Paper Type: REGULAR PAPER

Title: "How UGVs Physically Fail in the Field"

Authors: Jennifer Carlson (jcarlso1@csee.usf.edu), corresponding author;
Dr. Robin Murphy (murphy@csee.usf.edu)
University of South Florida

Address: Computer Science and Engineering Rm 342
University of South Florida
4202 E. Fowler Ave. ENB 118
Tampa, FL 33620

***Abstract*—This paper presents a detailed look at how unmanned ground vehicles (UGVs) fail in the field using information from ten studies and 15 different models in Urban Search and Rescue or military field applications. One explores failures encountered in a limited amount of time in a real crisis (World Trade Center rescue response). Another covers regular use of thirteen robots over two years. The remaining eight studies are field tests of robots performed by the Test and Evaluation Coordination Office at Fort Leonard Wood. A novel taxonomy of UGV failures is presented which categorizes failures based on the cause (physical or human), its impact, and its repairability. Important statistics are derived and illustrative examples of physical failures are examined using this taxonomy. Reliability in field environments is low, between 6 and 20 hours mean time between failures. For example, during the M1 PANTHER II study[6] 35 failures occurred in 32 days. The primary cause varies, one study showed 50% of failures caused by effectors, in another 54% of failures occurred in the control system. Common causes are: unstable control systems, platforms designed for a narrow range of conditions, limited wireless communication range, and insufficient bandwidth for video-based feedback.**

***Index Terms*—failure, failure analysis, field, taxonomy, UGV, mobile robot, meta-study**

How UGVs Physically Fail in the Field

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

***Abstract*— This paper presents a detailed look at how unmanned ground vehicles (UGVs) fail in the field using information from ten studies and 15 different models in Urban Search and Rescue or military field applications. One explores failures encountered in a limited amount of time in a real crisis (World Trade Center rescue response). Another covers regular use of thirteen robots over two years. The remaining eight studies are field tests of robots performed by the Test and Evaluation Coordination Office at Fort Leonard Wood. A novel taxonomy of UGV failures is presented which categorizes failures based on the cause (physical or human), its impact, and its repairability. Important statistics are derived and illustrative examples of physical failures are examined using this taxonomy. Reliability in field environments is low, between 6 and 20 hours mean time between failures. For example, during the M1 PANTHER II study[6] 35 failures occurred in 32 days. The primary cause varies, one study showed 50% of failures caused by effectors, in another 54% of failures occurred in the control system. Common causes are: unstable control systems, platforms designed for a narrow range of conditions, limited wireless communication range, and insufficient bandwidth for video-based feedback.**

***Index Terms*— failure, failure analysis, field, taxonomy, UGV, mobile robot, meta-study**

I. INTRODUCTION

The substantial risk to human life associated with Urban Search and Rescue (USAR) and modern Military Operations in Urban Terrian (MOUT) in have led professionals in each of these domains to explore the use of advanced technology, like robotics, to improve safety. Unmanned ground vehicles (UGVs) are appealing for USAR and the military because they can be sent either ahead of (or in place of) humans, do things humans cannot, and can send back a wide variety of data from their on-board sensors. Though a few UGVs have been used in real USAR (e.g. the World Trade Center rescue response) and military operations (such as demining in Bosnia), they are far from commonplace. The final acceptance of UGV technology in these new application areas will depend as much on their reliability, as on the capabilities of the robot platform (such as the ability to detect chemical or biological agents). A fragile robot that is in constant need of maintenance and repair is likely to be left behind to make room for more reliable equipment.

Recent studies on UGV performance in the field have shown a noticeable lack of reliability in real field conditions. In [4] the mean time between failures (MTBF) for field robots was a little over 6 hours. The analysis in [9] showed that tethered robots required assistance through the tether an average of 2.8 times per minute. Studies performed at the Test and Evaluation Coordination Office (TECO) at Fort Leonard Wood have found a MTBF less than 20 hours.

Email: jcarlso1@csee.usf.edu, murphy@csee.usf.edu.

This paper presents a detailed look at why UGV reliability is so low in the field. It serves as a meta–study taken from 10 studies of ground robots used for either USAR or MOUT applications. Two of the studies are focused explicitly on robot failures and explore the problems encountered over a two week period in a real crisis (WTC rescue response) and day to day use over a two year period of time. The other eight are field tests of robot platforms carried out by TECO.

The studies cover a wide variety of equipment, from small ($12.5 \times 6.5 \times 2.5$ inches) tracked vehicles capable of changing their geometry, to a modified M1 tank (over 60 tonnes). All of these platforms can be considered to be UGVs. A UGV is defined as *a ground-based mechanical device that can sense and interact with its environment*. It may possess any level of autonomy with respect to its human operator(s), from manual (where the human has complete control) to fully autonomous (where the robot can carry out assigned tasks on its own). The majority of the robots considered in all 10 studies are teleoperated, or *manually controlled by an operator at a distance that is too great for the operator to see what the robot is doing*[10].

Failure is a term that is given a wide range of definitions, based on the needs of the individual or group assessing the usefulness of technology for a new task. Reliability metrics like MTBF and availability are very sensitive to differences in the criteria used to determine if an event can be called a failure. Therefore, it is difficult to directly compare results across studies or to apply findings to new circumstances. 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 are included. This general definition of failure is used so that the findings below can be readily applied and compared to results from similar studies.

There are a variety of factors which make *field work* particularly challenging for robots compared to *lab conditions*. The field environment is defined as *an environment which has* not *been modified to ensure the safety of the robot or to enhance its performance*. Conditions which a robot may encounter in a field environment for USAR and MOUT include: dirt, standing water, rain, intense heat, intense cold, confined spaces (see Fig. 1), uneven surfaces (see Fig. 2), the presence of obstacles

Fig. 1. An Inuktun MicroVGTV inside a confined space training maze.

Fig. 2. An iRobot Packbot climbing a rubble pile used for USAR training.

with unpredictable movement, and hostile agents. Field robots used by the military and for USAR have to be packed and transported to remote locations. This adds the requirement that the robot itself be designed to be safely packed.

The rest of this paper is organized as follows. Sec. II provides a more detailed description of the studies from which this article draws, and an overview of related work. Sec. III establishes a novel taxonomy of UGV failures based on a synthesis of taxonomies from the fields of dependability computing, human-computer interaction, and robotics. In Sec. IV the robot platforms covered in this article are briefly described. Sec. V presents statistics on UGV failures in general, not referring to the categories defined in Sec. III. Sec. VI and provides a closer look at failures encountered in the field based on the taxonomy defined in Sec. III including examples which highlight the challenges involved in using UGVs in field environments. Sec. VII provides some general findings and discusses the implications of those findings for robot manufacturers, robot-related project managers, and other researchers working on developing fault-tolerance for UGVs.

II. Related Work

This article studies findings from two independent sources, the Center for Robot-Assisted Search and Rescue (CRASAR) and TECO, which have produced a total of ten studies on UGV performance in the field. In this section those sources will be described individually. Then a brief overview of similar studies on robot performance is provided followed by relevant work in Autonomic Computing.

A detailed analysis on the failures encountered by CRASAR while using robots in the World Trade Center (WTC) rescue operation was reported by Micire in [9]. Each time the robot was deployed, the robot's camera view was recorded. The data used in this analysis were taken from 5.5 hours of video that resulted from the initial two week rescue phase of the WTC response. The study covers four robots from two manufacturers representing three different models. During the response new techniques were developed for operating tethered robots. These allowed the person feeding the tether to assist the robot operator in overcoming problems. According to the analysis, such assistance was needed an average of 2.8 times per minute. Considering minor issues as well as more serious failures, 11% of the search time was lost on average due to traction slippage, 18% due to camera occlusion, and 6% due to lighting adjustments. Seven overall findings and a detailed taxonomy of the environmental and robot relationships are developed from these data.

The WTC study[9] also produced a taxonomy of UGV failures. Two levels of severity were defined: *catastrophic* and *significant*. A catastrophic failure *required that the robot be removed from the void it was searching* in order to be repaired[9]. A significant failure *"introduced sub-optimal performance"*[9] but did not keep the robot from continuing its mission. Five categories were defined for the common failures encountered while using the small Inuktun robots. Two categories were given for significant failures where the tether manager was required to assist the robot through the tether: *gravity assists* and *stuck assists*. The other three categories were defined as *track slippage, occluded camera*, and *incorrect lighting*. None of the other nine studies were as fine-grained in their data collection methods and involve a wider range of robot platforms, therefore this taxonomy will not be used in this article.

In [4] Carlson and Murphy performed an analysis on failures encountered during the day to day use of robots by CRASAR. The CRASAR failure data were collected over a period of two years and include data on thirteen robots representing three manufacturers and seven models. User logs, collected failure type, and frequency data served as the source of information for this study. The study was not limited to field environments, but included usage and failures which occurred in the lab as well. The failure data were analyzed by first categorizing each individual failure based on the robot's subsystems, then using standard manufacturing measures for the reliability of a product, like *mean time between failures* and *availability*. 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%. For the purposes of this article only field use and failures (including office and USAR domains) will be considered.

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[14]. TECO provides operational test and evaluation expertise to the Chemical, Engineer and Military Police Schools and assists in the development and execution of Advanced Warfighting Experiments (AWE). Each robotic platform was designed for a specified set of tasks within the Future Combat System (FCS). The overall goal of the studies was to evaluate the feasibility of using the robotic platform for those tasks. These studies were performed on the following: an All-Purpose Remote Transport System (ARTS) for clearing and demining; integration of Chemical, Biological, Radiological, and Nuclear (CBRN) sensor modules on existing robot platforms (URBOT and MATILDA); the delivery of non-lethal munitions from an existing robot platform (SARGE); a UGV-based rapid obscuration system; a D-7G bulldozer, Deployable Universal Combat Earthmover (DEUCE), and an M1 tank (each equipped with a standardized teleoperation system); as well as a variety of smaller platforms (URBOT, Urban, SOLEM, and TAR). All were carried out in mock military operations which can be considered to be high-fidelity field environments.

In addition to the 10 studies listed above, a recent workshop on robots used in museums produced two studies on the reliability of UGVs actively used for long periods of time. Nourbakhsh [13] 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. As compared to the scope of the CRASAR study [4], the analysis described by Tomatis, Terrien, Piguet, Burnier, Bouabdallah, and Siegwart in [19] was more narrow both in the applications and the robots analyzed and so is not included in this meta-study. However, their MTBF was 7 hours — similar to the 8.3 hour MTBF found in the CRASAR study[4].

Other efforts have concentrated on identifying the weaknesses of UGVs for field applications but have not provided quantitative failure data. In [2] Blitch provides a survey of the mobility problems which are keeping current robot technology from populating otherwise well-suited niches within USAR, especially the confined space access niche. He identifies tumble recovery, traction, and the (incorrect) assumtion of an obstacle-free working envelope as the key problems and presents some suggestions for solutions. Casper, Micire, and Murphy [5] present an overview of the USAR domain, listing tasks that robots are best suited for, followed by a discussion of the constraints which that application domain places on robotic technology. Sensors are identified as the area which requires the most improvement, though the lack of weather-proofing and the need for invertable chassis (also mentioned by Blitch) were also mentioned. In [11] Murphy, Casper, Hyams, Micire, and Minten discuss the same issues as Casper *et al*. [5] but provide some additional discussion on the need for *adjustible autonomy*, or the ability to change the allocation of control between the robot and its operator, in USAR scenarios.

While the new field of Autonomic Computing is concerned with general computer failures, not UGVs, it does offer insights in how to categorize failures. In [18] Sterritt and Bustard strongly advocate grounding new work using concepts and definitions long established in the older field of dependability computing, especially Laprie's work [8]. In 1985 Laprie[8] and his colleagues developed a set of concepts and definitions related to the dependability of computer-based systems. According to Laprie a fault is simply a cause, an error is a state, and a failure is an event. Specifically, failure is defined as *a deviation from the specified service as seen by the client*. The client may be a human user or another component of the computer system that is trying to use the service. An error is *a state within the system which can lead to a failure*. A fault is *anything which could cause the system to enter an error state*. Laprie defines two major fault classes, namely *physical faults* and *human-made faults*. Human-made faults are further subdivided into *design faults* and *interaction faults*. The taxonomy defines two levels for severity for failures. The consequences of *benign failures* are comparable to the benefits of the service they are preventing. *Malign* or *catastrophic* failures have a higher cost by one or more orders of magnitude than the service.

Laprie's dependability taxonomy is general enough that it can be applied to a large variety of systems, but in some aspects it is not suitable for analyzing UGV performance in the field. A UGV can suffer from an infinite variety of physical faults. Laprie's levels of severity are also difficult to apply since the benefit of a service and the cost of a failure tend to vary widely based on the situation, for example a military training exercise versus a real engagement. Nevertheless, the taxonomy used in this paper (see Fig. 3) is largely rooted in Laprie's. The definition for failure used in this article (the inability of the robot or the equipment used with the robot to function normally) is equivalent to his definition since "normal" can be interpreted to mean the behavior specified and agreed on by the service provider and the client. The definition of errors and faults can also be directly applied.

III. Taxonomy of Failures

In order to gain insight on how and why UGVs fail, single instances of failures cannot be treated as independent and unique events. Important common attributes must be identified and used to categorize failures into well defined groups. In order to provide foundation for such insights and to bridge the gaps between the *robotics*, *human-robot interaction*, and *dependability computing* communities, this section defines a classification or taxonomy which can be meaningfully applied to any failure that a UGV used in the field might encounter. The taxonomy is expected to be sufficient for mobile robots in general.

Failures are categorized based on the source of failure (or what Laprie[8] would call the *fault*) and are divided into physical and human categories, following dependability computing practice. Physical failures are further subdivided based on common systems found in all UGV platforms, these being *effector, sensor, control system, power*, and *communications*, following Carlson and Murphy[4]. Effectors are defined as *any devices that perform actuation and any connections related to those components*. Examples would be the motors, grippers,

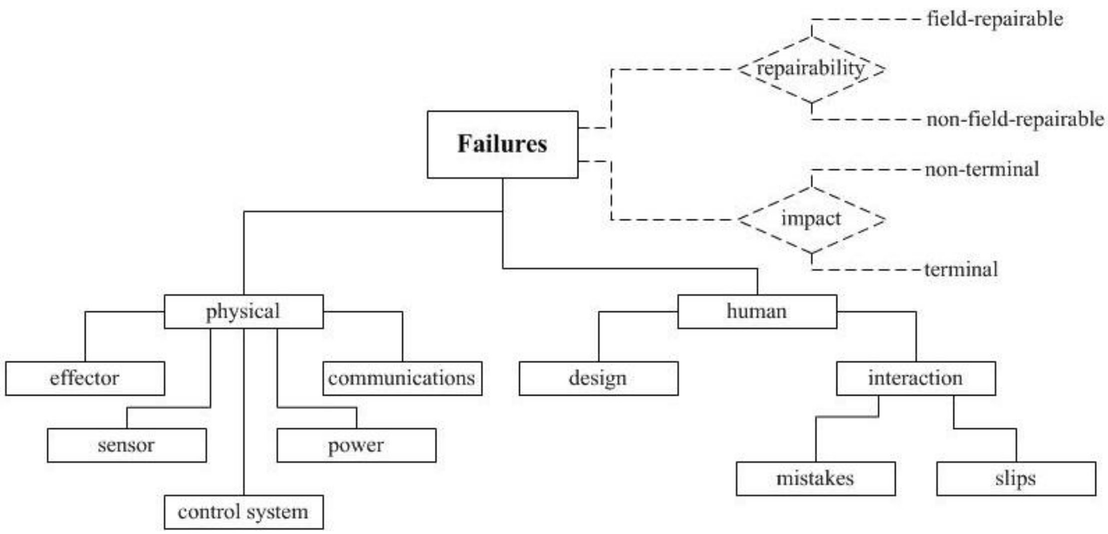

Fig. 3. The taxonomy of UGV failures used in this article. Classes are shown with solid lines, attributes with dashed lines.

and treads or wheels. The control system category includes *the on-board computer, manufacturer provided software, and any remote operator control units (OCU)*. For robots without on-board computers the control system category only covers the control unit.

Following Laprie's categorization, human failures (caused by human-made faults) are subdivided into design and interaction. Drawing from Norman's taxonomy [12] widely accepted in the human-computer interaction community, mistakes and slips are included under interaction. *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 falls into one of these classes and has two attributes. 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 *field-repairable* and *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 labeled as field-repairable.

This taxonomy is used in Sec. VI to classify and study UGV failures in the field. For the CRASAR study[4] and TECO's studies[14] the taxonomy was used to place each failure reported into a single leaf from Fig. 3. The classification of the data from the WTC study was more difficult. The WTC study recorded the operator's response to minor problems rather than the failures themselves. This makes it difficult to map the categories used in the WTC study to the failure taxonomy. Aside from four single instance failures, the data in the WTC study is presented in terms of the WTC study's categories. The details needed to classify the repeated failures themselves are missing. For this reason, the categories used in the WTC study appeared in several classes within the taxonomy. Each appearance was based on a different set of assumptions which held for a subset of the failures (of unknown size) that fell into that category. Human failures are beyond the scope of this meta-study but will be covered in detail in future work.

IV. Robots Surveyed

A total of 24 robots are considered in this article. They represent 15 different models from 7 manufacturers. They range from small ($12.5 \times 6.5 \times 2.5$ inches) tracked vehicles capable of changing their geometry, to a modified M1 tank (over 60 tonnes). Table I presents basic information on each robot model including the name, size, manufacturer, weight, communication method, traction method, and the studies which analyzed the performance of the model. The size of a robot is *man-packable*, *man-portable*, or *not man-portable*.[9] A man-packable robot is small enough to be safely carried by one or two people in backpacks. A man-portable robot cannot be easily carried into the field by a person but can be transported in a HUMMV or personal car and can be lifted by one or more people. Robots which are *not* man-portable require additional equipment to transport them, e.g. a heavy truck or trailer.

The smallest surveyed robots are Inuktun's MicroTracs and MicroVGTV platforms (Fig. 4). Though they have limited sensing capabilities (usually only video and 2-way audio, though more can be added), they have been shown to be very useful in a variety of USAR scenarios. Their small size enables them to go where humans and dogs simply cannot fit. They are also the most portable. The robot and all support equipment that is needed can easily be packed into a backpack and carried into the field.

The next size group includes the SOLEM and URBOT (Fig. 5) from Foster-Miller, the Urban and its successor the

TABLE I
THE ROBOTS AND SOME OF THEIR CHARACTERISTICS. ASV REFERS TO ALL SEASONS VEHICLES INC.

Model	Size	Manufacturer	Weight(lbs)	#	Comm.	Trac.	Studies
MicroTracs	man-packable	Inuktun	8	1	Tether	Track	CRASAR,WTC
MicroVGTV	man-packable	Inuktun	8	3	Tether	Track	CRASAR,WTC
SOLEM	man-packable	Foster-Miller	33	2	Both	Track	WTC,TECO
URBOT	man-packable	Foster-Miller	33	2	Both	Track	TECO
Urban	man-packable	iRobot	35	5	Wireless	Track	CRASAR,TECO
Packbot	man-packable	iRobot	42	2	Both	Track	CRASAR
MATILDA	man-packable	Mesa Associates	50	1	Wireless	Track	TECO
Talon	man-portable	Foster-Miller	85	1	Both	Track	TECO
ATRV-Jr	man-portable	iRobot	110	1	Both	Wheel	CRASAR
ATRV	man-portable	iRobot	260	1	Both	Wheel	CRASAR
SARGE	man-portable	Mod. Yamaha	298	1	Wireless	Wheel	TECO
ARTS	not man-portable	Mod. ASV	5800	1	Wireless	Track	TECO
DEUCE	not man-portable	Caterpillar	35500	2	Wireless	Track	TECO
D-7G	not man-portable	Caterpillar	59000	1	Wireless	Track	TECO
PANTHER	not man-portable	Mod. Army Tank	60000	2	Wireless	Track	TECO
Summary				24			

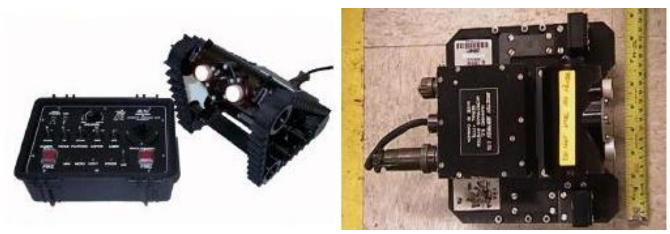

Fig. 4. Inuktun MicroVGTV (left) and Microtracs (right) robots. MicroVGTV photo courtesy of Inuktun Services Ltd.

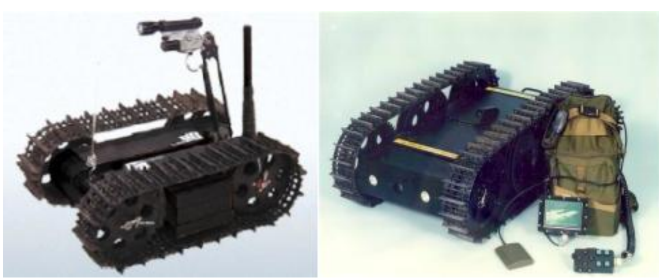

Fig. 5. Foster-Miller Solem (left) and URBOT built on a Solem base (right). Solem photo courtesy of Foster-Miller. URBOT photo courtesy of US Unmanned Ground Vehicles/Systems Joint Project Office (UGV/S JPO).

Fig. 6. iRobot Urban and ATRV (left). iRobot Packbot (right).

Fig. 7. Foster-Miller MATILDA (left) and Talon (right). MATILDA photo courtesy of UGV/S JPO. Talon photo courtesy of Foster-Miller.

Packbot (Fig. 6) from iRobot Corporation, and MATILDA (Fig. 7) from Mesa Associates. These robots are at the limit of what can be considered man-packable, requiring two people to carry the robot and its support equipment into the field. They are usually connected to the OCU through a wireless link and carry their own batteries and a computer inside their chassis. With an onboard computer and larger sensor payload capabilities, these robots can function with varied levels of autonomy. Though high levels of autonomy are still largely experimental and have rarely been used to date in the field, they can help ease the workload of the operator and enable one operator to control groups of robots.

The Talon (Fig. 7) from Foster-Miller, ATRV-Jr (Fig. 8) and ATRV (Fig. 6) from iRobot, and SARGE (Fig. 8) are all man-portable. These robots have a wider range and increased flexibility compared to the smaller robots due to their ability to carry more batteries, sensors, and effectors. The ATRV and SARGE are much larger platforms which can also be modified to carry smaller robots. These heterogeneous groups, usually referred to as marsupial, have the advantage of the larger range of the "mother" robot as well as the maneuverability of the "baby" robots.

The largest group including the ARTS, DEUCE (Fig. 9), D-7G, and PANTHER (Fig. 10) are all platforms which have been adapted from commercially available heavy construction equipment or military platforms. They each weigh in excess of 5 tons and require special equipment to transport. These platforms are too large for typical USAR or squad-level MOUT scenarios. TECO has equipped them with a standard teleoperation interface and tested them for demining and debris clearing tasks.

Fig. 8. iRobot ATRV-Jr shown here in camouflage (left). SARGE built on a Yamaha Breeze ATV base (right). SARGE photo courtesy of UGV/S JPO.

Fig. 9. ARTS built on a All Seasons Vehicles MD-70 base (left) and Caterpillar DEUCE (right). ARTS photo taken from ARTS study.[20] DEUCE photo courtesy of Caterpillar.

Fig. 10. Caterpillar D-7G and M1 PANTHER II built on a US Army M1 Tank base. Photos courtesy of UGV/S JPO.

V. FAILURE STATISTICS BY STUDY

This section begins with an overview of the information available from the 10 studies. Then important statistics are presented on UGV failures which have occurred in the field, gathered from the sources which provided sufficient information. In Sec. V-A the findings from the CRASAR study will be presented. The CRASAR study was not focused only on field failures. It was therefore necessary to extract new information from the source of the data presented in [4], rather than repeat the statistics reported in the published study. The source for the data from the other two studies which follow, the WTC study[9] and TECO's M1 PANTHER II study[6], was not available. The statistics presented in Sections V-B and V-C are taken directly or derived from those studies. Derived statistics were generated in the same manner as their counterparts in the CRASAR study.

Table II provides an overview of the information that was available from the 10 studies which have examined UGV performance in the field. For each study, whether or not a definition of failures was provided, whether or not the failures were classified, the granularity of the data collection process, the availability of basic failure information, and the format of the results are included. This table shows that key information (the definition of failure used, data collection methods, and numeric results) are missing from most of TECO's studies. The CRASAR study and the WTC study differed in both data collection and result generation methods. For this reason a true comparative study cannot be performed directly on this set of studies. A goal for this meta-study therefore was to gather as much information as it could from these studies, and to use that information to begin developing standard methods for collecting and analyzing UGV failure information. The taxonomy presented in Sec. III is the result of that effort.

A. *CRASAR Study Summary*

Table III summarizes the general findings from the data collected for the CRASAR analysis [4]. The data collected for that analysis covered both lab and field experience. Here, we consider only the portion of data (usage and failure records) generated in the field. The number of failures encountered, percentage of usage in the field over all the recorded usage, and the mean time between failures (MTBF) are included to describe the overall frequency of field failures. Note that the MTBF, calculated by equation (1), does not include idle time.

$$MTBF = \frac{\text{Number of Hours Robot Was in Use}}{\text{Number of Failures}} \tag{1}$$

Another important metric which the data provides is the impact of failures. *Availability* and *average downtime* were also calculated for field failures alone and have been included in this table. Availability is calculated from (3) using the Mean Time To Repair (MTTR) as defined in (2)).

$$MTTR = \frac{\text{Number of Hours Spent Repairing}}{\text{Number of Repairs}} \tag{2}$$

In this sense, availability reflects the minimum downtime for field failures, a stable statistic for a given set of failures.

$$Availability = \frac{\text{MTBF}}{\text{MTBF + MTTR}} \cdot 100\% \tag{3}$$

Whereas, the average downtime includes time which was not directly related to a repair. Therefore it will vary with the availability of the parts and skilled personnel required to carry out the repairs.

Table III shows that MTBF by itself does not paint a complete picture. The overall MTBF for Inuktun and iRobot are the same, but the other metrics of reliability like availability are quite different. The primary reason for this is that the Inuktun robots tend to suffer from minor failures, which often take less than a minute to fix. 70% of the recorded Inuktun failures were easy enough to be repaired in the field. iRobot's robots were more likely to suffer from serious failures that take hours to repair. Based on the findings in the CRASAR study[4], this is due to the maturity level of the platforms rather than differences between manufacturers. Within the same manufacturer, iRobot's ATRV-Jr (production level) has a high availability rate at 80% but the Urban (an early prototype) has only a 22% availability rate.

TABLE II

OVERVIEW OF FAILURE INFORMATION AVAILABLE FROM EACH STUDY.

Study	Failure Defined	Failures Classified	Granularity of Data Collection	Source Data Available			Results Presented As
				Usage	# Failures	Descriptions of Each Failure	
CRASAR[4]	Yes	Yes	Hours	Yes	Yes	Yes	Standard reliability metrics
WTC[9]	Yes	Yes	Seconds	No	No	No	Frequencies and percentages
ARTS[20]	No	No	Unknown	No	No	No	Summaries and recommendations
CBRNLOE[1]	No	No	Unknown	No	No	No	Summaries and recommendations
D7[3]	No	No	Unknown	No	No	No	Summaries and recommendations
DEUCE[7]	No	No	Unknown	Partial	Yes	No	Summaries and recommendations
PANTHER[6]	No	Yes	Unknown	Yes	Yes	Yes	Summaries and recommendations
SARGE[16]	No	No	Unknown	No	No	No	Summaries and recommendations
UGVROP[21]	No	No	Unknown	No	No	No	Summaries and recommendations
URBOT[15]	No	No	Unknown	No	No	No	Summaries and recommendations

TABLE III

THE PERFORMANCE OF INUKTUN AND IROBOT ROBOTS IN FIELD ENVIRONMENTS FROM THE CRASAR STUDY[4].

Manu.	# of Failures	% of Usage	MTBF(hrs)	Availability	Average Downtime (hrs)
Inuktun	34	94%	6.14	90%	177
iRobot	12	28%	6.27	36%	207
Overall	46	58%	6.17	64%	185

TABLE IV

FAILURES ENCOUNTERED AT WTC FROM [9].

Model	Attempt	#/Min.	% of Time
MicroTrac	1	1.8	0.7
MicroTrac	2	1.3	16.7
MicroTrac	3	1.3	4.0
MicroTrac	4	0.3	1.7
MicroTrac	5	4.7	21.3
MicroTrac	6	1.3	32.3
MicroTrac	7	0.0	13.3
MicroVGTV	8	0.1	0.7
Overall Average		1.4	11.7

B. WTC Study Summary

In the WTC study[9] analyzed the frequency and impact of failures on a much smaller scale and found that the robots at the WTC encountered minor problems (such as track slippage) quite often, 1.4 times per minute on average. Table IV presents overall statistics from this analysis. The data are broken down by *attempt*. An attempt is an event in which a robot is inserted into a void in order to search that void. Only attempts for which there are recorded data are included. They are numbered sequentially for simplicity. See [9] for more details on the conditions under which each attempt occurred. The WTC study reported the duration (in seconds) of the track slippage, occluded camera, and lighting incorrect failure modes. The number of occurrences was not provided for these classes, though it was for gravity and stuck assists. It is impossible to combine these two distinct measurements into a single statistic. Table IV therefore presents the best available analysis of unclassified robot field failures from the WTC study. It includes the frequency per minute of gravity and stuck assists; and the percentage of time spent in the track slippage, occluded camera, or lighting incorrect failure modes.

For the WTC study a failure was considered to be any event which hindered the progress of the robot in its search task. Considering the fact that Inuktun's robots were not designed to perform USAR search tasks, this is technically a broader definition of failure than the one used in this article. For the purposes of this meta-study, the WTC study's definition of failure is accepted. This is done in the interests of eventually producing robots which *can* be expected to perform better under the extreme conditions encountered during the WTC rescue response. In studying these results it is important to keep in mind that both physical and human failures are covered in this article. Many of these minor problems referred to in Table IV were not the fault of the robot or the physical components of the robot but of the humans that interacted with it.

An interesting attribute of the WTC study is that there were only four failures that would have been recorded without the detailed analysis that was performed on the 5.5 hours of video. In other words, if the WTC study had been performed at the level of granularity of the other studies only four failures would have been reported for the entire duration of the rescue response. The detailed video analysis uncovered 136 cases in which the tether manager had to assist the robot, and that one–third of the time was spent in failure modes (tracks slipping, camera occluded, etc.). These minor failures, which were not recorded by the other studies, had a significant impact on the robots' performance.

C. TECO Study Summary

Of all the studies carried out by TECO, only one provided details on each failure encountered during the performance evaluation period. During the 32 day period of the M1 PANTHER II study[6], a total of 35 failures occurred. Most of the failures were terminal in that testing was stopped for the robot that failed until it was repaired. However, several sensor failures were non-terminal. The average downtime was 7.31 hours overall, or 7.75 hours excluding non-terminal failures. According to TECO, despite the fact that they had two robots to work with, 7.2 days were lost due to unscheduled vehicle maintenance occurring on both robots simultaneously.

TABLE V
PROBABILITY THAT A FIELD FAILURE WAS CAUSED BY A COMPONENT TYPE FROM THE CRASAR STUDY[4].

Manu.	Effector	Control Sys.	Power	Sensing
Inuktun	0.50	0.34	0.03	0.13
iRobot	0.58	0.17	0.25	0.00
Overall	0.50	0.33	0.09	0.09

TABLE VI
PROBABILITY THAT A FIELD FAILURE WAS CAUSED BY A COMPONENT TYPE FOR M1 PANTHER[6].

Model	Effector	Control Sys.	Power	Sensing
PANTHER	0.11	0.54	0.09	0.26

VI. PHYSICAL FAILURES

This section covers in detail the failures reported in all ten studies which do not directly involve a human (that is, physical failures). The majority of failures found in CRASAR's study[4] and TECO's studies are physical failures. The failures encountered at the WTC are presented by category as defined in [9]. First the frequency of failures within the component categories that fall under the physical failure branch of the taxonomy are presented. In Sections VI-A through VI-E, examples of failures which fall into each category are provided. The examples selected will demonstrate not only how failures can be classified using the taxonomy in Sec. III, but also the challenges associated with using robots in field environments. Each section will also include some discussion of general traits of that physical failure category.

Table V presents data from the CRASAR analysis described in [4]. It shows the relative frequency of each of the component categories in the form of probabilities that a failure was caused by component(s) of each type. Note that this information was taken from the source data for [4], rather than the study itself. All lab usage and failures are again factored out of these statistics. The failures are grouped by manufacturer with the overall probabilities for each category shown at the bottom of the table. The Communications category is not included since none of the field failures included in [4] were communications failures.

Table V clearly shows that effectors are the biggest problem, followed by the control system, power, then sensors. Table VI for the M1 PANTHER II shows very different results. In TECO's study the primary problem was the new control system. The sensors that came with the control system were second with 26% of the failures, followed by the effectors. Only power has the same probability of causing problems on the smallest and the largest robots. The difference is probably due again to the maturity of the M1 Tank platform, which has been standard issue in the US Army for 20 years.

A. Effector

Across all the studies analyzed, the most common type of failure was effector failure (failure of components that perform actuation and their connections). Common failure sources in [4] were the shear pin and pinion gear in the geometry shifting mechanism on the VGTV. If the robot encounters resistance while shifting (low clearance in a confined space for example) the shear pin will break. The pinion gear is a problem because the area that houses it is open to the environment. Dirt and other debris get into that area and cause premature wear. According to TECO, the Urban platform suffers from similar problems (see the URBOT study [15] and Fig. 11). Open gearing for the articulating arms and the drive motor collected debris which caused them to stop working.

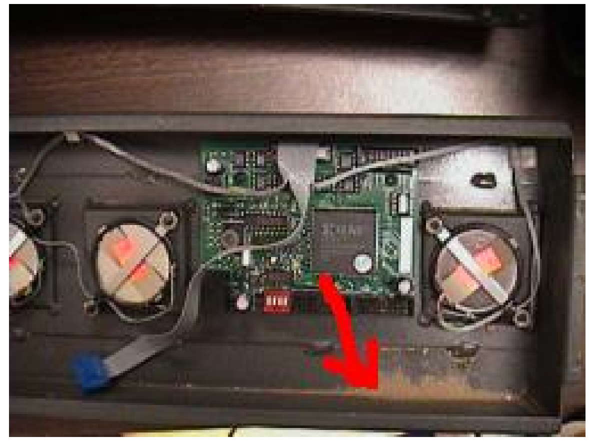

Fig. 11. Dirt found near sensitive equipment inside an iRobot Urban.

Fig. 12. Rock stuck in ARTS track mechanism. Photo taken from ARTS study.[20]

Tracked vehicles were more prone to failure than their wheeled counterparts. In [4] 96% of the effector failures occurred on tracked rather than wheeled vehicles. 57% of the effector failures were simply the tracks working off their wheels for various reasons. In the URBOT study [15] the Urban encountered the same problem. The study of the ARTS performed by TECO [20] mentioned two instances in which rocks became stuck in the track guides and sprockets (see Fig. 12). The PANTHER (a modified M-1 tank) also threw a track, a failure which took days to repair. Both the D-7 study [3] and the DEUCE study [7] mentioned repeated problems with track slippage.

The WTC study's track slippage category was used to catalog the amount of time the tracks were spinning while the

Fig. 13. Failure encountered at the WTC.

robot remained in place. If the tracks should have had enough friction with the ground surface under those conditions, then the failure would be considered to be an effector failure. Such a failure might have been induced by heat which in one documented case exceeded 122 degrees Fahrenheit and softened the track enough that it became loose. In another case an aluminum rod became lodged into the track mechanism where the space tolerance between the track and the platform was less than one-eighth of an inch (see Fig. 13). These are classified as terminal failures as the robots had to be pulled out of the void and repaired before they could be used again.

B. Sensor

The sensor category covers any failed sensors and any problems with their connections. By far the most common failed sensor in all of the studies was the camera. It was also the only sensor common to all of the robots' sensor suites.

In the WTC study [9] there were two categories of repeated sensor failures: occluded camera and incorrect lighting. Occluded camera is defined as a state in which the entire camera view is blocked by obstacles. Note that this is a conservative definition of occlusion and this failure was still found to occur during 18%, on average, of the total time the robots spent searching a void. Incorrect lighting included states in which the lights were completely off (in which case the operator could not see) or were in transition between intensities. Lighting was also a problem mentioned during TECO's ARTS study[20]. The camera's iris would not adjust enough to allow the operator to see to maneuver the robot.

The PANTHER study also uncovered sensor problems which do not tend to occur under lab conditions. Bumpy terrain, sudden changes in lighting, and rainy weather caused problems for the on-board cameras, making it difficult to control the robot from the remote OCU. They also mention other common external problems for cameras used in the field, namely lenses covered in moisture, dirt, or mud.

It is important to note here that the WTC study[9] and TECO in three separate studies cited the lack of depth perception as a problem encountered in the field. This appears to be an important feature lacking in current sensor suites, regardless of the payload capability, as the ARTS vehicle is considerably larger than the robots used at the WTC. The difference in payload capabilities led TECO to recommend the addition of cameras whereas the WTC study recommended software and/or cognitive science-based solutions.

C. Control System

Control system failures (problems caused by the on-board computer, manufacturer provided software, and OCUs) are the most common failures in the PANTHER study [6], second most common in the CRASAR study[4]. None were reported in the WTC study. In the PANTHER study [6] more than 6 days of the 32 available for testing were lost to diagnosis and maintenance of the teleoperation system.

The most common control system failures found in the CRASAR study[4] were cases in which the robot was simply unresponsive (60% of field control system failures) and the solution was to cycle the power. Since rebooting the robot and/or OCU solved the problem it was assumed that the problem was due to the control system, though the source of these problems remains unknown. These problems tend to occur as frequently in the lab as they do in the field.

The PANTHER study[6] revealed a wider variety of control system problems. The most frequently encountered source was the steering system. Symptoms ranged from sluggishness to a complete loss of steering, sometimes manifesting in only one direction at a time. The emergency stop switch failed multiple times. The UGVROP study [21] described similar problems but they seemed to be less frequent. The PANTHER's control system behavior was erratic and unstable in some cases. Reported failures include: uncontrolled acceleration, the RPMs shooting up to a critical level for no apparent reason, and a system shutdown when the operator tried to switch to teleoperation mode.

D. Power

Based on the analysis in the CRASAR study[4] and TECO's study of PANTHER [6] power failures produced few of the failures that occur in the field. The WTC analysis [9] revealed no failures due to batteries and their related connections during the two week rescue response. This was probably due to the fact that the robots were not used for an extended period of time. The longest period of time a robot was continuously used was a little over 24 minutes, therefore the batteries were not heavily taxed.

In the CRASAR study[4] half of the power failures on the robots are due to the battery and its connections. The PANTHER suffered repeatedly from low batteries and low fuel. One of the two DEUCE platforms tested by TECO [7] suffered from a recurring power failure due to clogged fuel filters, requiring a replacement roughly every six hours. TECO also had problems with the Urban platform [15] due mainly to the fact that the operators did not recognize that the drive system was frozen and would subsequently overtax the power system in a vain effort to get the robot to move.

E. Communications

The majority of communications failures in the field were found and described by TECO, with the WTC study [9] providing one example. This is due in part to the fact that wireless communications is a known problem in field environments, therefore wired robots are more commonly used (94% of usage from [4]) than the wireless robots (28% of usage).

Experience with the wireless SOLEM platform at the WTC provides a good example of why these robots are difficult to use remotely in field environments. The structural steel of the World Trade Center had a large impact on the range of the 2 watt transmitter the platform was carrying. Instead of the usual mile or more, the robot lost communication with the OCU in under 20 feet. Even up to that point the signal was not very stable, 23.8% of the 7 minutes the robot spent searching the rubble pile resulted in completely useless video. The robot finally completely lost contact with the OCU. It was never recovered.

Similar problems may have been found with the PANTHER platform tested by TECO [6]. 14% of the failures encountered during that study were due to video dropout, a good indication of communications failures. The study on the D-7 bulldozer [3] concluded that the non-line-of-sight control requirement for the platform could not be met due to the fact that the teleoperation equipment on the robot could not transmit video through any interposed materials or foliage. Video bandwidth and reliability limitations even impacted the performance of operators in line-of-sight scenarios. Occasional static was a problem for ARTS [20] operators as well.

Even if communication technology improves, new problems are likely to emerge. For example, the problem of sharing limited bandwidth among many wireless robots in the same area. Limited bandwidth is already a problem for the military, where rules on allowed frequencies are found to be a hindrance to improved signal strength and reliability. It is also a challenge to ensure that wireless transmitters adhere to the established rules. The ARTS for example would bleed over to radio frequencies it was not allowed to use.

F. Attributes

In the taxonomy defined in Sec. III two additional characteristics were defined in addition to the cause (or fault) of the failure: the *repairability* of the failure and the *impact* of the failure. Table VII presents overall metrics for comparing these attributes from the source data of the CRASAR study, using only data collected in the field. The percentage of failures, MTBF, and average downtime are included for each value of the two attributes. Note that this table divides repairability into *field-repaired* and *not-field-repaired* rather than field-repairable failures. Field-repaired failures represent a subset of the field-repairable failures, since in some cases the field work as done and therefore it was more convenient to bring the robot back to the lab for repair.

Based on Table VII, terminal failures are more common. This finding is supported by TECO's M1 Panther II study in which 95% of the recorded failures were terminal. The average downtime shows that terminal failures, based on the definition used in this article, need not be severe. For example, an unresponsive control system may take less than a minute to fix (reboot). This is still considered to be a terminal failure as the robot cannot continue its mission until the it is repaired. In fact, intermittent non-terminal failures tend to require more diagnosis time and additional technical knowledge of the robotic system. This can drive up the amount of time that a robot is out of service. Field-repaired failures were also more common, though their influence on the robot's reliability is noticeably less, since they require less than ten minutes to repair on average. Non-field-repaired failures, on the other hand, have a significant impact on a robot's availability, as they require weeks to repair on average. This is likely due to shipping overhead and the time it takes to create new custom parts by hand.

TABLE VII

COMPARISON OF FIELD-REPAIRED VS. NOT FIELD-REPAIRED AND TERMINAL VS. NON-TERMINAL FAILURES FROM THE CRASAR STUDY[4].

Attribute	%	MTBF(hrs)	Ave. Downtime(hrs)
Terminal	80%	7.7	103
Non-terminal	20%	31.6	511
Field-repaired	65%	9.5	0.14
Non-field-repaired	35%	17.7	553

G. Discussion

While this meta-study is a quantitative analysis of failures, some observations as to how to proactively reduce failures come to mind. Close partnerships need to be created between between robot manufacturers and groups of interested end-users. Field-testing is very important once a suitable prototype is ready. End-users will be needed to supply realistic scenarios and environments for the tests. Past experience has shown that end-users in the military and search and rescue fields will readily apply their acquired experience and expertise with other familiar tools to new technology. This usually results in new applications and modes of interaction which the original designers never anticipated. User-centered design techniques[17] from human-computer interaction (field-testing with end-users is one) can be used to help focus the engineering design team on the specific challenges which must be overcome to provide a new robot with more reliability than its predecessors.

Some of the problems mentioned above can be handled or at least reduced by the design team alone. The platform itself including the chassis and effectors should be designed for and tested in as many terrain, temperature, and moisture conditions as possible. Alternative solutions to the wireless communication problem need to be explored, for example the use of a combination of wired and wireless communication or repeaters to boost signal strength. Video quality over wireless can be improved by providing the robot with the ability to compress (using lossless compression techniques) the video stream before sending it. Ultimately, manufacturers must accept that their robots will suffer from failures in the field, and must design for maintainability. Regular maintenance tasks must be made as painless as possible for the end-user and any custom parts should be made readily available in the event that more serious failures occur.

VII. Conclusions

This article has explored the question of how UGVs fail in the field. A novel taxonomy of UGV failures was created to reconcile failure taxonomies from the dependability computing, human-computer interaction (HCI), and robotics communities. Following Laprie[8], failures were divided into physical and human categories, with human subdivided into design and interaction. The HCI classes of mistakes and slips fell under interaction failures. Physical failure component categories were taken from Carlson and Murphy.[4] The new taxonomy was then used to explore, in depth, how UGVs fail in the field based on 10 studies [1][3][4][6][7][9][15][16][20][21]. For each category, examples were provided which illustrate the nature of UGV field failures which fall under that category.

Fig. 14 provides a summary of the findings in terms of the taxonomy presented in Sec. III. For each leaf class, the probability for that category is displayed beneath the leaf in the diagram. If multiple source studies provided data, the probability is shown as a range. Ideally the probabilities of siblings should sum to 1.0 but since they are extracted from multiple sources, it is not suprising that they do not. As seen in Fig. 14, the probability of either the effectors or the control system is near 0.5. Sensor failures are less frequent with at most 26% of the failures, followed by power at 9%. Communications failures did not appear in any of the studies which provided enough details to derive relative frequencies of this type, but several of the rest reported examples of such failures.

The attributes are similarly marked with the probability that a given failure will have that attribute value. Field-repairable and terminal failures are more common than their opposites, with 80% and up to 95% respectively of the failures covered in this meta-study. The studies did not provide data on conditional probabilities between the classes and attributes, for example, given that the failure is due to an effector, what is the probability that it will be field-repairable.

Overall robot reliability in field environments is low with between 6 and 20 hours mean time between failures, depending on the criteria used to determine if a failure has occurred. Common issues with existing platforms appear to be the following: unstable control systems, chassis and effectors designed and tested for a narrow range of environmental conditions compared to the variety they will ultimately encounter in the field, limited wireless communication range in urban environments, and insufficient wireless bandwidth for video-based feedback. In order for robots to be accepted in primarily field applications, their reliability in field conditions must improve. In order to achieve this goal both the knowledge of the engineers (mechanical, electrical, and software) that designed and built the robots and the expertise of the end-users must be applied.

Currently a state of the art UGV cannot be expected to complete an entire shift (12 hours for USAR or 20 hours for the Department of Defense) without incident. This is important because it means that backup resources might be needed when the robot fails. It is also important that a new robot be tested in a realistic scenario as soon as possible, to learn what capabilities and shortcomings it has before it is needed.

The taxonomy presented in Sec. III was sufficient to cover all of the failures that were studied here, however future experience may show that new categories need to be added. In particular a structural category may be useful to cover the components which make up the body of the robot and do not fall under the other categories. The structure is commonly overlooked as a source of failures, though it cannot help but impact the mobility and survivability of a robot in the field. Well-designed structures aid the effectors and help to keep the robot balanced over a variety of terrain. They protect fragile components from exposure to dirt, water, mud, etc.

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. The authors would like to especially thank Dave Knichel, Chief of the Test and Evaluation Coordination Office at Fort Leonard Wood, for his assistance and Andrew Nelson for his helpful comments.

Appendix
Definitions

The following list of definitions covers the terminology used in this article and is provided as a reference.

availability Probability that a system will be error free at some given point in time. See (4).

$$Availability = \frac{\text{MTBF}}{\text{MTBF + MTTR}} \cdot 100\% \tag{4}$$

control system A robot subsystem that includes the onboard computer, manufacturer provided software, and any remote operator control units (OCU).

effector Any device that performs actuation and any connections related to those components.

error A state within the system which can lead to a failure.

failure The inability of the robot or the equipment used with the robot to function normally.

fault Anything which could cause the system to enter an error state.

field environment An environment which has *not* been modified to ensure the safety of the robot or to enhance its performance.

field-repairable failure A failure that can be repaired under favorable environmental conditions with the equipment that commonly accompanies the robot into the field.

mistakes Human failures caused by fallacies in conscious processing.

MTBF Mean Time Between Failures. See (5).

$$MTBF = \frac{\text{Number of Hours Robot Was in Use}}{\text{Number of Failures}} \tag{5}$$

MTTR Mean Time to Repair. See (6).

$$MTTR = \frac{\text{Number of Hours Spent Repairing}}{\text{Number of Repairs}} \tag{6}$$

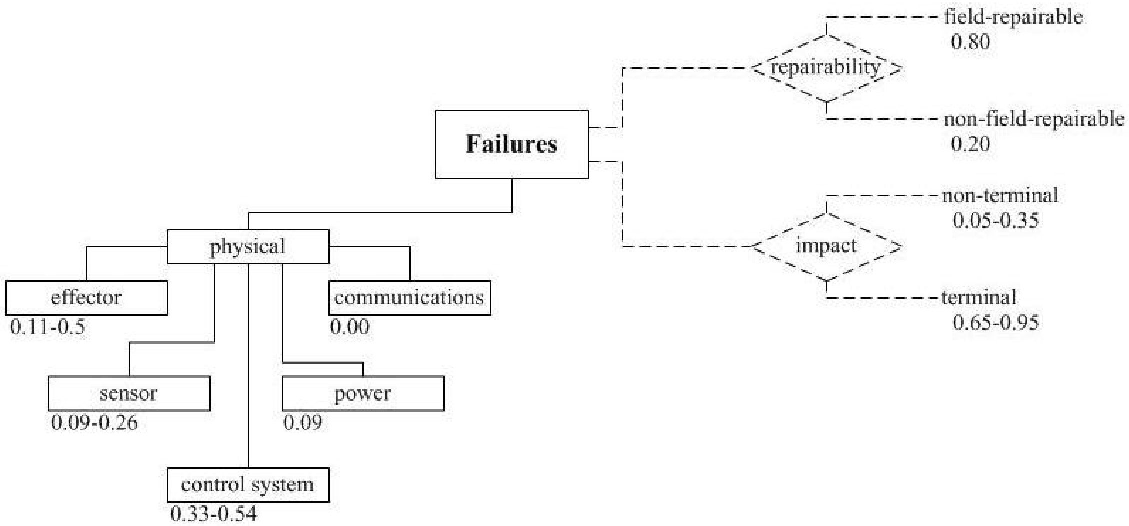

Fig. 14. Summary of classification results using the failure taxonomy from Sec. III including probabilities for each leaf class and attribute value. For example, the probability that a failure was an effector failure is at least 0.11 and at most 0.5 (depending on the source) and the probability that the failure will be field repairable is 0.80.

non-field-repairable failure A failure that is not field-repairable.

non-terminal failure A failure that introduces some noticeable degradation of the robot's capability to perform its mission.

slips Human failures caused by fallacies in unconscious processing.

teleoperated Manually controlled by an operator at a distance that is too great for the operator to see what the robot is doing.[10].

terminal failure A failure that terminates the robot's current mission.

UGV Unmanned Ground Vehicle. A ground-based mechanical device that can sense and interact with its environment.

REFERENCES

[1] Final report for the chemical biological radiological nuclear (cbrn) sensor module with robotic platform limited objective experiment (loe). Final report, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 2003. From archives at www.jointrobotics.com.

[2] 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.

[3] G. Boxley. Teleoparational d-7 dozer concept evaluation program executive summary. Executive summary, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 1999. From archives at www.jointrobotics.com.

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

[5] 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.

[6] F.W. Cook. Test report for the panther ii. Test report, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 1997. From archives at www.jointrobotics.com.

[7] F.W. Cook. Limited objective experimentation report for the deployable universal combat earthmover (deuce). Executive summary, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 2000. From archives at www.jointrobotics.com.

[8] 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.

[9] 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.

[10] Robin R. Murphy. *Introduction to AI Robotics*. The MIT Press, 2000.

[11] 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.

[12] Donald A. Norman. *The Design of Everyday Things*. The MIT Press, 2002.

[13] 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.

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

[15] 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.

[16] G. Piskulic. Military police cep for small robots in support of dismounted troops utilizing non-lethal weapons. Concept experimentation program (cep) report, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 2001. From archives at www.jointrobotics.com.

[17] J. Preece, Y. Rogers, and H. Sharp. *Interaction design: beyond human-computer interaction*. Wiley, 2002.

[18] R. Sterritt and D. Bustard. Autonomic computing – a means of achieving dependability. In *Proceedings IEEE International Conference and Workshop on the Engineering of Computer–Based Systems*, pages 247–251, 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.R. Walker and M.R. Scarlett. Test report for the all-purpose remote transport system (arts) executive summary. Executive summary, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 1999. From archives at www.jointrobotics.com.

[21] R. Weiss. Final test report for the unmanned ground vehicle rapid obscurant platform (ugvrop) chemical corps concept exploration program (ugvrop -cep). Concept experimentation program (cep) report, US Army, Test and Evaluation Coordination Office (TECO), Ft. Leonard Wood, 2001. From archives at www.jointrobotics.com.