A Cyber Forensics Ontology: Creating a New Approach to Studying Cyber Forensics

By

Ashley Brinson, Abigail Robinson, Marcus Rogers

From the proceedings of

The Digital Forensic Research Conference

DFRWS 2006 USA

Lafayette, IN (Aug 14th - 16th)

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/diin

Digital Investigation

A cyber forensics ontology: Creating a new approach to studying cyber forensics

*Ashley Brinson**, *Abigail Robinson*, *Marcus Rogers*

Department of Computer & Information Technology, Purdue University, 401 North Grant Street, Room 255, West Lafayette, IN 47907, USA

Keywords:
Ontology
Model
Cyber forensics
Curriculum
Certification
Specialization

ABSTRACT

The field of cyber forensics, still in its infancy, possesses a strong need for direction and definition. Areas of specialty within a professional environment, certifications, and/or curriculum development are still questioned. With the continued need to standardize parts of the field, methodologies need to be created that will allow for uniformity and direction. This paper focuses on creating an ontological for the purpose of finding the correct layers for specialization, certification, and education within the cyber forensics domain. There is very little information available on this topic and what is present, seems to be somewhat varied. This underscores the importance of creating a method for defining the correct levels of education, certification and specialization. This ontology can also be used to develop curriculum and educational materials. This paper is meant to spark discussion and further research into the topic.

1. Introduction

According to Noy and McGuinness (2001), ontology creates a common definition among a domain of information within a certain area. By doing this, common information structures can be formed, knowledge can be reused, assumptions within a domain can be made, and every piece can be analyzed.

There are two types of ontologies. One ontology starts with a capital ''O'' and the other starts with a lower case ''o''. The latter describes situations where classification schemes are being built. The former is a term borrowed from philosophy where Ontology is a systematic account of existence (Gruber, 2006). For the purposes of outlining cyber forensics tracks, a small ''o'' ontology was created for classifying data tracks.

This is important to the field of cyber forensics due to the fact that common areas for specialization and certification are still being developed. The creation of an ontological model may allow for these specific areas to be defined. It would be detrimental to the field of cyber forensics if it continues to be somewhat scattered and reactive. For example, it is beneficial that researchers continue to turn to other forms of forensic sciences in order to shape their own processes for cyber forensics. Researchers have not, however, examined what the individual needs to do at a higher level. This is unrelated to how the field should develop; this is a focus on what individuals in the field should do. How does an individual know how to go about preparing themselves for a career, such as practitioners, scientists, law enforcement, or other work in cyber forensics?

2. Ontological model

The proposed model, see Fig. 1, consists of a five layer hierarchical structure with the resulting final layer being specified areas for certifying and specializing, in most instances. It was determined that five layers would be needed to reach the correct layers for specializing and/or potentially certifying.

* *Corresponding author*. Tel.: +1 765 496 2021; fax: +1 765 496 3181.
E-mail addresses: abrinson@purdue.edu (A. Brinson), amrobins@purdue.edu (A. Robinson), rogersmk@purdue.edu (M. Rogers).
1742-2876/$ – see front matter © 2006 DFRWS. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.diin.2006.06.008

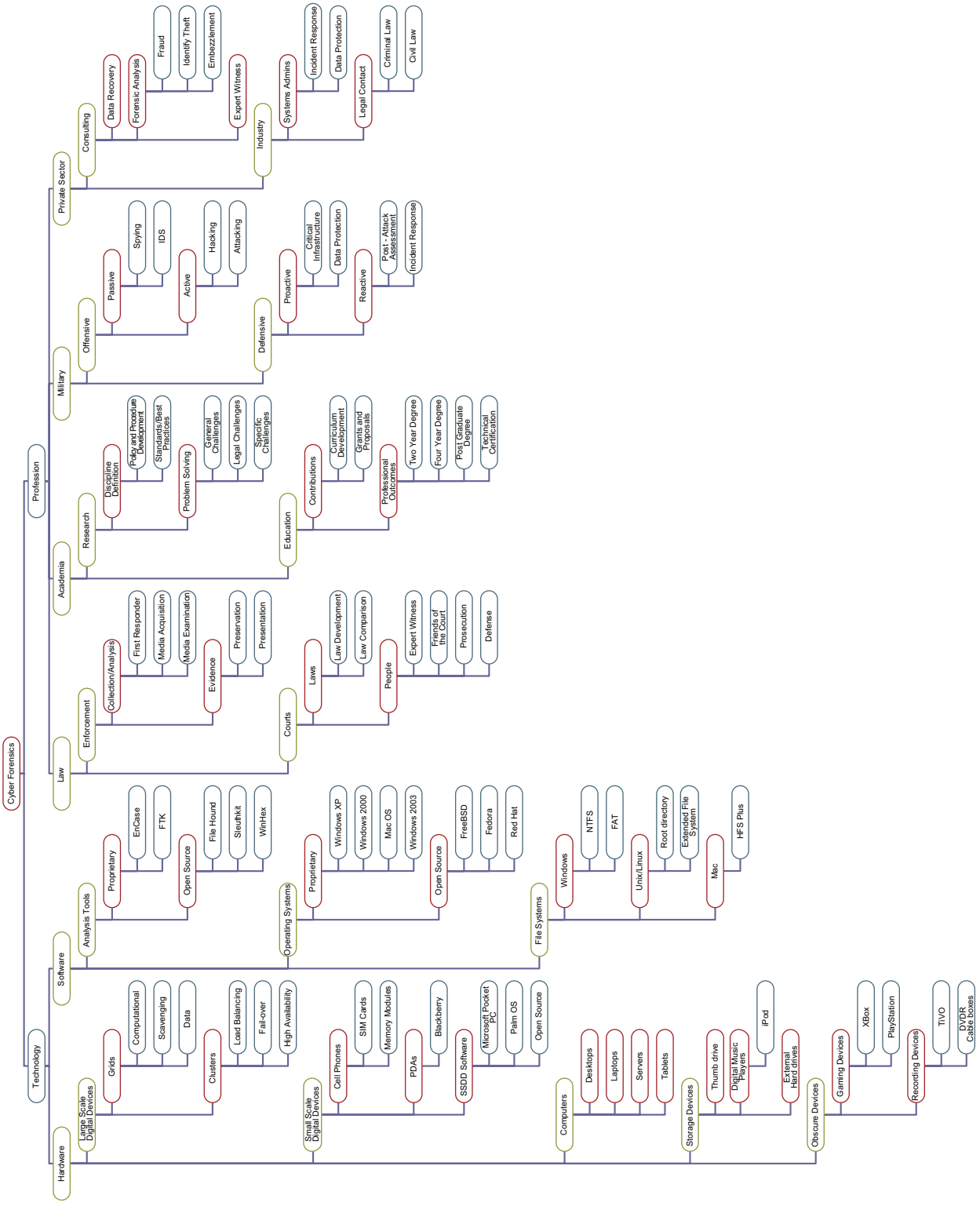

Fig. 1 – Cyber forensics ontology.

Anything prior to five could become too general and further than five gets too specific. In cases where the model does not fall to a complete fifth layer, varying reasons have determined that the fifth was not necessary. This will be explained within each particular section. This model can also be utilized for curriculum development within the field of cyber forensics. Both of these topics will be examined later.

The first main subtopics consist of technology and profession. When examining the topics at hand, specialization, certification, and education, all the relevant topics can fall into these subheadings. For the most part, the technology portion will examine areas of study within a topic as well as areas where certifications could be obtained. The profession side focuses on what professional specialty areas should be considered as well as areas of study for curriculum development. Technology is then broken down into hardware and software. This breakdown is logical because it keeps the technology that is being examined separate from the examining tools. The coinciding level on the profession side is broken down into the areas of law, academia, military, and private sector. These four areas are already recognized as the distinct areas of cyber forensics and therefore follow standard thinking (Rogers, 2004). Law focuses on prosecution, the military focuses on continuity of operations, and the private sector focuses on availability of service. Academia covers all areas and provides research across the board.

2.1. *Technology: hardware*

It is the authors' opinion that the hardware section of this model should be broken up into five different parts: Large Scale Digital Devices (LSDD), Small Scale Digital Devices (SSDD), computers, storage devices, and obscure devices. Based on the research by the authors, much more material was found on computer specific forensics than any other topic; therefore, it was decided that computers should be its own category. With all the new technology being developed with cell phones and PDA's, SSDD is an important area in computer forensics. Therefore, there is a need for a LSDD that deals with everything outside of computers and SSDD. The storage devices category is for hardware that is specifically used for storage and nothing else. The obscure devices category identifies hardware that has some of the same capabilities as a computer; however, that is not the main purpose of the device.

LSDD can be broken down into grids and clusters. The purpose of breaking down this category is to show the different areas within LSDD, not to show areas of specialization or certification. It is important to keep in mind that these can be topics for discussion in a LSDD course within a cyber forensics curriculum.

For the purpose of this ontology, SSDD was broken down into cell phones, PDA's, and SSDD software. PDA's and cell phones are very prevalent topics within cyber forensics right now because of the advances in technology. NIST has created a guideline for PDA forensics to attempt to bridge the gap between the recent technology phenomenon with the classical computer forensics (Jansen and Ayers, 2004). They are very small and versatile, which makes it difficult to deal with. SSDD software is comparable to an operating system on a computer, and needs to be understood as well.

The computer section in the model is broken down into four categories: desktops, laptops, servers, and tablets. However, these layers were not broken down any further. There are so many different possibilities and ways to classify computers; the authors found it too difficult at this time to break this category down any further.

Storage devices can be broken up into thumb drives, digital music players, and external hard drives. Thumb drives are becoming an issue because of their discreteness and small size. Digital music players are small in size, but can currently hold up to 60 GB of data, and as a hard drive it can hold documents and pictures (Marsico and Rogers, 2005). For the purpose of this ontology, an external hard drive varies from a thumb drive because of its large storage capacity. Currently, there are external hard drives that can hold terabytes of information. These areas are meant to be used in curriculum development rather than an area of specialization.

The obscure devices category includes any other device that contains a hard drive for storing information. For the purpose of this ontology, this category was only broken down into gaming devices and recording devices. It is the authors' opinion that this category can contain many different subcategories, and as technology advances, so will the category of obscure devices.

Some examples of grids include computational, scavenging, and data (Jacobs, 2003). Examples of clusters include load balancing, fail-over, and high availability (Turbolinux Cluster LoadBalancer, 2005). As stated before, these are not areas of specialization, rather topics for discussion in a course on LSDD.

Cell phones can be broken down into SIM cards and memory modules. These are two different ways to store information on a cell phone; however, these are not the only methods. This category could be broken up into several different ways, but the authors chose this way simply as an example. According to personal communication with studying cell phone manufacturers is also beneficial within SSDD, as some phones are compatible with others. This is also true for the PDA category. Blackberry is not the only brand of PDA, but it shows an acceptable level for certification and specialization. SSDD software incorporates Microsoft Pocket PC, Palm OS, and Open Source to name a few.

Two examples of gaming devices are Xbox and PlayStation. These were simply chosen by the authors as two instances of popular gaming devices for the purpose of the ontology; by no means is this an all inclusive list. The same is true for recording devices; TiVo is just one of many examples. These areas are too specific to be specialized in, however, they are important topics that one should be aware of.

2.2. *Technology: software*

The software section of this model contains three categories: analysis tools, operating systems, and file systems. In order to be able to analyze media for an investigation, one should be familiar with the different analysis tools (Rogers, 2006a,b). It is important to know the different operating systems when performing a media acquisition, to ensure the integrity

of the data. Along the same lines, it is also critical to understand the structure of the file systems. This makes it much easier when doing an investigation because the investigator knows where to search for evidence such as deleted data.

The authors felt it was most appropriate to break the tools down into proprietary tools and open source tools. Proprietary tools or open source tools alone is too broad of a topic to be certified or specialized in, as there are several tools for each category. It is important to know which tools are proprietary and which ones are open source, as well as the functions the tools can do.

The operating system category was also broken up by proprietary and open source. An open source OS gives users the freedom to run the program for any purpose, to study and modify the program, and to redistribute copies of either the original or modified program (Wheeler, 2005). It is also important to understand proprietary software. According to Wheeler (2005), Microsoft was the most prevalent OS vendor being used for web servers; it accounted for nearly 50%.

Since most operating systems provide their own file system, and the OS is dependent on this file system, it seemed logical to break down this category by operating system: Windows, Unix/Linux, and Mac OS. As stated before, it is important to know the different file structures for the different operating systems.

Two examples of proprietary tools are EnCase and File Tool Kit (FTK). These are by no means the only proprietary tools; the authors found these to be two of the more popular tools. This also provides guidance as to what level one should consider specializing, and also where one could get certified. Two open source tools are File Hound and Sleuthkit. It is valuable to know which tools are open source because these free tools might be enough to complete an investigation. This can save organizations a lot of money if they understand the functions of these tools, and are able to use them.

A few examples of proprietary operating systems are Windows 2000, Windows XP, Windows 2003, and Mac OS. These operating systems should be areas of certification and specialization. It would be impossible for someone to be specialized in all proprietary software; however, one could choose as many operating systems to become certified in as they would like. A few examples of open source operating systems are FreeBSD, Fedora, and Red Hat. These are only a few of the open source, and as stated before, open source is far too broad to be an area of specialization. Each operating system should be its own certification.

Currently, the two file systems for Windows are NTFS and FAT. Both of these would be good areas for specialization and certification. NTFS is a completely different file system from FAT, so it might be advantageous to be familiar with both (Rogers, 2006a,b). The Unix file system is based on the root directory, and the Linux file system is based on the extended file system (versions include EXT2, EXT3, XFS, JFS, and ReiserFS). The current Apple file system is the HFS Plus. There are different versions for the older Apple systems, but only HFS Plus was listed simply as an example. As explained before, it is important to understand the file system and the structure to aid in the investigation.

Switching gears, the profession half of this ontological model uses the four areas of cyber forensics to discuss their main functions and the resulting positions that come from those areas.

2.3. *Profession: law*

This category focuses on law enforcement and the involvement of the court and legal system within a cyber forensic investigation. This is not to say that law enforcement officers' duties are the same as a lawyer's involvement with cyber forensics. However, it does show the potential relationships between the two. This is the reason the sub layer of law is broken into enforcement and people. Both of these areas have specific research being done within them; however, as stated previously, looking at them from a higher level allows for a different interpretation. After breaking the enforcement side into two parts, collection/analysis and evidence, jobs roles and functions are seen. These are specific job roles that a forensic investigator would want to obtain. It is crucial that investigators understand what specific role they are looking for and how to achieve that. One person should not be doing all of the work so the roles must be separated. It also avoids tunnel vision that could impair the investigation. Separation of duties is important for investigators to keep an open mind. This ontological model organizes duties to be separated as such. Here, the roles have been defined as first responders, media acquisition, and media examination.

First responder duties and best practices have already been outlined by the National Institute of Justice (2001). Duties include securing and evaluating a crime scene, documenting, evidence collection, and finally packaging, transporting, and storing. However, it is recommended that the first responders do not actually do the media acquisition, and this is the reason for breaking it out into a new category with collection and analysis. First responders do not always have the best training or knowledge to do, for example, an image of a hard drive. First responders can be taught how a cyber investigation differs from a physical evidence investigation; however, they are still not technical experts. Someone with the proper training and tools should perform the acquisition itself. After media acquisition is complete, a media examination is done. This person should be the one to sort through all of the evidence, including utilizing software analysis tools to find evidence from the hard drive images. They should then document all findings into a report format.

The other equivalent layer focusing on law enforcement examines purely the evidence at hand. This includes everything that happens after the collection and analysis. An examiner's duties should focus on preserving evidence. If evidence needs to be accessed at a later date or utilized for presentation purposes, this is the role that would be responsible.

Besides law enforcement, it is also important to look at what the courts would need in the area of cyber forensics. It has been determined, by the authors, that the two main focuses within the legal system should be on the cyber forensic specific laws as well as the people involved in the jurisdictions. An issue of major concern within this same topic is getting the prosecution, defense, and when applicable, the jury, up to speed on technology. According to Rogers (2005), the average education level of a jury member is the seventh grade. Extra challenges are faced when technology is involved. It will

continue to be a hurdle in prosecuting and defending people in digital investigations if both parties do not fully understand the technology behind the case issues. For example, if the prosecution of a case were to come in and argue that they had used EnCase to find evidence off a defendant's home computer, the defense would need to know how EnCase works in order to refute anything that had been said. Their arguments would likely lie in questioning the reliability of the tool. This is why both parties need to understand the technology behind these investigations. Also, this is where expert witnesses and friends of the court come in to play. They have the expertise to explain technical topics. This layer of the model uses four areas: expert witnesses, friends of the court, prosecution, and defense are the people within the legal system that should be trained for cyber investigations. Human expert witnesses are important since the courts will not recognize software tools such as EnCase as an expert witness (Meyers and Rogers, 2004).

Less obvious than training the people, is looking at the actual laws that exist or need to exist such as penalties for cyber crimes. It has been stated in the proposed model, that existing relevant laws should be scrutinized for relevancy to digital environments and that new laws, specific to cyber forensics, need to be developed. This creates a need for legislation and government officials to also have understanding of cyber forensics investigations. These are the people that would develop and interpret the laws and therefore makes this an arena for them as well.

Overall, this law section can be viewed as a means to developing course material and what topics need to be considered from a law perspective. It also shows areas for professionals and something they could certify in. For example, a media acquisitionist could take a certification that requires them to have all of the latest knowledge and newest technologies. That certification would then need to be updated on a regular basis. It also shows the need for people, not obviously invested in cyber forensics, to actually have some understanding as well.

2.4. *Profession: academia*

Even though it may be easy to think of this area solely as curriculum development track within the ontology, it is still being looked at from that higher level for the individual's specializations. Here, it is important to look at the people involved in academia and what roles they should take to help shape this field as well as what they should be doing in terms of education. For this reason, this category has been broken into research and education.

Within research, the next level has been defined as discipline definition and problem solving because these areas still require so much work in order to keep all individuals working toward the same goal. That goal includes proper development of the field of cyber forensics. Discipline definition should include things that help shape the field of cyber forensics. This ontological model is an example of this. Basically, anything that falls into the subcategories of policy and procedure development and standards/best practices would be placed in discipline definition. As far as the relevance to this model is concerned, it shows an area where work needs to be done and people could focus their attention on providing more research for the areas mentioned throughout the entire model. Besides research for defining cyber forensics, solving problems that currently exist in the field should be important as well. Specific types of challenges have been outlined within the problem solving section. Resources also need to be dedicated to these areas. This is one of the major challenges in cyber forensics today (Rogers, 2004).

The other half of academia was defined as education. For this category, the model basically examines the inputs and outputs to education. This is an all-encompassing view since both educators and students are taken into consideration. The educators would fall into the contributions section that was created. This includes curriculum development and/or grant and proposals for money. These people play a vital role in the development of cyber forensics. Without money for the departments, none of this would happen in the first place. This section of the model also addresses potential degree tracks that should be applied at an institutional level. The reason it needs to be individual is because some schools could have the resources to do the full degrees, whether they are two, four year, or post graduate. Others may only have resources to do technical certifications. In any instance, the curriculum involved, should be tailored to exactly what they have the resources to produce. According to Armstrong and Russo (2004), all curriculums should be set around education on all aspects such as hardware, software, evidence, computer security, and others. This model can be used for gathering courses and course material and details can be seen later. A full degree would want to utilize all aspects of cyber forensics whereas a certification could just follow one of the tracks down the proposed model.

2.5. *Profession: military*

The military category focuses on what cyber forensic duties military personnel perform. It outlines duties that they perform as well as areas where they would need to educate themselves and/or train. The military has many needs including data protection, data acquisition, imaging, extraction, interrogation, normalization, analysis, and reporting (Giordano and Maciag, 2002).

The military category has been broken into offensive and defensive strategies. This was done to help delineate the different tactics that the military uses. With the offensive category, passive and active areas were examined. Passive includes any spying that is done or intrusion detection. Active includes hacking and/or actual information attacks on an enemy because the attacker is actively pursuing an intended target.

On the defensive end, proactive and reactive tactics were examined. The proactive area looks at ways to protect critical military infrastructure and protect data. The reactive area is defined by incident response and post attack assessment. The reason it is important to look at both of these areas is the great involvement that cyber forensics has in this area. The military does a lot of work here at home, but they also do a lot that crosses borders. Cyber forensics becomes more difficult to do when crossing these international lines.

2.6. *Profession: private sector*

Private sector was broken down into consulting and industry. As of now, there are no rules or regulations monitoring the consulting area of computer forensics, and this is definitely an area that needs attention. The industry also needs to understand the basics of computer forensics. If there is illegal activity taking place within a company, they should know how to handle the situation (Willer, 2001).

There are numerous categories that fit under consulting, however, three were chosen as the most prevalent based on researching private consulting firms: data recovery, forensic analysis, and expert witness. There are several different instances when companies lose important data that may not be backed up. In these cases, the organization can hire a consultant to recover all the lost data. This area appears as a small enough category to be a specialization itself, and needs no further sub-categorization.

Forensic analysis is a very broad area, and is meant to encompass the different types of cases consultants handle. Fraud, identity theft, and embezzlement are three examples, but by no means the only examples. It is very common for a consulting firm to be hired to find evidence and present it in a court of law. Therefore, the expert witness should be an area of specialization and certification for the court.

There are at a minimum two job roles in an organization that should be aware of the issues surrounding cyber forensics: systems administrator and a legal contact. The systems administrator needs to know how to handle situations in which there is a breach in security, as well as how to protect the data to prevent such breaches. An organization should also have a legal contact to deal with incidents that require lawful action.

In relation to computer forensics, and for the purpose of this ontology, the systems administrator should have two main duties: incident response and data protection. Should there be a situation where someone is attempting to hack the network, the systems administrator should know the appropriate steps to take. They should also know security measures to use to keep their data safe and protected. According to Willer (2001), having an incident response plan is critical to protecting evidence that may be on a computer. In case there is an issue within the company, there should be a legal expert to contact. There should be one that can deal with criminal law and one for civil law. These are areas that can be specialized in, but more than likely not areas for certification.

3. Certification areas

Throughout this paper, various areas of certification possibilities have been examined. While, it has been noted that particular certifications at the fifth layer, such as EnCase, FTK, Microsoft XP, or on the other side, first responder, would be good ideas; it should also be noted that one would not want to be certified in only one of these particular areas. They show places where an individual would benefit from having the excelled knowledge on one topic and would make them useful for investigations in this area. People could and should obtain certifications in all areas relevant to their specialization. For someone specialized in analyzing images, for example, it would be best to be certified in as many tools as possible. EnCase, FTK, File Hound, and Sleuthkit are all good examples as seen in the model. One current certification that was found and is inadequate was the Certified Computer Examiner Certificate. This certificate can be obtained by a simple written exam and testing media (Gregg, 2004). The written exam consists of basis overviews of topics crucial to cyber forensics. This is not enough to obtain certification and is also too broad. Meaningless certifications need to be avoided and that makes finding correct layers of certifying so critical. Individuals could ruin their credibility by obtaining certifications that are generalized. This would be obvious when they could not perform a specific job function.

4. Curriculum development

This ontological model can also be utilized for the purpose of curriculum development. This is done by following areas of the model to find topics to study within a potential course. For example, the third layer topics could become the potential courses. Underneath the hardware layer are the subtopics of large-scale digital devices, small-scale digital devices, computers, storage devices, and other miscellaneous devices. This means that the course itself would be on small-scale digital devices and topics within that course would be the sub layers to this SSDD category; examples include memory modules, SIM cards, PDAs, or the operating systems involved such as Palm OS. This model is very dynamic because it allows for additions to be made at any time to any location on the model. As more research topics are discovered, more course content can be added.

5. Conclusion

Cyber forensics is a prevalent part of criminal and civil investigations. As a newer forensic science, there is much research being done to create best practices, processes, and procedures by entities including the government, scientists, and educators. This is extremely important as proper field/discipline definition right from the beginning can help decrease problems later. However, the one area that seems to be lacking in this research is what exactly the people involved in cyber forensics are supposed to do to prepare them, not the discipline. How do they specialize or certify themselves? This research focused on creating an ontological model that addresses those issues, and additionally created a tool for curriculum development. Course ideas and content can be drawn from this model and after learning material contained within this model, a student should be able to specialize and/or certify their expertise.

6. Further research

As stated before, the purpose of this paper is to start discussion in the areas of certifying and specializing individuals

working in a cyber forensic environment. This area is of great importance and currently research here is lacking. Further research topics could elaborate on this model, modify it, or refute its contents. However, with any position that is taken, the bottom line should still be to produce a higher level of research that does not focus on processes or procedures for cyber forensic investigations but looks at what individuals need to be doing to prepare them for the work they want to do within cyber forensics.

REFERENCES

Armstrong H, Russo P. Electronic forensics education needs of law enforcement. Available from: http://www.ncisse.org/publications/cissecd/Papers/S4P02.pdf; 2004 [retrieved 09.01.06].

Giordano J, Maciag C. Cyber forensics: a military operations perspective. Available from: http://www.utica.edu/academic/institutes/ecii/publications/articles/A04843F3-99E5-632B-FF420389C0633B1B.pdf; 2002 [retrieved 12.01.06].

Gregg M. The certified computer examiner certification. Available from: http://www.gocertify.com/article/certifiedcomputerexaminer.shtml; 2004 [retrieved 20.01.06].

Gruber T. What is an Ontology? Available from: http://www-ksl.stanford.edu/kst/what-is-an-ontology.html; 2006 [retrieved 22.04.06].

Jacobs B. Grid computing: what are the key components?. Available from: http://www128.ibm.com/developerworks/grid/library/gr-overview/; 2003 [retrieved 25.04.06].

Jansen W, Ayers R. Guidelines on PDA forensics [NIST 800-72]. Gaithersburg, MD: National Institute of Standards and Technology; 2004.

Marsico C, Rogers M. Ipod forensics. Available from: https://www.cerias.purdue.edu/tools_and_resources/bibtex_archive/archive/2005-13.pdf; 2005 [retrieved 03.04.06].

Meyers M, Rogers M. Computer forensics: the need for standardization and certification. Available from: http://www2.tech.purdue.edu/cpt/courses/CPT499S/meyersrogers_ijde.pdf; 2004 [retrieved 09.01.06].

National Institute of Justice. Electronic crime scene investigation: a guide for first responders (NCJ 187736). Washington, DC: Office of Justice Programs; 2001.

Noy N, McGuinness D. Ontology development 101: a guide to creating your first ontology. Available from: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html; 2001 [retrieved 15.01.06].

Rogers M. Cyber forensics: are we there yet? Available from: http://72.14.203.104/search?q=cache:csMU5IWwVYoJ:www.cyberforensics.purdue.edu/docs/van2004.ppt+cyber+forensics+are+we+there+yet&;hl=en&gl=us&ct=clnk&cd=1&client=firefox-a; 2004 [retrieved 09.01.06].

Rogers M. Criminalistics. Presented at a CPT 499S lecture at Purdue University; January 18, 2005.

Rogers M. File systems. Presented at a CPT 499F lecture at Purdue University; 2006a.

Rogers M. Media analysis. Presented at a CPT 499F lecture at Purdue University; 2006b.

Turbolinux Cluster LoadBalancer. Available from: turbolinux.com http://www.turbolinux.com/products/middleware/lb10/docs/user_guide/index.html; 2005 [retrieved 19.04.06].

Wheeler D. Why open source software/free software? Look at the numbers! Available from: http://www.dwheeler.com/oss_fs_why.html [retrieved 19.04.06].

Willer L. Computer forensics. Available from: http://www.giac.org/certified_professionals/practicals/gsec/0854.php [retrieved 25.04.06].

Ashley Brinson is a graduate student at Purdue University studying Cyber Forensics and IT Management.

Abigail Robinson is a graduate student in the Cyber Forensics program at Purdue. She is also studying Information Security.

Marcus K Rogers, Ph.D., CISSP, CCCI is the Chair of the Cyber Forensics Program in the Dept. of Computer and Information Technology at Purdue University. He is an Associate Professor and also a research faculty member at the Center for Education and Research in Information Assurance and Security (CERIAS).