Attachment I: Data Analysis Methodology

1.0 Overview

The RF Feasibility Test Team conducted several hundred test runs using various combinations of vehicles, orientations and persons. This data produced thousands of "tag events" which were saved in a database and analyzed in a variety of fashions. This section details the underlying data acquisition and analysis methodology.

2.0 Terminology

It is important to detail some of the basics of RFID reader data. An RFID reader polls for tags from hundreds up to thousands of times a second. Thus, when a tag enters the field of an RFID reader, if detected, it may be read just once or hundreds to thousands of times. Therefore, a simple analysis of just the read occurrences does not reveal the whole picture. Ignoring the "quality" of the read leads to conclusions that may not bear out across scenarios different from the one tested. Additionally, the pass/fail nature of a read is not well suited to statistical analysis methods, forcing a simple comparison of average (mean) read occurrences. Thus, several related values could be analyzed to obtain a qualitative and quantitative picture.

The diagram in Figure I-1 illustrates some important concepts. In these diagrams, the horizontal axis is time. Each tick mark on the axis represents an attempt by a reader to read a tag. Successful reads are noted by an "X" above the axis.

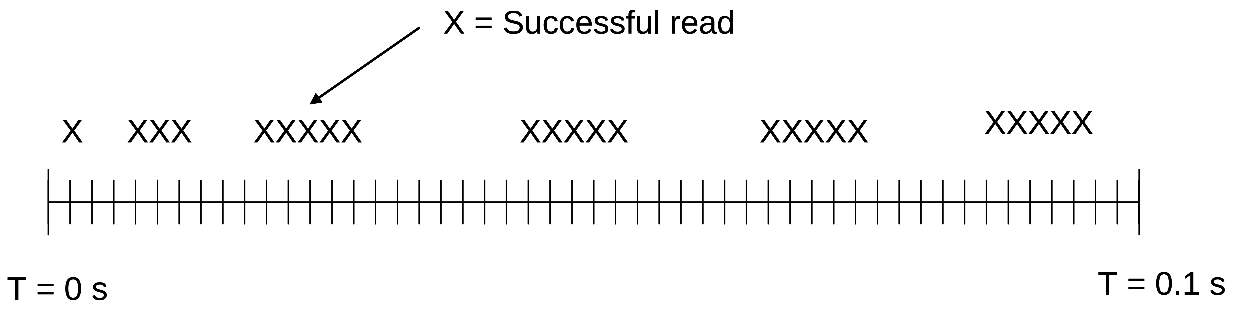

Figure I-1 – Illustration of RFID Reads over Time

Several terms require definition. The first is read score. This is a value indicating whether a tag has been read or not. Effectively, this measures whether a tag is seen - at all - across all read attempts. For each run executed, a pass-fail event was generated for each tag. In the diagram above, there are several successful reads for this tag, therefore the read score =1. read score is the most often quoted performance metric for RFID, sometimes referred to as readability or read rate. For the purposes of the RF Feasibility tests an average read score was calculated, which is the number of tags read divided by the total number of tags in the test. So, if a test run had 50 tags, and 36 were read, then the average read score for that run = 0.72 (36/50).

Another very useful statistic is the read count. This is the count of all the successful reads of a tag during a test run. If, for a given run, read count > 0 then it is a successful read, and the read score=1 for that run. In Figure I-1, the total successful reads, and hence the read count, is 24. This is one of the most valuable statistics in measuring readers, since higher read count's indicate robustness of reading.

Information withheld
under FOIA Exemptions
(b)(4) and (b)(5)

For the test setups, the team was able to retrieve the read score for all vendors' tags, but limited to read count information. The reason for this is that the vendors' application programming interfaces were tuned for read score reporting, but reporting of read count would impact performance. The vendors felt that read score is what ultimately matters from a "real world" perspective, hence the decision to leave out read count reporting.

3.0 Statistical Comparisons

This section details the analysis techniques that were used to determine performance, and hence vendor selection. Section 7 details the recommendation and the impacts on the concept of operations (CONOPS) that the recommendation entails.

In the analysis of the data, three measures of effectiveness were considered. The first is RFID performance. This basically translates to "which vendor's technology performed better under certain conditions?" The second measure is robustness of the system. How well could a vendor's product handle adverse handling and conditions? Considering these two areas allows for a vendor selection.

The third area involves CONOPS for RFID. Performance and robustness are not the sole drivers of a solution. Privacy, visitor acceptance, suitability for integration into existing operations and accuracy of data all impact the final decision.

For analysis of performance and robustness, both individual and aggregate measures of comparison were utilized. Individual measures are used to determine differences in technologies, as well as to provide a basis for extrapolation. Although speeds greater than 50 MPH were not tested, an inference of what performance would be obtainable is derived through the individual measures.

The team developed comparisons of two means for each statistic of merit. For each set of data, the mean (average) and the standard deviation s (the standard deviation is the positive square root of the variance $\sigma 2$) are calculated. Smaller standard deviations indicate that values will be more tightly clustered around the mean. For purposes of the analysis, a 90% confidence interval (CI) was adopted. This means that for two sets of statistics it can be determined, with 90% confidence, that they are different (or not).

Statistics for both [redacted] which are denoted by subscripts [redacted] and [redacted] have been generated. The analysis methodology tested the hypothesis that the distribution of the difference of the mean values, [redacted] does not include zero. In other words, if the difference in the population means allows for zero, one cannot accept that the difference between the two vendors' performance is statistically significant. In mathematical terms, the three applicable hypothesis statements are expressed as:

H_0: [redacted] > 0
H_1: [redacted] $= 0$
H_2: [redacted] < 0

A test for these hypotheses uses statistical tools for comparison of populations. To compare the population means, it is appropriate to use a common one-tailed t-test. This test incorporates the standard deviation, mean, and sample size to determine the difference in populations. The inclusion of a zero value in the range of possible differences determines if the population means are truly different – values of zero indicate that there is not a true difference in means. The following inequality is used to determine the ranges:

[redacted]

Figure I-2 – One tailed t-test of means (population mean and variance estimated)

For a CI of 90%, α = 0.05 and the t value for this is 1.645. The values of [redacted] are the total number of samples for [redacted] respectively, with [redacted] being the calculated averages of the [redacted] read scores, and [redacted] are the estimates of the read score population's variances. The calculated averages and variances from the test data sets are used construct the inequality values for the test cases. As a standard, the comparison always uses [redacted] as the statistical test.

If the inequality value intervals are purely positive, it can be stated that with 90% confidence that the [redacted] mean is higher than that of [redacted] If the interval contains zero, it is too close to call. Finally, if the interval is completely negative, it can be stated that the [redacted] population mean is higher than that of [redacted] The table in Figure I-3 illustrates the results of the test.

Description	Hypothesis
[redacted] outperforms	H_0: [redacted] > 0
Statistical Tie	H_1: [redacted] = 0
[redacted] outperforms	H_2: [redacted] < 0

Figure I-3 –Assessing Vendors Based on Hypothesis Testing