user should not have access but do not inhibit authorized access (Bishop, 2003). Unfortunately, disconnects can develop between policy and actual practice, which may present a major vulnerability that is not discovered until it is too late.

For instance, an employee of a government subcontractor who has access to sensitive and classified information with the ability to use a mobile device to extract data could have serious international consequences. If responsiveness to security protocol was measured for an organization compared with the sensitivity of data, measures could be produced so that an organization would know where best to modify its procedures, thereby reducing potential risk.

CONCLUSIONS

Given the increase of cyber attacks, known vulnerabilities, unknown vulnerabilities, and funding allocated for cybersecurity, defensive methods are a priority. Preventive methods would enable a system to increase its level of protection against an intruder prior to a successful attack. As the cyber warfare playing field expands globally adding increasing numbers of unknown or anonymous players, an increased need exists to consider alternative methods for defense. A system of prevention would greatly enhance our defensive capabilities, especially given the ability to use extant data would not only allow real application, but also contain costs.

REFERENCES CITED

Aldis, W. (2008). Health security as a public health concept: a critical analysis. *Health Policy and Planning*, 23(6), 369-375. doi: 10.1093/heapol/czn030

Amato, F., Lopez, A., Pena-Mendez, E. M., Vanhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, 11(2), 47-58. doi: 10.2478/v10136-012-0031-x

Bishop, M. (2003). What is computer security? *Security & Privacy, IEEE*, 1(1), 67–69. doi: 10.1109/MSECP.2003.1176998

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi: 10.1023/A:1010933404324

Broz, D., Wejnert, C., Pham, H. T., DiNenno, E., Heffelfinger, J. D., Cribbin, M., ... Paz-Bailey, G. (2014). HIV Infection and Risk, Prevention, and Testing Behaviors Among Injecting Drug Users—National HIV Behavioral Surveillance System, 20 U.S. Cities, 2009. *MMWR Surveill Summ, 63 Suppl* 6, 1–51.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards*. *Social Science Quarterly*, 84(2), 242–261. doi: 10.1111/1540-6237.8402002

Cutter, S. L. F., Christina. (2008). Temporal and spatial changes in social vulnerablity to natural hazards. *Profeedings of the National Academy of Sciences*, 105(7), pp. 2301-2306.

DiMase, D., Collier, Z. A., Heffner, K., & Linkov, I. (2015). Systems engineering framework for cyber physical security and resilience. *Environment Systems and Decisions*. doi: 10.1007/s10669-015-9540-y

Du, P. (2012). Smoothing Splines: Methods and Applications by WANG, Y (Vol. 68, pp. 1327-1328). Malden, USA: Blackwell Publishing Inc.

Dunham, K. (2008). *Mobile malware attacks and defense*: Syngress.

Freedman, D. A. (2009). *Statistical models: theory and practice*: Cambridge university press.

Gillies, D. (2010). B-splines. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), 237–242. doi: 10.1002/wics.77

Gostin, L. O., Waxman, H. A., & Foege, W. (2015). The president's national security agenda: Curtailing ebola, safeguarding the future. *JAMA, 313*(1), 27-28. doi: 10.1001/jama.2014.16572

Hastie, T., Tibshirani, R. J., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.): New York: Springer Science and Business Media.

Hoerl, R. W., & Snee, R. D. (2011). Statistical Engineering: Is This Just Another Term for Applied Statistics? *Physical and Engineering Sciences and the Quality and Productivity* (Vol. 18, pp. 3): American Statistical Association

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*: Springer New York.

Lanier, W. A., Johnson, E. M., Rolfs, R. T., Friedrichs, M. D., & Grey, T. C. (2012). Risk factors for prescription opioid-related death, Utah, 2008-2009. *Pain Med, 13*(12), 1580-1589. doi: 10.1111/j.1526-4637.2012.01518.x

Meng, X. (2009). Desired and Feared—What Do We Do Now and Over the Next 50 Years? *American Statistician*, 63(3), 9. doi: 10.1198/tast.2009.09045

Meynard, J.-B., Chaudet, H., Green, A., Jefferson, H., Texier, G., Webber, D., ... Boutin, J.-P. (2008). Proposal of a framework for evaluating military surveillance systems for early detection of outbreaks on duty areas. *BMC Public Health, 8*(1), 146.

O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M., & Mollison, D. (2000). Analyses of Infectious Disease Data from Household Outbreaks by Markov chain Monte Carlo Methods. *Journal of the Royal Statistical Society Series C (Applied Statistics), 49*(4), 517-542. doi: 10.1111/1467-9876.00210

Parker, R. D. (2011). Increasing Security through Public Health: A Practical Model. *Journal of Special Operations Medicine, 11*(4), 4-8.

Regier, M. D., & Parker, R. D. (2015). Smoothing using fractional polynomials: an alternative to polynomials and splines in applied research. *Wiley Interdisciplinary Reviews: Computational Statistics, 7*(4), 275-283. doi: 10.1002/wics.1355

Rencher, A. C., & Schaalje, G. B. (2008). *Linear Models in Statistics* (2 ed.). Hoboken, New Jersey: John Wiley & Sons.

Royston, P., & Altman, D. G. (1994). Regression Using Fractional Polynomials of Continuous Covariates - Parsimonious Parametric Modeling. *Applied Statistics-Journal of the Royal Statistical Society Series C, 43*(3), 429–467. doi: Doi 10.2307/2986270

Royston, P., Ambler G Fau–Sauerbrei, W., & Sauerbrei, W. The use of fractional polynomials to model continuous risk variables in epidemiology. (0300-5771 (Print)).

Rupert, M., Cannon, S. H., Gartner, J. E., Michael, J., & Helsel, D. R. (2008). *Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfires, Southern California, 2003-2006*: US Geological Survey.

Schmidtlein, M. C., Deutsch, R. C., Piegorsch, W. W., & Cutter, S. L. (2008). A Sensitivity Analysis of the Social Vulnerability Index. *Risk Analysis*, 28(4), 1099–1114. doi: 10.1111/j.1539-6924.2008.01072.x

Schmitt, M. N. (2015). The Law of Cyber Targeting *The Tallinn Papers*. Tallinn: NATO/CCD COE.

Schmitt, M. N., & Vihuil, L. (2014). Proxy Wars in Cyber Space. *Fletcher Security Review*, I(II), 55–73.

United Nations Security Council Resolution 1308 (2000) on the Responsibility of the Security Council in the Maintenance of International Peace and Security: HIV/AIDS and International Peace-keeping Operations. (2000). New York City: United Nations.

US. (2009). CERT Cyber Security TIP ST04-015. Retrieved June 11, 2012, from http://www.us-cert.gov/cas/tips/ST04-014.html

WHO. (2000). *Guidelines for Second Generation HIV Surveillance*. Geneva: World Health Organization and Joint United Nations Programme on HIV/AIDS.

WHO. (2011). *Guidelines on surveillance among populations most at risk for HIV*. Geneva: UNAIDS / World Health Organization.

Yang, Y., Longini Jr, I. M., Halloran, M. E., & Obenchain, V. (2012). A Hybrid EM and Monte Carlo EM Algorithm and Its Application to Analysis of Transmission of Infectious Diseases. *Biometrics*, 68(4), 1238-1249. doi: 10.1111/j.1541-0420.2012.01757.x

Zani, S. (2006). Data Analysis, Classification and the Forward Search: Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society, University of Parma, June 6–8 2005. Guildford; Ipswich: Springer London, Limited.

AUTHORS

R. David Parker (rdparker@hsc.wvu.edu) is an associate professor in the Department of Epidemiology at the School of Public Health at West Virginia University. His primary research areas include risk engagement to acquisition of HIV, sexually transmitted diseases (STIs), tuberculosis, and viral hepatitis. Additional areas of research include inter/national defense and health security, information security and risk modeling, and statistical engineering.

Michael D. Regier (mregier@hsc.wvu.edu) is an assistant professor in the Department of Biostatistics, School of Public Health at West Virginia University. His areas of expertise and interest include coarsened data, expectation-maximization (EM) algorithm, interpretable smoothers, marginal structural models, measurement error, missing data, model selection, and simulation studies.

EXCELSIOR COLLEGE
FOUNDED IN 1971