Cyberwarfare Doctrine and Strategy Execution using a Semantic Approach

Jeffrey J. Wiley, Frank P. Coyle
Lyle School of Engineering,
Southern Methodist University

Abstract

To the extent that cyber warfare is truly warfare, then it is a child class of conventional warfare and doctrine should be applicable and available for re-use. Cybersecurity professionals should leverage the inherited doctrine and approaches and adapt it. Military operations (i.e. defense) and military intelligence doctrine is of particular interest to an organization conducting cyber-defense. Defensive doctrine re-use suggests moving "outside the box" of a castle defense and into alternative technologies. Intelligence doctrine suggests a longer-term, proactive effort to apply existing and proven military doctrine within the cyberwarfare domain. This paper focuses on a proactive defense that leverages intelligence analysis of log data that has been semantically transformed.

1. Introduction

Cyberwarfare, and cyber-defense along with it, does not have a well-known and publicized doctrine and strategy for cyber professionals to follow. Too often, the individual cyber professional (aka defender) has to figure out how to defend the enterprise without much outside help. There aren't publications on how to strategically fight the fight; there generally are only tactical guides. This dearth of strategy forces the cyber-defender to focus on the "trees" and not the "forest", fighting tactically day-to-day. While this may generally win most of the time, defenders lose sight of the bigger picture. More accurately, they don't have a bigger picture. This allows the attackers to maneuver around traditional defenses and use strategies and related tactics that aren't currently defensible.

So, cyber-defenders need doctrine, strategy and related tactics to fight these battles and stay ahead of the enemy. Where does a cyber-defender look for doctrine and strategy? The authors suggest looking to the parent class of cyberwarfare, conventional warfare.

In this paper we describe a strategic approach to cyberwarfare that is modeled on, and re-uses, conventional warfare doctrine, as well as adapting traditional log analysis by creating a semantic transformation of computer system logs. Throughout, we borrow concepts from object-oriented programming where applicable.

Our strategy rides on the framework from the business world, again leveraging re-use. We suggest using a native semantic transformation of system logs that then enables a deeper, timelier log analysis as the authors envision. Having transformed log data into a semantic technology and performed the analysis, the resulting information can be used to make proactive information security decisions.

The remainder of this paper is structured as follows. We begin with a discussion of how conventional warfare and cyber warfare are related in a software "class" manner. Establishing this relationship, we next discuss the conventional warfare doctrines we see as directly available for re-use. The subsequent two sections discuss these doctrines in detail. Finally, the last half of the paper describes one strategy that synthesizes all of these concepts into a method of proactive information security decision-making based on intelligence analysis of semantic logs.

2. Cyber warfare is a class of warfare

Is cyberwarfare an instance of (child class) true warfare, true warfare referred to from here on as conventional warfare? The authors suggest the answer to this is, of course, yes! Noted military strategist General Carl Von Clausewitz claims "War is an act of violence, which in its application knows no bounds". [1] The two key points from this statement are that it is an act of violence and that these acts know no bounds. While in one sense it may be a stretch to call cyberwarfare an "act of violence", we can categorize cyber-attacks and crimes as violent. Because warfare knows no bounds, then cyberwarfare definitely qualifies. Search.Security.com claims that "Cyberwarfare is Internet-based conflict involving politically motivated attacks on information and information systems." The article goes on to quote Jeffrey Carr, author of "Inside Cyber Warfare," any country can wage cyberwar on any other country, irrespective of resources, because most military forces are network-centric and connected to the Internet, which is not secure. For the same reason, non-governmental

groups and individuals could also launch cyberwarfare attacks. [2] At the intersection of these two definitions, that war is an act of violence with no bounds, and that cyberwarfare is attacks on information and information systems, is the realization that cyberwarfare is a special type of conventional warfare, and thus, a child class of conventional warfare. Realizing that leads a cyber-defender to the need for doctrine and strategy.

2.1 Adversaries/enemies

Common to both classes warfare are the types of adversaries. Nation-states, organized crime organizations, and commercial/industrial entities can, and do, conduct both classes of warfare. A quick search of the internet will turn up multiple examples of each and need not be listed here.

2.2. Motivations

Political, economic, and cultural reasons motivate efforts in both classes as well. War is ultimately a political means to achieving a goal not attained otherwise. Many, many wars and battles have been fought over economic gain, to wit, the German attack on southern Russia to gain control of the Baku Oilfields in the Caucasus region. Lastly, there have been numerous wars motivated by cultural reasons. One of the most obvious was the Crusades from the Middle Ages. Similar examples exist for cyberwarfare motivations as discussed below.

2.3 Targets

Many of the targets of cyberwarfare are the same as conventional warfare. Political motivations have driven cyber-attacks of defensive networks like the ones targeting Georgia. Economic motivations drive cyber-attacks of industrial espionage. Other economic attacks target a country's infrastructure (e.g. SCADA), just like the Stuxnet worm attack on Iran.

3. Doctrine Inheritance and re-use

Given that cyberwarfare is indeed a child class of warfare, cyberwarfare professionals should leverage existing warfare doctrine instead of creating their own, or worse yet, not following any doctrine. In their book "The Future of War", George and Meredith Friedman state that "There is a deep chasm between the advent of technology; and its full implementation in doctrine and strategy."[3] It is this chasm between technology and doctrine / strategy that must be filled. There are two main doctrines inherited by cyber defense from conventional warfare. The two are the defense and intelligence. Each of these domains has their own field manual available to the public, Field Manual (FM) 3-0, Operations for the defense, and Field Manual (FM) 2-0, for Intelligence.

3.1 Defense

There are two inherited defensive doctrinal concepts cyberwarfare and cyber-defense should use: initiative and maneuver. FM 3-0 defines initiative this way: "Operational initiative is setting or dictating the terms of action throughout an operation." [4] Initiative is not though of very often in cyber-defense, certainly not for those practitioners of a traditional "castle defense." Too often, cyber-defenders are only waiting for something to happen and then trying to react to that/those events. From a warfare doctrine viewpoint, this is not the recommended approach.

Operations field manual states that "mobile elements maneuver constantly to confuse the enemy and prevent enemy exploitation." [4] The FM goes on to say that "Waiting for the attack is not a passive activity. Commanders conduct aggressive security operations and intelligence, surveillance, and reconnaissance. Such actions locate enemy forces and deny them information." [4] Maneuvering is a seemingly foreign concept to cyber professionals. Focusing on the results of maneuvering, namely confusing the enemy and denying them information, a cyber-defender can imagine creating a defensive strategy where the defender is in a better position to defend, using initiative and making it more difficult for the enemy.

3.2 Intelligence

In cyber-defense there isn't a concept of intelligence. Cyber professionals do not have to look very far to see the effective use of intelligence as many organizations use the concept of business intelligence. If it works for the business side of the house, it will work in cyber. Besides, intelligence works well in the military. All it takes is for cyber-defenders to broaden their perspective and begin to adopt doctrine and strategy. The Chinese strategist Sun Tzu said that "If you know the enemy and know yourself, you need not fear the result of a hundred battles." [5] His statement points to the strong need for intelligence in warfare, and in this case, cyberwarfare. Contrary to Sun Tzu's guidance, a majority of enterprises don't know themselves very well and they certainly don't have a good idea about the enemy.

FM 2-0 on intelligence says that "predictive intelligence enables the commander and staff to anticipate key enemy events or reactions and develop branch plans to counter them. [6] Predictive intelligence provides better situational awareness and

is based on answering key questions the organization has – key questions that are called intelligence requirements. Imagine following this doctrine and doing one's best to anticipate the enemy and negate their actions, again to the best of your ability. This situational awareness will not only help detect attacks, it will help deter them and definitely help prevent greater harm than otherwise.

Another area is in the arena of reconnaissance and surveillance, or R&S. In the cyberwarfare domain this is typically thought of as penetration testing. There is at least one other aspect of R&S that could additionally be used and that concept is conducting actual reconnaissance of the enemy. Reconnaissance is simply trying to gain information about the enemy. This recon does not have to take the form of actively probing a suspect network for example. That type of activity may not be legal or desired on the part of the cyber-defender. There are other, public, open-source ways to gather information as all cyber professionals know well.

A third doctrinal concept from intelligence to inherit is to establish an intelligence architecture, specifically the notion of intelligence reach and use of intelligence databases. [6] Intelligence reach refers to the use of multiple avenues as sources for information. In a cyber-defenders case, this means not only using your own internal logs, but open-source data and other databases available to you. Adding this information to your own intelligence store then allows predictive analysis to occur, all of which puts the defender in a more proactive, prepared state of readiness.

4. Cyber Defense Doctrine

In spite of the fact that we can categorize cyberwarfare as a child class of conventional warfare, and given that there are conventional warfare doctrines available for re-use by the cyber professional, traditional cyber defenses are becoming less effective at stopping attacks. The authors suggest that one reason is the lack of doctrine and strategy. The point here isn't to say that traditional cyber defense methods should be thrown out. No, traditional methods should still be used. These methods just need to be used in a larger context, a strategy-driven context using varied methods that will be discussed below. For example, Wiley and Coyle [7] state that the traditional castle defense is "quixotic" in nature as in Fig 1.

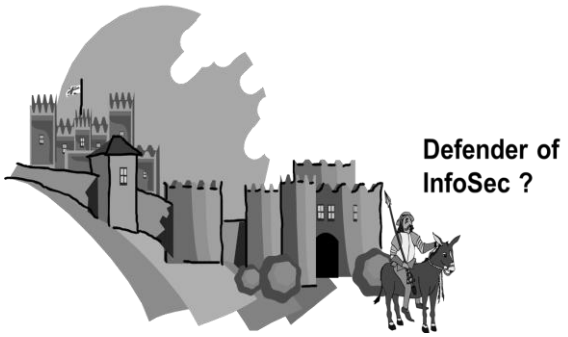

Figure 1. Traditional Castle Defense

This means that continuing to rely solely on a castle defense is a foolish pursuit of some ideal. This defense is not as effective as we would like and those that don't recognize that fact, and adapt their strategies accordingly, are like the literature Don. VeriSign Incorporated acknowledged this in a recent whited paper. "For organizations that continue to rely solely on firewalls, IPS, AV, and other signature-, reputation-, and basic behavior-based technologies, it is abundantly clear that compromises and infections will continue to grow. To effectively combat these attacks, it is imperative that organizations augment their traditional security defenses with technologies that can detect and thwart today's advanced, dynamic attacks. This requires capabilities for guarding against attacks being waged on the Web, and those being perpetrated through email, including spear phishing emails that use malicious attachments and URLs." [8] The following sections describe research by the authors that demonstrate how these doctrinal concepts may be applied in cyberwarfare and defense.

4.1 Proactive Defense (Mobility / Initiative)

Clausewitz asserts that while "the defensive is the stronger form of conducting war", it has a negative object. Thus, we "must only make use of it so long as our weakness compels us to do so". [1] Beraud, Cruz, Hassell, Sandoval and Wiley describe a proactive technique called the Network Maneuver Commander (NMC) as shown in Fig 2.

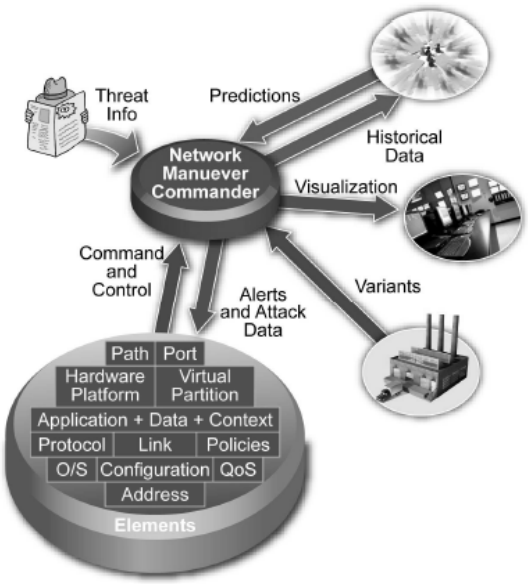

Figure 2. Network Maneuver Commander

Following Clausewitz's assertion and using a proactive defense as suggested by Beraud et al, the cyber-defender is able to seize operational initiative as the field manual describes. Furthermore, the Network Maneuver Commander concept is helps satisfy the mobility doctrine as well. The NMC has several main goals that dovetail to operations doctrine, namely to "increase the exposure of an attacker to detection as the attacker is forced to out-maneuver target reconfigurations, increase the uncertainty of the success of the attack, increase the survivability in the presence of attacks…"[9] Increasing the exposure of an attacker points to attribution of the attack to the attacking party. Confusing the attacker leads to their relative uncertainty. The last goal is self-explanatory.

4.2 Alternative Defensive approaches (Initiative)

An off-shoot of the NMC proactive defense is the MORPHINATOR project. This Raytheon Company project is currently funded by the U.S. Army. Morphing Network Assets to Restrict Adversarial Reconnaissance is a "prototype network to randomize its makeup to fool and foil attackers."[10] The Army has invested several million dollars for research and development, demonstrating the validity of an approach like this. Other research the authors have been involved in utilizing this type of doctrine is the Ontology Agent Decision-Making Support for Vulnerability Management concept. This concept allows a cyber-defender to draw keener inferences on system vulnerability than most system administrators would otherwise be capable of. Having inferred vulnerability status and applying tailored rule-sets, a vulnerability assessment can be conducted faster; resulting decisions on how to proceed are well-supported; and other information assurance decisions may be supported as well. [11] All of this allows a cyber-defender to gain the initiative in accordance with doctrine (see Fig 3).

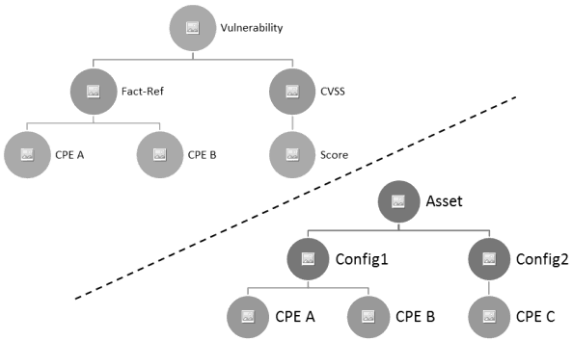

Figure 3, Vulnerability Ontology

5. Cyber Intelligence Doctrine

The concepts of intelligence doctrine described above can be followed in a number of ways, though only the following are discussed here. This is also the area where the use of semantic technologies is introduced as a potential strategy to aide in following these doctrines.

5.1 Intelligence Requirements

"If you know yourself but not the enemy, for every victory gained you will also suffer a defeat." [5] This Sun Tzu quote is from the same chapter as the earlier quote in Section III and has a different slant. Cyber-defenders need to know the enemy as well as themselves to truly gain the upper hand. The intelligence doctrine that directly relates to knowing your enemy is to develop intelligence requirements. FM 2-0 on Intelligence lists three types of validated information requirements are ordered in the following hierarchical structure: priority intelligence requirement (PIR), intelligence requirement, and information requirement. [5] The hierarchy spells out the necessity of answering each type. PIR must be answered. Intelligence requirements are collected if at all possible and information requirements are collected for general use and store population (reach). Though the VeriSign white paper has a slightly different take on intelligence requirements, VeriSign does have the same basic approach described here as they discuss critical intelligence requirements, priority intelligence requirements and requests for information. [8] Regardless of how the requirements pie is sliced, they are generally like the requirements as further defined in Table 1.

Table 1. Intelligence Requirements

Requirement Type	*Priority*	*Description*
Priority Intelligence Requirement	Must Collect	Purpose of the mission
Intelligence Requirement	Collect if possible	Nice to know information
Information Requirements	Optional collect	General information for later use

5.2 Reconnaissance & Surveillance

Reconnaissance & Surveillance (R&S) techniques are sometimes used by cyber professionals. As noted above, this typically comes in the form of penetration tests. There are some drawbacks with his doctrine concept, drawbacks that the reader is likely all too familiar with. For one, they take a lot of coordination and require high-level approvals.

The authors suggest integrating another doctrine concept in the R&S realm. This concept is that "every soldier is a sensor." [12] In cyber, that means that every employee is a sensor (along with mainstream sensors like IDS/IPS). Enterprises should develop a reporting system so that employees can report attacks (e.g. social engineering attempts) and that the report is investigated AND added to the enterprise intelligence store discussed below. In today's cyberwarfare environment, one of the most prevalent attack methods is social engineering. While employee training may enable an employee to recognize and negate this type of attack, what is lost is the valuable information on what happened, what was requested via what medium, etc.

5.3 Intelligence Architecture

"Thus, what enables the wise sovereign and the good general to strike and conquer, and achieving things beyond the reach of ordinary men, is foreknowledge." [5] Sun Tzu shows us that advance knowledge is the key to achieving cyber defenses "beyond the reach" of ordinary cyber professionals. Having intelligence requirements and an R&S reporting systems for employees set the stage for needing an intelligence architecture, and specifically, an intelligence store. This store obviously should include enterprise logs, etc., information generated internal to the organization. It should also include the answers to PIR and such, as well as surveillance reports.

6. Strategy execution

Of particular interest to the authors in this area is the use of semantic technologies to gather this information and as a specific type of data store. There are many semantic possibilities that broaden the ability of cyber professionals to analyze intelligence. For instance, the use of triple stores and an inference engine would allow a cyber-defender analyst to do more with information, beyond a simple pattern matching. Many modern-day event management or analysis tools use complex algorithms and heuristics to produce so-called intelligence. The authors believe, by and large, this is still simply pattern matching within a large dataset. The algorithms and heuristics allow the identification of something of interest out of millions of transactions or log entries. What these tools do not do is ingest the semantic information and then not only match items of interest, but then infer relationships across the dataset to reveal new and previously undiscovered intelligence. To begin with, there are two supporting concepts of our strategy that are discussed.

6.1 Big data

Big Data is more than just multiples of petabytes of data. It is BIG because it's the union of three different universes of data [13]. Structured data has just what its name implies: structure. This structure is typically defined by a schema that tells you what kind of data to expect. If you have an Oracle or MySQL database, you have structured data. Semi-structured data (e.g. XML) is data that carries with it some descriptive meta-data. Meta-data is data that says something about what you've got, and for XML, that data is the element and attribute names. Unstructured data can be server logs, audio and video streams, paragraphs of text, bit streams of all kinds with no inherent meta-data, semantic or structural design. The challenge of Big Data is how to make connections between these different universes, each with its own technology subculture [13].

6.2 Hedgehog Concept

Jim Collins suggests in the business world that there is a "hedgehog" concept to follow. The Hedgehog Concept is centered on what you are deeply passionate about, what drives your economic engine, and what you can be the best in the world at [14]. Collins asserts that if a business can define and follow its own hedgehog, then that business can achieve breakthrough performance. Fig. 4 below shows Collins' concept. It is in the context of big data, proactive defense, and the hedgehog concept that the authors look to provide recommendations for

proactive, predictive log analysis using semantic technologies.

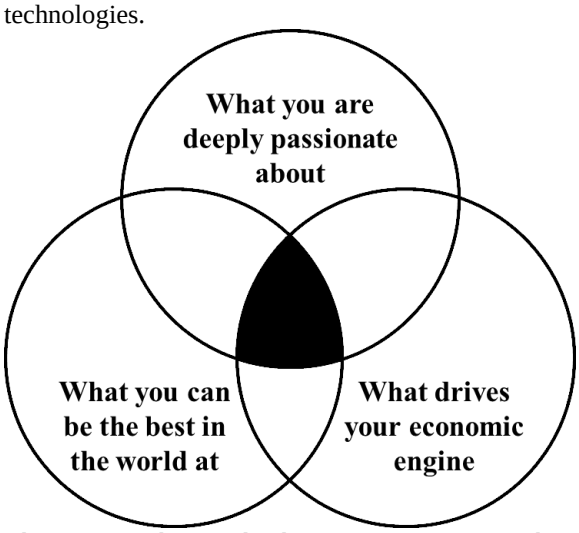

Figure 4. "The Hedgehog Concept" *Good to Great*, by Jim Collins, copyright © 2011. Used with permission

6.3 Log Review

There are issues with the current modus operandi for log review. One issue is that if log data is ever reviewed at all, it is almost always in a reactive mode. The typical scenario where some log analysis occurs is that something anomalous or nefarious happens. In the ensuing investigation, log review happens in the course of this investigation. This is reactive, not proactive, log review (defense) and does little to improve security. Not only does this approach not prevent an attack, it only discovers what happened long after the attack has occurred.

6.4 Log format

Log data is generally in an unstructured (e.g. flat file) format that is not conducive to quick analysis. There are exceptions and advancements in this area to some extent. For example, Microsoft Windows event logs can be viewed and exported in a semi-structured (eXtensible Markup Language - XML) format. Still, these formats are difficult to perform meaningful queries on. Simply matching a date-time group or a process ID (PID) is not indicative of a meaningful query. Syslog has an unstructured output, though it is well defined in RFC 5424. This format is a HEADER SP STRUCTURED-DATA [SP MSG] [15]. The header has several defined fields such as priority, hostname, process and message identification, and time. If there is any structured data, that data would come next after the defined "space" delimiter. The log message then follows at the end. Windows event logs, viewable within the Event Viewer management console, present similar information though that information can be viewed by default in two formats, friendly and XML.

6.5 Human review

Many software tools exist to assist the review of log data as mentioned above. Despite the existence of these tools, the log analysis and creation of information / intelligence is the sole domain of human review only. We mean by this that the intelligence produced still requires a human in the loop to review – there is no machine analysis other than the aforementioned pattern matching. This makes log analysis error-prone (missing a logged event, e.g.) and subject to the technical expertise of the analyst. Due to the varying degrees of expertise, log analysis often is found lacking the completeness needed for a proactive defensive environment.

6.6 Timeliness

The discussions of the previous three paragraphs lead to another problem. The problem is that log analysis is extremely time-consuming for a human analyst. An analyst must process unstructured data that does not lend itself to queries and analysis. This lengthy process will always take longer to conduct. Furthermore, the review and elimination of human error in the analysis adds to the timeline as well. Finally, since this analysis is likely reactive as suggested above, the utility of the analysis report is in question.

7. Semantic strategy for log analysis

There are semantic technologies that can be used in each of the four areas discussed above. The technologies and tactics described below are not the only useful tools and methods, just the ones within the scope of this paper. Other methods will be the subject of future research.

7.1 Proactive Log Review

Proactive log review is more than just a matter of timing, though reviewing log data and analyzing it before something undesirable occurs is one part. The analysis should be automated or semi-automated to reduce the time required; in a format that lends itself to analysis (e.g. semantic structure); and both machine- and human-readable. Semantic technologies bring all of these criteria to the table. Our original Big Data Venn diagram now modified and modeled after Collins' Hedgehog Concept [14], shows in Fig. 5 the intersection of the three data universes. The shaded area in Fig. 7 shows where the authors suggest that semantic technologies can be brought to bear on integrating and synthesizing this data. This integration should provide for a greater, more in-depth log analysis.

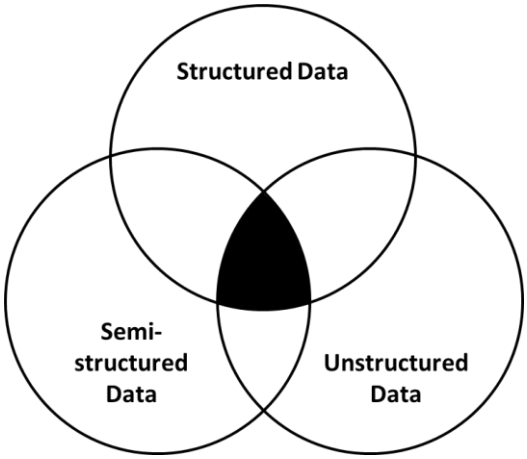

Figure 5. Big Data Hedgehog

7.2 Semantic Log Format(s)

Log data should be converted to a semantic structure. There are a couple of methods to do this. One method is simply taking the unstructured log format and to transform that output into a semantic structure. The other method is to transform the application(s) and provide an alternative output natively, one that is already semantic in structure. The ways to do this and the merits of either/both are the subject of future research.

For this paper, a manual transformation of the unstructured log data was performed, done on both Windows event log data exported in XML and syslog data sent to a text file. Table 2 and Table 3 below show the progress at each conversion step in the Windows event log process. The conversion from XML to CSV (semi-structured to alternate semi-structured) might have been unnecessary, though it was performed because of time limitations and lack of advanced familiarity with the online tool. Table 4 shows the similar CSV output with regards to the syslog file.

Once the log files are in a semi-structured format, the transformation to a semantic structure is simple enough using any one of a number of web-based conversion utilities. In this effort, the web-based conversion was performed using the "Any23 - Anything to Triples - Live Service Demo" at any23.org [16].

Table 2. Windows Log in XML

```
- <Event xmlns="http://schemas.microsoft.com/win/2004/08/events/event">
- <System>
  <Provider Name="Microsoft-Windows-Winlogon" Guid="{DBE9B383-7CF3-4331-91CC-A3CB16A3B538}" EventSourceName="Wlclntfy" />
  <EventID Qualifiers="32768">6000</EventID>
  <Version>0</Version>
  <Level>4</Level>
  <Task>0</Task>
  <Opcode>0</Opcode>
  <Keywords>0x80000000000000</Keywords>
  <TimeCreated SystemTime="2013-04-29T18:41:02.000000000Z" />
  <EventRecordID>16682</EventRecordID>
  <Correlation />
  <Execution ProcessID="0" ThreadID="0" />
  <Channel>Application</Channel>
  <Computer>Dad-PC</Computer>
  <Security />
  </System>
- <EventData>
  <Data>SessionEnv</Data>
  <Binary>D9060000</Binary>
  </EventData>
  </Event>
```

Table 3. Windows Log CSV Format

ns1:EventID	ns1:Task	SystemTime	ns1:EventRecordID	ThreadID	ns1:Data	Name2
4672	12548	2012-06-11T20:50:19.	201880	4132	S-1-5-18	SubjectUserSid
4672	12548	2012-06-11T20:50:19.	201880	4132	SYSTEM	SubjectUserName
4672	12548	2012-06-11T20:50:19.	201880	4132	NT AUTHORITY	SubjectDomainName
4672	12548	2012-06-11T20:50:19.	201880	4132	0x3e7	SubjectLogonId
4624	12544	2012-06-11T20:50:19.	201879	4132	S-1-5-18	SubjectUserSid
4624	12544	2012-06-11T20:50:19.	201879	4132	DAD-PCŞ	SubjectUserName
4624	12544	2012-06-11T20:50:19.	201879	4132	GONZOWILEYNET	SubjectDomainName
4624	12544	2012-06-11T20:50:19.	201879	4132	0x3e7	SubjectLogonId
4624	12544	2012-06-11T20:50:19.	201879	4132	S-1-5-18	TargetUserSid
4624	12544	2012-06-11T20:50:19.	201879	4132	SYSTEM	TargetUserName
4624	12544	2012-06-11T20:50:19.	201879	4132	NT AUTHORITY	TargetDomainName
4624	12544	2012-06-11T20:50:19.	201879	4132	0x3e7	TargetLogonId
4624	12544	2012-06-11T20:50:19.	201879	4132	5	LogonType
4624	12544	2012-06-11T20:50:19.	201879	4132	Advapi	LogonProcessName
4624	12544	2012-06-11T20:50:19.	201879	4132	Negotiate	AuthenticationPackageName
4624	12544	2012-06-11T20:50:19.	201879	4132		WorkstationName
4624	12544	2012-06-11T20:50:19.	201879	4132	-	TransmittedServices
4624	12544	2012-06-11T20:50:19.	201879	4132	-	LmPackageName
4624	12544	2012-06-11T20:50:19.	201879	4132	0	KeyLength
4624	12544	2012-06-11T20:50:19.	201879	4132	0x210	ProcessId
4624	12544	2012-06-11T20:50:19.	201879	4132	C:\Windows\System32\services.exe	ProcessName
4624	12544	2012-06-11T20:50:19.	201879	4132	-	IpAddress
4624	12544	2012-06-11T20:50:19.	201879	4132	-	IpPort

Table 4. syslog CSV Format

PRIVAL	VERSION	TIMESTAMP	HOSTNAME	APP-NAME	MSGID	MSG
<34>	1	2011-06-11T20:50:19.	mymachine.com	su	ID47	BOM'su root' failed for jwiley on .dev/pts/8
<34>	1	2011-06-11T20:50:19.	mymachine.com	admin	ID48	
<34>	1	2012-06-11T20:50:19.	mymachine.com	su	ID49	BOM'su root' failed for jwiley on .dev/pts/10
<34>	1	2012-06-11T20:50:19.	mymachine.com	su	ID50	BOM'su root' failed for jwiley on .dev/pts/11
<34>	1	2012-06-11T20:50:19.	mymachine.com	admin	ID51	BOM'su root' failed for jwiley on .dev/pts/12
<34>	1	2012-11-11T20:50:19.	mymachine.com	jwiley	ID52	
<34>	1	2012-11-11T20:50:19.	mymachine.com	jwiley	ID53	
<34>	1	2012-11-11T20:50:19.	mymachine.com	jwiley	ID54	
<34>	1	2012-12-11T20:50:19.	mymachine.com	jwiley	ID55	
<34>	1	2012-12-11T20:50:19.	mymachine.com	jwiley	ID56	

At this juncture, a discussion of triples is in order. A triple is a statement about some "thing", comprised of a subject, predicate and an object – not unlike the required grammatical construct of an English sentence. There are several different versions of triples, though the latest recommendations for their representation (e.g. Turtle) may be found on the W3C website. A simple example of a triple is shown below:

<http://example.org/#spiderman>
<http://www.perceive.net/schemas/relationship/enemyOf>
<http://example.org/#green-goblin> . [17]

NOTE: this triple would normally be on one continuous line and is wrapped here due to formatting constraints. The triple says that

"Spiderman is an enemy of the Green Goblin." Spiderman is the subject, "enemyOf" is the predicate and the object is the Green Goblin. For further exploration of the subject of triples, visit the W3C website.

7.3 Logs that SPARQL

Now that the logs have been converted to their respective triple stores, the results can be seen in the figures below. Table 5 shows the Windows event log transformed into a triple store.

Table 5. Windows Log Triple Store

```
<nsl:EventID> <http://vocab.sindice.net/csv/columnPosition>
"0"^^<http://www.w3.org/2001/XMLSchema#integer> .

<nsl:Task> <http://vocab.sindice.net/csv/columnPosition>
"1"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Systemtime> <http://www.w3.org/2000/01/rdf-schema#label>
"Systemtime" ;
   <http://vocab.sindice.net/csv/columnPosition>
"2"^^<http://www.w3.org/2001/XMLSchema#integer> .

<nsl:EventRecordID> <http://vocab.sindice.net/csv/columnPosition>
"3"^^<http://www.w3.org/2001/XMLSchema#integer> .


<http://any23.org/tmp/Threatid> <http://www.w3.org/2000/01/rdf-schema#label> "ThreatID" ;
   <http://vocab.sindice.net/csv/columnPosition>
"4"^^<http://www.w3.org/2001/XMLSchema#integer> .

<nsl:Data> <http://vocab.sindice.net/csv/columnPosition>
"5"^^<http://www.w3.org/2001/XMLSchema#integer> .


<http://any23.org/tmp/Name2> <http://www.w3.org/2000/01/rdf-schema#label> "Name2" ;
   <http://vocab.sindice.net/csv/columnPosition>
"6"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/>tmp/row/0> a <<http://vocab.sindice.net/csv/Row> ;
   <nsl:EventID> "4672"^^<http://www.w3.org/2001/XMLSchema#integer> ;
   <nsl:Task> "12548"^^<http://www.w3.org/2001/XMLSchema#integer> ;
      <http://any23.org/tmp/Systemtime> "2012-06-
11T20:50:19"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <nsl:EventRecordID> "201880"^^<http://www.w3.org/2001/XMLSchema#integer> ;
      <http://any23.org/tmp/Threatid>
"4132"^^<http://www.w3.org/2001/XMLSchema#integer> ;
   <nsl:Data> "S-1-5-18"^^<http://www.w3.org/2001/XMLSchema#string> ;
      <http://any23.org/tmp/Name2>
"SubjectUserSid"^^<http://www.w3.org/2001/XMLSchema#string> .


<http://any23.org/tmp/> <<http://vocab.sindice.net/csv/Row> <http://any23.org/tmp/row/0> .
<http://any23.org/tmp/row0> <<http://vocab.sindice.net/csv/rowPosition> "0" .
```

Table 6 shows the syslog converted into a triple store. These two examples are very much similar in structure due to the use of the same conversion tool. Regardless of the conversion tool, the triple store in the end will be just as easy to work with in the analysis phase. Now that the logs have been converted to their respective triple stores, a log analyst can then query the files using a query engine like SPARQL. In a sense, these queries are only a small improvement over simple matching within unstructured data. The real gain is that the analysis is going on in a semantic domain where further improvements can, and will be, made and suggested in future works. A log-based schema could also be developed so that the triple store can be generated according to that schema. The schema brings into the analysis more meaning about each triple. As the example above showing a triple, the predicate relationship "enemyOf" is defined in the schema at that uniform resource identifier (URI). Making use of existing schemas and/or creating new ones contribute to this semantic process of log analysis.

Table 6. syslog Triple Store

```
<http://any23.org/tmp/Prival> <http://www.w3.org/2000/01/rdf-schema#label> "PRIVAL" ;
   <http://vocab.sindice.net/csv/columnPosition> "0"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Version> <http://www.w3.org/2000/01/rdf-schema#label> "VERSION" ;
   <http://vocab.sindice.net/csv/columnPosition> "1"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Timestamp> <http://www.w3.org/2000/01/rdf-schema#label> "TIMESTAMP" ;
   <http://vocab.sindice.net/csv/columnPosition> "2"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Hostname> <http://www.w3.org/2000/01/rdf-schema#label> "HOSTNAME" ;
   <http://vocab.sindice.net/csv/columnPosition> "3"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/App-name> <http://www.w3.org/2000/01/rdf-schema#label> "APP-NAME" ;
   <http://vocab.sindice.net/csv/columnPosition> "4"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Procid> <http://www.w3.org/2000/01/rdf-schema#label> "PROCID" ;
   <http://vocab.sindice.net/csv/columnPosition> "5"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Msgid> <http://www.w3.org/2000/01/rdf-schema#label> "MSGID" ;
   <http://vocab.sindice.net/csv/columnPosition> "6"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Structure-data> <http://www.w3.org/2000/01/rdf-schema#label> "STRUCTURED-DATA" ;
   <http://vocab.sindice.net/csv/columnPosition> "7"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/Msg> <http://www.w3.org/2000/01/rdf-schema#label> "MSG" ;
   <http://vocab.sindice.net/csv/columnPosition> "8"^^<http://www.w3.org/2001/XMLSchema#integer> .

<http://any23.org/tmp/row/0> a <http://vocab.sindice.net/csv/Row> ;
   <http://any23.org/tmp/Prival> "<34>"^^</34><http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Version> "1"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Timestamp> "2011-06-11T20:50:19"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Hostname> "mymachine.com"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/App-name> "su"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Procid> "-"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Msgid> "ID47"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Structure-data> "-"^^<http://www.w3.org/2001/XMLSchema#string> ;
   <http://any23.org/tmp/Msg> "BOM'su root' failed for jwiley on
.dev/pts/8"^^<http://www.w3.org/2001/XMLSchema#string> ;

<http://any23.org/tmp/> <http://vocab.sindice.net/csv/row> <http://any23.org/tmp/row/0> .
```

Without schemas to add structure and definition, analysis would be no further than a simple matching function that occurs today in current analysis tools.

7.4 Log analysis efficiency

This was a very rough and high-level test of log analysis. The authors' intent was to demonstrate that, while manual and somewhat rudimentary, log analysis can be improved through the use of semantic technologies. Having brought in these semantic technologies, further progress and research as discussed below can now be undertaken. Even with the limited nature of this demonstration, several observations were noted. The Windows event log in XML format was 1.5MB in size. The conversion to a CSV file was performed using the import feature resident in Microsoft® Excel. This import only took a few seconds and resulted in a new file size of 2.0MB. The online web conversion from CSV to a triple store took approximately 30 seconds and resulted in a triple store file of 31.8MB. While this is a 20:1 growth in file size, no solid conclusions can be made about this size growth since there was not a suitable sample size. In addition, secondary storage devices are growing larger and becoming cheaper, presumably negating any drawbacks to any size growth conclusions that would be made given suitable data. The file size growth and processing time for the syslog file was very similar to the Windows event log. Finally, SPARQL queries will,

and did, vary based on the size of the log files and the relative complexity of the query itself.

8. Future work

8.1 Cyberwarfare Doctrine

Intelligence doctrine is the more fertile field between cyberwarfare inherited defense and intelligence doctrine. This area of doctrine contains many more concepts that are candidates for re-use in cyberwarfare. We feel that using semantic technologies are a key, mostly overlooked future avenue for implementing intelligence doctrine. The ability to analyze information that is in a machine-readable format enables organizations to implement doctrine without a huge investment in resources (money and/or workforce). Multiple data stores may be brought together providing additional information that can lead to a better situational awareness and that leads to being more proactive. Analysts can tune their inference engines to do multi-faceted searches across the data and learn more about their enemy.

8.2 Log Schema and Ontology

Log analysis needs not only a log-based schema, but it requires a log-based ontology as well. While the schema is important and gives needed structure and definition to the analysis effort, the log-based ontology provides the ability to reason on the semantic structured data. This reasoning is where the log analysis should be taken to new heights and where more can be done with less. Or, as author F. Coyle says, "Trying to figure out how to get it done with less effort."

8.3 Inferencing Engine

To further progress down the line on semantic log analysis, the analysis needs to develop an inference engine to conduct the reasoning and make inferences across the data. Granted, developing an engine may not be absolutely necessary since there are already models published that do so. The Apache Jena Project, open-source Java framework, is one that the authors are familiar with [18].

8.4 Semantic Hedgehog Concept for Log Analysis

In information technology / proactive defense / log analysis / semantic web domain, Collins' Hedgehog concept would be comprised of the following: a semantic log data / schema, a log-based ontology and an inference engine, as shown in Fig. 7 below.

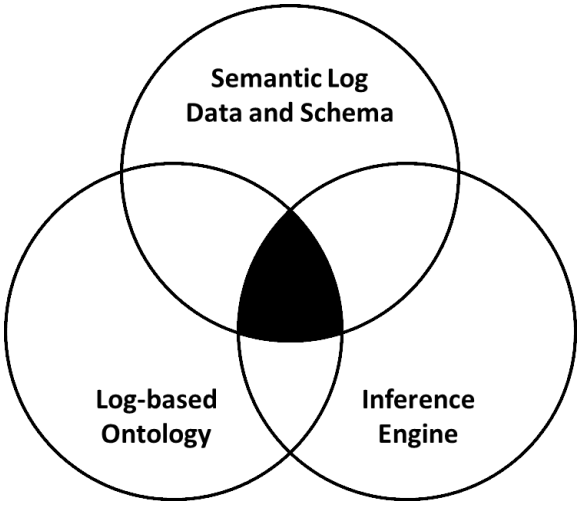

Figure 6. Semantic Hedgehog for Log Analysis

9. References

[1] General Carl Von Clausewitz, "On War," Translated from the original German by Colonel J.J. Graham. LaVergne, TN: Wildside Press, 2009.

[2] "What is Cyber Warfare," Search.Security.com. May 20, 2010. www.searchsecurity.techtarget.com/definition/cyberwarfare?vgnextfmt=print Accessed April 9, 2013.

[3] George and Meredith Friedman, "The Future of War," http://www.phibetaiota.net/tag/friedman/ Accessed January 15, 2013.

[4] United States Department of the Army, "Field Manual 3.0 - Operations," Washington, DC., February 2008.

[5] Sun Tzu, "The Art of War," Edited by Dallas Galvin. Translated from the Chinese by Lionel Giles. New York: Barnes & Noble Books, 2003.

[6] United States Department of the Army, "Field Manual 2.0 - Intelligence," Washington, DC., March 2010.

[7] J. J. Wiley and Frank Coyle, "Semantic Hedgehog for Log Analysis," 2012 International International Conference for Internet Technology and Secured Transactions (ICITST), pp. 748–752, 10-12 December 2012.

[8] Verisign™, "Establishing a Formal Cyber Intelligence Capability," https://www.verisigninc.com/en_US/forms/idefensecyberintel.xhtml Accessed April 5, 2013.

[9] P. Beraud, A. Cruz, S. Hassell, J. Sandoval, J. Wiley, "Cyber defense Network Maneuver Commander," 2010 IEEE International Carnahan Conference Security Technology (ICCST), pp.112-120, October 2010.

[10] Kevin McCaney, "Army's MORPHINATOR: A shape-shifting approach to network defense," Government Computer News, August 3, 2012. http://gcn.com/articles/2012/08/03/army-morphinator-cyber-maneuver-network-defense.aspx?sc_lang=en Accessed August 12, 2012.

[11] James Halbert, Blake Middleton, and Jeffrey Wiley, "Ontology Agent Decision-Making Support for Vulnerability Management," unpublished.

[12] United States Department of the Army, "Field Manual 2.91-6 - Soldier Surveillance and Reconnaissance," Washington, DC., October 2007.

[13] F. Coyle, "The Semantic Web and the Big Data Revolution", http://cs2-nogravity.blogspot.com/p/semantic-web-and-big-data-revolution.html, accessed 27 July 2012.

[14] J. Collins, "Good to Great", New York: HarperCollins, 2001, ch.5.

[15] RFC 5424, The Syslog Protocol, March 2009.

[16] www.any23.org, accessed 26 July 2012.

[17] http://www.w3.org/TR/turtle/, accessed 27 July 2012.

http://jena.apache.org/, accessed 30 July 2012.

