Assignment 6
Gunnar Yonker

1.

An anomaly in the context of intrusion detection is a data such as a pattern that is considered abnormal or an outlier when compared to the larger dataset of “normal” actions. For intrusion detection this could be something like network traffic being very active during the hours of 7am to 5pm, and little to no traffic during the hours of 1am to 4am. A tool like NIDS could then compare an increase of traffic during the hours of 1am to 4am to what is considered typical to then know that the increase of traffic is an anomaly and can be flagged as potentially malicious traffic. Another example of an anomaly would be a large influx of traffic from an unexpected IP address, this IP address would be considered an anomaly because it is unexpected when compared to the usual IP address connections. The IP address could also be compared to the database of known malicious addresses and then deemed malicious, so it is flagged by the NIDS.

2.

Z-score is one of the most popular models for anomaly detection based on a set of data and the z-score stands for the number of standard deviations. Z-score works based on standard deviations, the more standard deviations away a given piece of data is, the more likely that it is an anomaly. A z-score is calculated using the mean and standard deviation of a dataset. The z-score is then calculated by using the following equation where x is the data point in question: (x – mean)/Standard Deviation

Interquartile range (IQR) is another statistical model that can be used for anomaly detection where the mid-point is the median of the dataset, and each mid-point is called a quartile. The IQR is the region that is between the 1st and 3rd quartiles. An outlier is determined by being either before the 1st quartile or after the 3rd quartile. If it is, this is determined to be an outlier and can be labeled as anomaly for a detection program to flag.

One last statistical model that builds off of the IQR is the boxplot, which functions very similar to the IQR but has a better visual representation of the data with outliers. When creating a boxplot for anomaly detection, the box is the IQR and then the whiskers extend out left and right from the box by 1.5*IQR. An outlier/anomaly is considered for this data set if the datapoint is outside of either of the whisker values. Therefore, if a datapoint is under or over either of those values then it would be considered an anomaly and could be flagged as potentially malicious.

3.

Activity profiling is used to describe an activity management and what kind of improvements can be made. All of the aspects of activity profiling include managerial, operation, social, and technological aspects that can be improved to promote efficiency.

One anomaly metric used for profiling of anomaly detection is precision. Precision answers the question of what proportion of identified anomalies are true anomalies? The precision for profiling of anomaly detection is represented by the equation (true positive)/(true positive + false positive). An example of this would be in a situation where a NIDS identified 4 true positive anomalies, and 3 false positive anomalies.

4/(4+3) = (4/7) = .57

Anomalies can also be profiled by using recall, which is the question of what proportion of true anomalies were identified? This metric calls into question how many true positive anomalies were detected compared to the amount of true positives and false negatives. The equation for this metric is: (true positive)/(true positive + false negative). An example of this continuing off the above example would be a NIDS detecting 4 true positive anomalies, 3 false positive anomalies, and 2 false negatives.

4/(4+2) = (4+2) = .67

The two above metrics can be combined into an anomaly metric called an F1 score. The F1 score takes into account the precision and the recall calculations to assign an F1 score which is the overall performance of the anomaly detection system. This calculation is completed by using the equation (2*(recall*precision))/(precision+recall). An example of calculating the F1 score of an anomaly detection system could be applied using the values calculated above in the situation that a NIDS had identified 4 true positives, 3 false positives, and 2 false negatives. By calculating the F1 score there would be a numerical assignment to the overall performance of the NIDS based on the data analyzed through these values.

(2*(.57*.67))/(.57+.67) = 0.76/1.24 = 0.61

These metrics are appropriate for detecting anomalies because they can provide a numeric representation of the performance of the anomaly detection system. These metrics consider what anomalies were identified and what proportion of the anomalies are true anomalies. There are other metrics such as false positive and false negative rates that can be calculated to assess the performance of anomaly detection.

4.

a. I agree with the statement that signature based detection may have a higher rate of false positive rates than anomaly based detection. I think that this is possible because setting a specific rule could result in “good” behavior still being detected as a threat even when it was intentional and not a malware threat. Whereas, anomaly based detection recognizes the pattern of a user to see what is normal and what that user typically does. When that user starts to do something abnormal and off the normal pattern, it results in a flag. There are a few reasons that a user might start doing something abnormal, but I feel like they are less likely to be a legitimate change in the user’s behavior compared to the chance of it being due to an intrusion of malware. An example of this situation would be that a false positive could occur with signature based detection through a rule such as the user not being able to run a .exe file for an installation. The user may be trying to install a program they need such as Microsoft Outlook and this would be flagged as abnormal due to running that type of file and would result in a false positive as the user is intentionally trying to install this program and it is not an intrusion where an attacker is installing a malicious program. I believe that it is more likely that a situation like that may happen where a user unintentionally commits abnormal behavior due to a specific rule stating that the behavior is abnormal. I think that anomaly based detection would have less false positive rates because the user would most likely stick to their typical normal pattern of how they operate on their system. If suddenly that user at 2am starts renaming and digging through administrative/system files, this would be considered abnormal and flagged as such. I think that it is more likely that when the detection system finds a user breaking away from their normal pattern, that it will be a true positive. Signature based detection sets one rule that applies to many different users/situations as more of a blanket, resulting in more false positives. Anomaly based detection operates on a more case to case basis for each user having a normal pattern. One user may not have the same pattern as the other, but each would still have what their “normal” pattern is and when they stray from that to a certain degree it will be flagged as abnormal behavior resulting in fewer false positives than signature based detection.

b. I disagree with this statement because NIDS mainly detects network traffic abnormalities based on recognized patterns but does not inspect the traffic itself if it is encrypted data. This is where HIDS would have an advantage because after the data was received to the system through a NIDS where detection may have been avoided. For example, an encrypted data packet goes through a NIDS to the system undetected because it would not be analyzed. Then the destination host becomes infected with malicious code that runs a program at 2am that starts copying over system data to a remote host. NIDS would not detect this situation, but HIDS would be able to detect this difference in the pattern of activity on the user’s system. In this situation, NIDS did not detect this intrusion, but HIDS was able to detect it, therefore this statement would not be true. Another example would be if HIDS was able to read the hash of a file, comparing it to a known database of malicious hashes to detect an intrusion and notify the administrator. This same file could be contained in a packet that freely passes through the NIDS because it has an approved spoofed IP address, during a time which it is expected and normal, and is encrypted so that the NIDS is unable to analyze what is contained. This is a situation where HIDS is able to detect an attack that NIDS was unable to detect. Using NIDS and HIDS together would help to combat this type of attack, but neither HIDS nor NIDS can “do it all”.

c. I disagree with the statement that anomaly based intrusion detection has better use on HIDS than on NIDS because I believe that anomaly-based intrusion detection has its own uses on both HIDS and NIDS, not that one is better than the other. I also think that for the best security setup, if possible, it is best to use HIDS and NIDS together so that any encrypted malicious data that passes through NIDS can be detected on the system using HIDS. An example of how anomaly based intrusion detection is useful on NIDS if that a network could anticipate a certain amount of traffic during business hours, and any amount of traffic deemed as an outlier compared to normal could be treated as an anomaly thus triggering a flag for that traffic to be reviewed. This could be in a situation where a business expects outside traffic coming in at a steady stream between 7am to 7pm but expects outside traffic coming in at a steady stream between 7am to 7pm. Then if there was a sudden spike in traffic at 1am, that would be considered an anomaly based on the expected amount of traffic at that time going through the NIDS. With HIDS, anomaly based intrusion detection is also useful because it can help to detect abnormal patterns that may be carried out by malicious programs/codes. An example of this would be that the system never accesses higher privileged files or does so very rarely, but at abnormal hours a program runs that starts to rewrite important system files. HIDS would detect this pattern as an anomaly and flag it as such so that an administrator would be notified. Anomaly based intrusion detection has it’s own use for NIDS and HIDS to best prevent against intrusion of malware, only using it on HIDS or NIDS will not guarantee more success over the other. Use of both NIDS and HIDS, if possible, using anomaly based intrusion detection would provide a more secure defense.