Assessment Framework for Evaluation of Active Defense of In-Vehicle Networks

Nathaniel Toll, Northeastern University Global Resilience Institute

Akash Patel, Northeastern University Global Resilience Institute

Robert Knake, Northeastern University Global Resilience Institute

Stephen Raio, US Army Research Lab

Daniel E. Krych, US Army Research Lab

Approved for release: September 9, 2019

Introduction

With vehicles becoming more and more connected and reliant on in-vehicle networks and connected systems to operate safely, defense methods have evolved to include active defenses to potential adversaries. The next step after implementation of an active defense on an in-vehicle network is to develop a way to quantify the active defense system's ability to protect the in-vehicle network from adversaries that threaten the confidentiality, integrity and availability of the network and the safe operation of the vehicle.

There have been previous works that focus on penetration testing of vehicles [1] [2] [3], as well as numerous guides to penetration testing of enterprise networks [4]. The majority of these focus on how to discover and use exploits for privilege escalation, lateral movement and enumeration. This framework does not look at the specifics of how to conduct a penetration test on a vehicle, or an in-vehicle network, using active defense techniques, but rather proposes a system for measuring the success of any deployed systems in defending the in-vehicle network. Measuring the impact of active defense against an adversary will help decision makers understand the effectiveness of active defense and determine how it will fit into the overall schematics, security and risk management of vehicle design and operations. First, we outline our assumptions for the testing, then describe how the testing should be implemented, and finally describe our system of metrics for measuring the success of active defense against adversarial tactics.

Assumptions for Testing

In order to create valuable metrics for evaluating active defense, first we must outline what assumptions we make about the environment in which the defense will be tested.

1. We assume that the testing takes place on an in-vehicle control network using CAN bus, which is on a tabletop and simulating the sensors found in a real vehicle.
2. The assessment framework assumes a testing environment that utilizes multiple electronic control units (ECUs) and sensors with some of the sensors producing and reading simulated data.

3. The ‘red team’, or adversary, has a skillset commensurate with the anticipated threat actor(s) for the system under evaluation. The metrics proposed in this framework are inherently entangled with the capability of the ‘red team’. Therefore, in order to produce relevant results, the evaluation must proceed utilizing the most capable adversary expected to be encountered by the system. Alternatively, evaluation can be performed with ‘red teams’ of various skill level to determine the effectiveness on the active defenses in relation to adversary capability.
4. The ‘red team’ has physical access to the network. The goal of testing is to examine the active defense effectiveness and not that of the physical security measures which could be put into place to stop an adversary from accessing the network. Although, an active defense solution could include logic based security measures to stop an adversary from connecting an unauthorized node to the bus.
5. The active defense framework measures are implemented on the network autonomously so that the ‘blue team’ will not need to interact with the active defense during the time that the defenses are tested by the ‘red team’.

Methodology of Assessment

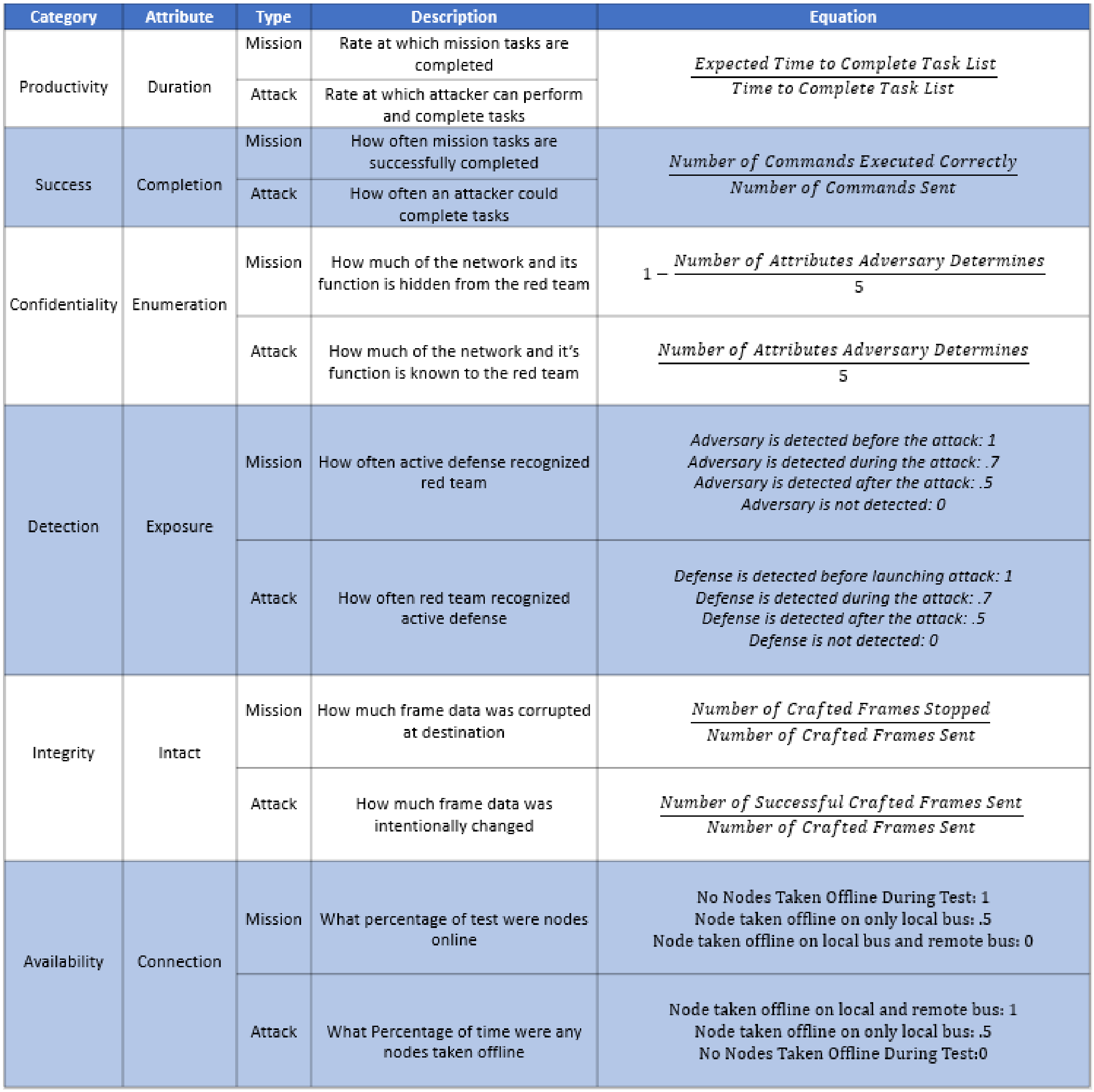
The assessment can be framed by the National Institute of Standards and Technology (NIST) Technical Guide to Information Security Testing and Assessment recommendations [5]. The testing utilizes both target identification and target analysis as the adversary tries to enumerate sensors and ECUs and identify defenses in use on the CAN bus. Testing also incorporates target vulnerability validation because the adversary will attempt to use security flaws in the CAN bus protocol to reverse engineer messages or deny service to the CAN bus. The testing can be classified as an internal and overt assessment. Internal, because of the assumption that the adversary has physical access to the in-vehicle network, and overt because of the nature of the testbed, the team defending the network is aware of the testing but will not be actively trying to change technical methods of defense.

The proposed metrics are designed to measure effectiveness in an assessment that includes the red team attempting to complete tasks that are major threats to in-vehicle networks. An offensive and defensive task list, which encompasses all of the categories and their descriptions, should be defined in order to properly employ the metrics and evaluate the system. In our example, some of the tasks the red team attempts to complete are: reverse engineering frames and replaying them to replicate a specific set of vehicle operations, detecting the presence of defenses on the network, crafting and sending frames to achieve targeted results, and attempting to deny service to one or more nodes on the network. The blue team is not be able to act dynamically, but rather will rely on the deployed defense to act autonomously. The effectiveness of the deployed defenses are then quantified through the following metrics.

Metrics to Quantify Effectiveness

To quantify the effectiveness of the active defense framework against simulated adversaries, a methodology similar to that outlined in [6] and [7] is be proposed. [6] describes a way to quantify the success of moving target defenses on an enterprise network by evaluation within the four categories shown in Table 1 of [6] which are productivity, success, confidentiality, and integrity. This paper also adds an availability and detection category to the existing four. The categories are quantified in an attribute and evaluated, both with and without the active defenses in place. This allows for comparison and quantification of success compared to an undefended network by subtracting the metrics for each run and interpreting the final result. A minimum of three consecutive runs is suggested for a final result, in order to capture the changes in the Mission and Attack side as they may both adapt to the others actions and responses and learn. It is also possible to continue running the test with different

Table 1*: Metrics for quantifying offensive and defensive effectiveness when deploying active defense techniques on vehicle networks.*



Category	Attribute	Type	Description	Equation
Productivity	Duration	Mission	Rate at which mission tasks are completed	$\frac{Expected\ Time\ to\ Complete\ Task\ List}{Time\ to\ Complete\ Task\ List}$
		Attack	Rate at which attacker can perform and complete tasks	
Success	Completion	Mission	How often mission tasks are successfully completed	$\frac{Number\ of\ Commands\ Executed\ Correctly}{Number\ of\ Commands\ Sent}$
		Attack	How often an attacker could complete tasks	
Confidentiality	Enumeration	Mission	How much of the network and its function is hidden from the red team	$1 - \frac{Number\ of\ Attributes\ Adversary\ Determines}{5}$
		Attack	How much of the network and it's function is known to the red team	$\frac{Number\ of\ Attributes\ Adversary\ Determines}{5}$
Detection	Exposure	Mission	How often active defense recognized red team	*Adversary is detected before the attack: 1* *Adversary is detected during the attack: .7* *Adversary is detected after the attack: .5* *Adversary is not detected: 0*
		Attack	How often red team recognized active defense	*Defense is detected before launching attack: 1* *Defense is detected during the attack: .7* *Defense is detected after the attack: .5* *Defense is not detected: 0*
Integrity	Intact	Mission	How much frame data was corrupted at destination	$\frac{Number\ of\ Crafted\ Frames\ Stopped}{Number\ of\ Crafted\ Frames\ Sent}$
		Attack	How much frame data was intentionally changed	$\frac{Number\ of\ Successful\ Crafted\ Frames\ Sent}{Number\ of\ Crafted\ Frames\ Sent}$
Availability	Connection	Mission	What percentage of test were nodes online	No Nodes Taken Offline During Test: 1 Node taken offline on only local bus: .5 Node taken offline on local bus and remote bus: 0
		Attack	What Percentage of time were any nodes taken offline	Node taken offline on local and remote bus: 1 Node taken offline on only local bus: .5 No Nodes Taken Offline During Test:0

combinations of defenses and calculating the differences between metrics which would show how one combination preforms over another or how a defense can be changed to better account for adversarial tactics. We modify the categories and definitions of types, descriptions, and attributes of these categories to better fit the automotive network. First, in the context of our framework, 'Mission' is defined as standard actions by authorized nodes on the CAN bus. 'Attack' is defined as actions by an adversary on the CAN bus. Each category is reviewed and discussed in the context of the in-vehicle network.

Each metric category is weighted to represent its importance to the protection of the in-vehicle network. The weights chosen here, and scores for *Detection* and *Availability* are suggested as a starting point for a general vehicular network, but should be adjusted according to the importance of each category in the environment being evaluated. These scores are multiplied by their respective weights and then summed to form a final score. The reasoning for each proposed weight is discussed in the description of the categories. See Table 1 for a summary of what is defined in the following sections.

Productivity:

Weight: 5

In [6], *Productivity* is described as "the rate at which mission tasks are completed". This metric, applied to in-vehicle networks, measures the time it takes for either adversary or legitimate user to execute a given series of commands over the course of the assessment. An example task list could be: start engine, turn on headlights, release break. The time it takes either a user to complete the task (Mission) or the time it takes an adversary (Attack) is quantified in the *duration* attribute. The completion of the task list is evaluated as a percentage (represented in decimal form) of the shortest expected time it should take to complete the task list divided by the actual time it took for the list to be completed. The attribute *duration,* in the attack type, represents the difficulty for an adversary to reverse engineer the active defense employed. For example, if unable to reverse engineer the active defense frames and complete the task (Attack), the adversary receives a score of 0. The legitimate users score (Mission) should be closer to 1 as they do not have to reverse engineer the frames. A weight of 5 is assigned to *Productivity* because time to complete the tasks is less important than the completion of the tasks themselves, which is captured in the *Success* metric.

$$\text{Mission and Attack :}\ \left(\frac{Expected\ Time\ to\ Complete\ Task\ List}{Time\ to\ Complete\ Task\ List}\right) = Duration$$

Success:

Weight: 15

[6] describes *Success* as "how often mission tasks are successfully completed" and, from an adversary's view, "how successful an attacker may be while attempting to attack a network". This is tested on the in-vehicle network through the same process as the *Productivity* category: commands are sent by the authorized user and the adversary attempts to reverse engineer and use commands to control aspects of the network. The difference in the *Success* category is the quantification attribute. The attribute *completion* is the number of commands executed correctly divided by the number of commands sent. This metric, therefore, enumerates the adversary's ease of not just reverse engineering, but their ability to successfully manipulate the network. If they are able to quickly engineer the frames (as represented by *duration*), then they may be able to easily use them to send authentic commands. This metric also quantifies if a defense method has an effect on the execution of legitimate commands (Mission). A weight of 15 is assigned to *Success* because the execution of commands is important for both the red and blue teams in Mission or Attack success.

$$\text{Mission and Attack: } \frac{Number\ of\ Commands\ Executed\ Correctly}{Number\ of\ Commands\ Sent} = Completion$$

Confidentiality:

Weight: 20

In [6], *Confidentiality* seeks to measure the amount of mission information exposed in a way that an adversary could eavesdrop or intercept. For *Confidentiality,* in the context of the in-vehicle network, Mission is described as "How much of the network and its function is hidden from the red team" and for Attack, "How much of the network and its function is known to the red team". In order to exploit a network, it is important for an adversary to be able to enumerate the network and also understand the protocols used through analyzing observed traffic. This knowledge, or lack thereof, could greatly increase or hinder an adversary's capability on the network. To quantify *Confidentiality*, we use the metric *enumeration* which measures the adversary's ability to determine five different attributes: the architecture of the bus the adversary is given access to, the architecture of all buses, the functions and protocols of the nodes on the bus the adversary is given access to, the functions and protocols of all the nodes on all buses, and the state of the vehicle itself. State of the vehicle refers to the adversary's ability to fully interpret the data and determine what the vehicle is doing at that moment (e.g., speed, rpm, light status). There are many different things the adversary could do to determine these vehicle attributes, including analyzing message and data rates on the bus and comparing them to what is written into standards or analyzing the utilization of each node. Additionally, vehicle state is its own attribute because there may be instances where all other attirbutes are known, but the true vehicle state is not, such as the case where deception is leveraged as a defense. *Confidentiality* is given a score of 20 because an adversary knowing the structure of the in-vehicle network and the function of the nodes, has the ability to better exploit the nodes or network via crafted frames.

Mission: $1 - \frac{Number\ of\ Attributes\ Adversary\ Determines}{5} = Enumeration$

Attack: $\frac{Number\ of\ Attributes\ Adversary\ Determines}{5} = Enumeration$

Detection:

Weight: 20

Detection is used in conjunction with *Confidentiality*, and represents the ability of the active defense to identify an adversary, and the ability for an adversary to recognize the active defense. The Mission type description is “how often active defense recognized the red team” and Attack type description is “how often the red team was able to identify active defense”. The attribute *exposure* is used and represented as a value between 0 and 1. For the Mission and Attack there are three components to the metric. First, a component of *Detection*. On the Mission side, if an adversary is detected before, or at the onset of an attack , a score of 1 is assigned because detection at this point allows for the maximum ability to respond, block or monitor the adversary. If the adversary is detected during the attack, a score of .7 is assigned because action can be taken in response to the adversary’s actions as they are ongoing, but the adversary has already attempted or completed some malicious activity. If the adversary is detected after the attack, then a score of .5 is assigned because action cannot be taken to stop or monitor their actions, but knowledge of thier presence can initiate incident response and other follow-on measures as appropriate. Here we define ‘before, or at the onset of an attack’ as an adversary being present and/or performing passive reconnaissance. The scores are skewed toward detection, because detection, even after the fact, is disproportionately more valuable than failing to detect. On the Attack side, the same numbers are used for *Detection*, but the definition changes to the adversary detecting the defense versus the defense detecting the adversary. The number scores assigned for Attack represent the usefulness of deployed defense knowledge by the adversary. The scores for Attack are: 1, when the adversary knows a defense is in place before launching the attack and can modify tactics to be more successful, .7 when the adversary detects the defense during their actions and can, therefore, modify their tactics to increase their success, .5 when they discover a defense after an attack which could help in countering it in future malicious actions, and 0 if the adversary does not detect the defense.

Over several runs of attacks and defenses, the defender may learn and adapt to the adversary and vice versa. Consider even if the attack is successful the first run, but the adversary learns and adapts their methodology to improve their speed or accuracy. Similarly, a defender may note a simple way to detect and block a unique, repetivitve attack. A weight of 20 is assigned to *Detection* because detection itself is less important than the ability for commands to be successfully executed, but is important because the known presence of an adversary can help kick off more defenses or alert the operator to disregard certain vehicle commands and take physical precautions to ensure safe operation.

Mission: *Adversary is detected before the attack: Exposure = 1*
Adversary is detected during the attack: Exposure= .7
Adversary is detected after the attack: Exposure = .5
Adversary is not detected: Exposure = 0

Attack: *Defense is detected before launching attack: Exposure = 1*
Defense is detected during the attack: Exposure = .7
Defense is detected after the attack: Exposure = .5
Defense is not detected: Exposure = 0

Integrity:

Weight: 10

[6]’s category of *Integrity* has an attribute of *intact* and is described for Mission type as “mission information is transmitted without modification or corruption” and for Attack type, “the accuracy of information viewed by an attacker”. In the context of testing active defense, to evaluate the metric of *Integrity,* the adversary’s ability to spoof/modify legitimate messages and craft messages is quantified. A measure of intactness closer to 1 for either Mission or Attack indicates more success. Mission will have a higher score when crafted frames are rejected by the network, while Attack will score higher if the crafted frames have the desired effect and are not identified and/or rejected. This number represents how difficult it would be for an adversary to spoof legitimate traffic with their own crafted frames, versus replaying a known set of commands. *Integrity* has a weight of 10 because, in this context, it is more dependent on *Productivity* and *Success* attributes. An adversary that scores highly in *Productivity* and *Success* will also most likely score well in *Integrity*, making it more of an addition to those and making 30% of the total score an adversary’s ability to reverse engineer, replay, and craft frames.

Mission: $$\frac{Number\ of\ Crafted\ Frames\ Stopped}{Number\ of\ Crafted\ Frames\ Sent} = Intact$$

Attack: $$\frac{Number\ of\ Successful\ Crafted\ Frames\ Sent}{Number\ of\ Crafted\ Frames\ Sent} = Intact$$

Availability:

Weight: 30

For an in-vehicle network assessment framework we also add *Availability* to the metric. *Availability,* in the context of this framework, is defined as the time that all nodes on the network are available to receive traffic. The description for Mission is, “the time that all sensors in the network are online” and for Attack, “the time any sensor on the network is not able to receive or send traffic”. The attribute for *Availability* is called *connected*. The testing method looks at whether an adversary, attempting a denial of service attack, is able to take down any nodes on any CAN bus which would interfere with their ability to receive and process traffic. The node taken offline cannot be the same on which they are operating. A weight of 30 is assigned to *Availability* because it is a critical aspect of nodes and systems to be online for a vehicle to function correctly and safely. If even one node is brought offline, it could affect the operation of the entire vehicle. It also is a different type of attack that relies on a different type of message creation than what is captured in the other metrics, leading to its higher weight.

Mission: $No\ Nodes\ taken\ offline\ during\ test: Connection = 1$

$Node\ taken\ offline\ on\ only\ local\ bus: Connection = .5$

$Nodes\ taken\ offline\ on\ local\ bus\ and\ remote\ bus: Connection = 0$

Attack: $Node\ taken\ offline\ on\ local\ and\ remote\ bus: Connection = 1$

$Node\ taken\ offline\ on\ only\ local\ bus: Connection = .5$

$No\ nodes\ taken\ offline\ during\ test: Connection = 0$

Summary of Metrics

These six attributes (P*roductivity*, *Success*, *Confidentiality*, *Detection*, *Integrity* and *Availability*) taken together, provide a metric score which quantifies the effectiveness of active defenses. The closer the score is to 100, the better for either the blue team or red team. For example, a score closer to 100 for the Attack type (red team) would mean an adversary was able to quickly reverse engineer frames and replay them, detect active defense measures, craft frames of their own and take parts or all of the network offline. For the Mission category (blue team) a high Mission score would mean that all of the legitimate traffic was able to be sent and read, including quick correct task list execution, they were able to identify an adversary on the network and stop denial of service of any of the nodes. Ideally, the Mission score would be close to 100 and attack score close to 0. Enclosed in Appendix A is a sample score sheet to use during a test to keep track of scoring.

On a macro level, the success of an active defense needs to be quantified by what features of the vehicle the adversary could ultimately control or influence in a way that negatively effects the safety of its operation. Decision makers need to determine how the output of this framework and analysis contribute to the ability for an adversary to influence vehicle operations and how that affects the overall risk threshold of their mission.

Conclusion and Further Work

We have proposed a framework to quantify active defense in a testbed environment based on six categories and their attributes, measured in a 'red team' versus 'blue team' testing scenario. One additional application could be to individually test different combinations of the techniques that, used in concert, make up an active defense. This testing, compared with baseline results of a system without defenses in place, would help quantify the effectiveness of each component and find the most effective component or combination of techniques. The next step would be to move evaluation from the testbed environment to a real-world vehicle. Future work could also consider the limitations of our assumptions and how alterations could bring about a more advanced quantification of success in penetration testing against active defense techniques on in-vehicle networks. This would ultimately include physical defenses and quantifying the difficulty of achieving the access needed to begin to attack a network employing active defense. Another avenue of research is that of proposing a risk structure for in-vehicle networks, specifically those with active defense techniques employed, to determine what level of risk, using metrics like the ones proposed in this paper, are acceptable for different vehicle networks and vehicle mission utilizations. Overall, the use of these metrics allow the employers of active defenses to quantify their ability to thwart adversaries who have network access and are attempting to exploit their access to maliciously control or influence in-vehicle networks.

References

[1] S. Talebi, "A Security Evaluation and Internal Penetration Testing of the CAN-bus," Chalmers University of Technology, Goteborg, Sweeden, 2014.

[2] C. Smith, The Car Hacker's Handbook, San Francisco: No Starch Press, 2016.

[3] R. Currie, "Developments in Car Hacking," SANS Institute, Rockville , 2016.

[4] T. Wilhelm, Professional Penetration Testing, Burlington, MA: Syngress, 2015.

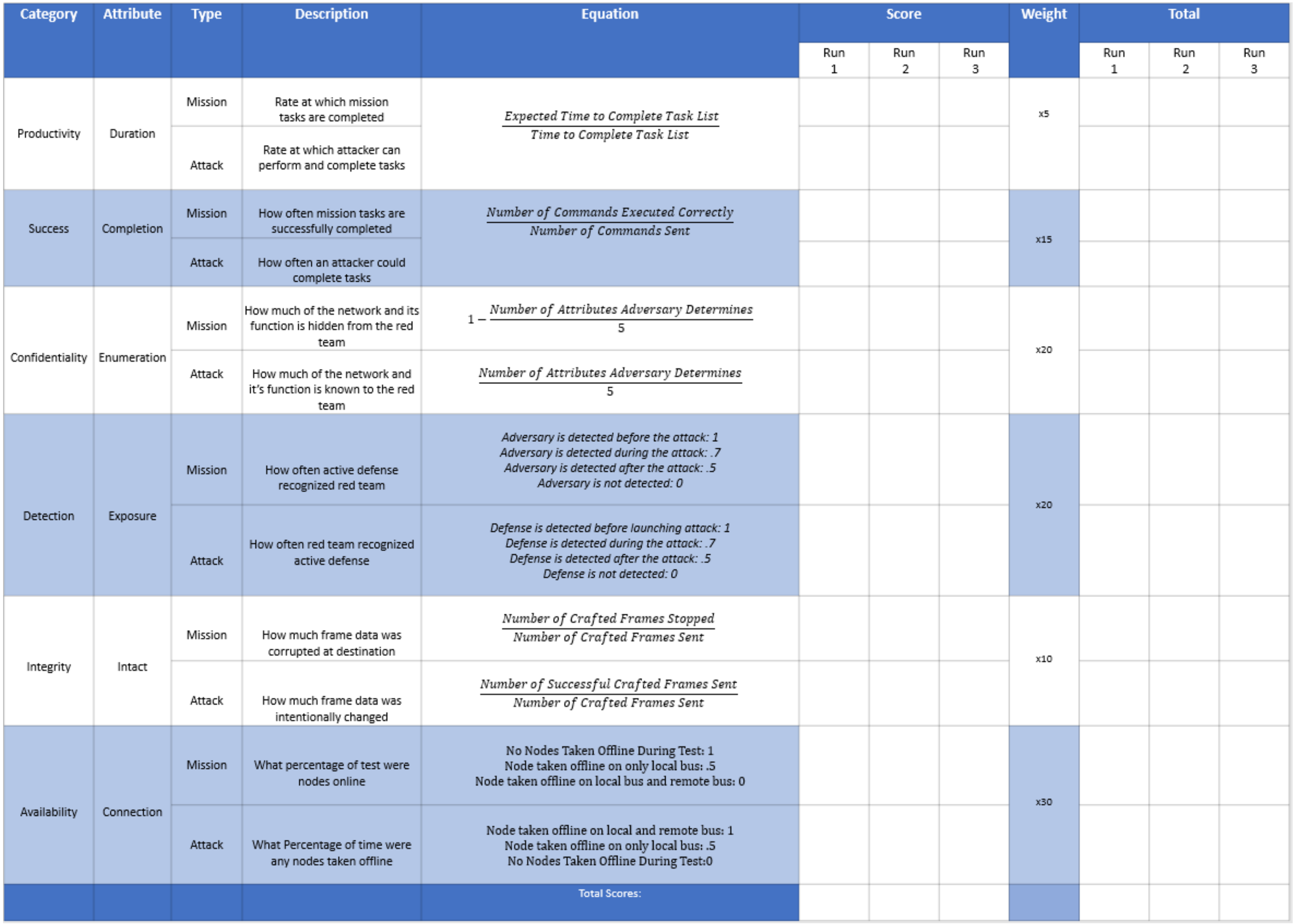
[5] K. Scarfone, M. Souppaya, A. Cody and A. Orebaugh, *Technical Guide to Information Security Testing and Assessment,* Gaithersburg, MD: National Institute of Standards and Technology, 2008.

[6] J. Taylor, K. Zaffarano, B. Koller, C. Bancroft and J. Syversen, "Automated Effectiveness Evaluation of Moving Target Defenses:Metrics for Missions and Attacks," in *MTD '16 Proceedings of the 2016 ACM Workshop on Moving Target Defense*, Vienna, Austria, 2016.

[7] K. Zaffarano, J. Taylor and S. Hamilton, "A Quantitative Framework for Moving Target Defense Effectiveness Evaluation," in *MTD '15 Proceedings of the 2015 ACM Workshop on Moving Target Defense*, Denver, CO, 2015.

[8] S. Abbott-McCune and L. A. Shay, "Techniques in Hacking and Simulating a Modern Automotive Controller Area Network," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, Orlando , 2016.

Appendix A: Score Sheet for Assessment Framework



Category	Attribute	Type	Description	Equation	Score			Weight	Total		
					Run 1	Run 2	Run 3		Run 1	Run 2	Run 3
Productivity	Duration	Mission	Rate at which mission tasks are completed	$\frac{\text{Expected Time to Complete Task List}}{\text{Time to Complete Task List}}$				x5			
		Attack	Rate at which attacker can perform and complete tasks								
Success	Completion	Mission	How often mission tasks are successfully completed	$\frac{\text{Number of Commands Executed Correctly}}{\text{Number of Commands Sent}}$				x15			
		Attack	How often an attacker could complete tasks								
Confidentiality	Enumeration	Mission	How much of the network and its function is hidden from the red team	$1 - \frac{\text{Number of Attributes Adversary Determines}}{5}$				x20			
		Attack	How much of the network and it's function is known to the red team	$\frac{\text{Number of Attributes Adversary Determines}}{5}$							
Detection	Exposure	Mission	How often active defense recognized red team	*Adversary is detected before the attack: 1* *Adversary is detected during the attack: .7* *Adversary is detected after the attack: .5* *Adversary is not detected: 0*				x20			
		Attack	How often red team recognized active defense	*Defense is detected before launching attack: 1* *Defense is detected during the attack: .7* *Defense is detected after the attack: .5* *Defense is not detected: 0*							
Integrity	Intact	Mission	How much frame data was corrupted at destination	$\frac{\text{Number of Crafted Frames Stopped}}{\text{Number of Crafted Frames Sent}}$				x10			
		Attack	How much frame data was intentionally changed	$\frac{\text{Number of Successful Crafted Frames Sent}}{\text{Number of Crafted Frames Sent}}$							
Availability	Connection	Mission	What percentage of test were nodes online	No Nodes Taken Offline During Test: 1 Node taken offline on only local bus: .5 Node taken offline on local bus and remote bus: 0				x30			
		Attack	What Percentage of time were any nodes taken offline	Node taken offline on local and remote bus: 1 Node taken offline on only local bus: .5 No Nodes Taken Offline During Test:0							
				Total Scores:							