Informing a Defensive Strategy by Analyzing Reactive Cyber Intrusion Detection

V.V. Kulkarni and S.C. Whetstone

The Challenge: The Department of Defense Information Network (DoDIN) is under constant attack from adversaries ranging from independent hackers to sophisticated nation-states. Breaches in computer networks can occur through phishing campaigns, exploiting vulnerabilities via hacking tradecraft, malicious insiders, and surreptitious physical access to network infrastructure.

Breaches in computer networks can occur through phishing campaigns, exploiting vulnerabilities via hacking tradecraft, malicious insiders, and surreptitious physical access to network infrastructure.

The Department of Defense Information Network (DoDIN) contains more than 4 million computers and 3 million users. The DoDIN has access points around the world, including on military bases, ships, aircraft, and cellular devices. Many sub-networks in the DoDIN transfer data over the same physical channels as the Internet, including fiber optic cables and satellite. With so many entry points, the DoDIN is under constant attack from adversaries ranging from independent hackers to sophisticated nation-states. Breaches in computer networks can occur through phishing campaigns, exploiting vulnerabilities via hacking tradecraft, malicious insiders, and surreptitious physical access to network infrastructure. Network administrators and defenders traditionally focus on protective defense, i.e., preventing initial network compromise. Unfortunately, persistent adversaries will inevitably find ways to breach protective defenses. It is therefore crucial that network defenders receive training on reactive defense: the ability to detect, respond to, and restore networks after initial compromise. This article describes how IDA's analytical framework for operational cybersecurity assessments of Combatant Command training exercises conducted between fiscal years 2014 through 2016 informed a defensive strategy.

The data collected during these assessments provide insights on both the attacker and defender actions. The attackers act as part of the assessment team and provide information regarding the individual actions, the linkages between actions, and their perception of success. The attack thread is the instrument used to probe and measure the detection capability of the cyber defenses. The defenses and defenders are under evaluation, and the data collection captures detection of the attacks, defensive responses, and operational effects.

Analytical Focus—Cyber Attack Threads

An attack thread is *a series of steps taken by an adversary encompassing the intrusion and subsequent exploitation of a network*. Attack threads end either with the adversary conducting an information effect on an objective node or with the network defenders stopping the adversary before they reach an objective node.

What factors are important in detecting an attack? Experience with assessments suggests two factors: type of access and type of tool used.

Detection Factor—Logical Access

Logical access to software is either authenticated or unauthenticated. Authenticated access involves presenting a credential, which the software checks and validates before granting access. Credentials come in many forms, such as usernames and passwords, or tokens and PINs. Unauthenticated access involves accessing software without presenting credentials. Adversaries gain unauthenticated access by techniques such as SQL injection, malicious file uploads, booting a workstation from an unauthorized DVD, buffer overflows, and other malformed requests.

Detection Factor—Tool Type

Cyber attackers use tools to perform actions. These tools are either native or foreign. A native tool is one that the network owners authorize for use on the network. Since many operating systems in the DoDIN run Microsoft Windows, many of Microsoft's command line tools and software packages are permitted and are considered native tools (Powershell for instance). Foreign tools are tools that network owners have not authorized for use on the network. Foreign tools include scanners, malware, viruses, beacons, and command and control software.

Cybersecurity principles suggest how detection will vary with these two factors. Defenses typically attempt to identify suspicious or unusual actions and alert defenders to investigate them. Authenticated access and use of native tools by definition are normal actions and thus typically generate no alerts, making attack actions with these characteristics harder to detect, as illustrated in Figure 1.

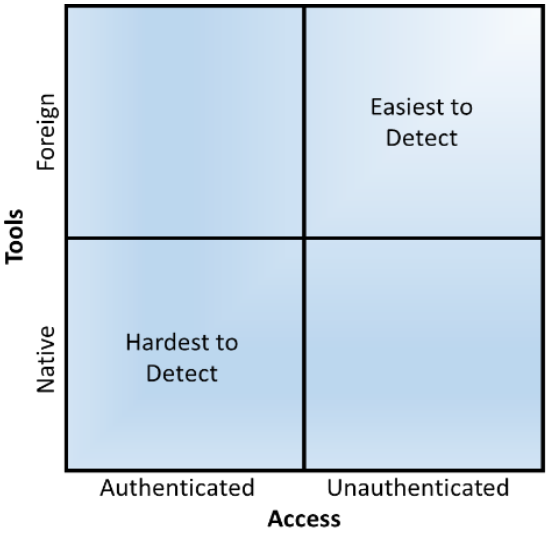

Figure 1. Network defenders have a higher likelihood of detecting the Red Team when they use unauthenticated access and foreign tools in their attack threads

This simple model suggests a two-fold defensive strategy for improving detection. First, force the adversary to operate in the portion of the space where detection is easier by using more foreign tools and unauthenticated access.

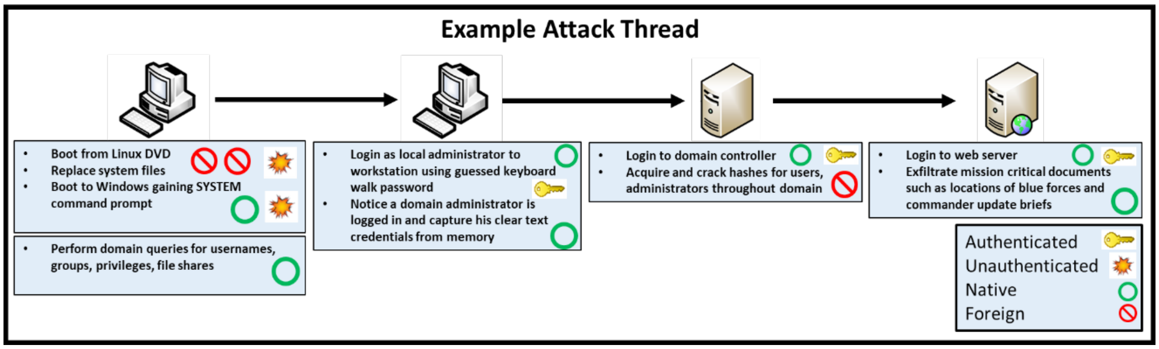
Example Attack Thread
Boot from Linux DVD
Replace system files
Boot to Windows gaining SYSTEM command prompt
Perform domain queries for usernames, groups, privileges, file shares
Login as local administrator to workstation using guessed keyboard walk password
Notice a domain administrator is logged in and capture his clear text credentials from memory
Login to domain controller
Acquire and crack hashes for users, administrators throughout domain
Login to web server
Exfiltrate mission critical documents such as locations of blue forces and commander update briefs
Authenticated
Unauthenticated
Native
Foreign

Figure 2. An example Red Team attack thread

Second, structure the defenses to reduce the portion of the space where actions are difficult to detect by improving detection of native tools and authenticated accesses by unauthorized users or adversaries.

The operational assessments use DoD Cyber Red Team attacks to stimulate and gather data about the defenses. Figure 2 shows an example attack thread with the Red Team starting from physical access to a workstation (without possessing login credentials), escalating their privileges to domain administrator, and subsequently stealing mission-critical documents from a file server undetected.

Strategy Implementation—Attack Thread Characterization Metrics

The details of the attack threads provide insight on how well the network defenses are forcing an adversary to rely on detectable actions during cyber-attacks. The ratio of actions within an attack that used unauthenticated accesses and foreign tools to the total number of actions in that thread captures the ability of the attacker to operate where detection is less likely.

$$metric\ 1 = \frac{\#\ of\ unauthenticated\ accesses}{total\ \#\ of\ accesses}$$

$$metric\ 2 = \frac{\#\ of\ actions\ using\ a\ foreign\ tool}{total\ \#\ of\ actions}$$

The approach is to compute the metric for each observed attack thread. For example, the attack thread depicted in Figure 2 contains five accesses, two of which are unauthenticated, yielding a score of 2/5 or 40%. Of 10 total actions, 3 use foreign tools, yielding a score of 3/10 or 30%. These metrics provide insight on the success in forcing attackers to operate where defenses can detect them. The metrics enable a comparative analysis to determine how changes in the network, perhaps over time or when changing a detection device, affect the attacker actions.

Next consider the ability to measure the defenses themselves.

Defensive Performance—Logistic Regression

A logistic regression predicts the probability to detect future attacks based on the ratios of unauthenticated accesses and use of foreign tools. This analytical approach uses a binary response: the defenders either detected or did not detect the attack. Combining this binary response with the two metrics previously defined as the continuous variables in a logistic regression allows us to model the conditional probability that the defenders will detect an attack as a function of the fraction of unauthenticated accesses and foreign tool use. The conditional probability of detecting an attack thread given factors x and y is:

$$P(Detect|x,y) = \frac{e^{f(x,y)}}{1 + \ e^{f(x,y)}}$$

$$f(x,y) = \ \beta_0 + \beta_1 x + \beta_2 y + \beta_3 (x - \bar{x})(y - \bar{y})$$

In this model, the variables x and y represent metric 1 (unauthenticated access) and metric 2 (foreign tool use). β_0, β_1, β_2, β_3 are constants. $\bar{x}$and $\bar{y}$represent the averages of the two metrics over all attack threads. Each attack thread is represented by three values (x, y, Detected? (yes or no)), and statistical software determines the four constants in the logistic regression model.

This model contains only two factors: the percentage of unauthenticated accesses in an attack thread and the percentage of foreign tools used. Other factors affect detection rates as well. Network defenders may receive tips from intelligence reports for example. Additionally, network defenders may know the Red Team's IP address space prior to an assessment, skewing detection rates. Nonetheless, the two factors provide useful insight into observed defensive performance and how to safeguard an information network.

As an empirical model, the regression requires a data set of diverse attack threads. The analyst can then construct a two-dimensional space with contour lines showing how the probability to detect an attack varies with the characteristics of the attack, as illustrated in Figure 3. The operational assessments of training exercises in fiscal years 2014 through 2016 provided a data set to apply the methodology and provide insights to the DoD network defenses.

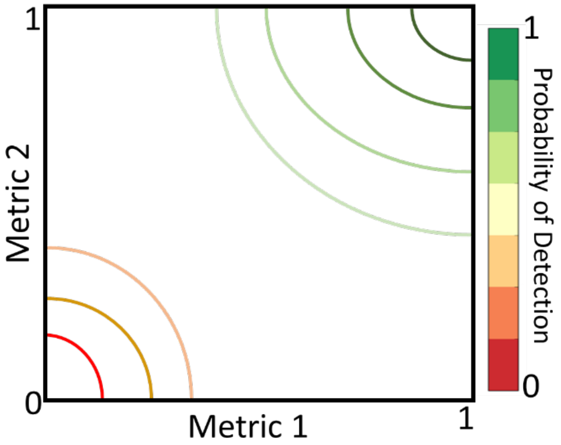

Figure 3. Conditional Probability to Detect defines an operational space with contours identifying regions of greater or lesser chance of detection. (Qualitative sketch only)

Application of the Framework

Logistic regression analysis of attack threads during training exercises confirmed insights on the detection performance of the observed cyber defenses. Although the specific performance values for the DoD networks are classified, observations regarding applicability of the framework are not.

The probability of detection varied with lower probabilities to detect in the lower left quadrant of the operational space and higher probabilities in the upper right quadrant. This observation is consistent with expectations of difficulty in detecting actions with different types of access and tools. The logistic regression supports the two-fold defensive strategy for improving detection by denying an adversary the ability to operate where defenses are weak, and it identifies specific areas where defenses are less likely to detect an adversary.

The framework analytically confirmed the anecdotal observation that the network defenses are improving, but not enough to stop the cyber Red Teams. The logistic regression quantitatively confirmed an improving probability of detection over the period of the assessments, and that the Red Teams remained successful by adjusting their tactics to operate in the region of the operational space where the defenses were less likely to detect them.

In addition, the analytical framework can provide insights on how network design affects detectability. For example, assume that a network includes publically accessible websites that do not require authentication but store sensitive information. The network defenses would not necessarily alert on such accesses or detect attacks against those assets. The logistic regression should show a low probability of detection for attacks having a high fraction of unauthenticated access, which contradicts expectations from the simple model for difficulty in detection. Defenders with such knowledge could adjust their network design to minimize such publicly accessible websites or alter operational procedures to specifically monitor for such attacks to improve the probability of detection.

Integrating over the conditional probability to detect yields the total probability of detecting an attack:

$$P(Detect) = \iint P(Detect \mid x, y)\, p(x, y)\, dx\, dy$$

where $p(x, y)$ is the probability for an adversary to operate at point (x, y). In other words, $p(x, y)$ is the probability of a given attack thread to have metric 1 = x and metric 2 = y. This probability distribution is strongly dependent on adversarial tools, techniques, and procedures.

A simple approach is to assume the adversary is equally likely to operate anywhere in the operational space of Figure 3. In this scenario, $p(x, y) = 1$. The total probability to detect an attack can be calculated for different sets of attack threads to compare across time. Doing so shows that in fiscal years 2014 through 2016, the probability of a defender detecting an attack has risen. An analyst also could develop a tailored profile $p(x, y)$ from intelligence data for specific adversaries to estimate the expected defensive performance.

The analytical framework also leads to a general set of principles and recommendations for improving detection of attacks.

Force the adversary to use more foreign tools and unauthenticated access

Cyber adversaries seek valid credentials in order to blend in as authorized users. They may obtain credentials by cracking hashes stored on disk, extracting clear-text credentials from memory, locating clear-text password files, guessing, and keylogging. Cracking hashes is trivial when users have weak passwords such as keyboard walks. Network administrators must routinely perform password audits to ensure that users have strong passwords.

Additionally, network administrators must remove all or encrypt all clear-text password files found on workstations and servers. If adversaries cannot acquire credentials, they must try unauthenticated access and are therefore easier to detect. Network administrators should also restrict the use of certain native tools. Publicized breaches of organizations show that adversaries commonly use Windows native tools such as Powershell, Procdump and PsExec. Although network administrators use these tools as well, normal users will not. Therefore network administrators should restrict the use of these tools.

Increase the detectability of native tools and credential misuse

Network defenders can configure host-based security system rulesets to flag upon any executable. They should therefore be aware of users using native tools that are not necessary in their day-to-day routines. For example, hackers execute Procdump on the Windows background running process lsass.exe in order to access clear text credentials from running memory. An innocent user would not do such a thing, and therefore network defenders can spot anomalous behavior performed by native tools. Furthermore, network defenders should be aware of which users routinely use remote login tools such as PsExec and which do not. Anomalous use of PsExec could be an indication of an adversary attempting to move laterally in the network with valid credentials.

Dr. Vikram V. Kulkarni is a Research Staff Member in IDA's Operational Evaluation Division. He holds a Doctor of Philosophy in physics from Rice University.

Dr. Shawn C. Whetstone is a Research Staff Member in IDA's Operational Evaluation Division. He holds a Doctor of Philosophy in nuclear engineering from the University of Michigan.