Proceedings of the 2004 IEEE
International Conference on Robotics & Automation
New Orleans, LA • April 2004

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— **Mobile robot reliability must be guaranteed before they can be employed in hazardous domains like mine clearing or nuclear waste handling, but recent studies of robots used in urban search and rescue and military scenarios have shown a mean time between failures (MTBF) in the field of 6 to 20 hours. This paper extends previous work characterizing robot failures by including recent data and organizing failures according to a novel taxonomy which includes human failures. Failure type and frequency data were collected from 15 robots representing three manufacturers and seven models over a period of three years, in a variety of environments. Standard manufacturing measures for product reliability were used. The results show that overall MTBF and availability have improved since the previous analysis but are still low. The MTBF across all robot types was 24 hours and availability was 54%. The control system was the most common source of failures (32%), followed by the mechanical platform. Statistical analysis shows that the time between failures, time to repair, and downtime vary widely. For this reason the means reported here are not reliable predictors for future failures, but still provide information on the overall frequency and consequences of mobile robot failures.**

I. Introduction

Great benefit could be derived from robots used to replace living beings in hazardous tasks and environments (e.g. surveying an area for chemical or biological hazards, bomb disposal, or nuclear waste cleanup). However, before robots can be employed in such dangerous domains, a certain level of reliability must be guaranteed. Recent studies of mobile robot performance in urban search and rescue (USAR) and military operations in urban terrain (MOUT) have shown a significant lack of reliability, with a MTBF in the field of between 6 [2] and 20 hours [18].

Information on how and when mobile robots fail helps to identify the weaknesses of current mobile robot technology, which in turn, illuminates the challenges which robot manufacturers and developers of fault-tolerant control systems must meet to improve robot reliability. Data on how mobile robots fail can also be used to provide a realistic starting point for fault modeling in model-based fault tolerance systems, such as [9] and [20]. In addition, potential users of mobile robot technology can benefit from an unbiased, quantitative assessment of current technology. This can aid them in balancing the capabilities a mobile robot will bring to their application domain, against the actual cost of maintaining the equipment.

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 (CRASAR) 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 using basic statistical analysis methods. Sec. V presents the frequency and impact of failures, indoor research versus field robot reliability, and the relative frequency of failures in common robot subsystems. Human-robot interaction failures, and the repairability of the failures are also included. The expected probability of failure associated with each leaf in the taxonomy tree is provided. In Sec. VI the paper concludes that the MTBF has improved but overall reliability is still low. Additionally, the MTBF in field robots is far lower than in research 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 [10]. In 2002, 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 overall MTBF of 8 hours (6 for field robots) and an 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 2002 study 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 a brief examination of human-robot interaction failures.

The results from eight studies conducted by Test and Evaluation Coordination Ofiice (TECO — part of the Maneuver Support Center at Fort Leonard Wood) have been posted to the Department of Defense Joint Robotics Program[15]. The overall goal of these studies was to evaluate the feasibility of using the robotic platform for its assigned tasks in the Future Combat System (FCS). The studies were performed on a wide variety of platforms including 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 reported 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 [14] describes a set of four autonomous robots used for a period of five years as full-time museum docents. Their robots reached a MTBF of between 72 to 216 hours. In [19] 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 mobility problems. Casper, Micire, and Murphy [4] present a discussion of the constraints which the USAR application domain places on robotic technology. In [12] Murphy, Casper, Hyams, Micire, and Minten peruse 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 a robot or equipment used with a 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*[13], and *dependability computing*[7] communities.

Failures are categorized based on the source of failure and are divided into *physical* and *human* categories, following dependability computing practices. Physical failures are subdivided into classes based on common systems found in all robot platforms. These are *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 *on-board computer, manufacturer provided software, and any remote operator control units (OCU's)*.

Human failures (also called human error) are subdivided into *design* and *interaction* subclasses. The interaction subclass represents the failures of interest to the human computer interaction (HCI) and human-robot interaction (HRI) communities. Following HCI practice, it is further refined into mistakes and slips. *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 ability 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, types of data 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 in this analysis. These 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.

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 how many robots of each model were examined, the robot's size, communication method(s), whether it is a tracked or wheeled vehicle, and the general application for which it was designed. The size of a robot is either *man-packable* or *man-portable*[10]. 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 persons.

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 terrain (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. This allows them to raise or lower the camera tilt unit and change the track profile.

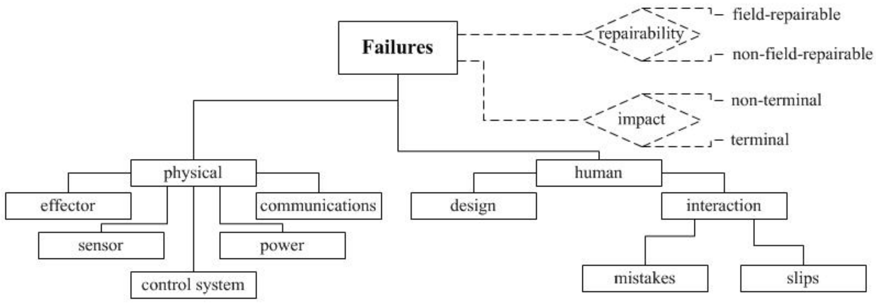

Fig. 1. The taxonomy of mobile robot failures used in this analysis. Classes are shown with solid lines, and attributes with dashed lines.

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			

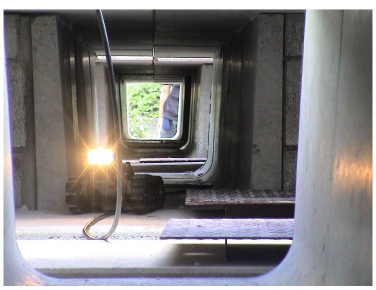

Fig. 2. Field X man-packable inspection robot in a confined space.

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 the Model C is not durable enough for such operations [17].

E and F models manufactured by Field Y are larger, wheeled robots with differential steering. Model E has a footprint of 78 by 62 cm 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 longer battery life and can carry smaller robots such as the Models C and D [11].

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.

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

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

Fig. 4. An Indoor Z research robot.

1999 and went through several major modifications during the next two years. The D model is the newest, it was introduced in 2001.

B. Data Collection

User and failure logs served as the sources of data for this analysis. A total of 171 failures were recorded over a three year period between June 21, 2000 and 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 repairer. 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
- 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 [16]. The mean time between failures (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).

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

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

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

The usage logs do not cover the entire three year 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 were 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 an arbitrary failure was caused by a component of class C (e.g. sensor, effector, mistake) is simply (4).

$$P(c|failure) = \frac{\text{\# Failures Caused by c}}{\text{Total \# Failures}} \mid c \,\epsilon\, C \qquad (4)$$

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 failure 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.

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

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

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 mobile robot failures is considered. The last subsection is limited to the repairability of the physical failures, since repair 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 repercussions (or impact) measured by availability and downtime.

1) Failure Frequency: Table II shows failure frequencies for the different manufacturers. The total number of failures recorded, the overall frequency of failures (in failures per hour), and the mean time between failures (MTBF), in hours are included. Overall statistics provided at the bottom of the table.

The statistical analysis showed that the MTBF (active usage time, not idle time) data had a high variance, resulting in extremely wide confidence intervals for the mean. For example,

TABLE II

OVERALL FREQUENCY AND MTBF BROKEN DOWN BY MANUFACTURER. ABOVE ARE THE RESULTS OF THE 2002 ANALYSIS, AND BELOW THE 2003 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

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 2002 ANALYSIS, AND BELOW THE 2003 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

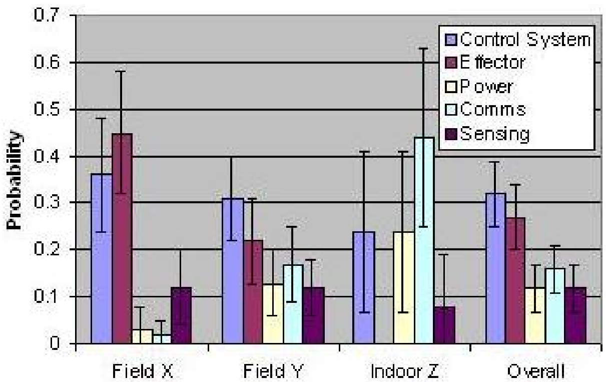

Fig. 5. Probability that a failure was caused by a component type.

statistical variance of the means in Table II 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 of failure frequency.

Comparing these results with those found in the 2002 analysis shown in Table II shows that the overall MTBF has improved by almost a factor of three. Each manufacturer's MTBF also gained ground, with Indoor Z showing by far the most improvement. Based on the results 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 right-hand side of the figure. The sample probabilities are shown as bars with 95% confidence intervals indicated.

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 a Model C, a timing delay which hung the boot process on the same Model C, and electrical problems in Model B's OCU. In 2002, 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% or more 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 from Field X are tracked vehicles. 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-amplifier on the Model E.

The communications failure 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 failures 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 due to its simplicity (compared to the other subsystems), maturity, and the fact that it is the least affected by environmental hazards.

3) Indoor research versus field robots: In order to compare indoor 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 indoor research robots, while only usage and failures in the field were considered for field robots. The percentage of usage in the target environment over all the recorded usage is 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 2002, the gulf between field and indoor research robots has increased dramatically. Again, this appears

TABLE IV

AVERAGE DOWN TIME AND AVAILABILITY. ABOVE ARE THE RESULTS OF THE 2002 ANALYSIS, AND BELOW THE 2003 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

to be due to the innovative capabilities of field 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 (almost 16 hours). One likely reason for this is that field environments are more challenging. Another reason is increased use over the past year of the larger (Models E and F), more reliable (68 and 12 hours MTBF resp.), platforms for research work in the lab. In 2003 these platforms had a greater influence on the overall results. Models C and D, which fail more frequently (0.6 and 2.5 hour MTBF resp.), are the primary contributors to Field Y's target-environment results. These models are typically used in the field due to their mobility and size.

4) Impact: Table IV shows the collective influence of these failures as measured by *availability* and *average downtime*. The projected availability of the robot is included as a percentage. This metric, also called reliability, is the probability that the robot will be free of failures at a particular point in time. The average downtime, or the average amount of time between the occurrence of the failure and the completion of the repair, is also included. Failures are again grouped by manufacturer and then summarized at the bottom of the table.

As with MTBF (see Sec. V-A.1) the *downtime* and *time to repair* (used to calculate availability) varied widely, and their means 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 2002 results (also shown in Table IV) average downtime in 2003 was 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 2002 were also used in the 2003 analysis, it is again unlikely that the reliability of the robots themselves have improved. Changes in operator and technician behavior are a more likely cause. 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 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],[10], 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 operator's assigned task 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. The studies did not document the time of each failure, therefore the MTBF was 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 than in the field experiments. Considering the difficulty of navigating a collapse site as large and compact as the WTC disaster, compounded by fatigue and the 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 could have been repaired in the field may not have been. The failures are grouped by manufacturer and 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 effect of repairability

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][10]	Search Small Voids	55 min	15+	0.06	40%	60%
Overall		79 min	18	0.28	39%	61%

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

is the difference in availability of Field X robots over the past year. The analysis performed in 2002 showed that 70% of their failures were field repairable and their availability was 84%. In 2003 analysis, only half were field repairable and the availability dropped to 17%. The failures which contributed to 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 2002 results is also due to a difference in the relative frequency of field repaired failures (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 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 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 comprise 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 analysis.

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 three times better than the average found during the 2002 analysis[2]. Maturity still appears to have an influence on a platform's overall reliability. 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 effect on their reliability. The additional data 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. The improvements in MTBF and average downtime are likely due to the additional logs (providing a better estimate of robot failure characteristics) and to improved operator and technician behavior. By identifying each robot's common failures, downtime was reduced. Commonly failed parts could be ordered in advance and more reliable robots were used in place of fragile platforms. 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 occurrence intervals between future failures. The control system was the most common source of failures (32%) 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 examined 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 not well studied (at least in quantitative form) in the literature.

The results of the statistical analysis present additional opportunities for future work. Isolating the factors responsible for

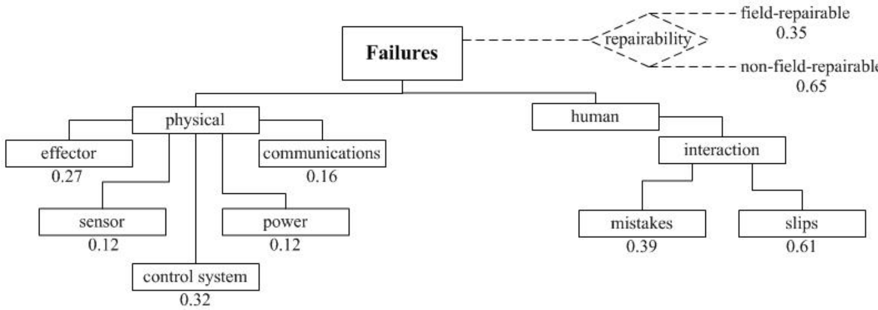

Fig. 6. Summary of classification results using the failure taxonomy from Sec. III including probabilities for each leaf class and attribute value.

the large variance in time between failures will lead to a deeper understanding of the conditions in which robots fail. A similar analysis for repair time may lead to an objective, quantitative measure of the severity and (with downtime) impact of a given failure.

Though this analysis relies on manual logging, it has still expanded knowledge on how mobile robots fail, a topic which is noticeably lacking in the robotics literature. The information presented here can be used as estimations and general assessments in a variety of applications. Diagnosis methods which use probing (gathering additional information through sensors or near-by robots) to isolate failures, as in [8] and [6], can use probabilities to rank hypotheses for testing, thereby reducing overhead by testing the most likely hypotheses first. Researchers and program managers who already work with mobile robots, as well as potential adopters of robotics technology, can anticipate needing two robots for every one intended for use, based on the 54% availability rate. The results also suggest that more mature robot systems should be tested for suitability for a new application before complex experimental platforms.

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.

[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. de Kleer and O. Raiman. Trading off the costs of inference vs. probing in diagnosis. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1736–1741, 1995.

[7] 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.

[8] Matt Long and Robin R. Murphy. Multi-agent diagnosis and recovery from sensor failures. October 2003.

[9] 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.

[10] 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.

[11] Robin R. Murphy. Marsupial robots for urban search and rescue. *IEEE Intelligent Systems*, 15(2):14–19, March 2000.

[12] 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.

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

[14] 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.

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

[16] 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.

[17] G. Piskulic. Military police/ engineer urban robot (urbot) concept experimentation program (cep) report, engineer module. Concept experimentation program (cep) report, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 2000. From archives at www.jointrobotics.com.

[18] Roxana Tiron. Lack of autonomy hampering progress of battlefield robots. *National Defense*, May 2003.

[19] 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.

[20] 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.