The cyber simulation terrain: Towards an open source cyber effects simulation ontology

Kent O’Sullivan
*Australian Centre for Cyber Security, University of New South Wales*

Benjamin Turnbull
*Australian Centre for Cyber Security, University of New South Wales*

Follow this and additional works at: [https://ro.ecu.edu.au/isw](https://ro.ecu.edu.au/isw)

Part of the Information Security Commons

DOI: [10.4225/75/57a84e3bbefbc](https://ro.ecu.edu.au/isw/60)
THE CYBER SIMULATION TERRAIN: TOWARDS AN OPEN SOURCE CYBER EFFECTS SIMULATION ONTOLOGY

LT Kent O’Sullivan, Dr Benjamin Turnbull
Australian Centre for Cyber Security, University of New South Wales, Canberra, Australia
kent.osullivan@defence.gov.au, b.turnbull@adfa.edu.au

Abstract
Cyber resilience is characterised by an ability to understand and adapt to changing network conditions, including cyber attacks. Cyber resilience may be characterised by an effects-based approach to missions or processes. One of the fundamental preconditions underpinning cyber resilience is an accurate representation of current network and machine states and what missions they are supporting. This research outlines the need for an ontological network representation, drawing on existing literature and implementations in the domain. This work then introduces an open-source ontological representation for modelling cyber assets for the purposes of Computer Network Defence. This representation encompasses computers, network connectivity, users, software, vulnerabilities and exploits and aims for interoperability with related representations in common use. The utility of this work is highlighted against a functional use-case depicting a realistic operational network and mission. Finally, a future research direction is defined.

Keywords
Cyber, Simulation, Cyber Effects, Ontology, Cyber Terrain, Virtual Terrain, CESO, CST, Cyber Simulation Terrain, Cyber Effects Simulation Ontology, Land Power, Futures, Cyber-Dependence Paradox, Cyber Resilience

Disclaimer
The views expressed are the authors’ and not necessarily those of the Australian Army or the Department of Defence. The Commonwealth of Australia will not be legally responsible in contract, tort or otherwise for any statement made in this publication.

INTRODUCTION
On the modern battlefield, the integrity of the cyber systems supporting operations is paramount. A commander who requires precision fires to support mission objectives relies heavily on their networked Command, Control, Communications, Computing, Intelligence, Surveillance and Reconnaissance (C4ISR) capabilities to deliver these effects. Compromise of these systems by an opposing force will deny a commander unimpeded use of their capabilities and lead to probable mission failure. Understanding the interrelation of these systems with the physical and human systems of the battlespace is essential to developing an understanding of their potential vulnerabilities and developing resilience strategies. The effect of a cyber attack is nuanced, and its impact on assets and capabilities is often not immediately apparent against mission objectives. We seek to address these challenges with this work.

BACKGROUND

The new digital networked battlespace presents a paradox. To operate effectively in this future environment, military forces are exploiting C4ISR networks to provide unprecedented levels of command and control (Ormrod 2014a). A military force’s reliance on these systems to operate effectively can quickly become dependence and bring vulnerability (Ormrod, 2014b). The British identify networked C2 as a key enabler of their ability to effectively conduct manoeuvre warfare (DJFCD, 2012). The US Army also recognises this paradox (DJFCD, 2012, p.v) and identify that their superior networked C2 capabilities are a likely target for enemy forces seeking to disrupt their command and control and neutralise their technological advantage (USATDC, 2014).
The significance of this exposure creates an imperative to investigate means to prevent cyber attacks from occurring and plan for disaster recovery, preparing for the probability of operating in an information degraded environment (Scott, 2013). Understanding these problems and preparing for their eventuality is working towards creating organisational cyber resilience. Cyber resilience is premised on:

> The current philosophy of trying to keep the adversaries out, or the assumption that they will be detected if they get through the first line of defense, is no longer valid. Given the sophistication, adaptiveness, and persistence of cyber threats, we can no longer assume that we can completely defend against intruders and must change our mindset to assume some degree of adversary success and be prepared to “fight through” cyber attacks to ensure mission success even in a degraded or contested environment (Goldman, McQuaid, Piccotto, 2011, p1).

The United States Air Force extends this approach and actively advocates for the need to contextualise the impact of cyber attack from a mission perspective, stating:

> The time has come to think of cyberspace in a new light; not only must we defend against any attack, we must be able to “fight through” any attack, accomplish our missions and retain the ability to respond–thus giving us mission assurance in the face of future attacks or other disruptions (USAFSC, 2009, p4).

One of the first prerequisites of a “fight through [approach to cyber resilience is] ...to map USAF network to USAF missions with end-to-end forensics approach ” (USAFC. 2009, p11). Work related to these documents, if any, is not in the public domain. Consequently, there is little research that adequately models the cyber domain from a mission assurance and security impact perspective.

**Planning For A Resilient Future**

There are a near-infinite number of possible scenarios for which a resilient organisation must prepare. A futures view of developing resilience requires decision makers to consider the possibilities, enumerate the plausibilities and deduce the probable issues that they will encounter, enabling the decision maker to determine what actions need to be taken to guide their organisation towards a preferable future state (Hajkowicz, 2015). Anticipating problems and developing contingencies increases resilience.

Computer networks and cyberattacks interact with the battlefield and decision makers as a collection of inextricably linked systems of systems. The complexity of the systems, issues of system to decision maker trust, the inherent uncertainty of complex interactions and the resulting emergent phenomena intersect to form a wicked problem for decision makers (Ormrod, 2014b), making the enumeration of futures extremely challenging.

To understand the complexities of these systems, we must clearly identify the entities, properties and relationships of the systems. To achieve this, we apply a network approach. Philosophically, networks are simply a collection of concepts with defined interrelations. Complex systems are are collections of entities that interact - fitting the network model. The benefit of applying a defined network structure to deconstruct complexity is that it formalises the representation and enables analysts to validate the Emergent Phenomena of the network. Emergent Phenomena are collective behaviours observed in network interactions. These phenomena offer the best window of insight into possible futures. The generation of phenomena is highly dependent on the underlying network structure (Caldarelli, Catanzaro 2012). We have elected to use ontological structures to reliably represent the complex systems of cyber-physical-cognitive interaction on the battlefield as networks.

An ontology is an explicit specification of a shared conceptualization (Gruber, 1995). It is a unifying framework that unites multiple viewpoints to solve problem solving (Uschold, Gruninger, 1996). A ‘strong’ ontology is judged by its real-world semantics, use of logical axioms and machine readability (Obrst, 2010). They are built by collecting and integrating numerous subject-predicate-object triples. Each triple should define a single fact within the ontology. The effectiveness of this triple to represent only a single fact reflects the triple’s semantic strength. In the context of the big data problems associated with modelling complex systems of systems a semantically strong ontology will optimise searching, promote automation and efficiency of scale.

The Cyber Simulation Terrain (CST) proposed in this paper is the explicit definition of the concepts, properties and relationships of the complex system that is organisational cyber-infrastructure. It is a component of the Cyber Effects Simulation Ontology (CSEO) - an effort to model the effects that cyber attacks have on a computer system as part of a larger, more complex interconnection of systems. The transparent semantic strength of the CSEO and CST will permit validation of observed emergent phenomena, enhancing the credibility of simulations modelling the interactions of cyber effects on land forces in combat enabling future organisational shaping towards resilience.
**RELATED WORK**

Representation of computer networks to facilitate analysis of vulnerabilities, attack vectors and mission-node criticality has a long history. Initially work was focused on graphing approaches to the problem (Phillips, Swiler 1998; Swiler, Phillips, Ellis 2001; Sheyner, et. al, 2002), utilising probabilistic statistical models to deduce current and future states and attack paths. Limitations of scalability and fidelity contributed to supersession by information fusion approaches. These approaches seek to efficiently exploit data collected by intrusion detection sensors to generate probabilistic models of likely attack vectors and future network states. The INFERD (Sudit et. al, 2006) and ECCARS (Sudit, Stotz, Holender, 2005) models were among the first major steps to achieving an effective information fusion solution.

TANDI evolved from similar ideas, choosing to isolate the logical network topology from the predictive model (Holsopple, Yang, Sudit, 2006). Independent topology or ‘terrain’ models are a core feature of many of the subsequent information fusion approaches. VTAC (Argauer 2008), FuSIA (Holsopple, Yang, 2008), CAMUS (Goodall, D’Amico, Kopylec, 2009) and the High-Level Information Fusion for Tracking and Projection of Multistage Cyber Attacks (Yang et. al, 2009) all utilised independent terrain models.

Independent terrain models are also used in mission impact analysis as part of the CSIM (Jakobson, 2011a; Jakobson 2011b; Jakobson 2013). They have been influential on the development of the Cyber-ARGUS framework that models the command and control impacts of cyber attacks (Barreto 2012; Barreto 2014) and the modelling of cyber situational awareness (Machado, Barreto, Yano, 2013; Machado, Yano 2014). It has demonstrated utility in supporting cyber attack simulation, present in the CyberSim Modular Cyber Attack Simulator (Moskal et. al, 2013), MASS (Moskal et. al, 2014) and CASCADES (Wheeler, 2014; Kreider, 2015).

All of these approaches have evolved out of the independent Cyber-Virtual Terrain model that was put forward by Fava et. al (2007). Their paper applied an independent terrain model to complex multistage cyber attacks. The catalyst for the development of independent terrain models, however, is the Virtual Terrain (VT) model (Holsopple et. al, 2008). The VT model encapsulates the Cyber Virtual Terrain and defines the core elements of terrain models that endure across its evolutions.

The VT core evolved in the Virtual Terrain version 2 (VT.2) to include the capacity to represent incomplete network information, include a more accurate depiction of internet connectivity and represent routers exclusively as traversal nodes. (Moskal et. al, 2013; Moskal et. al, 2014). The most recent iteration of the VT is the Dynamic Virtual Terrain (DVT) (Wheeler, 2014). Like VT.2, the DVT is intended to support cyber-attack simulations. The DVT is a significant revision of the VT and VT.2. Its key aim is to facilitate the simulation of Moving Target Network Defence Measures (MTNDM) such as IP address hopping and port hopping. The Cyber Terrain (CT) developed by Jakobson in 2011 (Jakobson 2011a; 2011b) has evolved from the original VT to support mission planning and resilience building.

Each of these terrains was assessed for their suitability for inclusion into the CESO. However, lack of publicly available schema has prevented a detailed suitability analysis. Based on the publicly available information, a consistent issue is a lack of granularity. CVT, VT, VT.2 and DVT simplify representation at the cost of fidelity. The mapping of nodal interdependencies by the VT family of models is not sufficient to realistically represent the network. CT maps the dependencies in a granular manner but then abstracts much of the nodal detail that makes the VT family of models useful. DVT has stepped further away from reality, centralising many functions of the network - modifying and abstracting the terrain structure to represent MTNDMs. Existing terrain models also do not address the actual flow of information across the network, sessional communication between nodes, wireless as its own use-case, data spilt or virtualisation.

The CST adopts the CT approach to dependency mapping with the philosophy of the VT family of terrains - gearing towards the simulation of cyber attacks. It will do this at greater levels of granularity and maximise interoperability with linked efforts such as the MITRE standards and the Structured Threat Information eXpression (STIX) (Barnum, 2012). The semantic strength of this ontological structure will enable validation of results generated by simulations and analysis, maximising the efficacy of the predicted future states it represents.

**THE CYBER SIMULATION TERRAIN**

The purpose of the CST is to accurately model the assets and systems across a computer network. The representation must be generic enough to encompass corporate networks also extensible enough to model bespoke and operational systems. Modelling a computer network has several benefits; a comparison between network sensors and predefined models allow for rogue and masquerading device detection, an understanding of impact from hypothetical scenarios across realistic infrastructure, allows for a greater and immediate understanding of the impact of a cyber-security incident, and to understand the impact of a cyber-security
incident on business processes or missions. Finally, a detailed computer and network model can be used for research-based simulation.

Although government publicly advocates endpoint protection (Australian Signals Directorate, 2012) and there are multiple commercial implementations (FireEye; CYLANCE; Bromium; SentinelOne; Bit9; EMC 2015), most of the benefits of these tools relate to blocking infection, reporting endpoint state and patch status, and remote forensic acquisition. There is little work on modelling the network state and shape over time against known states. Changes to the expected behaviour of objects or network shape, caused by misconfiguration, device masquerading or rogue devices, can be easily detected through comparisons to models representing the predefined ‘normal’ network.

A comprehensive cyber security model can also be used to test hypotheses against a network for impact analysis. An example of this was the announcement of the OpenSSL Heartbleed vulnerability (Codenomicon, 2015). Immediately after its publication and public disclosure, there was little understanding of the impact of the exploitable vulnerability across the network. In some cases, it took several days for providers to fully appreciate the effects to internal systems, operations and data. A network model allows for new vulnerabilities and exploits to be tested against a particular network system to determine immediate impacts and ongoing leverage points. These can be injected into the model and their impact tested without endangering an operational environment. The CST is designed to accomplish these aims.

The CST Schema is open source and available at:

https://github.com/AustralianCentreforCyberSecurity/Cyber-Simulation-Terrain

Public availability makes the ontology unique in that it is available for critique and analysis. The repository also contains several use-cases that highlight its intended use. The schema has been implemented using the Resource Description Framework (RDF) (RDF Working Group, 2014a) Turtle (TTL) (RDF Working Group 2014b) syntax with minimal elements of the Ontology Web Language (OWL) (OWL Working Group, 2012). Queries are performed using the Sparql Protocol And Rdf Query Language (SPARQL) (SPARQL Working Group, 2015). The CST Schema has been designed to achieve granularity in content and structure in representation. Figure 1 depicts the schema. The blue nodes represent the concepts or objects in the ontology. The red represent properties (will be leaf nodes) and relationships (will be traversal nodes). The visualisations in this paper were created using the easyRDF converter (Humfrey, 2015) and the RDFGravity Visualisation Tool (Goyal, Westenhaler, 2015).

Schema

The CST is an interconnection of nodes. Nodes can be specialised as computers, routers, IDS sensors or Domain Controllers. Nodes connect to Subnetworks through a Network Interface Controller (NIC) (including wireless and virtual nodes). Computer MAC and IP addresses are associated with their NIC, permitting the representation of a computer belonging to multiple physical, virtual and wireless networks via multiple interfaces. Subnetworks connect to each other and the internet through a router. A router will have defined routes that control the traffic flow across the network between subnets. The internetwork is the network of routers - a conceptual addition that assists in modelling the interconnection of subnets.

Computers will have associated software and services. These concepts are arranged hierarchically in a similar manner to the Virtual Terrain service tree concept. Computers are associated with an installed version of software. That installed version will be associated with a parent software type class to facilitate categorisation and querying. The more granular relationships are to the service that software projects when running on the computer and any known vulnerabilities associated with a piece of software. Zero-days are represented as an exploit related to a vulnerability, but will have no links to a CVE or other published vulnerability. Software versions are associated with their CPE ID (MITRE, 2015a), vulnerabilities with their CVE identifier (Martin, 2001) and weakness types are associated with a CWE (MITRE 2015b) number. Metrics of criticality and exploitability associated with vulnerabilities utilise the CVSS (FIRST CVSS SIG, 2015) scores. The intention of integration with these standards is to leverage the existing resources, maximise the interoperability of the CST with the wider CESO as one of the modules. The vulnerability association is also the interface with the ‘red team’ elements of the CESO, an abridged implementation of STIX, also aims to leverage existing work and maximising interoperability.
Installed software runs either local or remote services. Local services run internally to the node, and remote services project their Service Name, Port Number and Protocol to the NIC connected to their computer forming an implicit host-based firewall - if a service has a port open on the network card, it is assumed to be listening. If it is not listening it will not be shown. Services also have an associated privilege level. These privilege levels map to the privilege levels associated with users. If a user has a privilege level that matches the service, they can interact with it. Users can be associated with a computer either locally or as part of a domain. Local users can access only a computer with which they have an explicit link. Domain users belong to a domain that can have multiple member computers. Domain users can access any computer in a domain to which they are authenticated. Domains have a domain controller that connects to a single subnet but can control domains across multiple subnets. There can be multiple domains per subnet and multiple domains per computer, allowing us to represent a user who might have a secret and an unclassified domain account on the same computer and accounting for data-spill use-cases. Users are owners of their data. Data is stored locally on computers on a disk. The disks can be encrypted or unencrypted. If encrypted, a user has to authenticate to access the data.

Intrusion detection and antivirus is implemented at the host level. A host-based signature IDS is represented as a running service on a node. The IDS links to a signature database that will consist of the unique identifiers matching vulnerabilities and exploits. The IDS has configurable rules that are a subtype of the software configuration concept linked to the installed version of the software. These rules allow the IDS to be ‘tuned’ by selecting the signature and action on detection. MTNDMs can be represented in the CST by its linkage to the CESO Event Ontology.
**Example Use Case**

The use-case implemented below is drawn from the paper defining the CESO (Ormrod, Turnbull, O'Sullivan, In-Press). It depicts the process required for a Joint Fires Team (JFT) to call for Offensive Support (OS), the call for fire to be handled by a Joint Fires Communication Centre (JFCC) and have the mission fired by an Artillery Battery. Figure 2 visualises the required cyber infrastructure to conduct this task. It does not demonstrate the full capabilities of the terrain but is a proof-of-concept of the most commonly used features.

The use-case depicts three subnetworks connected by routes on the same router - the JFT Network, the JFCC Network and the Artillery Battery Network. All computers on these networks are members of the ‘Joint Fires’ Domain and are accessible by all authenticated users. Computers are running a mix of local and remote software and services (services are fabricated for this use-case). The JFT has Targeting and Request for Fires Communications software. They use this to communicate with the JFCC, who use their Mission Control software to push the fire mission to the Artillery Battery HQ. The communications between the JFCC and the Artillery Battery HQ are encrypted. However, the software that they are using is an older version of OpenSSL that is still vulnerable to Heartbleed. The Battery HQ uses their Battery Management Software to push the fire missions to individual Artillery Troops who use their Fire Control Software to prosecute the fire mission.

In addition to effectively representing this use-case, the CST can also elicit information about the network. Figure 3 shows the results of queries run against the use-case. The first query can quickly produce information about computer and network attributes. The second query checks for the presence of known vulnerabilities,
detecting Heartbleed on the JFCC and Battery HQ computers. Query three then returns all known information about the vulnerability.

These are only a small subset of the queries that the CST uses to determine possible, plausible and probable futures. The emergent phenomena(potential future state) in this use-case is the potential relationship that an actor attacking the network has with the Heartbleed vulnerability. The attacker could compromise the confidentiality of that link, use it to intercept communications and learn of the impending fire mission, and warn the targets of the fire mission, potentially giving the enemy time to evade or prepare counter-battery fire in response to the attack. These fascinating and relevant second and third order effects are not knowable from looking at information in isolation. A decision maker who learns that a probable future has their OS capability rendered ineffective by an enemy at a time of their choosing will likely wish to take actions (such as patching the vulnerability) to move towards a preferable future where their OS capability remains effective.

CONCLUSION AND FUTURE WORK

In this paper we have outlined the requirement and initial implementation of a publicly-available ontological schema designed to assist in Cyber Security Research. There are several areas of future work arising from this; the highest priority of these being the continual improvement and refinement of the ontological schema. Beyond this, it is expected that the ontology will evolve over time as new concepts arise, new network designs emerge, and additional information is required. Some of these changes will be minor, and some may require backwards-incompatible changes. As these occur testing, harnesses and deployments will be updated.

The biggest area of future work is in the development of systems to populate, reason on and visualise this ontology. There are also multiple defined use-cases that require further development. There are precedents for the automated detection and ingestion of data sources and several associated challenges (Moir, Dean, 2015; Grove et al., 2013) that will guide these efforts. The first stage of development will be to support automated simulation development and analysis.

REFERENCES


