Follow-up Analysis of Mobile Robot Failures

Jennifer Carlson, Robin R. Murphy, Andrew Nelson
Center for Robot-Assisted Search and Rescue
Computer Science and Engineering
University of South Florida
Tampa Florida

***Abstract*— This paper extends previous work characterizing mobile robot failures by including recent data and organizing the failures according to a novel taxonomy which includes human failures. Failure type and frequency data were collected from fifteen robots representing three manufacturers and seven models over a period of 3 years, in a variety of environments. The data was analyzed using standard manufacturing measures for product reliability. Statistical analysis shows that the time between failures, time to repair, and downtime vary widely. For this reason the means reported here are not statistically significant. The results do show that the overall mean time between failures (MTBF) and availability have improved since the previous analysis but are still low. The MTBF across all robot types was found to be 24 hours and availability was 54%. The analysis showed that the control system was the most common source of failures (32%), followed by the mechanical platform itself. It also showed that different components caused failures at different relative rates depending on the body type of the robot.**

I. Introduction

Data about how robots fail is needed for a variety of reasons. Many model-based fault-tolerant systems, such as [7] and [15], require the type and frequency of failures. Common failures and the conditions in which they are most likely to occur are important information for manufacturers as well. Such data can be used in both the design and testing phases of product development to improve the reliability of the next generation of robot platforms. Failure data are also needed to allow researchers and program managers to better estimate and control development time.

The extensive use of mobile robots over the past three years at the University of South Florida (USF) has produced a reasonable database of mobile robot failures and their characteristics. The Center for Robot-Assisted Search and Rescue currently has twenty-one robots from six manufacturers. CRASAR spends more than 200 hours per year using the robots in the field.

This paper examines the user logs and collected failure type and frequency data of the most heavily used robots at CRASAR. The failures were categorized using a newly developed taxonomy of robot failures described in Sec. III. Standard manufacturing measures for the reliability of a product were also used to examine the data (Sec. IV in terms of the *mean time between failures*, *availability*, and *average downtime*. These results were further examined in Sec. V using basic statistical analysis methods for physical failures (frequency, component, research vs. field, impact), human failures, and repairability. The expected probability of failure associated with each leaf in the taxonomy tree is provided. The paper concludes in Sec. VI that the MTBF has improved but a huge gap still exists between research and field robots.

II. Related Work

Previous work by CRASAR includes a detailed analysis on the failures encountered while using robots in the World Trade Center (WTC) rescue operation reported by Micire in [8]. In 2003, this work was expanded by adding an analysis on failures encountered during the day to day use of robots by CRASAR [2]. The findings showed an average MTBF of 8 hours (6 for field robots) and availability of less than 50% (64% for field robots). The effectors were the most common sources of failures (42%) for field robots. Overall the control system was the second most frequent source of failures at 29%. This paper extends the study completed a year ago in 2003 with the addition of a complete taxonomy of mobile robot failures, inclusion of an additional year's worth of logs, statistical analysis of the results, and the addition of human failures.

The results from eight studies conducted by TECO, part of the Maneuver Support Center at Fort Leonard Wood, have been posted to the Department of Defense Joint Robotics Program[12]. The overall goal of the studies was to evaluate the feasibility of using the robotic platform for its assigned tasks in the Future Combat System (FCS). These studies were performed on a wide variety of platforms: small mobile platforms, several bulldozers, and a modified M1 tank. TECO has reported a MTBF of less than 20 hours, similar to the 24 hours found here.

In addition to the 10 studies listed above, a workshop on robots used in museums produced two studies on the reliability of mobile robots actively used for long periods of time. Both studies were focused on presenting their respective platforms and briefly mentioned the MTBF in order to help categorize the performance of those systems. Nourbakhsh [11] describes a set of four autonomous robots used for a period of five years as full-time museum docents. Their robots reached a mean time between failures of 72 to 216 hours. In [14] Tomatis *et al.* described a robot used for a shorter period of time, and reported a MTBF of 7 hours.

Other efforts have concentrated on identifying the weaknesses of robots in field applications but have not provided quantitative failure data. In [1] Blitch provides a survey of the mobility problems. Casper, Micire, and Murphy [4] present a discussion of the constraints which the USAR application domain places on robotic technology. In [9] Murphy, Casper, Hyams, Micire, and Minten discuss the same issues as Casper *et al*. [4] but provide some additional discussion on the need for adjustable autonomy.

III. Taxonomy of Failures

For the purposes of this paper, a *failure* is defined as *the inability of the robot or the equipment used with the robot to function normally*. Both complete breakdowns and noticeable degradations in performance are included. In order to gain insight into how and why mobile robots fail, a taxonomy was developed and is illustrated in Fig. 1. This taxonomy draws from the *robotics*[2], *human-computer interaction*[10], and *dependability computing*[6] communities.

Failures are categorized based on the source of failure and are divided into *physical* and *human* categories, following dependability computing practices. Physical failures subdivided into classes based on common systems found in all robot platforms, these being *effector*, *sensor*, *control system*, *power*, and *communications*. Effectors are defined as *any components that perform actuation and any connections related to those components*. This category includes for example, motors, grippers, treads, and wheels. The control system category includes *the on-board computer, manufacturer provided software, and any remote operator control units (OCU)*.

Human failures (also called human error) are subdivided into *design* and *interaction* subclasses. *Mistakes* are caused by fallacies in conscious processing, such as misunderstanding the situation and doing the wrong thing. *Slips* are caused by fallacies in unconscious processing, where the operator attempted to do the right thing but was unsuccessful.

Each failure, regardless of physical or human, has two attributes, *repairability* and *impact*. The severity of the failure is evaluated based on its impact on the robot's assigned task or mission. A *terminal* robot failure is one that terminates the robot's current mission, and a *non-terminal* failure is one that introduces some noticeable degradation of the robot's capability to perform its mission. The repairability of the failure is described as either *field-repairable* or *non-field-repairable*. A failure is considered field-repairable if it can be repaired under favorable environmental conditions with the equipment that commonly accompanies the robot into the field. For example, if a small robot which is transported in a single backpack encounters a failure, the tools required for the repair would have to fit in the backpack along with the robot and its support equipment in order for the failure to be classified as field-repairable.

IV. Methods

This section describes the equipment used, the methodology for data collection, which data were collected, and the calculations used to generate the results presented in Sec. V.

A. Robots

Of the twenty-four mobile robots used at USF over the past three years, fifteen were considered by this analysis. These fifteen robots represent seven different models made by three manufacturers. Thirteen of the robots serve in field domains. Field robots are expected to work outdoors, though generally not in rain or snow. They are intended to be able to handle rougher terrains, tolerate dirt and dust, even multi-story falls. The two indoor robots are the more traditional research robots, with small, narrow wheels suitable for operating on smooth flat surfaces.

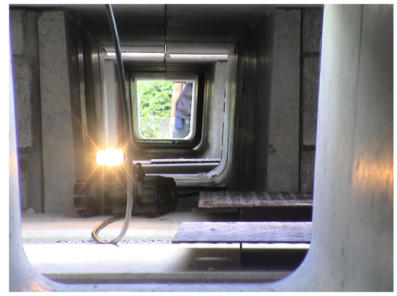

Fig. 2. Field X man-packable inspection robot in a confined space.

To maintain focus on how and how often robots fail rather than which robots fail, the paper labels the three manufacturers by *X*, *Y,* and *Z*, and the models are labeled with *A*. . . *G*. Table I includes the label for the robot's manufacturer and model as well as the number of each type of robot used in the lab, the robot's size, communication method(s), whether it is a tracked of wheeled vehicle, and the general application for which it was designed. The size of a robots is either *man-packable* or *man-portable*[8]. A man-packable robot can be safely carried by one person. A man-portable robot is larger than a man-packable robot but can still be transported in an automobile and can be lifted in and out by one or more people.

Robot Models A and B were designed for chemical and nuclear inspection, though they were used for urban search and rescue (USAR) and military operations in urban terrains (MOUT). Models C and D were specifically designed for MOUT, while E and F were designed for general outdoor research. Model G was intended for indoor research.

Field X A and B model robots are the smallest robots examined and are no larger then 15.5 by 30.5 cm, see Fig. 2. Both are tracked vehicles and do not have onboard computers. Both have a microphone, speaker, a motor-driven manual-focus CCD camera, and a camera tilt unit with halogen lighting. Model B robots also have the ability to adjust the shape of their chassis to raise or lower the camera tilt unit and change the track profile.

Field Y's C model was a precursor of the D model robots. Both are about the size of a large backpack, see Fig. 3. They are tracked vehicles with onboard computers and carry multiple cameras and lighting. The Model C robots also have a set of 13 sonar range sensors. Both were developed for MOUT operations though only the Model D is durable enough for such operations.

Field Y E and F models are larger, wheeled robots with differential steering. Model E has a footprint of 78 by 62 cm

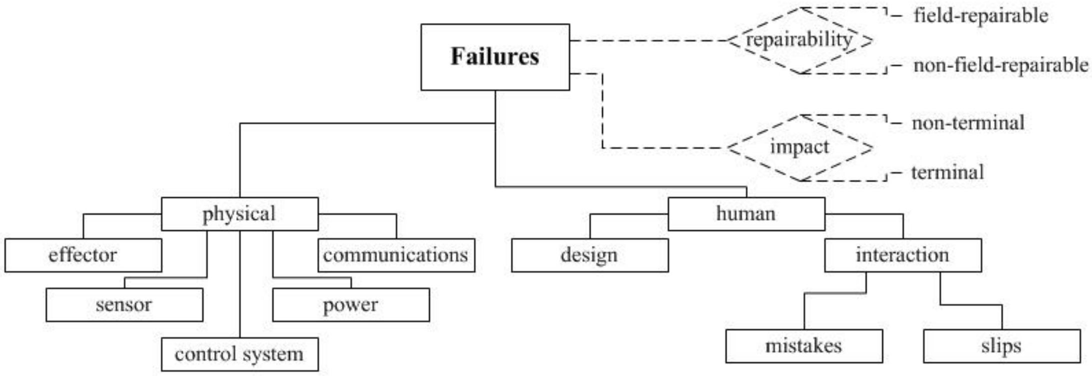

Fig. 1. The taxonomy of mobile robot failures used in this analysis. Classes are shown with solid lines, and attributes with dashed lines.

TABLE I
THE ROBOTS AND SOME OF THEIR CHARACTERISTICS.

Model	Size	Manufacturer	#	Comm.	Drive	Purpose
A	man-packable	Field X	1	Tether	Track	inspection
B	man-packable	Field X	3	Tether	Track	inspection
C	man-packable	Field Y	3	Wireless	Track	MOUT
D	man-packable	Field Y	4	Both	Track	MOUT
E	man-portable	Field Y	1	Both	Wheel	outdoor research
F	man-portable	Field Y	1	Both	Wheel	outdoor research
G	man-portable	Indoor Z	2	Wireless	Wheel	indoor research
Summary			15			

Fig. 3. A man-portable general purpose field robot (top), and a MOUT field robot exploring a rubble pile (bottom), all from manufacturer Field Y.

compared to 104 by 81 cm for Model F. Both carry onboard computers and multiple cameras. The E model robots are small enough to be used for both indoor or outdoor research projects. The larger, Model F, robots are less maneuverable, but have a much longer battery life and can carry smaller robots like the Models C and D.

Model G robots, shown in Fig. 4, are cylindrical in shape, with a 53 cm diameter. Both are wheeled robots with synchronous, non-holonomic drive systems. These robots have two onboard computers with a sensor suite which can include tactile, ultrasonic, and basic vision systems.

Another important factor to consider when comparing robot models is their maturity. The Field X robots are the most mature; over ten years of experience with similar platforms proceeded the design of these robots. The G model was developed in 1996. Both E and F models have been in production for about six years. The C model robots were first developed in 1999 and went through several major modifications during the next two years. The D model is the newest, it was first produced in 2001.

B. Data Collection

User and failures logs served as the sources of data for this analysis. A total of 171 failures were recorded over a period of three years, specifically June 21, 2000 through January 10, 2003. Prior to February 2002, informal records were kept including changes to the robots and information about ongoing repairs. Starting in February 2002 formal failure and user logs were kept. The user logs were entered by robot operators and the failure logs were recorded by the person who performed the repair. Since then over 2100 hours of usage have been logged, including 500 hours of field work. The following information was gathered for quantitative analysis:

- which robot was involved
- who repaired it

Fig. 4. An Indoor Z research robot.

- the date the failure was discovered
- the date the failure was fixed
- the total repair time
- which component failed
- where the failure occurred
- where the repair was performed

C. Calculations

All the formulas used for reliability analysis of the data were taken from the IEEE standards presented in [13]. The mean time between failures or MTBF is calculated by equation (1). This metric provides a rough estimate of how long one can expect to use a robot without encountering failures. Another metric used in this analysis is the failure rate, which is simply the inverse of MTBF. Availability is calculated using (3), where the Mean Time To Repair, *MTTR* is defined as in (2).

$$MTBF = \frac{\sum_{i=2}^{n} \text{Hours Usage Between } F_i \text{ and } F_{i-1}}{\text{\# Failures}} \quad (1)$$

$$MTTR = \frac{\text{\# Hours Spent Repairing}}{\text{\# Repairs}} \quad (2)$$

$$Availability = \frac{\text{MTBF}}{\text{MTBF + MTTR}} \cdot 100\% \quad (3)$$

It may be recalled that the usage logs do not cover the entire time-frame in which the failures occurred. In an attempt to remedy this discrepancy, logs were added for every failure which did not already have a corresponding entry in the usage logs. The estimated usage hours for the added logs was calculated based on the average duration of recorded usage logs for that type of robot.

Other values included in this analysis were calculated using standard formulas. For example, the probability that a failure was caused by a component type c is simply (4).

TABLE II

OVERALL FREQUENCY AND MTBF BROKEN DOWN BY MANUFACTURER. ABOVE ARE THE RESULTS OF THE 2003 ANALYSIS, AND BELOW THE 2004 ANALYSIS.

Manu.	# Failures	Failures/hr	MTBF (hrs)
Field X	37	0.17	6.03
Field Y	44	0.16	6.13
Indoor Z	16	0.05	19.50
Overall	97	0.12	8.29
Field X	58	0.12	8.74
Field Y	89	0.06	15.77
Indoor Z	25	0.01	91.81
Overall	172	0.04	23.99

$$P(c|failure) = \frac{\text{\# Failures Caused by c}}{\text{Total \# Failures}} \quad (4)$$

The statistical analysis of the results consisted of calculating the confidence intervals for the mean-based results and the probability-based results. The mean-based results (MTBF, MTTR, and Average Downtime) were analyzed using the standard equation (5) for the 95% confidence interval where m represents the sample mean. Confidence intervals for the component probabilities were similarly calculated using equation (6) where s represents the sample probability. Due to the inclusion of estimated usage times, it should be noted that the 95% confidence intervals for MTBF are approximations.

$$m - 1.96\sqrt{\frac{\sum x - m}{n}} \leq \mu \leq m + 1.96\sqrt{\frac{\sum x - m}{n}} \quad (5)$$

$$s - 1.96\sqrt{\frac{s(1-s)}{n}} \leq \mu \leq s + 1.96\sqrt{\frac{s(1-s)}{n}} \quad (6)$$

V. RESULTS

This section examines the physical and human failures recorded to date. It is organized to coarsely follow the taxonomy presented in Sec. III. Physical failures are examined first, followed by human failures, and then the repairability of failures is considered. The last subsection is limited to the repairability of the physical failures, as this information was not documented for the human failures.

A. Physical Failures

Physical failures are considered in terms of their frequency, the probability that the cause was a particular type of component, and their impact measured by availability and downtime.

1) Failure Frequency: Table II shows how frequently failures occur with the robots. It shows the total number of failures recorded, the overall frequency of failures (in failures per hour), and the mean time between failures (MTBF), in hours. The failures are grouped by manufacturer, with overall statistics provided at the bottom.

The statistical analysis showed that the time between failures (active usage time, not idle time) took on a broad range of values, resulting in extremely wide confidence intervals for the mean. For example, statistical variance of the means in Table II

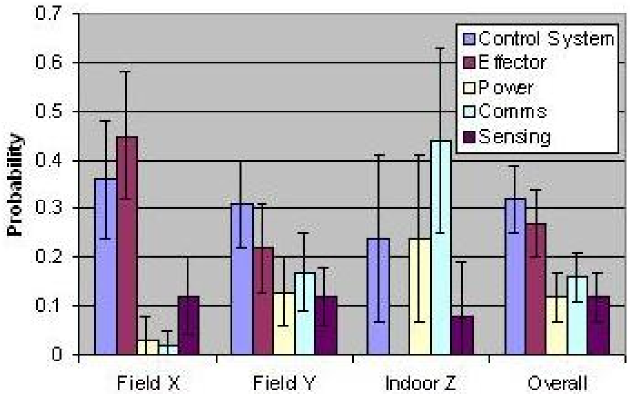

Fig. 5. Probability that a failure was caused by a component type.

TABLE III

COMPARISON OF THE PERFORMANCE OF RESEARCH AND FIELD ROBOTS. ONLY FAILURES IN THE TARGET ENVIRONMENT ARE INCLUDED. THE UPPER TABLE SHOWS THE RESULTS OF THE 2003 ANALYSIS, AND BELOW THE 2004 ANALYSIS.

Manufacturer	Type	% of Usage	Failures/hr	MTBF (hrs)
Field X	Field	94%	0.16	6.14
Field Y	Field	28%	0.16	6.27
Indoor Z	Research	100%	0.05	19.50
Field X	Field	80%	0.10	10.27
Field Y	Field	24%	0.21	4.57
Indoor Z	Research	94%	0.01	149.08

lie between 294 and 8,465 hours. The result of this variance is that the MTBF's are *not* very reliable predictors for the time that the next failure will occur given the time of the last failure. It also means that the differences in MTBF between the manufacturers are not statistically significant. They do still provide a good summary of the information found in the logs and a general assessment.

Comparing these results with those found in 2003's analysis shown in Table II shows that the overall MTBF has improved by almost a factor of three. Each manufacturer's MTBF also improved, with Indoor Z showing by far the most improvement. Based on the results which are presented in Sec. V-A.4, it is unlikely that this resulted from an actual improvement in the reliability of the robots. Instead, an additional year's worth of usage logs and the discovery of archived information on when the Indoor Z robots were used prior to logging, provided better records (and subsequently estimates) of actual usage time.

2) Component: Figure 5 was generated using the component categories defined in Sec. III. As in the previous table the failures are grouped by manufacturer with the overall probabilities for each category shown at the bottom of the figure. The sample probabilities are shown as bars with double 'T' lines showing the confidence intervals which resulted from the statistical analysis.

The most common source of failures is the control system. In most of these cases the robot was unresponsive and the solution was to cycle the power; the source of these problems remains unknown. Other examples of control system failure include a corrupted hard drive on an Model C, a timing delay which hung the boot process on the same Model C, and electrical problems in Model B's OCU. In 2003, effector failures were the most common followed by the control system. The difference between effector and control system relative frequencies is significant only if a 50% confidence interval is used. Both are significantly more common than the other categories.

Tracked vehicles continue to be more susceptible to effector failures then their wheeled counterparts. This is reflected in the fact that Field X is the only manufacturer for which effector failures is the most common. All of the robots examined in this study that were manufactured by Field X are tracked. Overall, thrown tracks are the most common form of effector failure. Other examples of effector failures are Model B's pinion gear becoming stripped or the same gear's pin breaking, and the failure of a motor-amp on the Model E.

The communications category has become more common due to increased use of wireless robots over the past year. The predominant failure is communication loss. According to the data, the least common sources of failure for these robots are sensing and power failures. This is due in part to the fact that the manufacturers purchase mass produced sensors. Conversely, the robot's effectors, control, and power systems are custom built. The most common failed sensor is the camera. It is also the only sensor which appears in every robot's sensor suite. Power may be more reliable than the other systems since it is the least affected by environmental hazards.

3) Research versus field robots: In order to compare research and field robots it is important to consider only failures which occurred in the environment for which each robot was designed. To accommodate this, only in-lab usage and failures were considered for research robots, while only usage and failures in the field were considered for field robots. For reference, the number of in-lab or in-field failures is copied from Table II. The percentage of usage in the target environment over all the recorded usage is also included. The performance in terms of failure metrics is captured in the overall frequency of failures and the mean time between failures (MTBF).

In comparison to 2003, the gulf between field and indoor robots has increased dramatically. Again, this appears to be due to the innovative capabilities of these robots, and the inherent difficulty in constructing robots which can operate in unstructured, outdoor environments. Robots manufactured by Field Y in particular have a much lower MTBF in the field compared to their combined field and lab MTBF. The likely reason for this is that field environments are more challenging. Another reason is that the less reliable Field Y platforms were used more often in the field then the less fragile (but larger) platforms and that only 28% of Field Y usage was in the field. This year, therefore, the more reliable platforms had a better chance of influencing the overall results.

4) Impact: Table IV shows the collective impact of these failures as measured by *availability* and *average downtime*. The projected availability of the robot is included as a percentage of

TABLE IV

AVERAGE DOWN TIME AND AVAILABILITY. ABOVE ARE THE RESULTS OF THE 2003 ANALYSIS, AND BELOW THE 2004 ANALYSIS.

Manu.	Availability	Average Downtime (hrs)
Field X	84%	195
Field Y	24%	353
Indoor Z	94%	61
Overall	47%	243
Field X	17%	49.6
Field Y	57%	12.1
Indoor Z	99%	0.3
Overall	54%	23.2

TABLE VI

FREQUENCY AND IMPACT OF REPAIRABILITY.

Manu.	Field Repaired		Not Field Repaired	
	%	Ave.Downtime(hrs)	%	Ave.Downtime(hrs)
Field X	52%	0.18	47%	92.2
Field Y	36%	0.34	62%	19.9
Indoor Z	0%	N/A	100%	0.3
Overall	35%	0.28	65%	37.3

time. This metric, also called reliability, should be interpreted as the probability that the robot will be free of failures at a particular point in time. The average downtime is also included. Average downtime is the average amount of time between the occurrence of the failure and the completion of the repair that fixed it. In the table failures are again grouped by manufacturer and then summarized.

As with MTBF (see Sec. V-A.1) the *downtime* and *time to repair* (used to calculate availability) varied widely, and the results are also not reliable predictors for future failures. Due to a large MTBF and small MTTR, Indoor Z's availability is above 99%, almost double that of either of the field groups. This is likely due to the fact that these robots are used exclusively indoors and rarely venture out of the safely controlled lab environment.

In comparison with the 2003 results (also shown in Table IV) average downtime is considerably lower. The overall average downtime has improved by a factor of ten. For all but Field X, this has resulted in an increase in availability. Since the majority of robots analyzed in 2003 were also used in the 2004 analysis, it is again unlikely that the reliability of the robots themselves have improved due to changes in operator and technician behavior. By learning each robot's common failures, downtime can be reduced as commonly failed parts can be ordered in advance and more reliable robots can be used in place of more fragile platforms. From a more global point of view the result is the same, the human-robot system has become more reliable over time.

B. Human Failures

The failure logging procedure used for the past year and a half records only physical failures, but other studies performed in previous work covered both physical and human failures. Two field events, a set of field experiments with Hillsbourgh County Fire Rescue[5] and the WTC rescue response[3],[8], were analyzed in previous work by CRASAR. Those studies recorded the number and type of failures encountered as well as the duration of the tasks performed. In each study a mixture of human and physical failures were documented.

Table V isolates the human failures and categorizes them based on the taxonomy presented in Sec. III. The field event and the task the operator was asked to perform with the robot are included followed by the total duration of that task, total number of failures, MTBF in hours, percentage of mistakes, and percentage of slips. The results are broken down by event with overall values provided at the bottom.

In studying these results it is important to keep in mind that the data set, time frame, and range of environments are very limited as compared to the results examined throughout the rest of this analysis. The studies did not document the time of each failure therefore the MTBF had to be calculated as the total usage time divided by the total number of failures, instead of the equation specified in Sec. IV-C. It is also important to note that in the WTC studies, for some forms of human failures, the duration was recorded rather then the number of individual failures. For the purposes of this analysis, each duration value recorded was considered to be a single failure. Therefore the number of failures used in this analysis represents the minimum that actually occurred.

Table V shows that human failures occurred more often during the actual USAR response as compared to the field experiments. Considering the difficulty of navigating a collapse site as large and compact as the WTC disaster, compounded by fatigue and risk to personal safety, this result is expected. On the other hand, the ratio of mistakes to slips is similar despite these differences. More data is needed to determine if this is a universal attribute of human-robot interaction.

C. Repairability

Table VI compares the rates of physical failures that were field-repaired and those that were not. For each the percentage of failures and average downtime are included. It should be noted that these results are based on *field repaired* failures rather then *field-repairable* failures as defined in Sec. III. In theory, field-repaired failures are a subset of the field-repairable set, as some failures which might have been repaired in the field may not have been. Average downtime is the average amount of time between the occurrence of the failure and the completion of the repair that fixed it. The failures are grouped by manufacturer and then summarized at the bottom of the table.

As expected, the average downtime for field repaired failures is very low compared to those that were not field repaired, with the exception of Indoor Z for whom all repairs were performed in the lab. Based on Table VI, not field repaired failures occur more frequently. This is probably the main reason for the overall 54% availability. A good example of the impact of repairability is the difference in availability of Field X robots over the past year. The analysis performed in 2003 showed that 70% of their failures were field repairable and their availability was 84%. In this year's analysis, only half were field repairable and availability dropped to 17%. The failures which contributed to

TABLE V

HUMAN FAILURE ANALYSIS RESULTS.

Field Event	Task	Duration	# Failures	MTBF (hrs)	% Mistakes	% Slips
Field Experiments[5]	Climb Stairs	24 min	3	0.13	33%	67%
WTC[3][8]	Search Small Voids	55 min	15+	0.06	40%	60%
Overall		79 min	18	0.28	39%	61%

this decline were typically severe and very difficult to diagnose. These factors are likely to have reduced the positive impact that experienced operators and technicians had on the average downtime for other common failures. Field Y's improvement over the 2003 results is also due to a difference in the relative frequency of field repaired failures, which is up by 22%.

D. Composite

Fig. 6 provides a summary of the findings in terms of the taxonomy presented in Sec. III. The probability that a given failure is of a given class is displayed beneath each class leaf (node) in the taxonomy tree. The ranges of the confidence intervals for the component categories are not included because they are more difficult to interpret in this form. Instead, only the sample probability used to generate those intervals is presented. The probability of a failure being caused by the control system or the effectors or is near two thirds. Communications failures are less frequent with 16% of the failures, followed by sensing and power at 12%. Of the human failures, slips are more common with 61% of documented failures and mistakes comprising 39%. Since the physical and human failure results came from different sources, the relative frequency of physical versus human failures cannot be determined from this study.

The field-repairable attribute is similarly marked with the probability that a given failure will have one or the other attribute value. Not field repaired failures are more common than field repaired failures, with 65% of the failures covered in this analysis. Note that this categorization is not equivalent to field-repairable and non-field-repairable failures as defined in the taxonomy (see Sec.III). Procedures for using the robots in the field are currently under development and have not been completed to a point where this categorization is can be consistently applied. For now, which failures were and were not repaired in the field provides a deterministic estimator for this attribute. Design failures (under human failures) and the terminal versus non-terminal attribute have not been consistently recorded and are therefore excluded from this figure.

VI. CONCLUSIONS

Over the past year an additional 1082 hours of robot usage (241 of those in the field) and 75 failures have been recorded. The additional data have shown that the MTBF is actually three times better than the average found during the 2003 analysis[2]. This may be due to maturing products. For example, Model D built by Field Y shows much better reliability (availability near 90%) than its prototype Model C (below 40%). On the other hand, changes made to the mature Field X platforms that seemed so dependable a year ago, have had a significant impact on their reliability. They also show that the gulf between field and research robots is wider than expected. Field robots fail more often by a factor of 10, probably due to the demands of field environments. Though MTBF and average downtime have improved compared to 2003, the reliability is still low, with an overall availability of 54%. The current complexity level of the systems and difficulty in maintaining quality control for these low volume products are suspected to be the underlying causes of the observed low overall availability.

Physical failures occurred, on average, once every 24 hours and human failures occurred once every 17 minutes of robot usage time. Statistical analysis shows that the time between failures, the time to repair, and the downtime vary widely therefore none of the differences between related means can be considered to be reliable predictors for future failures. The control system (32%) was the most common source of failures with effectors as the second most common at 27%. Based on the statistical analysis either could be more common and both occur more often then the other categories of failures.

More work is needed to understand human failures. This analysis could only examine under 20 failures. Only the frequency, MTBF of 17 minutes, and type of failure, 61% were slips and 39% were mistakes, could be determined from the information gathered. Reliable data collection methods, like those in place for physical failures, need to be developed and implemented. Design failures, in particular appear to be missing (at least in quantitative form) from the literature.

The results of the statistical analysis presents additional opportunities for future work. Isolating the factors responsible for the large variance in time between failures will lead to a deeper understanding of the conditions in which robots fail. A similar analysis for the repair time may lead to an objective, quantitative measure of the severity and (with downtime) impact of a given failure.

ACKNOWLEDGMENTS

The work reported in this paper was supported by grants from the Department of Energy RIMCC grant and the Office of Naval Research Grant N00773-99PI543.

REFERENCES

[1] John G. Blitch. Adaptive mobility for rescue robots. In *Proceedings of SPIE. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement II*, volume 5071, pages 315–321, 2003.

[2] J. Carlson and R. Murphy. Reliability analysis of mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2003.

[3] J. Casper and R. Murphy. Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Systems, Man, and Cybernetics*, 2002.

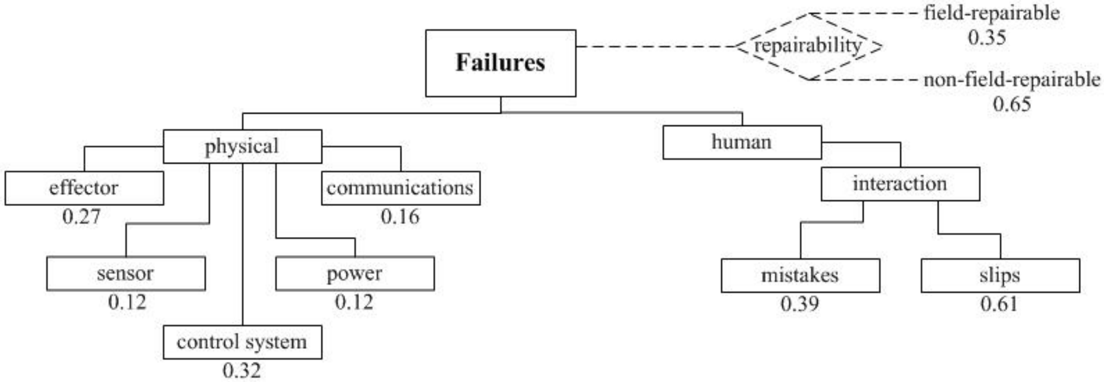

Fig. 6. Summary of classification results using the failure taxonomy from Sec. III including probabilities for each leaf class and attribute value.

[4] Jennifer Casper, Mark Micire, and Robin Murphy. Issues in intelligent robots for search and rescue. In *SPIE Ground Vehicle Technology II*, volume 4, pages 41–46, 2000.

[5] Jennifer Casper and Robin Murphy. Workflow study on human-robot interaction in usar. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2002.

[6] J-C. Laprie. Dependable computing and fault-tolerance: Concepts and terminology. In *Twenty-Fifth International Symposium on Fault-Tolerant Computing*, pages 27–30, 1995. Reprint, originally published in 1985.

[7] R. Mackey, M. James, Han Park, and M. Zak. Beam: Technology for autonomous self-analysis. In *2001 IEEE Aerospace Conference Proceedings*, pages 2989–3001, 2001.

[8] Mark Micire. Analysis of the robotic-assisted search and rescue response to the world trade center disaster. Master's thesis, University of South Florida, July 2002.

[9] Robin R. Murphy, Jennifer Casper, Jeffery Hyams, Mark Micire, and Brian Minten. Mobility and sensing demands in usar. In *IEEE International Conference on Industrial Electronics, Control and Instrumentation*, October 2000.

[10] Donald A. Norman. *The Invisible Computer*. The MIT Press, 1999.

[11] Illah R. Nourbakhsh. The mobot museum robot installations. In *Proceedings of the IEEE/RSJ IROS 2002 Workshop on Robots in Exhibitions*, pages 14–19, 2002.

[12] Department of Defense. Department of defense joint robotics program. http://www.jointrobotics.com.

[13] Institute of Electrical and Inc. Electronics Engineers. Ieee recommended practice for the design of reliable industrial and commercial power systems. IEEE Std 493-1997, 1998. IEEE Gold Book.

[14] N. Tomatis, G. Terrien, R. Piguet, D. Burnier, S. Bouabdallah, and R. Siegwart. Design and system integration for the expo.02 robot. In *Proceedings of the IEEE/RSJ IROS 2002 Workshop on Robots in Exhibitions*, pages 67–72, 2002.

[15] R. Washington. On-board real-time state and fault identification for rovers. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings*, pages 1175–1181, 2000.