Special Issue: DARPA Subterranean Challenge, Advancement and Lessons Learned from the Finals (DARPA SubT Final)

Regular Article

Inspiring Field Robotics Advances through the Design of the DARPA Subterranean Challenge

Viktor L. Orekhov[1], Angela C. Maio[2], Roshan P. Daniel[3] and Timothy H. Chung[*,4]
[1]Booz Allen Hamilton
[2]DEVCOM Army Research Laboratory
[3]Koandah
[4]Microsoft Corporation

Abstract: As the latest Defense Advanced Research Projects Agency (DARPA) “Grand Challenge,” the Subterranean Challenge was a robotics competition that sought to stimulate innovation and investment in solutions that can rapidly map, navigate, and search complex environments, including human-made tunnel systems, urban underground spaces, and natural cave networks. The program hosted a series of evaluations, namely, three Circuit Events and a Final Event, which assessed each competing team’s approaches in representative subterranean environments. This paper details the careful planning and intentional decisions that went into the design of the competition elements of the Final Event of the DARPA Subterranean Challenge. Intended to offer both insights and motivations, this paper comprehensively describes the official rules, scoring objectives, artifact selection, environment setup, and scenario configurations, all in the context of driving towards advancing key technologies of interest to DARPA and to the field robotics community.

Keywords: subterranean robotics, robot teaming, emergency response, perception, GPS-denied operation

1. Introduction

The DARPA Subterranean Challenge was a high-impact worldwide robotics competition designed to inspire the discovery and development of resilient field robotics technologies, specifically in diverse underground environments. With emergency response and security mission applications in mind, the Subterranean (SubT) Challenge sought innovative technologies to rapidly map, navigate, and search complex settings, such as human-made tunnels, urban underground systems, and natural cave networks. Advancing such technologies, leveraging anticipated breakthroughs in four key tech areas, namely autonomy, mobility, networking, and perception, was envisioned to dramatically impact and positively inform how underground operations will be conducted in future scenarios. To ensure

*This work was conducted while this author served as the DARPA Subterranean Challenge Program Manager at the Defense Advanced Research Projects Agency (DARPA), Tactical Technology Office.

Received: 21 June 2022; revised: 4 October 2022; accepted: 8 February 2023; published: 3 April 2023.

Correspondence: Viktor L. Orekhov, Booz Allen Hamilton, Email: orekhov_viktor@bah.com

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI: https://doi.org/10.55417/fr.2023018

alignment and relevance of these developed technologies, deliberate design of every facet of the competition was critical to offer a fertile testing foundation to facilitate these desired technology advances. These insights align with the community's heightened interest in reproducibility and benchmarking in robotics through competitions (Amigoni et al., 2015; Dias et al., 2016; Brancalião et al., 2022).

This paper details the careful planning and intentional decisions that went into the design of the competition elements of the DARPA Subterranean Challenge. Intended to offer both insights and motivations, this paper comprehensively describes the official rules, scoring objectives, artifact selection, environment setup, and scenario configurations, all in the context of driving towards advancing key technologies of interest to DARPA and to the field robotics community. In-depth design discussions and associated summary analyses are provided herein to highlight DARPA's technical approach to competition design; detailed review of specific competition results and competitor performance assessments is deferred to companion papers. Team approaches, contributions to the state of the art in field robotics, and lessons learned during the Circuits Stage of the competition are presented in Orekhov and Chung (2022); Tranzatto et al. (2022); Hudson et al. (2022); Ohradzansky et al. (2022); Scherer et al. (2022); Lu et al. (2022)

This paper is organized first by highlighting the operational contexts and associated mission priorities that served as real-world inspiration sources for the SubT Challenge competition design. Next, given its centrality to the competition itself, the scoring objective—to find and correctly report as many artifacts within the limited time—is discussed at length in Section 2. Section 3 presents the artifacts and the rationale for each artifact's inclusion in the various competition events. Section 4 describes the concepts of operations and limitations placed on human intervention in each competition. In Section 5, we provide a description and characterization of the competition environments, whether Systems Competition courses or Virtual Competition worlds, and share the importance of the environment in designing for technology development outcomes. The impact of both the artifact design and world design manifests in the design of the competition scenarios, which are outlined in Section 6. In Section 7, we describe the ground truth datasets, reference datasets, and open-source software tools that have been publicly released in support of the field robotics community's continued development and evaluation of subterranean technologies. We conclude by presenting a summary of insights and closing remarks in Section 8 to share with the broader SubT community.

1.1. Operational Motivations

Whereas many technology-focused competitions are focused primarily on furthering research in a specific area, or incentivizing specific commercial or entertainment applications, the DARPA SubT Challenge's inspirations are aligned with and geared toward addressing near-term operational considerations balanced with discovering and maturing advanced capabilities. By accelerating these technologies, but with immediate operations in mind, the SubT Challenge sought long-game impacts with near-term benefits.

The underground environments themselves represent highly dangerous and unpredictable conditions, especially in the operationally relevant contexts of interest to DARPA and the Department of Defense. In these settings, the nature of the mission dictates the urgency, relevance, and type of information required by incident commanders or decision makers. Although maps of an unknown environment are absolutely desirable prior to conducting the mission and critical enablers to support the tasks, envisioning future capabilities highlight that maps—which only nominally provide a static, spatial layout—may be insufficient to provide key information to best equip incident commanders to conduct risk assessments relative to deploying human team members. A semantic understanding of the environment is especially important for time-sensitive missions in complex and/or large-scale environments in which metric map data alone would be too cumbersome for an incident commander to review. This idea that an enhanced understanding of the operating environment through *actionable situational awareness* can be achieved through advanced technologies was the inspiration for the SubT Challenge. Decision makers can more effectively execute and direct resources with the benefit of semantic understanding of the environment at high degrees of accuracy, including better awareness of the type of equipment (e.g., rebreathers, shoring equipment); number and deployment of personnel; and even expertise areas needed to respond and address a specific incident or mission needs.

In DARPA's active and broad-reaching engagement with end-users, the most significant priorities in operational settings can be distilled into the following categories:

- correctness (e.g., cannot deploy to respond to false positive cues),
- (geo)metric data in absolute frames (i.e., topological or semantic information is not enough),
- timeliness of information.

These key insights and needs were directly captured in the SubT Challenge's definition and implementation of "actionable situational awareness," as described in Section 2.

1.2. Challenge Structure

The SubT Challenge was organized into two parallel competitions (Systems and Virtual). In the Systems Competition, teams developed and demonstrated physical systems to participate in live competitions on physical, representative subterranean courses. These teams focused on advancing and evaluating novel physical solutions in realistic field environments. In the Virtual Competition, teams developed software and algorithms using virtual models of systems, environments, and terrain to compete in simulation-based events.

Each competition comprised three Circuit Events and a Final Event. The Circuit Events (Tunnel Circuit, Urban Circuit, and Cave Circuit) represented three underground subdomains: human-made tunnel systems, urban underground, and natural cave networks. The Final Event comprised elements of all three subdomains. The subdomains' distinct characteristics and associated course design considerations are further discussed in Section 5. The concepts of operation and run structure for each competition are further described in Section 4.

In the Final Event, each team conducted three runs. The Preliminary Round consisted of two scored runs, each 30 minutes in duration. The Prize Round consisted of one scored run, 60 minutes in duration. The final ranking was based solely on the Prize Round run score. The course was populated with 20 artifacts for the Preliminary Round and 40 artifacts for the Prize Round.

2. Scoring Function

For public-facing competitions, a clear and straightforward scoring metric is critical (a) for the public audience, who benefit from an intuitive understanding of the evaluation metrics, and (b) for the competing teams, who need the assurance of knowing they will be scored fairly and consistently. In practice, however, selecting and implementing such a metric that balances clarity with technical richness represents a challenge in and of itself.

In selecting a unified scoring metric, DARPA sought to simultaneously advance key capabilities across the following four key technology areas.

- Autonomy: ability to map, navigate, and search in complex and dynamic environments without substantive human interventions.
- Perception: operating under varied and degraded conditions with the dynamic range to accommodate dust, fog, mist, water, smoke, low-light, obscured, and/or scattering environments.
- Networking: robust communications solutions that address the limited line of sight, effects of varying geology, and radio frequency (RF) propagation challenges in subterranean environments.
- Mobility: systems with demonstrated endurance and robustness to navigate mobility-stressing and dynamic terrain features including constrained passages, sharp turns, large drops and climbs, inclines, steps, falling debris, mud, sand, and water.

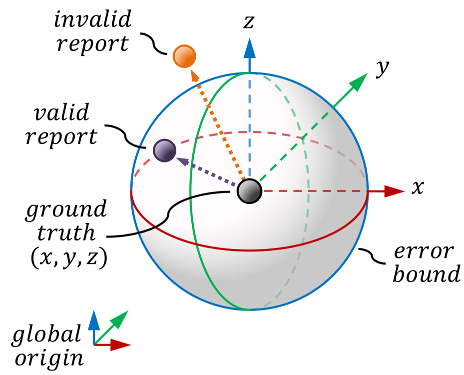


Figure 1. Accuracy-based scoring of artifact reports.

While each of the four technology areas are critical to solving the challenges of operating underground, each team chose varying strategies in balancing the emphasis placed on each tech area. For example, a team with an especially effective autonomy solution may have chosen to place less emphasis on their networking solutions because they were able to rely on their robots remaining operational beyond communications range.

With the possibility of widely varying approaches, an effective scoring metric must fairly and consistently evaluate each team's performance against the stated operational objectives and avoid unintentionally favoring a particular technology strategy. More importantly, the scoring metric needs to incentivize teams to develop technologies that are operationally relevant. For these reasons, DARPA chose to ground the scoring metric on achieving a high-level mission objective that is inherently operationally relevant.

2.1. Artifact Reports

The main scoring objective for the SubT Challenge was to search for, detect, and provide spatially referenced locations of objects of interest, a.k.a. artifacts, placed in the environment. Teams earned one point for each valid artifact report. To be designated a valid artifact report, the artifact type was required to be correctly determined *and* the artifact's reported location was required to be less than or equal to five (5) meters (Euclidean distance) from the ground truth location (as illustrated in Figure 1.)

Artifacts were distributed throughout the competition course in a manner which rewarded teams that were able to rapidly explore and maneuver through more of the course elements. The teams possessed no *a priori* knowledge of the expanse, length, topology, or terrain of the competition courses. The total number of artifacts, but not the number of each type, was disclosed to the competitors. Each team was given a fixed number of artifact report attempts to discourage spurious guessing. Any duplicate reports were considered invalid and counted against the total number of reports. Teams attained the highest score by finding the most artifacts within the limited duration of a single run.

For the Systems Competition in the Final Event, each run in the Preliminary Round had 20 artifacts, with 25 total artifact report attempts, allowing for a maximum of 20 points possible per run. Each run in the Prize Round had 40 artifacts, permitted 45 total artifact report attempts, with a maximum of 40 points possible for the single run. For the Virtual Competition, each simulated run had 20 artifacts with 25 total artifact report attempts in both Preliminary and Prize Rounds.

In case of a tie, the team rank was determined per the official *SubT Challenge Competition Rules* (2021), i.e., by the earliest time that the last artifact was successfully reported. The tiebreaker was intentionally selected to reward teams that achieved the secondary goal of providing "rapid" situational awareness rather than other possible metrics (e.g., closer accuracy of reports, furthest distance traveled).

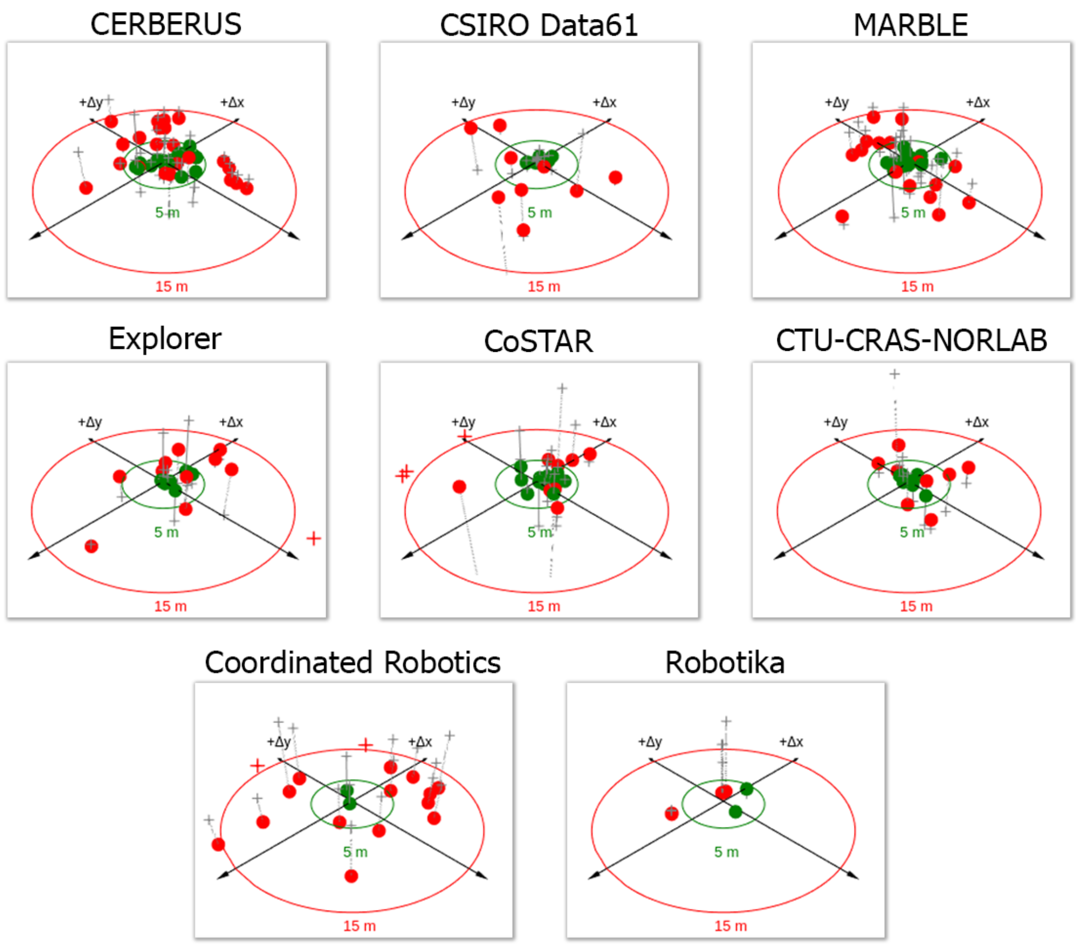


Figure 2. Artifact report accuracy and precision by team.

2.2. Error Threshold

The 5-meter error threshold for artifact localization was selected to ensure that the localization accuracy provided first responders actionable situational awareness upon entering the previously unknown environment. Figure 2 shows the 3D accuracy and precision of submitted artifact reports by team for the Prize Round of the Systems Competition. Each plot shows the error of submitted artifact reports relative to their respective ground truth locations. Each marker is plotted on the X-Y plane with vertical lines indicating the Z error. Successful reports are shown with green markers while unsuccessful reports are shown with red markers.

As part of post-event analyses, DARPA examined the potential role of this error threshold on the developed technology solutions. Figure 3 shows the score for each team when their submitted artifact reports were re-scored across a range of error thresholds, both above and below the competition setting of 5 meters. It is important to note that these adjusted scores are hypothetical and do not necessarily represent what would have happened if a different error threshold had been announced at the beginning of the competition. If the threshold were initially set much lower, teams would have likely invested greater resources into their perception and localization solutions in ways that are difficult to predict. Nevertheless, these results provided an insight into the relative accuracy of each team's localization and the sensitivity of their approaches to the error threshold.

The results indicate a low sensitivity to increasing the error threshold. As the threshold is increased from 5 to 10 m, no teams change their rank and the greatest increase in score is only

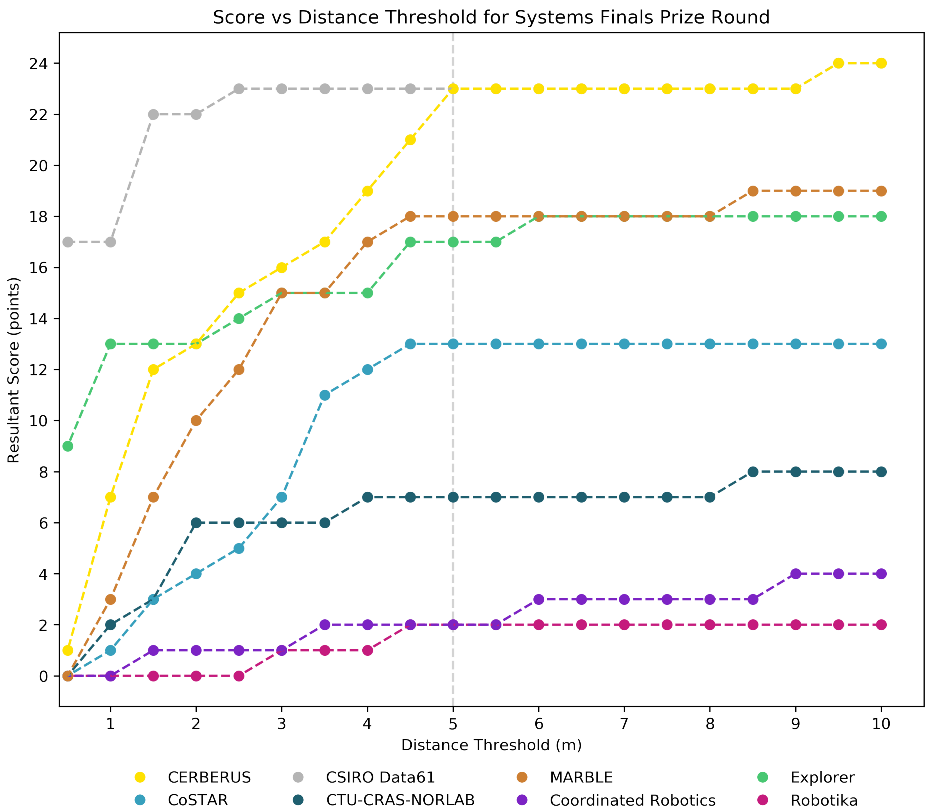


Figure 3. Scoring sensitivity to the artifact report error threshold radius.

2 points. As the threshold is decreased, however, the results indicate a possibility that the final rankings of the competition could have been impacted. When the results are re-scored with an error threshold of 4.5 m, for example, the top two ranked teams reverse their positions. Further reducing the threshold to 2.5 m results in the reversals of third and fourth place as well as fifth and sixth place. The rapid drop-off in the resultant scores as the threshold is reduced indicates that the teams' solutions were likely conditioned to the published 5 m threshold. It is impossible to know what new innovations would have been developed to respond to a more restrictive threshold. However, such analysis provides insights into the resulting solutions as demonstrated at the end of the competition and their potential ability to address different mission requirements for artifact reporting error.

2.3. Alternate Evaluation Metrics

In selecting a scoring approach for the SubT Challenge, many evaluation metrics and scoring functions were considered. Some of the most relevant alternatives are discussed in the following section to showcase the diversity of options and also the difficulty of the design process (Piazza et al., 2022).

Function of Time. Instead of scoring all artifacts with equal value, the points scored for a given artifact could have depended on the time of its reporting. For example, an artifact successfully reported earlier in the run could have been worth more points in order to more explicitly incentivize faster exploration and earlier reporting. This option was not selected because it would disproportionately increase the value of artifacts that were closer to the entrance of the competition course, even though they were arguably easier to reach and report. Instead, DARPA chose to incentivize rapid exploration and reporting by designing large-scale competition courses that could only be fully covered by teams that were able to deploy their systems and explore the environments rapidly. Furthermore, time-based incentives were included as the tiebreaker mechanism and ultimately made the difference between first and second place.

Function of Distance. The points scored for a given artifact could have varied based on distance from the entrance. This approach would have the effect of rewarding teams that reached deeper into the course while maintaining low drift in their perception and localization solutions. This option was not selected because proximity to the entrance is often not correlated to the actual path that a robot needed to take to reach a given artifact due to the complex topology of the competition courses. This option may have unduly rewarded depth-first exploration algorithms versus more thorough exhaustive search algorithms. Furthermore, in an emergency response scenario, all artifacts are operationally relevant; a survivor further away is not more operationally relevant than a survivor closer to the entry point.

Varying Error Threshold. Instead of varying the value of the points scored, the error threshold could have been increased for artifacts that were further from the entrance. This approach could account for the tendency of localization to drift over distance. However, this option was also not selected because distance from the entrance does not directly correlate to the path taken by a given robot in complex environments, especially in the context of exploration. Further, its selection would have intrinsically designated a nominal target drift rate set by DARPA, rather than encouraging teams to push to achieve the best integrated localization capabilities possible.

Mapping. Instead of scoring teams based on finding and reporting artifacts, teams could have been evaluated on the quality of their maps. The desire for high-quality maps and map evaluation metrics was frequently emphasized in DARPA's engagements with operational stakeholders. However, map evaluation metrics would require a prescribed map representation and inconsistently reward solutions due to variations in sensing modalities, map representations, and resolution of the data. The most relevant evaluation metrics for maps are also often dependent on the operational use case. For example, an approximate 2D topographical sketch semantic labeling of key features can be, in some cases, more operationally relevant than a high-accuracy and high-resolution 3D point cloud representation. Instead of directly scoring maps, artifacts were distributed throughout the competition course in a manner which rewarded teams that were able to rapidly explore and maneuver through more of the course elements. In this way, localization of artifacts not only served as a surrogate for accurate mapping, but also extended and accentuated the utility of the developed systems beyond mapping to provide timely and actionable situational awareness.

Unfortunately, these mapping metrics were not viable as an objective scoring function for the SubT Challenge. The direct map comparison approach requires 3D point cloud maps (some teams provided only 2D data) and the metrics are sensitive to teams' individual design choices for map representations and map density. This mapping analysis approach also does not allow direct correspondence of map points to ground truth points; hence, points may be mistakenly classified as inliers even when representing a scan of an entirely different part of the course.

In practice, the artifact-based scoring function worked in concert with other intentional design decisions including artifact design (types, placement), world design (topology, scale, complexity, RF mitigation), and scenario design (time limits, personnel rules, dynamic obstacles) to appropriately incentivize and reward teams for achieving breakthroughs across all four of the technology areas.

3. Artifacts

As the main scoring objective for the competition, artifact design played a key role in driving operationally relevant technology advances. The artifacts intentionally varied in their size, color, placement, and detection signatures (e.g., visual, thermal, chemical) and represented an operational metaphor for objects that would be of interest in representative scenarios involving search and rescue operations, military defense operations, scientific exploration, or safety monitoring.

Three artifacts were common to all three subdomains and appeared in all three Circuit Events (a.k.a. the Tunnel Circuit, Urban Circuit, and Cave Circuit). Two additional artifacts were specified for each Circuit Event that were event-specific and did not appear in the other Circuit Events.

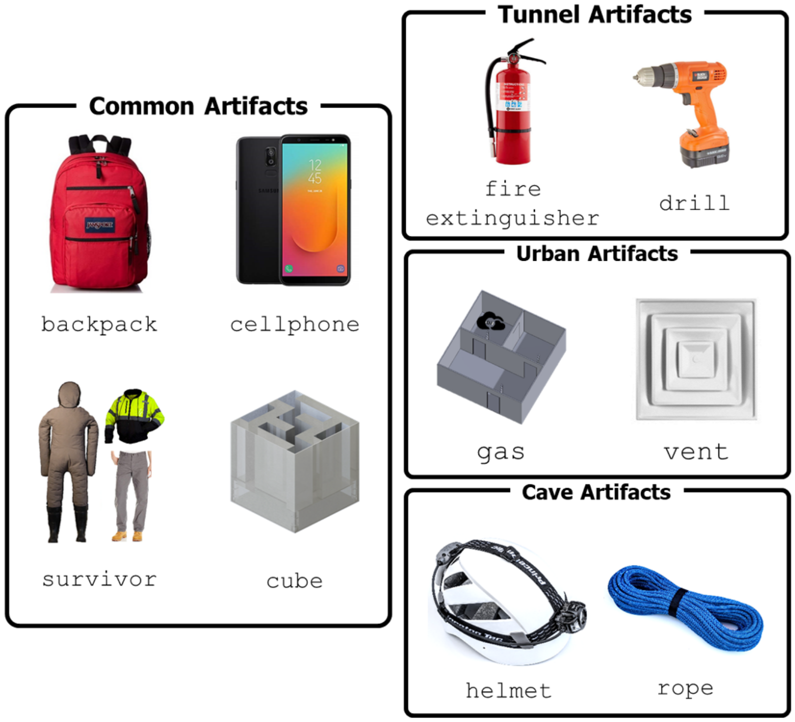


Figure 4. The common and subdomain-specific artifacts that were used in the Final Event.

Thus each Circuit Event included a total of five artifact types: the three common artifacts and two event-specific artifacts. The Final Event included a total of ten artifact types: the original three common artifacts, all six event-specific artifacts, and a Finals-specific artifact that was common to all three subdomains. The ten Final Event artifacts were: survivors, cell phones, backpacks, drills, fire extinguishers, vents, gas, helmets, ropes, and cubes as illustrated in Figure 4.

- The survivor artifact was a Smartdummy Thermal Manikin produced by LION and intended to represent both human shape and body temperature through heating elements in the head, torso, and limbs. The manikin was dressed in a high-visibility yellow jacket, grey work pants, and leather work boots. A voicebox played a continuous recording of human speech using two male voices and two female voices.
- The cell phone artifact was a Samsung Galaxy J8 J819M/DS which served as a surrogate for hand-held electronic devices such as radios and surveying equipment, which when discovered are indicators of human presence and activity. During the run, the screen was on full brightness and was playing a full-screen video with audio. The phone's 2.4 GHz WiFi was operating as an access point with a visible SSID, and the phone's Bluetooth radio was on and in discovery mode.
- The backpack artifact was a red JanSport backpack that represents a typical, adult sized backpack used for transporting personal items and equipment. The backpack was weighted to aid in holding the backpack in place and padded with packing material.
- The drill artifact was a Black & Decker GC960 cordless drill and represented a multitude of hand tools (manual or powered).
- The fire extinguisher artifact was a First Alert FE2A10GR and could represent finding either safety equipment or possibly a dangerous gas canister.
- The gas artifact was implemented as a room filled with CO_2 to simulate a range of hazardous air quality conditions including a gas leak, poor ventilation, or fumes and smoke. A CO_2 emitting

device maintained a concentration of at least 3000 parts per million (ppm) within the room; however, a robot would likely need to enter the room to detect the elevated concentration.
- The vent artifact was a Grainger 4MJV3 three-cone square ceiling diffuser that represented a typical supply register commonly found in human-occupied or -working environments. Finding this artifact represented identifying potential areas with fresh air or an escape route to the surface. The vent artifact was actively heated to present a thermal signature that was at least 10 °C above ambient.
- The helmet artifact was a Petzl BOREO caving helmet with a Princeton Tec Apex (APX550-BK) headlamp. Finding this artifact could indicate nearby human presence or represent a partially obstructed survivor awaiting rescue. The headlamp was turned on in the “low spot” setting.
- The rope artifact was a coiled 35 m length of 9.9 mm climbing rope commonly used for traversing vertical sections of caves. Finding the rope artifact represents identifying areas where humans may have traversed or the location of a vertical passage.
- The cube artifact was a 3D representation of the SubT Challenge logo and served as the tenth and final artifact introduced at the Final Event. The cube artifact was approximately 200 mm x 200 mm x 200 mm and was 3D-printed using a translucent plastic material. The cube was illuminated from within by RGB LEDs that rotated through the color spectrum and included a Bluetooth radio that was on and in discovery mode.

The specific details of all artifacts, including part numbers, dimensions, assembly information, and localization points, were provided to all competitors in the *SubT Challenge Artifact Specification Guide* (2021) to facilitate teams’ preparation and practice with replicas of actual competition artifacts.

Virtual models of each artifact were also available as part of the Virtual Competition infrastructure, including the three-dimensional structure, visual, and thermal representations in simulation. The gas artifact was simulated by a plugin providing a boolean value indicating the presence or absence of gas in a room. The virtual models were utilized by teams across both Competitions for training and testing artifact detection algorithms.

3.1. Sensing Modalities

The ten artifact types were intentionally selected to motivate multimodal sensing approaches and to prevent an over-reliance on a single sensing modality. As highlighted in Table 1, all artifact types (with the exception of gas) could be detected and/or classified by more than one sensing modality. This design choice not only provided alternate detection methods but also a means by which to

Table 1. Mapping of sensing modality (for detection and classification) to the artifact type.

		Visual Cameras	Infared/ Thermal	Lidar	Acoustic	Radio Frequency	Gas/ Chemical
Common	**Survivor**	✓	✓	✓	✓		
	Cell Phone	✓	✓		✓	✓	
	Backpack	✓		✓			
Tunnel	**Drill**	✓					
	Fire Extinguisher	✓		✓			
Urban	**Vent**	✓	✓	✓			
	Gas						✓
Cave	**Helmet**	✓	✓				
	Rope	✓		✓			
Finals	**Cube**	✓	✓	✓		✓	

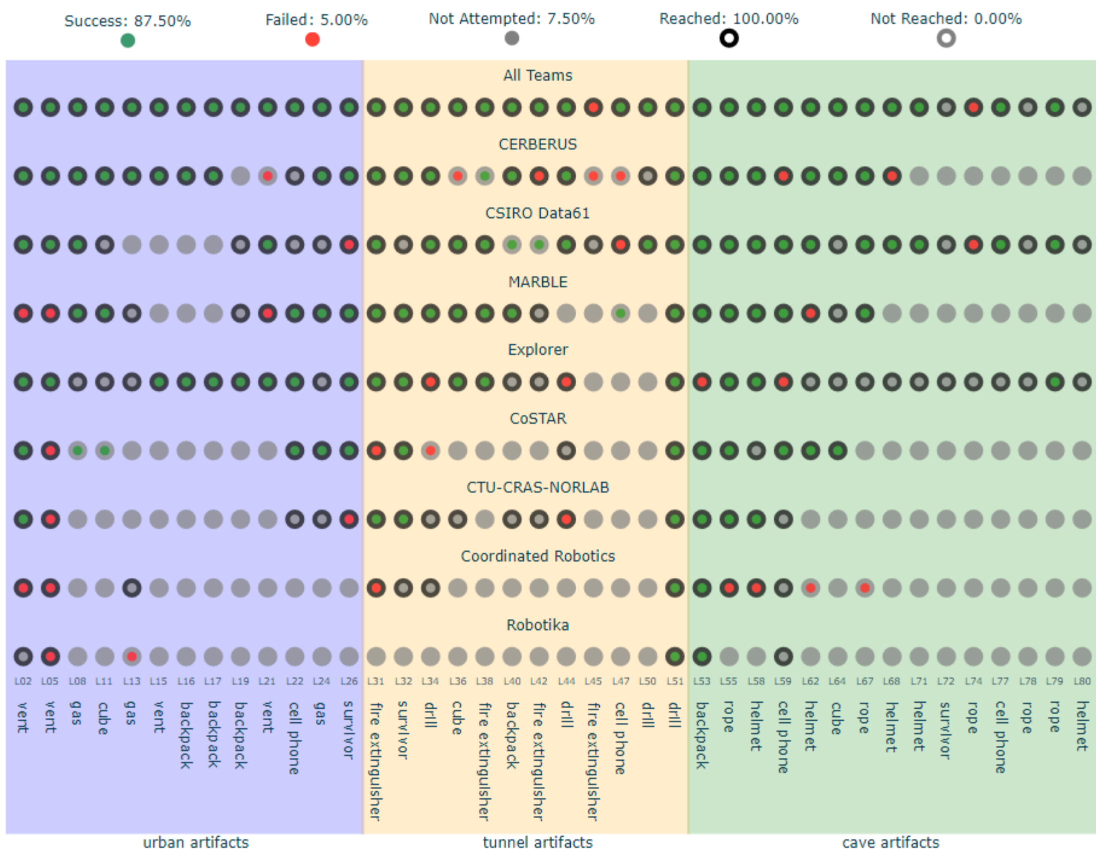


Figure 5. Aggregate artifact scorecard across all Systems Competition teams.

refine the localization of a detection for teams that were able to detect an artifact with multiple sensing modalities.

Beyond motivating multimodal sensing approaches, the selection of artifacts had additional impacts on the systems design. The incentive to include more sensing modalities introduced additional size, weight, power, and cost (SWaP-C) tradeoffs, especially for UAVs with already limited payloads and flight times. In response, some teams chose to take a heterogeneous approach not just in their mobility strategies but also in their perception payloads and functions within the team strategy.

3.2. Artifacts Scorecard

Figure 5 shows the aggregate artifact scorecard for the Prize Round of the Systems Competition which includes the artifacts that each team reached, attempted, and successfully scored. The top row shows the aggregate performance across all teams. Also shown are the distribution of position estimate errors for all submitted artifact reports for each of the 40 artifacts. All of the artifacts were reached by at least one team. 87.5 % of the artifacts were successfully reached and scored by at least one team (35 out of 40).

4. Concepts of Operations

Much of the competition design for the SubT Challenge was shared across both the Systems and Virtual Competitions except for necessary and intentional differences in run operations. The Systems Competition motivated a high level of autonomy by limiting human involvement, whereas the Virtual

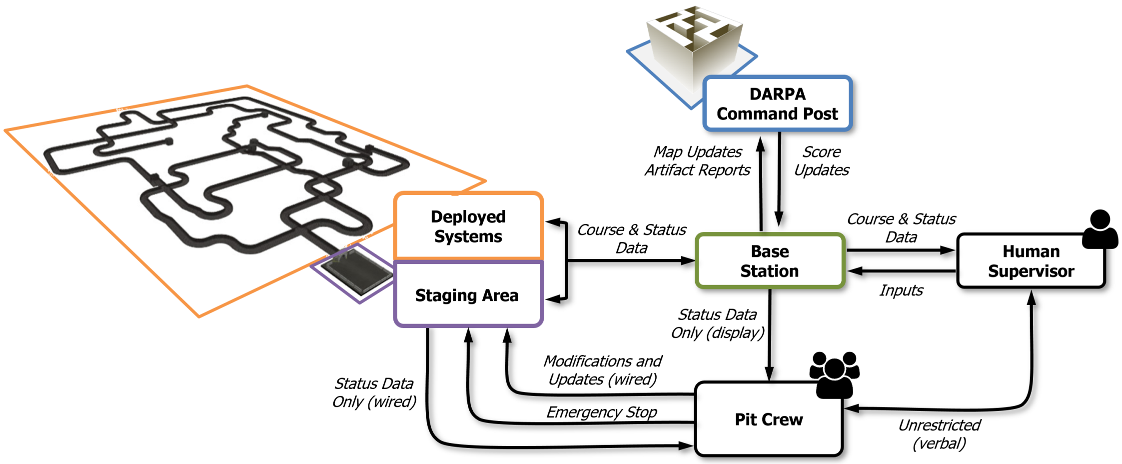


Figure 6. SubT Challenge Systems Competitor Concept of Operations illustrates how information may be shared among competitor systems and competition infrastructure.

Competition required full autonomy to push the limits of software solutions. In this section, we describe the concept of operations for official runs in each competition and their design motivations.

4.1. Systems Competition

As the operational scenario suggests, DARPA was interested in approaches that are highly autonomous without the need for substantive human interventions; capable of remotely mapping and/or navigating complex and dynamic terrain; and able to operate with degraded and unreliable communication links. The teams had no *a priori* knowledge of the expanse, length, topology, or terrain of the competition courses.

Figure 6 illustrates an annotated concept of operations and how information was allowed to be shared between the systems, team Base Station, team personnel, and DARPA Command Post. The competing team began its run in the Staging Area, which was immediately outside of a known entrance to the otherwise unknown underground course. At the beginning of a run, the team deployed its robotic systems into the course where they explored, mapped, and searched for artifacts. Relevant observation data were transmitted to the team's Base Station, which was defined as one or more computers or controllers that served as the interface between the systems, the DARPA Command Post, and the Human Supervisor. The Base Station was responsible, either automatically or with supervisor monitoring, for communicating with the deployed systems and relaying artifact reports and map updates to the DARPA Command Post. DARPA defined these interfaces in the Interface Control Document (2021) and provided a Test Scoring Server (2021) and Test Mapping Server (2021) to ensure consistency and compatibility of these data exchanges.

Two categories of data were delineated: *status data* and *course data*. Status data are primarily derived from proprioceptive sensors for the purposes of calibration and internal health monitoring. Status data may also include exteroceptive sensor measurements that are collected within the Staging Area for the purposes of calibration. Course data are primarily derived from exteroceptive sensors that acquire information directly or indirectly from the competition course. Course data specifically include any information related to mapping and/or artifacts.

4.1.1. Human Supervisor

The team was permitted to have a single Human Supervisor at a Base Station external to the course but within the Staging Area. The Human Supervisor was permitted to monitor and manage the communications with their deployed systems as they choose. Only the Human Supervisor was permitted to use wireless communications with the systems during the competition run. The Human Supervisor was permitted to view, access, and/or analyze both course data and status data.

The Human Supervisor role was especially critical at the Final Event as the complexity of the environment increased (due to the presence of all three subdomains), the number and complexity of the deployed systems increased, and the multimodal nature of artifact detections increased (given all artifact types were in play).

For most teams, the role of the Human Supervisor included commanding high level missions, providing manual intervention when needed, deciding when to deploy communication nodes, coordinating the next system deployments with the Pit Crew, monitoring incoming sensor streams, reviewing artifact detections for viability, and sending artifact reports to the DARPA Command Post. Though observed throughout earlier challenge events, the Final Event results also demonstrated that while the Human Supervisor is typically thought of as a "mission enabler," they can also often be the "weak link" in the human-robot team setting. For example, in some cases, the user interface and cognitive load of the Human Supervisor were the limiting factor to the team's performance.

As intended, the restriction of only one Human Supervisor has significantly driven investments into enhanced autonomy and improved user interface design. Throughout the SubT Challenge, the teams continued to learn and improve their Base Station interfaces, as well as pushed towards more reliable autonomy that can be trusted to deploy with fewer human interventions.

4.1.2. Pit Crew Personnel

Up to four additional team personnel were permitted in the Staging Area to serve as a "Pit Crew" to assist with operations tasks such as physically deploying the systems, performing repairs, modifying software or firmware, and changing batteries. Pit Crew personnel were permitted to view and access status data but were not permitted to view or access course data.

While teams were permitted up to nine Pit Crew personnel in the Circuits Stage, DARPA decided to further restrict teams to only four Pit Crew personnel for the Final Event in order to further incentivize greater autonomy in the deployments and to encourage teams to envision deployments that require fewer support personnel. Such small-crew operations are reflective of typical rescue team compositions, as found in mine rescue teams (Mine Rescue Teams, 2022) and military operations (U.S. Dept. of the Army, 2019).

4.1.3. Preliminary Round and Prize Round

Each team conducted three runs. The Preliminary Round consisted of two scored runs, each 30 minutes in duration. The Prize Round consisted of one scored run, 60 minutes in duration. The final ranking was based solely on the Prize Round run score. The course was populated with 20 artifacts for the Preliminary Round and 40 artifacts for the Prize Round. The Preliminary Round used the following six artifact types: survivor, cell phone, backpack, fire extinguisher, vent, and rope. The Prize Round used the following ten (10) artifact types: survivor, cell phone, backpack, fire extinguisher, drill, vent, gas, rope, helmet, and cube. A team could earn a maximum of 20 points per run in the Preliminary Round and 40 points per run in the Prize Round.

4.2. Virtual Competition

In the Virtual Competition, a fully autonomous mode of operation was selected, both to push the boundaries of autonomy algorithms and to enable rapid solution evaluation against a variety of unique scenarios (Choi et al., 2021). Competitors uploaded solutions by selecting their team configuration from a set of robots provided by DARPA in the SubT Tech Repo (2018), then providing software for each robot. Experiments were executed on Amazon Web Services (AWS) cloud machines, where the SubT Simulator loaded each scenario and robot, providing sensor data to the solution software at runtime. Each solution operated with no human intervention or prior knowledge of the scenarios, aiming to navigate, map, and find artifacts. Meanwhile, DARPA's competition

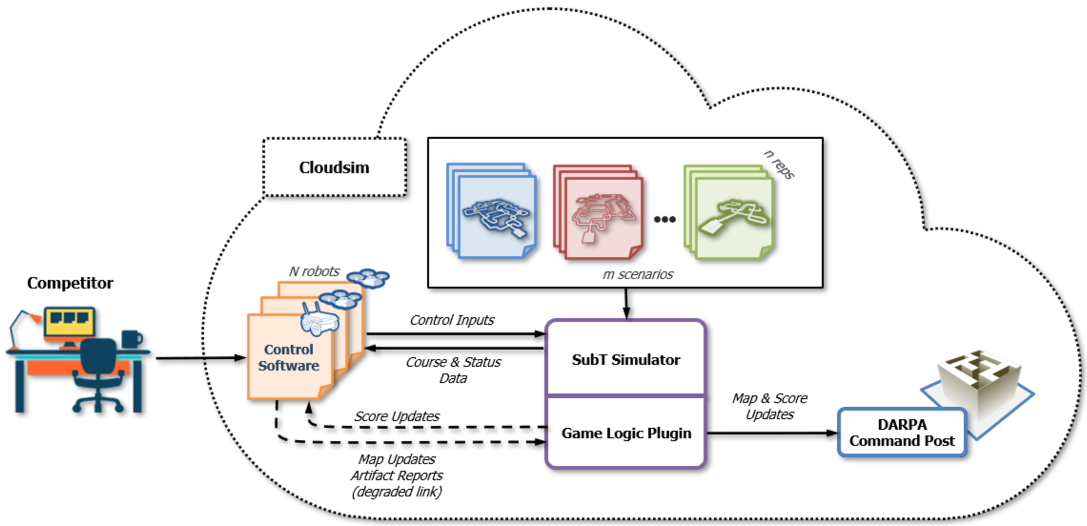


Figure 7. Virtual Competition Concept of Operations illustrates the interactions and information exchange between competitor solutions and testbed infrastructure.

infrastructure software analyzed solution performance through artifact reports as well as mapping, traversal, distance and other metrics. Figure 7 illustrates the workflow for competitors utilizing the SubT Virtual Testbed Repository (2018) infrastructure.

For the Preliminary Round of the Final Event, the twelve qualified competitors' solutions were run three times each through three competition worlds, totaling 108 runs. The top nine competitors who advanced from the Preliminary Round and submitted solutions to the Prize Round were tested against eight competition worlds with three replications each, totaling 216 runs. Winners were determined by summing scores across all runs in the Prize Round.

4.2.1. Autonomous Operation

The Virtual Competition was designed to push the boundaries of autonomy by requiring fully autonomous software solutions, i.e., competitors' software made all decisions with no human intervention at run time. Competitors selected a team of virtual robots, detailed in the following section, and submitted a prebuilt software container associated with each robot. The software was then submitted as a Docker image, an executable unit of software that includes dependencies. Hence, when the simulations were run, all the code within executed automatically to read the robot's sensor data, explore the virtual world, find artifacts, and communicate to report their locations.

The competition environments were unknown to competitors during solution submission. The solutions were only provided course and status data from each robot's simulated sensors and were isolated from each other except for networking attempts sent through the degraded communication model (as described in Section 5.4.2.) The SubT Virtual Testbed provided primarily Robot Operating System (ROS) message interfaces between competitors' software and the simulation and scoring infrastructure.

- Sensor data were provided through ROS messages; each robot's sensor data were only accessible by its corresponding solution software.
- Communication across robots was facilitated through a communications client with unicast, multicast, and broadcast options. The communication client also provided the means to transmit artifact reports for scoring and receive score updates.
- A basic controller was provided for each robot to interface between wheels, rotors, or legs and the competitor's linear and angular velocity commands. Aerial robots could optionally be controlled via thrust commands and legged robots via joint commands.

- Active payload components such as gimbals, droppable communication radios, and marsupially deployed robot pairs were provided with ROS-based controllers.
- As a virtual representation of initial gate calibration, each robot's ground truth pose relative to the artifact origin was available while the robot was inside the staging area.
- A periodically-published ROS message relayed competition run status and time remaining to all robots.
- Each robot was able to directly convey mapping data to the scoring infrastructure via ROS messages.

4.2.2. Robot Selection

Virtual competitors built their robot teams by choosing from the platform and sensor configurations available from the SubT Tech Repo. Each robot possessed a set of distinguishing attributes depending on type of platform, choice and placement of sensors, and associated cost in "SubT Credits" based on its estimated real-life cost. Competitors selected and composed their multi-robot teams within a total budget of 1,000 "SubT Credits."

The SubT Tech Repo offered competitors a total of 34 different robot platforms (Figure 8) with 108 sensor configurations for the Final Event. The availability of robot models reflected several priorities of the competition and provided an opportunity to gain insights about the requirements and design choices surrounding the competitors' software solutions. DARPA set the cost of each robot platform and configuration, the total cost limit for each team, and the maximum quantity of deployed robots of each unique platform (five robots). These limitations aimed to compel competitors to optimize the capabilities of their solutions by balancing team size, robot mobility, sensor fidelity, and heterogeneity. Additional insights may be gleaned by improving upon and further expanding the robot models in the SubT Tech Repo as well as adjusting the many possible tuning knobs such as robot cost and platform limits, robot performance characteristics, and sensor configurations.

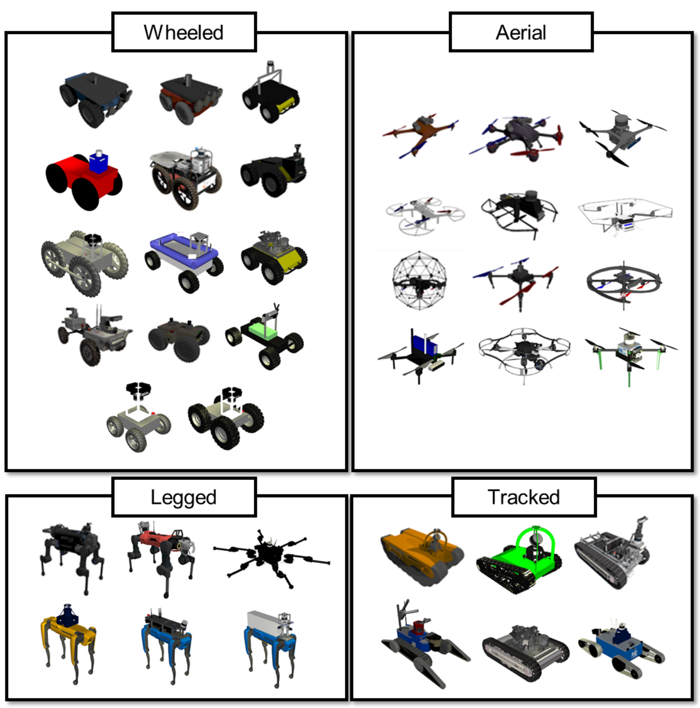


Figure 8. Virtual robot models (based on physical Systems platforms) created throughout the SubT Challenge and incorporated into the SubT Tech Repo.

Realism. The robot models included 38 robot platforms modeled after physical robots utilized by teams in the Systems Competition, matching the real robots' validated robot motion and sensor specifications.

Heterogeneous Mobility. Methods of locomotion varied by platform, facilitating assembly of heterogeneous robot teams. The SubT Tech Repo offered the choice between twelve wheeled UGVs, twelve multirotor UAVs, five tracked UGVs, and four legged UGVs for the Final Event.

Multimodal Sensing. Robots were outfitted with a variety of payloads including 2D and 3D scanning lidars, point lidars, visual and thermal cameras, RGBD and time-of-flight depth cameras, rotating gimbals, inertial measurement units (IMUs), barometers, magnetometers, and deployable communication radios. Payload components such as lidars, cameras, and gimbals were characterized based on real sensor datasheets with increased cost in SubT Credits for models with superior specifications (e.g., wider field of view, higher resolution).

4.2.3. Cloud Simulation

The Virtual Competition employed a cloud-based simulation infrastructure for scored competition runs to ensure that competitor-submitted solutions were completely autonomous with no external interaction and prior knowledge of scenarios was prohibited. Cloud machines provided a level playing field across all teams and eliminated access to ground truth simulation data and other exploitations. Cloudsim, the software system that ran and managed simulation instances on AWS, was developed and deployed as both a practice and competition tool. Its use allowed the Virtual Competition to perform batch competition runs through many diverse competition worlds, to scale with the number of competitors, and to recruit developers from around the world (Choi et al., 2021; Courchesne, 2021).

Teams submitted solutions to the cloud-based simulator via Docker images associated with each selected robot. The SubT Virtual Testbed infrastructure was also containerized as a publicly-hosted Docker image to allow teams to mimic the structure of a competition run on their local machines. Across all competition events, the SubT Virtual Testbed Docker image was downloaded approximately 7,000 times, indicating many instances of teams updating their local testing environment for practice.

Each cloud-based simulation utilized *(2 + N)* cloud instances, where N is the number of deployed robots. Figure 9 illustrates an example setup for a simulation with two robots, which was distributed across multiple cloud instances with different containers of software for both the simulation infrastructure and the competitor's solution. The cloud instance running the simulation (left) included the environment model, robots, sensors, and scoring and logging infrastructure. The

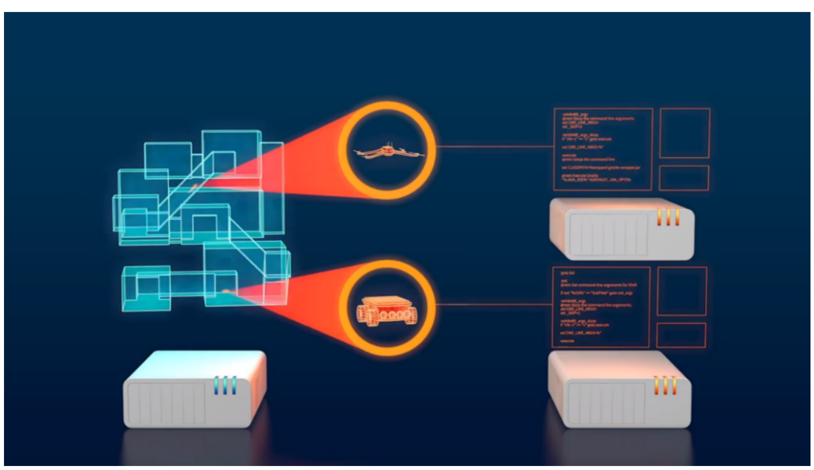


Figure 9. Illustration of cloud instances required for a simulation run with two robots: an instance containing the simulator and scoring infrastructure (left), and an instance for each robot equipped with a competitor's solution software (right) (SubT Challenge Virtual Competition: Cloud-hosted Simulation, 2020).

robot instances (right) ran the competitor's solution software and message bridges. The message bridges provided the solution software with only select data from the robot's sensors and successful communications with nearby robots (determined by the simulated communication model). Another instance, not pictured, analyzed and stored teams' mapping data for each simulation.

Cloud-based execution of the Virtual Competition events facilitated parallel simulation of all scenarios, enabling faster solution evaluation. In total, the Prize Round of the Final Event utilized 1,824 cloud instances. Batch simulation of both rounds of the Final Event totaled 7,781 hours of run time, with each run averaging 36 hours. Thus concurrently running simulations reduced evaluation time drastically when compared to sequential execution of each run.

5. Worlds

In the sequel and in Section 6, we present the multi-faceted considerations that contributed to designing the competition courses for the Final Event. A similar discussion of the Circuits Stage event competition courses is provided in Orekhov and Chung (2022). We introduce the following terminology used throughout: *worlds*, to refer to the actual competition environments (physical or simulated), and *scenarios*, to describe how various challenge elements and artifacts were arranged within the worlds.

5.1. Design Priorities

As a public-facing competition, the SubT Challenge carried the dual responsibility of a) serving as a large-scale experimental testbed to evaluate subterranean technologies and b) educating the general public about an underappreciated domain and the associated technical challenges. Public-facing events provide an opportunity for DARPA to showcase how research investments impact innovation, encourage people outside of the robotics community to apply their fields of study to subterranean challenges, and inspire younger audiences to consider studying robotics and specifically subterranean applications. Because of this dual role, the design of the competition courses was a critical component in achieving the right balance between the "science" and the "show" components of designing the Final Event.

The high-level priorities for course design were aimed at achieving a testbed listed below.

- Representative. Serve as a representative environment that captures the key environmental challenges that make operating in subterranean environments difficult for robots and humans.
- Predictable. Present common subterranean features so that teams have a fair chance of preparing well and are able to focus their efforts where they will be most effective and generalizable.
- Unknown. Maintain a level of surprise to ensure teams do not "study to the test" and instead develop generalized and robust solutions that are able to effectively operate across diverse subterranean environments.
- Realistic. Present an appropriate level of difficulty that facilitate differentiation between approaches without being too hard (no team covers the entire course) or too easy (most teams cover the entire course).
- Fair. Provide a standardized course with reproducible conditions so that all teams are able to showcase their solutions without bias due to specific elements of the course design or the order in which runs are conducted.

5.2. Subdomains

The inclusion of all three subdomains in competition courses was intended to encourage teams to develop robust solutions that are capable of operating in any subterranean environment rather than point solutions that are only effective in a specific type of environment. While subterranean

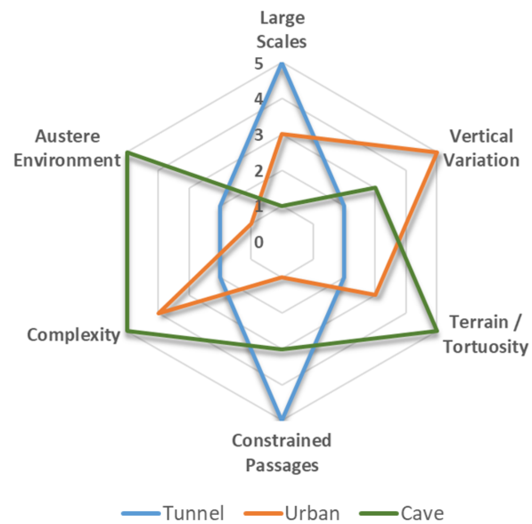


Figure 10. Variation of environment characteristics by subdomain.

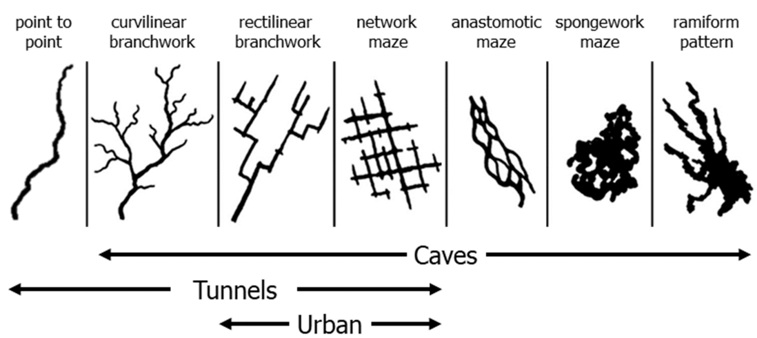


Figure 11. Common topology types by subdomain [Adapted from Palmer (1991)].

environments can often exhibit a common set of characteristics including confined spaces, limited visibility, poor air circulation, unpredictable communications, obstructions, moisture, vertically, and expansive spaces, each subdomain presents a varied combination of these characteristics as shown in Figure 10.

On a more macro level, the topology of subterranean environments exhibits a similar "similarity with distinctions" as shown in Figure 11. To capture the varying topologies observed across subterranean environments, the Final Event course and virtual worlds included long point-to-point passages, maze-like areas with a grid structure, dead ends, rectilinear intersections, curvilinear intersections, and areas that resembled the more organic and freeform topologies found in naturally occurring caves.

In the Final Event, the single Systems Competition course combined segments of all three subdomains. In contrast, the Virtual Competition courses comprised both single-subdomain worlds and combinations, taking advantage of the numerous worlds to compare teams' performance across subdomains.

5.3. Systems Competition Course Overview

Figure 12 shows the competition course for the Final Event of the Systems Competition. The course included three distinct subdomains which were interconnected and even overlapped each other. The

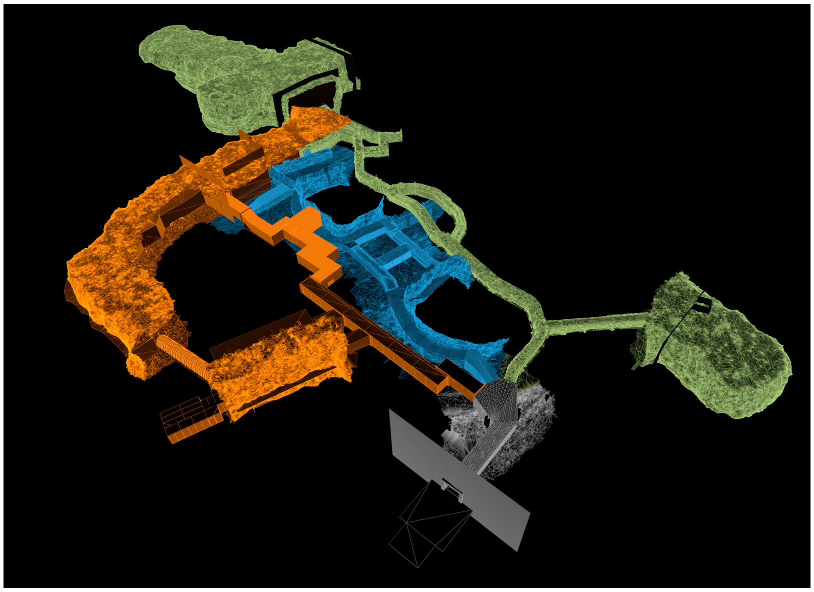


Figure 12. Final Event Competition Course.

layout of the course provided teams with (1) direct access to all three subdomains at the beginning of the course, (2) crossover points midway through each subdomain to the adjacent subdomain, and (3) provided a crossover point between all three subdomains in the back of the competition course. These deliberate design decisions were intended to give teams an opportunity to demonstrate their ability to operate in all three subdomains, encouraged the full utilization of the competition course, rewarded the ability to transition and operate across different subdomains, and gave DARPA insights into deployment strategies and each team's level of confidence in tackling each subdomain.

5.3.1. Venue Selection

The Final Event sought to incorporate challenge elements from all three subdomains including human-made tunnel systems, urban underground, and natural cave networks into a single integrated competition course. After an extensive site search process, the Louisville Mega Cavern in Louisville, Kentucky was selected as the venue for the Final Event. This wholly underground venue, once a limestone mine, provided a commercial site in which DARPA could conduct sustained operations, including infrastructure installation, course installation, and competition execution while providing the use of the venue's natural terrain, walls, ceilings, and caverns. The Final Event course incorporated a one-of-a-kind, custom-fabricated modular structure to accurately represent all three subdomains in a composite course. Dedicated access and complete control of the site and course design were critical to the successful installation, maintenance, and execution of the competition activities.

Despite extensive utilization of the venue terrain and walls, much of the course still needed to be fabricated. Due to the costs associated with designing and fabricating a high-fidelity course, the Systems Competition only had one competition course. Conversely, the Virtual Competition did not have the same logistical and practical constraints which enabled the development of multiple competition worlds, each with its own unique characteristics.

5.3.2. Network Topology Analysis

Table 2 presents a network topology analysis comparing the Final Event competition course with the courses in the Tunnel Circuit and Urban Circuit. Five key metrics were used in the analysis: nodes, edges, cycles, cross sections, and levels. An "edge" denotes a path between decision points

Table 2. Course network topology analysis of Systems Competition courses.

	Tunnel SR	Tunnel EX	Urban Alpha	Urban Beta	Final Event
Nodes/Decision Points (#)	71	43	68	105	120
Mean Node Degree	2.5	2.3	2.0	2.1	2.4
Mean Node Degree (excl. Dead Ends)	3.0	2.6	3.2	3.1	2.5
Max Node Degree	4	4	8	6	5
Dead Ends (#)	16	8	38	51	12
Dead Ends (%)	23	19	56	49	10
Edges (#)	90	49	67	108	142
Mean Edge Length (m)	22	36	11	12	7.26
Max Edge Length (m)	79	97	29	33	22.5
Total Edge Length (m)	1958	1762	751	1276	1030.9
Simple Cycles	76130	25	0	10	36067
Avg. Cycle Length (m)	631	206	n/a	265	418
Max Cycle Length (m)	1105	288	n/a	428	579
Avg. Nodes Per Cycle	30.9	8.6	n/a	14.4	57.7
Max Nodes Per Cycle	49	13	n/a	23	78

(nodes). Nodes were manually placed at all intersections and occasionally along regular intervals of longer segments that could constitute a "decision" to continue moving along the same direction. In large open areas, nodes were placed along paths that would be required to fully explore the area around obstructions (e.g., shelves in warehouse) or with the limited field of view and illumination of common sensing payloads (e.g., completely dark cavern). Simple cycles include all cycles in which no node is repeated (except the start and end node). Number of levels does not include intermediary levels, landings, half-height rooms or their roofs, or mezzanines. All distances have been rounded to the nearest meter.

These results were used to compare the competition courses used in the Circuits Stage and to inform the design of the Final Event Course to ensure the network topology was representative of real-world subterranean environments. A rich area for future work could include a more thorough analysis of the impact of network topology on autonomy and component technologies. For example, what exploration approaches are more effective for environments with higher mean node degree, what effect do dead ends have on mapping, or what combination of metrics should an incident commander consider in selecting the number of robots to deploy. The simulation infrastructure described in Section 4.2 could be an effective resource for this future work.

5.3.3. Cross Section Analysis

The Final Event course intentionally varied cross sections within each subdomain and across subdomains to recreate cross section characteristics commonly found in subterranean environments. Tunnel Environments often have fairly consistent cross sections based on the mining or boring method used when the environment was created. Urban environments often have rectangular cross sections and can vary greatly in their cross-sectional area, ranging from small doorway passages and crawl spaces to large open areas such as warehouses. Cave cross sections also vary greatly in area, from very constrained human-crawlable passages to large cavernous areas. The cross sections often include slopes perpendicular to the direction of travel, organic shapes based on geological properties, and obstacles in the form of stalagmites, stalactites, and boulders. The Final Event course included both constrained passages with human-crawlable cross sections as well as larger open spaces that included large ledges, mezzanine levels, and vertical shafts.

Table 3 provides a cross-sectional analysis of the Final Event competition course and compares it to the courses used in the Tunnel Circuit and the Urban Circuit. Figure 13 shows the 113 representative cross sections that were taken and used in this analysis. Figure 14 shows a composite

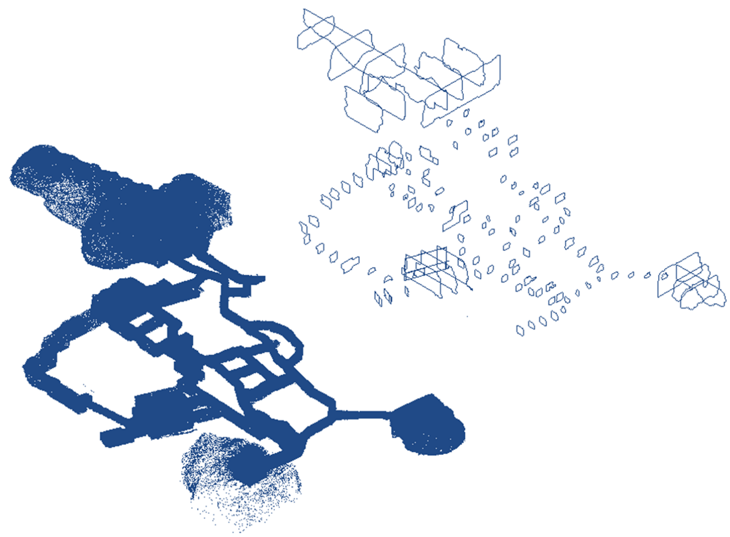


Figure 13. Representative cross sections of the Final Event competition course.

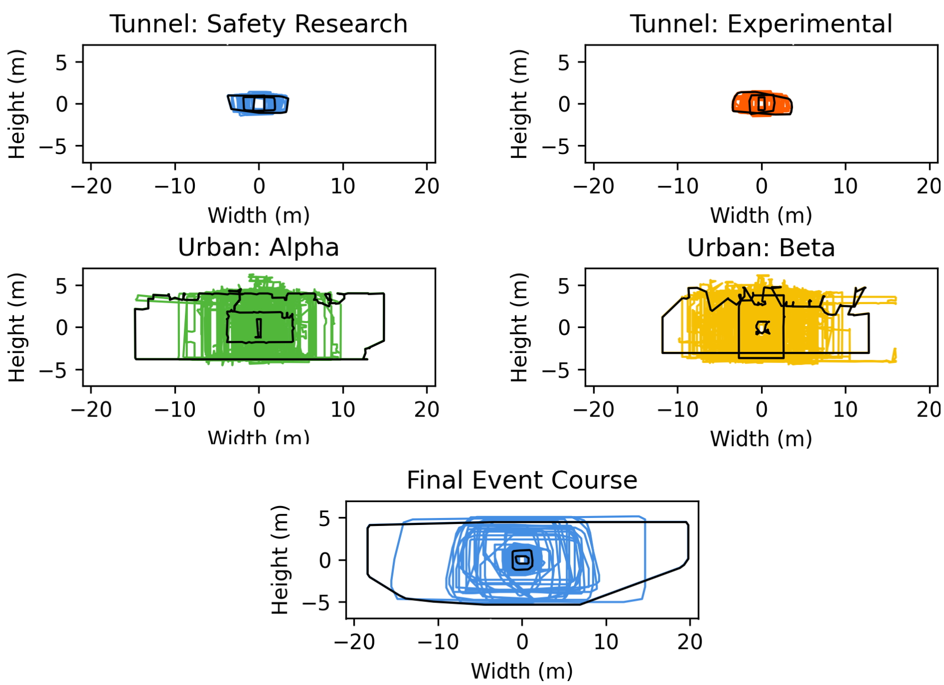


Figure 14. Composite cross sections of competition courses from the Tunnel Circuit (top), Urban Circuit (middle), and Final Event (bottom).

of all the cross sections across the Tunnel Circuit, Urban Circuit, and Final Event competition courses for comparison. Table 3 provides a comparison of cross sections across the Circuit Events and the Final Event.

As the previous events were intentionally held in real-world environments, the cross-sectional analysis of the Tunnel and Urban Circuit was used to inform the design of the Final Event course. These metrics were further informed via common phenomenological characteristics of the numerous subterranean environments visited by the organizers over the course of the SubT Program. The vast

Table 3. Comparison of cross sections across the competition event courses.

	Tunnel SR	Tunnel EX	Urban Alpha	Urban Beta	Final Event
Max Cross Section (m^2)	12.9	15.4	217.1	145.4	329.3
Mean Cross Section (m^2)	6.3	6.1	40.2	42.3	25.2
Min Cross Section (m^2)	2.3	1.1	0.8	1.4	1.2

majority of the competition course (65.49 %) was considered "constrained" (<7 m^2), 20.35 % was considered "open" (between 7–75 m^2), and 14.16 % was considered "cavernous" (>75 m^2).

5.3.4. Course Difficulty

In considering course difficulty, the intent of the competition course was to be representative of real-world environments, which are inherently very difficult, while still providing all teams a chance to showcase their technologies. The approach taken for the Final Event course was to begin with relatively easier sections as baselined against the capabilities that teams had already demonstrated in the previous Circuits Stage events. The course difficulty was then substantively increased such that the furthest areas of the course surpassed where the current state-of-the-art could reasonably perform and approached the difficulty of the more challenging real-world environments that were used as inspiration for the SubT Challenge. In this way, all teams were able to showcase some aspects of their technology while the highest performing teams were able to meaningfully differentiate their capabilities from that of other competitors. In post-event discussions and public presentations, each team has been able to highlight significant accomplishments independent of their final rank.

Of course, an intentional and gradual increase in difficulty is not realistic to expect in all real-world environments. The most difficult bottleneck could be located anywhere and in some cases could be the entrance itself. Nevertheless, for a competition setting, the carefully designed course provided a means to evaluate technologies and approaches while still presenting the teams with a range of difficulty that approached the difficulty of real-world environments.

Figure 15 provides a coded map (scale of 1 to 5) of the intended difficulty across the competition course, representing the as-designed difficulty anticipated across all mobility types. Figure 16 provides a map of the specific challenge elements which contributed to achieving the desired course difficulty.

In practice, course difficulty is a function of many variables: topology, terrain, obstacles, lighting, particulates, environmental conditions, distance from the entrance, and artifact placement. Teams took varying approaches to addressing and solving all of these variables, so it is important to note that the perceived difficulty of a given section could be dependent on a particular team's approach, in particular by mobility type. Wheeled robots have difficulty with high obstacles (e.g., stairs), legged robots struggle with slippery surfaces, tracked robots can suffer with high-centering, and aerial robots struggle with hanging obstacles and constrained passages. Given the wide variety of mobility platforms used by the teams, special care was taken to consider the fairness of the course across all mobility types by distributing these challenges and avoiding early choke points that would dramatically disadvantage a particular approach.

5.3.5. Course Design Methods

Once an area of the venue was selected, a professional surveying team surveyed key control points, captured a 3D point cloud, and generated 2D terrain profiles to aid in the design of the course. This effort was an essential step to ensure that the modular sections fabricated off-site would line up as expected during the installation on-site. All of the course design was completed using digital 3D Computer Aided Design (CAD) tools that was spatially referenced to the 3D ground truth scan and georeferenced to global coordinates. Using the digital design, the team was able to survey key locations and provide physical markers at key intersections that significantly enabled the rapid installation of the course segments as soon as they arrived on-site. After installation, the course was

Figure 15. Design intent for course difficulty across the Final Event competition course.

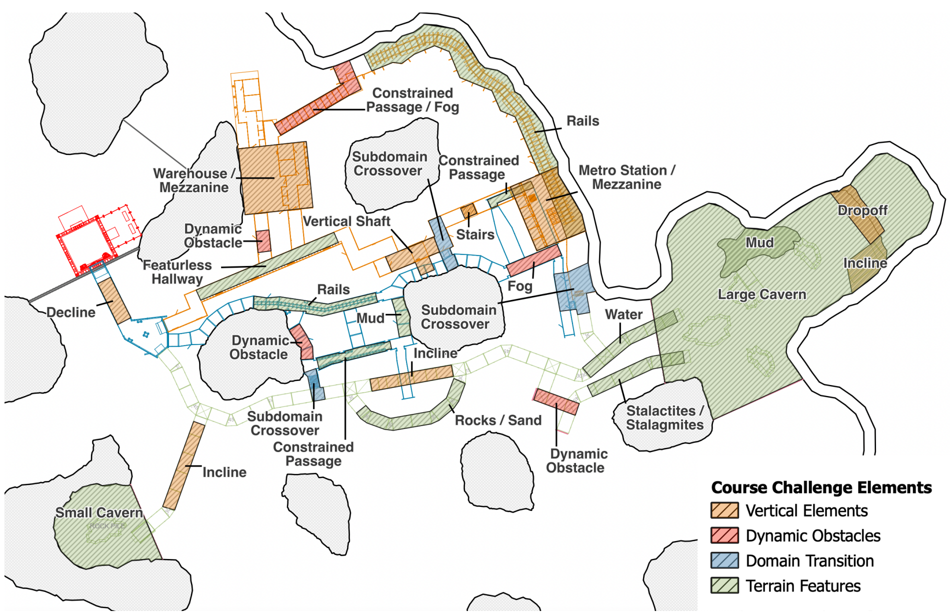


Figure 16. Distribution of Challenge Elements on the Final Event Competition course including terrain obstacles, vertical elements, dynamic obstacles, and subdomain crossovers.

professionally surveyed and scanned to produce a detailed 3D point cloud of the course to serve as the final ground truth data.

The course was broken up into 121 segments, which included the Staging Area, 22 urban segments, 20 tunnel segments, and 78 cave segments. The higher quantity of cave segments is due primarily to the construction method utilizing shorter sections; each subdomain had approximately equal total lengths of fabricated sections. The urban subdomain was 374 meters (1227 ft), the tunnel subdomain was 222 meters (728 ft), and the cave subdomain was 292 meters (958 ft) for a total length of 888 meters (2913 ft) of fabricated sections. Including the utilization of the venue terrain, the total competition course length reached just over 1030 meters.

Each of the individual 121 segments was inspired by and intentionally designed to replicate segments of either previous Circuit Event courses or other locations that the DARPA team had visited throughout the SubT Challenge program. A Finals Course Callouts (2021) design document was created to detail the design of each segment and included design parameters, fabrication details, inspiration references, and pictures of both the physical build and the virtual model of the segment.

5.3.6. Course Construction Methods

The approach for course construction was to leverage the local venue terrain, walls, and cavernous areas; fabricate most of the course off-site using modular structures that could be packaged into standard tractor trailers; and then rapidly construct the course on-site in the weeks leading up to the event. The Final Event course made use of the local venue terrain as much as possible to contribute to a high course fidelity while saving on material and installation costs. All of the tunnel subdomain used the venue terrain with additional aggregate and obstacles (e.g., gravel, rail, fire hose) brought in to increase the difficulty of some areas. Most of the urban subdomain relied on the installation of scaffolding to provide the level areas, steps, and drop-offs typical of urban environments. The organic terrain of the cave subdomain was achieved by prefabricating (carving foam) the segments to produce slopes, inclines, and obstacles inspired by naturally occurring caves. Various types of aggregate (pea gravel, large rocks, sand, mud) were added in different areas of the cave subdomain to capture the wide range of aggregate found in different caves. Ceiling obstacles in the form of hanging cables, tarps, mesh, door thresholds, and stalactites were included to replicate common UAV hazards.

The modular construction of the course consisted of three main approaches for the underlying structure of each segment: scaffolding, theater flats, and prefabricated pods. The urban segments predominately utilized scaffolding to establish a relatively level and stable floor for rooms and hallways. The tunnel segments largely utilized theater flats placed on natural ground terrain and often wrapped existing limestone walls and pillars to contribute to course realism. The cave segments nearly exclusively utilized prefabricated pods to enable the complex terrain and wall features to be fabricated, coated, and painted in advance of the course installation on-site.

The underlying structure was then covered with layers of varying materials to provide resilience to damage, fire proofing, RF mitigation, and realistic aesthetics. The base layer in most cases consisted of 1/2" plywood that was treated with fire-retardant and RF-retardant coatings. The primary flame-retardant materials used were Fire Stop E84 latex paint additive, FR-1 FlameX paint additive, and Rose Brand INSPECTASHIELD fire retardant. Additional layers provided additional 3D structure to create realistic wall textures, ground terrain, and obstacle features. The top-most layer then provided the scenic realism.

5.3.7. Course Design Tradeoffs

The practical considerations of time available for course fabrication, time on-site for installation, and budget constraints led to design tradeoffs primarily across three categories: course fidelity, course length, and course features.

Course fidelity was identified as critically important in order to effectively challenge and evaluate the teams' solutions, especially in the tech areas of mobility, perception, and networking. Realistic terrain, aggregate, fabrication materials, thematic props and obstacles, lighting, environmental conditions, and RF propagation characteristics all contribute to the challenges experienced by robotic systems underground. While the real-world environments of the Circuits Stage events included all of these elements intrinsically, the course fidelity of a fabricated course was a significant driver of cost.

For course length, the previous Circuits Stage events served as an effective baseline to ensure that the Final Event course was long enough to support the 60-minute run duration and representative in the ways that course length challenges mobility and networking in subterranean environments. The Urban Circuit course lengths were 751 m for Alpha and 1276 m for Beta while the Tunnel Circuit course lengths were 1762 m for Experimental and 1958 m for Safety Research. The intent for the Finals course was to have a course length no less than the shortest Circuits course (i.e., 750 m) while being at least as difficult as any of the previous competition courses. In the end, the length of the Final Event course was just over 1,030 m.

Several important course features were also significant drivers of cost: multiple levels, large cavernous areas, dynamic obstacles, radio frequency blocking, damage resilience, fire proofing, and environmental hardening. Fire proofing and reasonable damage resilience were firm requirements due to safety considerations and limited time between runs for course repair and reset. Other features such as dynamic obstacles, fog machines, adjustable lighting, and realistic props were selected as effective investments that would create a dynamic environment and provide means to adjust course difficulty as needed. The two features that presented the greatest design tradeoffs were vertical elements and RF mitigation.

5.3.8. Vertical Elements

Vertical elements such as multiple levels and cavernous elements were important to include, but were limited by the available budget. Fabrication and installation methods to create multiple levels while maintaining safety for both robots and humans proved to be very costly. Cavernous areas also presented significant material cost and installation cost, so the course leveraged two areas within the venue that naturally provided a cavernous area without significant cost. Additional areas were designed to specifically include vertical elements and provide opportunities for UAVs to showcase their capabilities. The Staging Area was placed so that teams were presented with a significant decline immediately at the beginning of the course, two additional cavernous areas (i.e., urban warehouse, metro platform) provided areas that could only be fully explored with UAVs, a vertical shaft in the urban subdomain was only accessible by UAVs, three sets of stairs and a large drop-off (1.71 m) from the metro platform were included in the urban metro area, and a very step inclined area was included in the cave subdomain that was tall enough to pass over one of the tunnel segments. The larger cavern in the cave subdomain also included a very high ledge (4.5 m high) that had a long incline to ascend and large drop-off in the form of a large cliff.

5.3.9. Radio Frequency Mitigation

Radio frequency (RF) mitigation was another exercise in tradeoffs. Previous events had the benefit of intrinsic RF blocking of thick walls made of rock, coal, or reinforced concrete. Constructing a course with the same RF blocking characteristics would not have been feasible. The approach taken was to construct a course and apply RF mitigation measures in select areas such that the teams' networking approaches would be sufficiently impacted to differentiate between networking approaches that were more effective from others. This approach included strategically using the rock pillars and walls present within the venue, intentionally placing the Staging Area in a location where a pillar blocked much of the RF signal, installing RF mitigation curtains around the Staging Area, and applying more of the RF mitigation measures (i.e., RF paint) in sections closer to the Staging Area.

Most of the fabricated course segments were coated in three layers of sealed RF paint on the underlying structure (i.e., plywood), and in some cases were additionally covered with medium muslin fabric that had been painted with three layers of sealed RF paint. The curtains around the Staging Area consisted of two layers. The first layer was a heavy muslin material covered on both sides with three layers of hand-rolled RF paint and sealed with a layer of stock gray, flame resistant additive paint for a total of eight layers of paint. The second layer was positioned 12" from the first and was made of a single piece of flame-resistant treated copper fabric. The RF paint used was CPC-54 RF Shielding Paint from LBA Technology.

In order to validate the results, RF testing of the course area was conducted prior to installation and again after the course had been fully installed. Tests were conducted at 900 MHz, 2.4 GHz,

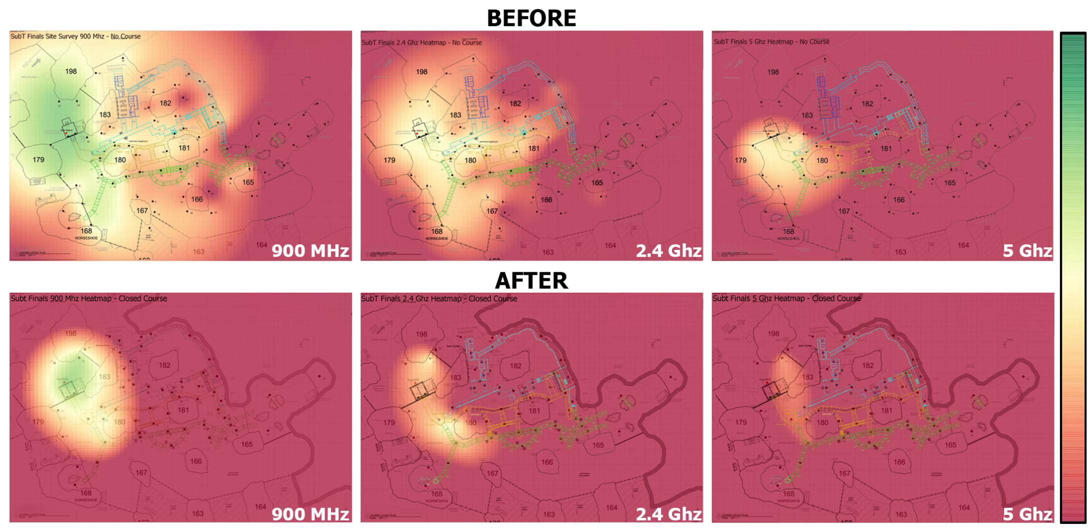


Figure 17. RF propagation as measured before and after the installation of the competition course at different frequencies.

and 5 GHz between the Staging Area and various points within the course area. Figure 17 shows the results of this testing and demonstrates the effectiveness of the RF mitigation measures taken during the fabrication and installation of the competition course. Some areas of the course saw a reduction of as much as −40 dB.

5.3.10. Course Design Outcomes

The scale and scope of the course build is difficult to capture and represent, but the following statistics provide a sense of the efforts undertaken to construct the SubT Challenge Final Event competition course:

- 2,500+ sqft in-course signage (enough signage to cover 4 highway billboards),
- 98,000+ sqft of lumber (enough plywood to cover a football field twice),
- 23,000+ lnft of steel tubing (over 4 miles),
- 2,650 gallons RF paint (enough paint to cover 22 747 airplanes),
- 1,500 gallons scenic paint (enough paint to cover 75 single family homes),
- 1,700 individual props (40 times the number used in Broadway's *Phantom of the Opera*).

The numerous tradeoffs and design decisions led to the successful creation of a competition course that had the length to be challenging in a 60-minute run, presented high fidelity to stress-test mobility and perception, and included the most important features to stress-test autonomy and networking. Some teams were able to navigate more than 80 % of the course, but no team was able to reach all areas of the course. All teams experienced networking losses and all of the teams with droppable breadcrumb nodes relied on them to extend the reach of their communications further into the course. No team that reached more than 50 % of the course was able to maintain continuous communication to their Base Station. The course fidelity was identified by teams, media (Ackerman, 2022; Montgomery, 2022), and stakeholders as being incredibly realistic and successfully differentiated approaches to mobility and perception.

5.4. Virtual Competition Worlds Overview

For the Virtual Competition, subterranean worlds were modeled not only to be representative of real-world environments, but also to highlight specific challenges to facilitate comparisons of teams' solutions. Scalable simulation in the cloud allowed DARPA to test solutions against a diverse array of scenarios with varied topology, terrain, obstacles, and length. To facilitate participation and preparation for teams to enter the Virtual Competition, multiple classes of virtual worlds were designed and released, including some worlds released prior to the competition in addition to those worlds developed for the competition itself. For the Final Event, these various classes of virtual worlds (with numbers of worlds indicated in parentheses) included the following.

- Qualification World (1), which allowed prospective competitors to demonstrate readiness and sufficient familiarity with the SubT Virtual Testbed for competition;
- Practice Worlds (3), which provided publicly visible simulation environments with scales, complexity, and artifact placements that were intended to be representative of the competition environments;
- Preliminary Round Competition Worlds (3), which represented the hidden set of test environments in which competitor-submitted solutions were evaluated to determine advancement into the Prize Round; and
- Prize Round Competition Worlds (8), which represented the hidden set of test environments in which competitor-submitted solutions were evaluated for the Final Event.

Teams were also able to utilize worlds from all prior Circuit Events, including prior Circuit Competition Worlds, as test scenarios in advance of the Final Event. In total, sixty-five worlds were available on the SubT Tech Repo (2018) including the Competition Worlds which were publicly released after the Final Event. Each world contained twenty artifacts for robots to discover and report for scoring. In total, the worlds on the SubT Tech Repo had:

- Over 200 km of traversable length,
- 7,782 nodes and 8,218 edges,
- 1300 artifacts,
- 94 dynamic obstacles.

5.4.1. World Design Methods

Virtual subterranean environments were created for the SubT Simulator using two main approaches: (a) building from synthetic mesh tiles, or (b) generating meshes from scans of real-world underground environments.

Over 250 synthetic tiles were designed as modular representations of core elements of the three subdomains (inspired by physical site searches and visits by the DARPA team) to allow reconfiguration into new test scenarios, including procedural generation for a given tile count and subdomain type. The tiles functioned as building blocks to create varied world layouts using straight, curved, intersecting, and vertical shaft segments with varied lighting and ground features (e.g., rails, debris). Figure 18 shows several examples of tile variations for each subdomain.

Tunnel tiles were designed to reflect human-made mine segments with consistent cross-sectional area except for select constrained tiles with passages as small as 1-meter diameter.

Urban tiles represented two main environment styles, namely a factory/power plant style and a subway style. The plant-style tiles included service rooms, stairwells and ramps, large open rooms with pillars, and multi-story rooms connected by elevator shafts and stairwells with landings. The subway-style tiles had straight and curved tunnel sections with rails, subway platforms, and inclines/declines. The subway rails proved to be particularly challenging for competitors' solutions, since the ground robots could only successfully traverse the rails at carefully executed angles and speeds. Ground vehicles were also limited in traversal of different levels; they could climb and descend ramps to circumvent some of the stairwells but could not conquer stairs or elevator shafts.

Cave tiles were created to allow assembly of rectilinear branchwork, curvilinear branchwork, and anastomotic topologies, including variable cross-sectional size, shape, and elevation changes. Cave elements included stalactites and stalagmites, lava tubes, and large caverns with entrances on multiple levels.

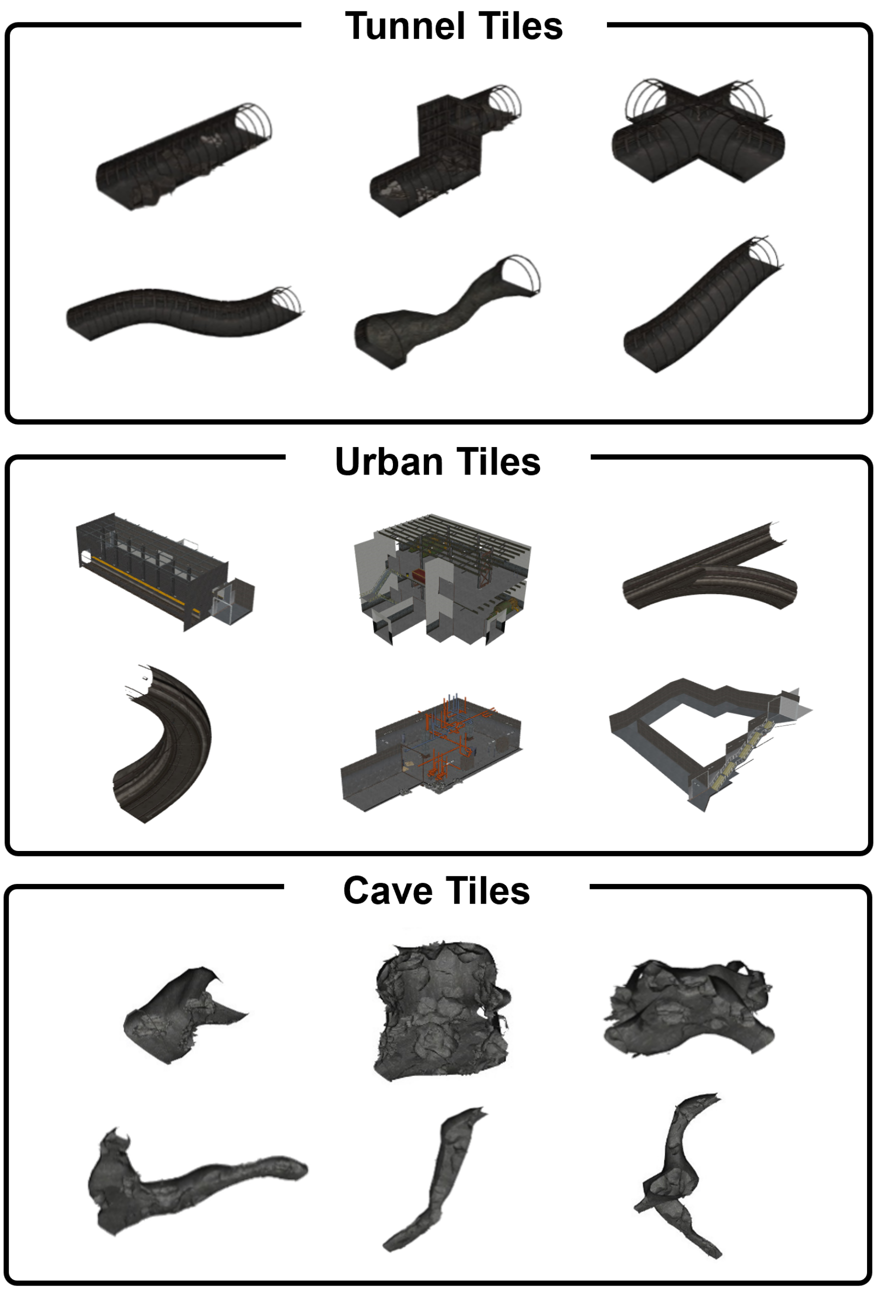


Figure 18. Examples of virtual environment tiles for tunnel, urban, and cave subdomains.

Transition tiles released for the Final Event made multi-subdomain worlds possible by bridging between the varied cross-sectional shapes of each subdomain.

The variety of tiles facilitated design of a multitude of worlds, but desire for further realism as well as additional terrain and feature variation motivated development of worlds based on scans of real-world environments. To generate a fully enclosed world mesh, the point clouds from high-fidelity scans needed to be connected by triangles and/or polygons as described in the Gazebo API (2020). In practice, automated conversion using software such as CloudCompare was often augmented or replaced by extrusion of 2D maps or site drawings. Manual feature placement was also necessary

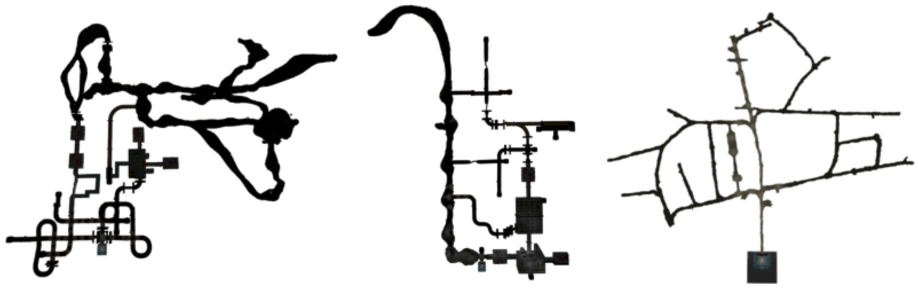


Figure 19. Virtual Competition worlds from the Preliminary Round of the Final Event.

to ensure terrain features were not lost; the scan data and walkthrough videos were referenced for this purpose. One notable exception to this process was the modeling of the Systems Competition Finals course, which had natural components of the Louisville Mega Caverns site converted from scan data and built sections of the course modeled concurrently in the course design process.

The worlds designed for the Final Event each provided unique challenges for autonomy technologies, including both worlds composed of synthetic mesh tiles and scans from real-world underground environments. The worlds presented diverse subdomain composition, path length and cross-sectional area, complexity, and verticality. The Preliminary Round included two worlds comprising tiles from all three subdomains and one world modeled from a scan of the Edgar Mine in Idaho Springs, Colorado (utilized as the site of the SubT Integration Exercise (STIX) Event for the Systems Competition) as shown in Figure 19.

The eight Prize Round Competition Worlds (Figure 20) further varied in composition to facilitate comparison of solution performance across subdomains and in scenarios directly modeled after real-world environments. Half of the Prize Round worlds were built from synthetic tiles and the other half were modeled from physical environment maps and real-world scan data. All three subdomains (Tunnel, Urban, and Cave) were represented in separate worlds as well as combined "mashup" worlds with elements from all three subdomains. Two of the worlds were modeled using scan data of the Systems Competition courses from the Tunnel Circuit and Urban Circuit events, but each virtual representation combined the courses by removing physical blockages. This manipulation effectively doubled the searchable area in the virtual worlds compared to their course counterparts in the Circuits stage, which are analyzed further in Sections 5.3.2 and 5.3.3. The scanned cave world was also formed from datasets of two separate caves which were artificially connected. The last real-world model was generated from scan data augmented with 3D modeling of the Systems Competition Finals course, which was performed concurrently with course design. The use of both synthetically generated virtual worlds and those based on real-world environments further reinforced the SubT Challenge's vision for unifying and synergistically fostering advances in both the Virtual Competition and the Systems Competition.

5.4.2. Communication Modeling

To reflect the difficulty of radio communication in subterranean environments, the virtual environments included a model of communication degradation. Thus communication became an important software development challenge requiring innovative solutions to manage transmission of large amounts of data, updates of shared information, and mesh networking.

Degraded communication presented as message drops, where the probability of successful transmission was calculated through a cost function weighing the attempted message size and data rate, the distance between radios, and the topology of the environment in the path between the source and destination. The effect of environment topology was quantified by a graph-based representation of each world. For each connected edge, a component of communication cost was assigned based on expected visibility, such as intersections and turns were assigned higher costs than straight

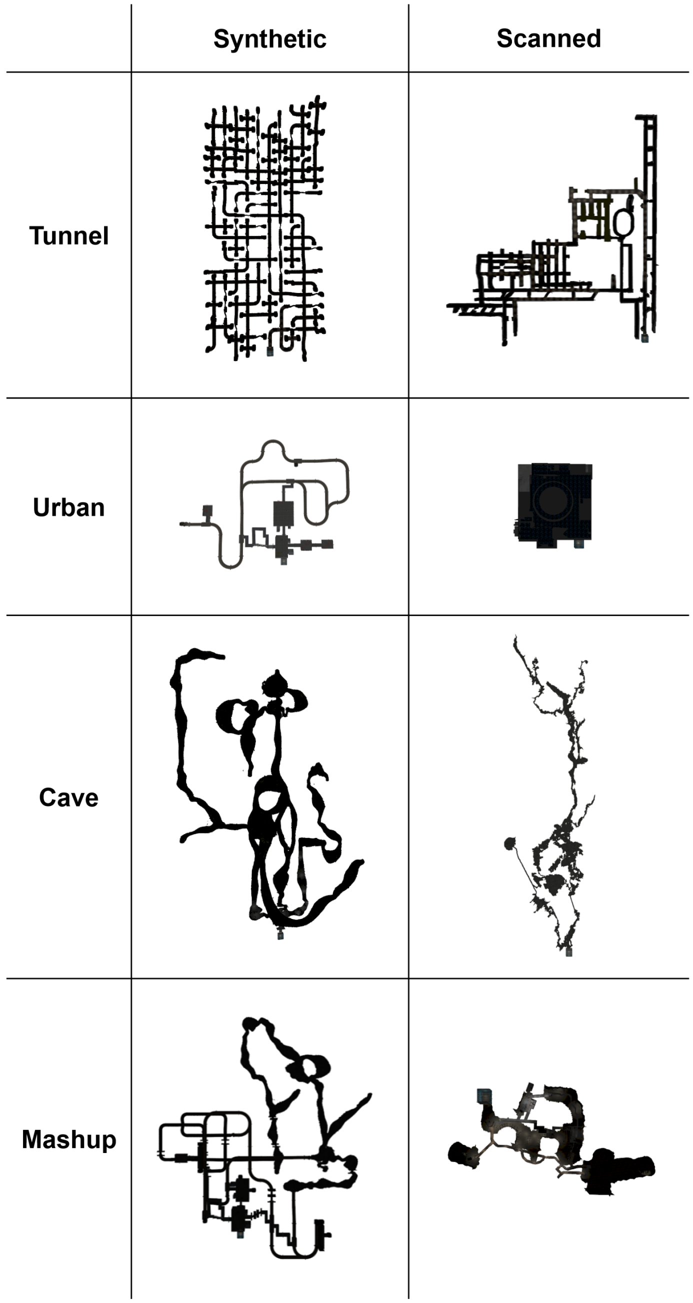


Figure 20. Virtual Competition worlds from the Prize Round of the Final Event.

edges. This assignment was designed to emulate the behavior of environmental features absorbing or scattering RF signals when radios are not in line-of-sight. The visibility cost was assigned as the sum of costs of edges in the path between any two nodes.

To improve communication range, competitors utilized deployable mesh radios, or breadcrumbs, which were incorporated into the communication model. When breadcrumbs were deployed, the cost of communication between robots was modified by the path of breadcrumbs between them. The visibility and range costs were assigned as the greatest cost between any single hop in the path, with an added distance penalty for the number of hops.

Table 4. Topology analysis of Virtual Competition worlds.

World	Type	Traversable Length (km)	Dynamic Obstacles (Count)	Simple Cycles (Count)	Wheeled-UGV Accessible (%)
Hyper Hashtag	Synthetic Tunnel	4.2	5	213	84
SubT Central	Synthetic Urban	1.4	4	44	100
Bear Claw	Synthetic Cave	3.0	5	39	34
Blast from the Past	Scanned Tunnel	3.7	8	76204	100
Gamma	Scanned Urban	2.0	2	155	33
Lantern Fish	Scanned Cave	2.5	0	718	4.3
Locomotive	Synthetic Mashup	3.2	5	180	97
SubT Lair	Scanned Mashup	1.0	3	36067	Unavailable

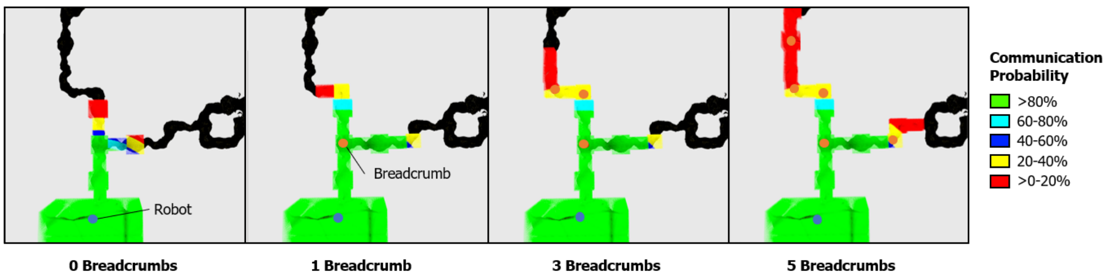


Figure 21. Demonstration of extension of virtual communications range using radio breadcrumbs.

Figure 21 visualizes the probability of successful communication in an area of a cave world before and after breadcrumb deployment. The colors represent the visibility cost to communicate between a robot in the Staging Area and a robot located at each colored point in space. When breadcrumbs were dropped intelligently to combat expected losses due to range and visibility, the effective communication range was extended.

Further details on the software implementation of the communication model are documented in the SubT Virtual Testbed Repository (2018).

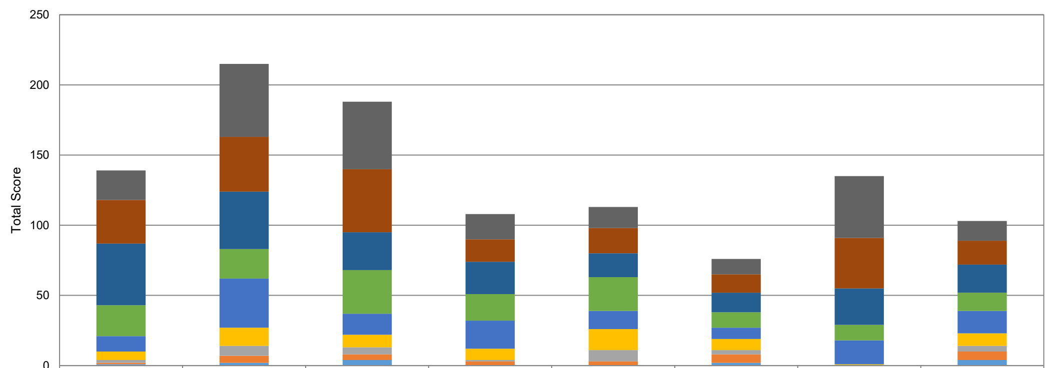
5.4.3. World Analysis

The metrics in Table 4 provide topology analysis of each Prize Round world. Simple cycles, nodes, and edges were conservatively estimated for synthetic worlds based on overall tile connectivity and did not include loops within individual tiles, so the number of cycles could be estimated to be higher depending on how loops were defined. The worlds varied in scale from one to over four kilometers in traversable length. In five worlds, the majority of the course was reachable with wheeled ground vehicles, although often via less efficient routes, such as by traveling a longer route with gradual elevation changes to avoid stairwells and vertical drops. The cave worlds required flight through steep or vertical shafts to reach a majority of the course.

Compared to the Systems Competition courses, the larger scale of the virtual worlds reflects the Virtual Competition's emphasis on the autonomy technology area. Although the virtual environments did not include high-fidelity gravel terrain or slip modeling, the vast expanse of traversable area and varied topology sufficiently challenged solutions' high-level decision making and multi-robot coordination.

5.4.4. World Design Outcomes

Figure 22 displays the distribution of scored points by world, summed across all Prize Round runs (three replications per team). Although eight of the nine teams found at least one artifact in each scan-based world, scores summed across synthetic worlds (1, 2, 3, and 7) were 69 % higher than



	(1) Hyper Hashtag	(2) SubT Central	(3) Bear Claw	(4) Blast from the Past	(5) Gamma	(6) Lantern Fish	(7) Locomotive	(8) SubT Lair
Dynamo	21	52	48	18	15	11	44	14
CTU-CRAS-NORLAB	31	39	45	16	18	13	36	17
Coordinated Robotics	44	41	27	23	17	14	26	20
BARCS	22	21	31	19	24	11	11	13
Robotika	11	35	15	20	13	8	17	16
SODIUM-24 Robotics	6	13	9	8	15	8	1	9
MARBLE	2	7	5	1	8	3	0	4
Flying Fitches	1	5	4	3	3	6	0	6
COLLEMBOLA	1	2	4	0	0	2	0	4

Figure 22. Scores in each Prize Round world across all Virtual Competition teams.

in scan-based worlds. Much of this difference can be attributed to navigation difficulties in the more complex worlds with rougher terrain. Teams' algorithms, such as artifact classifiers and local navigation, were trained and iterated largely in synthetic practice environments leading up to the Final Event. The resulting difference in performance across world types underscores the importance of testing solutions against a wide variety of environments, including those developed from real-world data sets.

6. Scenarios

The primary scenario of interest for the competition was providing rapid situational awareness to a small team of operators preparing to enter unknown and dynamic subterranean environments. Potential representative scenarios involve rescue efforts in collapsed mines, post earthquake search and rescue in urban underground settings, and cave rescue operations for injured or lost spelunkers.

In the preceding sections, we presented how the scoring function was intentionally designed to deliver actionable situational awareness, how the artifacts were designed to be operationally relevant and to motivate multi-modal sensing approaches, and how the competition courses were designed to have a high level of fidelity and relevance for evaluating the four technology areas of autonomy, perception, networking and mobility.

Each of these significantly contributed towards ensuring a high degree of realism, but on their own, are not enough to drive innovation that directly addresses the time sensitivity, small teams, or the unknown and dynamic environments described in the primary scenario of interest. In this section, we discuss the scenario design decisions, namely, the artifact placement, challenge elements, and dynamic obstacles.

6.1. Artifact Placement

More than any other consideration, artifact placement had the greatest impact on the competition due to the main scoring objective being the need to search for, detect, and provide spatially referenced locations of the artifacts. Artifacts were deliberately distributed throughout the competition course in a manner that rewarded teams that were able to rapidly explore and maneuver through more of the course elements. The placement of the artifacts was not known in advance of a run by competitors and was varied between rounds.

Through careful placement of artifacts, the course directly and indirectly rewarded teams for course coverage, difficult traversals, mapping quality, rapid exploration, multi-modal sensing, effective networking, multi-agent teaming, and exhaustive search algorithms. Several guidelines and motivations contributed to the final placement of artifacts in both the Systems Competition course and the Virtual Competition worlds.

- Place one of each artifact type close to the course entrance to give all teams a chance of demonstrating detection of all ten artifact types.
- Place subdomain-specific artifacts in their respective subdomains; place common artifacts in all three subdomains.
- Distribute artifacts throughout the course such that artifact reports can serve as a surrogate for course coverage and mapping quality.
- Place artifacts in areas that were farthest from the entrance to reward course coverage, rapid exploration, and effective networking to exfiltrate the data back to the Base Station.
- Place artifacts in realistic locations and orientations.
- Vary the vertical placement of artifacts to reward wide field of view perception approaches and exhaustive search of large spaces (e.g., vents located low to the ground, on walls, and on the ceilings).
- Reward significant traversals (e.g., steep incline, water hazard), especially UAV traversals (e.g., vertical shafts, mezzanine levels), and navigating dynamic obstacles (fog and dynamic obstacle).
- Use the new cube artifact as an "Easter egg" to reward reaching one of the interesting locations within each subdomain.
- Avoid repeating two artifacts of the same type from being immediately adjacent to each other.
- Place artifacts at least 10 meters apart to avoid reports from nearby artifacts unintentionally scoring another artifact.

Additional practical and logistical considerations were included for the Systems Competition.

- Skew the placement in the Preliminary Round closer to the entrance due to the shorter run times, and farther from the entrance for the Prize Round to increase overall difficulty.
- Avoid repeating locations from earlier rounds (especially same type) to prevent guesses from previous run locations awarding a point.
- Place artifacts within view of infrastructure camera locations for monitoring and production.
- Avoid placing RF-emitting artifacts too close to each other to prevent interference.
- Place powered artifacts (survivor, vent, gas, cube) close to power distribution boxes.
- Place active artifacts that may need monitoring or resetting (gas, cell phone) close to access panels.

Figures 23 and 24 show examples of artifacts as they were installed on the Systems and Virtual Final Event competition courses, respectively. Figure 25 shows the locations of the Prize Round artifacts for the Systems Competition, and Figures 26 and 27 show the locations of the Prize Round artifacts for the Virtual Competition.

The decision to place 40 artifacts for the Systems Competition Prize Round run was a tradeoff between practical limits of reconfiguring the course overnight and the desire to have a sufficient artifact density to appropriately reward teams for traversing significant sections of the course. With an approximately 1 km course length, 40 artifacts represent an average of 1 artifact per 25 m. The Preliminary Round runs had 20 artifacts on the course to reduce the burden of installing and resetting artifacts between the Preliminary Round runs. This also reduced the overall number of artifact locations that needed to be professionally surveyed to establish a reliable ground truth dataset (i.e., 80 total unique locations). The Prize Round included a higher artifact density than the Preliminary Round runs because it was more critical for teams to be rewarded for significant traversals with the prize money on the line. It would be disappointing for a team to navigate a

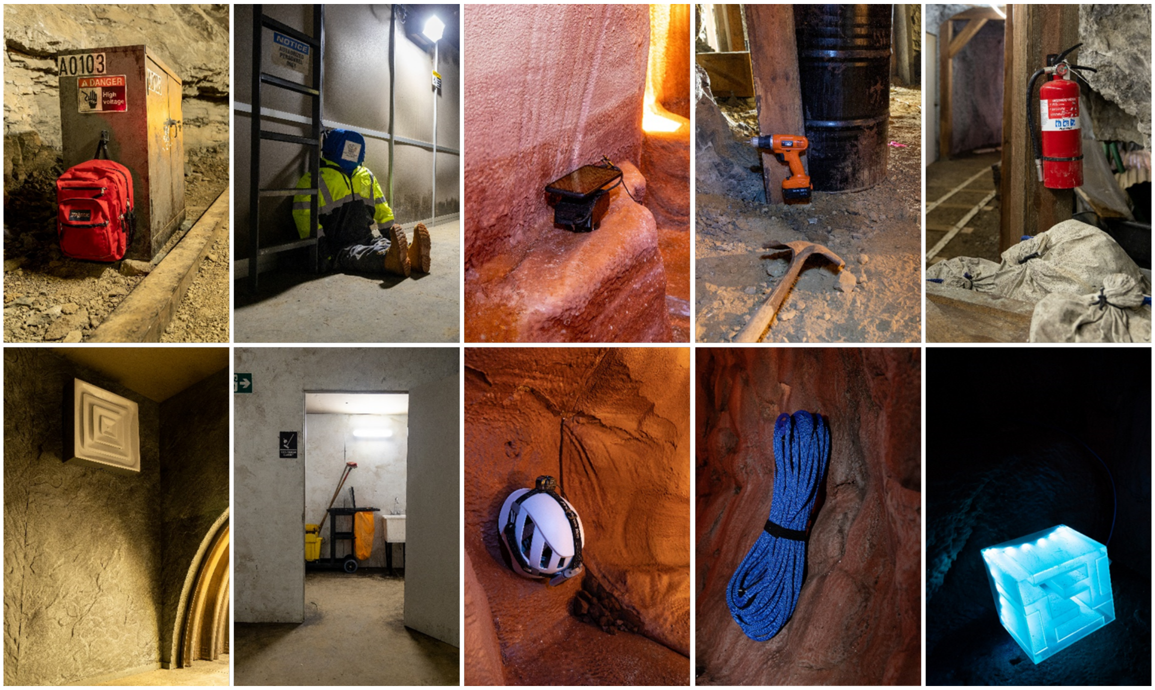


Figure 23. Examples of artifacts as installed within the Final Event course.

Figure 24. Examples of artifacts as installed within the Final Event virtual worlds.

Figure 25. Prize Round artifact locations for the Systems Competition.

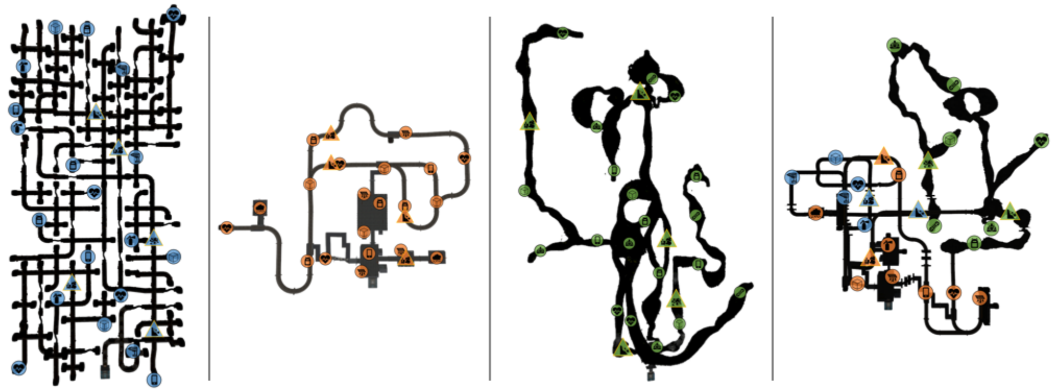


Figure 26. Prize Round artifact and obstacle placement in synthetic virtual worlds.

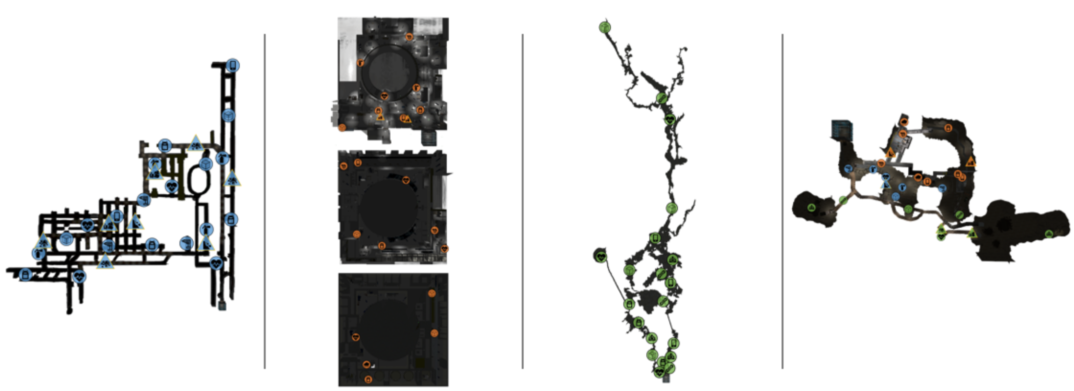


Figure 27. Prize Round artifact and obstacle placement in virtual worlds based on real underground environments.

vertical shaft or a challenging obstacle without being rewarded in the form of an artifact on the other side. More artifacts also placed a greater burden on the Human Supervisor to sort through all of the artifact detections to decide which reports to submit to the scoring server.

6.2. Challenge Elements

The Final Event competition courses were designed to drive innovation across the four technology areas by presenting teams with a range of technical challenge elements.

- **Austere Navigation**. Multiple levels, inclines, loops, dead-ends, slip-inducing terrain interfaces, and sharp turns.
- **Degraded Sensing**. Constrained passages to large openings, lighted areas to complete darkness, wet to dusty conditions, and scattering environments including fog, mist, and smoke.
- **Severe Communication**. Limited line-of-sight, radio frequency (RF) propagation challenges, and effects of varying geology.
- **Terrain Obstacles**. Mobility-stressing terrain features and obstacles including constrained passages, sharp turns, clutter, collapsed structures, large drops/climbs, inclines, steps, ladders, and mud, sand, and/or water.
- **Dynamic Elements**. Shifting terrain, falling debris, and/or other physical changes to the environment including atmospheric effects.
- **Endurance Limits**. Large-scale courses that require aggregated endurance of 60 minutes to be mission-relevant.

Figure 16 shows the distribution of some of the major challenge elements on the Final Event Systems Competition course.

6.3. Dynamic Obstacles

The course and virtual worlds also included dynamic elements to test the agility of the system autonomy to reason, react, and recover from the possibility of a changing map or changing environment. Both Systems and Virtual Competitions mirrored these dynamics obstacles to further highlight opportunities for comparison between physical and simulated solutions.

Fog Machines. Two fog machines were installed within the course and activated to change the environmental conditions during each run. Two methods were used to trigger the fog machines: an initial release of fog at run start and a gate-tripped release once a robot crossed a predetermined threshold within close proximity to the fog machine. The fog machine settings and durations were tuned to achieve a density that reduced visibility enough that navigation for a human was very difficult (<1 meter visibility).

In the Preliminary Round, the initial fog that was released at run start for each run was reduced because the shorter 30-minute duration of the runs meant that there was less time within a run for

Figure 28. Fog machine before (left) and after (right) being activated.

Table 5. Fog machine settings by round.

	Preliminary Round		Prize Round	
	Run Start	**Gate**	**Run Start**	**Gate**
Tunnel Fog	0.5 s	5.0 s	6.0 s	N/A
Urban Fog	1.0 s	6.0 s	8.0 s	N/A

Figure 29. Urban dynamic obstacle before (left) and after (right) being deployed.

Figure 30. Dynamic obstacles included fog (left), falling rocks (middle), and tunnel collapses (right).

fog to dissipate. Instead, gate triggers were used to activate the fog at higher levels only for teams that reached far enough into the course.

The Prize Round had an increased emphasis on consistency across teams due to the final ranking being based solely on the Prize Round scored run. To ensure consistency, fog was only released at run start, but at much higher durations. Large fans were used between runs to clear out any remaining fog via access panels strategically located close to the areas where fog was released.

Ceiling Collapse. New to the Final Event Systems Competition were the ceiling collapse dynamic obstacles which created mobility blockages by dropping simulated debris from the ceiling. Three dynamic obstacles, one in each subdomain, were placed on the course and were designed to blend into their respective environments. The dynamic obstacles relied on a robot to pass below the obstacle before being triggered. This trigger system was implemented using the emergency stop transponders and two transponder gates that were located on either side of the dynamic obstacles. Passing under one of the two gates would arm the obstacle and passing the other gate would then trigger the collapse. In the Prize Round, all three dynamic obstacles were reached and triggered, although no one team triggered all three dynamic obstacles within the same run.

Dynamic obstacles were also modeled in the Virtual Competition to challenge solutions' abilities to sense and adapt. Each obstacle was activated by entry or exit of a robot model within a detection volume surrounding the obstacle model. The following three categories of obstacles, each illustrated in Figure 30, were placed in the virtual worlds.

Virtual Fog. Emitters of obscurant particles were placed to emulate foggy conditions and were activated continuously while at least one robot was inside the detection region. The particles obscured visual sensors to an approximate visibility of four meters. Approximately 65% of lidar and depth sensor returns were scattered using a white noise model between emitted particle locations and true returns.

Virtual Terrain. Dynamic falling rocks were modeled to challenge local navigation and traversability planning by ground robots. Rocks of varied sizes, all less than one meter diameter, fell within ten meters in front of a robot entering the detection region. The behavior could be triggered up to five times, causing obstacle density to increase upon successive robot entries. As such, the terrain became more difficult to navigate with each robot passage through the area and required local planners to react.

Virtual Collapses. Passageways were blocked by simulated ceiling collapses in order to challenge dynamic global navigation and multi-robot coordination. Three collapse types were modeled with mining debris, subway tunnel debris, and large boulders for tunnel, urban, and cave subdomains, respectively. The collapses were triggered to appear after the first robot had entered and exited the detection region, blocking the robot's return path and subsequent robots' passage through the area. For the cave-themed collapse, boulders blocked ground-based traversal while allowing aerial robots to fly over the debris. The other two obstacle types blocked the entire passage from any type of traversal.

Collapse obstacles were often placed to obstruct major passageways, but locations were carefully selected to always include alternate unblocked path(s) connecting the Staging Area to either side of the obstacle. Navigation solutions were thus forced to adapt by finding different routes to either return to communications range or send additional robots to explore the area.

6.4. Systems Course Reconfiguration

A key element of the SubT Challenge was that the subterranean environment was *a priori* unknown to the competitors, requiring competitors to develop technologies to rapidly and remotely explore in the face of such lack of prior knowledge. For the Systems Competition, course reconfiguration between runs served to limit the viability of using prior knowledge and provided an opportunity to calibrate the difficulty of the course between competition days.

- Artifact Locations. The most important aspect of reconfiguration was moving all of the artifacts to new locations and never reusing an artifact location.
- Obstacles and Terrain. Figures 31 and 32 show the modifications made to the course between the Prelim 1 and Prelim 2 rounds as a means to change the course and fine-tune the difficulty based on how teams were performing.
- Environmental Conditions. Figure 33 shows the lighting distribution across the competition course and locations where lighting was removed between the Preliminary Round and Prize Round to increase the difficulty of the course. The fog machine settings were also adjusted for the Prize Round to increase the difficulty for all teams.

7. Community Contributions

The SubT Challenge offered a unique opportunity to generate and curate data products and resources for the field robotics community's use of open datasets to develop common benchmarks, support reproducible robotics research, and collectively enhance the integration and evaluation of relevant technologies including through the commitment and release of open-source software tools and testbeds. Several resources have been publicly released and are available on the various SubT Challenge resource outlets and repositories, as described in detail below.

7.1. Ground Truth

In an effort to provide broader opportunities to leverage the unique access to the SubT Challenge-enhanced test environments, DARPA made substantial efforts to capture high-quality data relevant

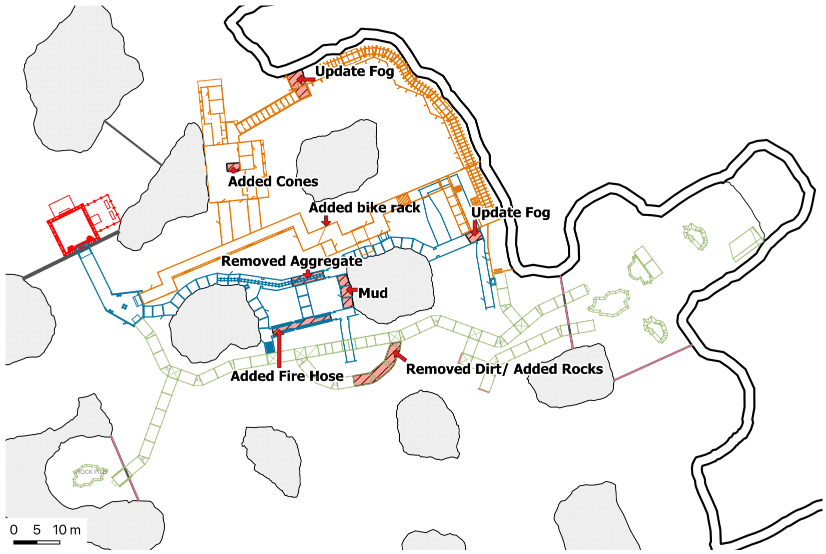


Figure 31. Configuration changes made between the Prelim 1 and Prelim 2 runs.

Figure 32. Pictures of Configuration changes: (top row) mud, bike rack, fire hose, (bottom row) rocks, cones, and removing gravel from between the rails to increase their height.

to (a) the physical environments themselves via 3D lidar-based scans, videos, and immersive scenes, as well as (b) the SubT Challenge competition scenarios, artifact locations, and configurations, for further testing or validation. These extensive ground truth data products benefited from survey-grade measurements and high-precision scans, offering a rich and validated foundation for future research endeavors.

Available ground truth scan data include the Tunnel Circuit, Urban Circuit, Final Event, and all Virtual Competition environments (SubT Ground Truth Datasets, 2021). Each of these respective repositories include (where available) the following.

- Spreadsheet listing each artifact; its type; and its (x, y, z) location in the relevant DARPA coordinate frame for each course and configuration;
- Spreadsheet listing reference frame fiducial coordinates for each course;
- Map of artifact locations and associated artifact types for each course configuration;
- Survey-grade high resolution 3D point cloud scans of the courses;

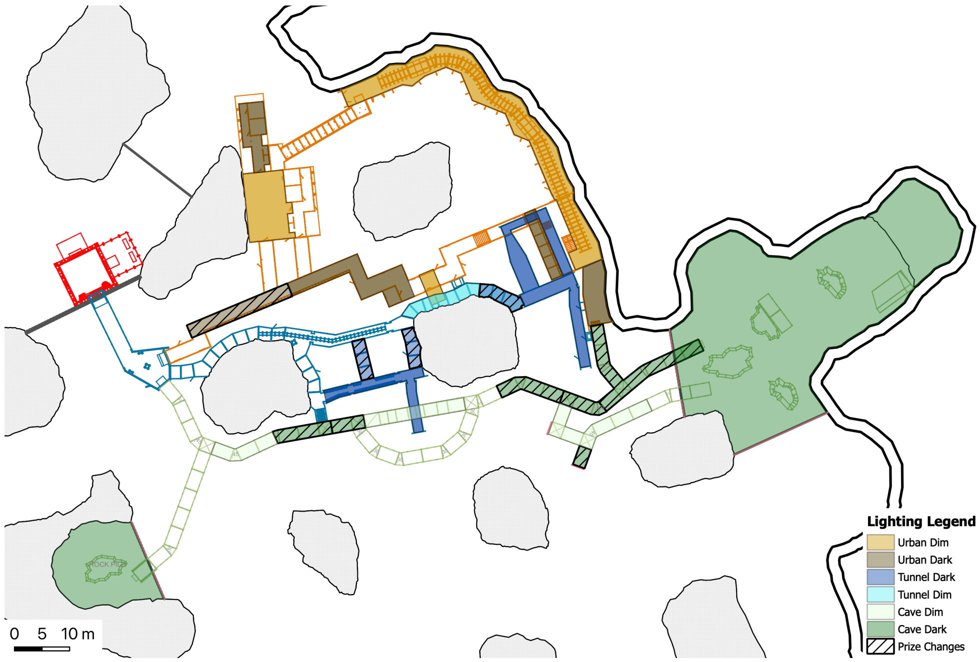


Figure 33. Lighting levels and changes made between Preliminary Round and Prize Round.

- Virtual flythroughs of the point-cloud data for each course;
- Course walkthrough videos showing the courses and relevant challenge elements; and
- Virtual tours of each course based on data collected with a Matterport™scanner.

7.2. SubT Reference Datasets

Subterranean environments present a starkly different set of environmental characteristics than those found in typical SLAM datasets: poor to no lighting, varied levels of roughness and irregularity in structure, sometimes significant changes in topography, wetness, dirtiness, and no access to GPS. The SubT Reference Datasets Repository includes datasets collected from the competition courses using robotic platforms equipped with an array of onboard sensors and running relevant robotic mapping and localization algorithms, along with analysis tools intended for benchmarking SLAM algorithms in subterranean environments (SubT Reference Datasets Repository, 2021).

In addition to the datasets, Rogers et al. (2020) provides detailed descriptions of the data collection procedures, proposes an absolute-accuracy analysis metric for map evaluation, provides a set of open-source support tools to evaluate mapping approaches against this metric, and presents a baseline comparison of common SLAM algorithms.

7.3. SubT Virtual Testbed

To support the efforts in both competitions, DARPA developed the SubT Virtual Testbed, which includes an extensible Ignition Gazebo-based simulation environment, automated testing and assessment tools, and associated software support infrastructure provided as government-furnished equipment (GFE). The SubT Virtual Testbed was developed to emulate as many of the same subterranean environments and characteristics as were presented in the physical competition as realistically as possible. Upon initial announcement of core elements ahead of the Challenge Kickoff at Competitors Day, frequent releases incorporated continuous integration of updates and enhanced features to the SubT Virtual Testbed. DARPA made the SubT Virtual Testbed and

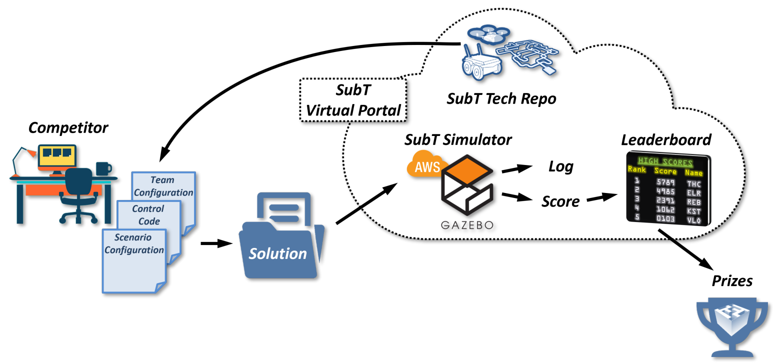


Figure 34. The SubT Virtual Testbed, comprising the SubT Simulator, SubT Virtual Portal, and the SubT Tech Repo, offers integrated cloud-based robotics simulation testbed infrastructure.

associated resources available publicly, including extensive documentation and tutorials. All SubT Challenge competitors, including those in the Systems Competition, were encouraged to leverage these cloud-based simulation and virtual environment resources to enhance and accelerate their technology development efforts.

Included in the SubT Virtual Testbed is the SubT Virtual Portal, which is the web-based front-end, and the SubT Simulator, which is the cloud-enabled scalable simulation capability. Altogether, the SubT Virtual Testbed represents significant digital infrastructure investments which can be tailored to advance robotics and autonomy development beyond subterranean scenarios.

More details, as well as source code, documentation, and tutorials, can be found at the open-source SubT Virtual Testbed Repository (2018).

8. Concluding Remarks

In addition to advancing the state of the art throughout its duration, the DARPA Subterranean Challenge itself represented an innovative approach to inspiring robotics technology breakthroughs. The overall structure of the program offered a macroscale cadence that incentivized iterative improvement over time, and the careful mission-focused design of the competition drove teams to deliver both quantifiable and operationally impactful technologies as a result. Further, the use of both Systems and Virtual Competitions provided a unified approach to gather valuable insights helping to span the immense design space for resilient multi-robot teams in underground settings. With these contributions in mind, this paper detailed the purposeful inspiration and meticulous design that went into creating and implementing the DARPA SubT Challenge competition.

Key insights stemming from the design process of the SubT Challenge revealed how intricately dependent the effectiveness of real-world solutions are on the formulation of the mission objective, the nature of environmental settings, and the specific details of any given scenario. Recognizing the near-impossibility of testing all variations in an exhaustive fashion, the design of the competition rules, metrics, and courses aimed to coherently and collectively stimulate all tech areas–autonomy, networking, mobility, perception—to elicit robust solutions to the problem. The result of this principled, albeit resource-intensive, competition design approach included measurable technological innovations and enhanced operational knowledge within the field robotics community, traceability of systems solution performance to mission-relevant requirements, and a foundational approach for testing new advances in integrated robotics technologies.

Based on the presented lessons learned, there are a number of avenues for future study and enhancements. As described, the SubT Challenge intentionally sought solutions where no prior knowledge about the course(s) was available, with the Final Event competition design combining all three subdomains into an integrated course further limiting possible advanced planning. One potential effort would be to more deeply investigate how additional information about the environment, such as average cross-sectional areas or general graph topological statistics, being available in advance would potentially influence the design and deployment strategies of the robot teams.

Another compelling research area is the development of a rigorous framework for optimizing the composition of heterogeneous robot teams; such work would require a richer and more analytic understanding of how sophisticated teams of robots collaborate in the context of complex and dynamic scenarios, well beyond the task allocation algorithms and alliance formation approaches available today.

In the presented competition context, the role of human teammates was intentionally constrained, with only one Human Supervisor permitted for the Systems Competition and none at all for the Virtual Competition. While various approaches for user interfaces and interactions were discovered and refined throughout the competition, there is significant opportunity to further investigate human-multi-robot teaming paradigms specifically focused on management and execution of complex time-sensitive tasks amidst complex operating environments. In addition to enhanced decision support tools more effectively integrating machine-learned perception or robot behavior models, exploration of interactive A.I. “copilots” and rapid multi-modal data summarization pipelines may dramatically increase mission performance through these future A.I.-powered interfaces.

Finally, in the context of enhancing how robotic systems are tested and assessed, a future investment could streamline the collection and integration of competitor-generated data (as opposed to those data captured only from DARPA’s instrumentation), which would directly enable computation of additional metrics not directly measurable from the challenge infrastructure alone. Given the myriad of alternative evaluation metrics, such as described in Section 2.3, additional data could be more explicitly requested or required to be provided by competitors to support deeper performance assessments leveraging these integrated data.

The DARPA Subterranean Challenge represented a continuation of DARPA’s legacy of using Grand Challenge competitions to spur innovation in robotics. As described in this paper, the SubT Challenge is particularly notable for its unique and innovative approach to tackling and quantifiably advancing core technology challenges in robotics for real-world operations in complex environments through its iterative and deliberate design of the competition. Though the full extent of its impact may continue to emerge in the years to come, the DARPA Subterranean Challenge has undoubtedly advanced the state of field robotics in the near-term and has energized a resurgent field robotics community (Figure 35) to usher in a new wave of robotics innovation.

Figure 35. The DARPA Subterranean Challenge: inspiring the next generation of field robotics innovation

Acknowledgments

The authors would like to thank everyone who contributed to the vision and execution of the SubT Challenge including the other competition leads; Ryan Halterman, Margaret Rowland, Elissa Rupley, and DJ Tyree; and the many team members across the following organizations; Defense Advanced Research Projects Agency (DARPA), Booz Allen Hamilton, Strategic Analysis, Inc., NIWC Atlantic, NIWC Pacific, DEVCOM Army Research Laboratory, Hargrove, Open Robotics, Big Picture, and Next/Now.

ORCID

Viktor L. Orekhov https://orcid.org/0000-0002-0626-9137
Angela C. Maio https://orcid.org/0000-0002-8141-0129
Roshan P. Daniel https://orcid.org/0000-0002-1147-790X
Timothy H. Chung https://orcid.org/0000-0001-7959-3555

References

Ackerman, E. (2022). Robots conquer the underground: What darpa's subterranean challenge means for the future of autonomous robots. *IEEE Spectrum*, 59(5):30–37.

Amigoni, F., Bastianelli, E., Berghofer, J., Bonarini, A., Fontana, G., Hochgeschwender, N., Iocchi, L., Kraetzschmar, G., Lima, P., Matteucci, M., et al. (2015). Competitions for benchmarking: Task and functionality scoring complete performance assessment. *IEEE Robotics & Automation Magazine*, 22(3):53–61.

Artifact Specification Guide (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/official_docs/blob/main/SubT_Finals_Artifacts_Specification.pdf.

Brancalião, L., Gonçalves, J., Conde, M. Á., and Costa, P. (2022). Systematic mapping literature review of mobile robotics competitions. *Sensors*, 22(6).

Choi, H., Crump, C., Duriez, C., Elmquist, A., Hager, G., Han, D., Hearl, F., Hodgins, J., Jain, A., Leve, F., Li, C., Meier, F., Negrut, D., Righetti, L., Rodriguez, A., Tan, J., and Trinkle, J. (2021). On the use of simulation in robotics: Opportunities, challenges, and suggestions for moving forward. *Proceedings of the National Academy of Sciences*, 118(1):e1907856118.

Courchesne, A. (2021). On quantifying the value of simulation for training and evaluating robotic agents. Master's thesis, Université de Montréal.

Dias, J., Althoefer, K., and Lima, P. U. (2016). Robot competitions: What did we learn? *IEEE Robotics & Automation Magazine*, 23(1):16–18.

Finals Course Callouts (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/official_docs/blob/main/Finals_Course_Callouts.pdf.

Gazebo API (2020). Retrieved April 6, 2022, from https://gazebosim.org/api/gazebo/3.2/pointcloud.html.

Hudson, N., Talbot, F., Cox, M., Williams, J., Hines, T., Pitt, A., Wood, B., Frousheger, D., Surdo, K. L., Molnar, T., Steindl, R., Wildie, M., Sa, I., Kottege, N., Stepanas, K., Hernandez, E., Catt, G., Docherty, W., Tidd, B., Tam, B., Murrell, S., Bessell, M., Hanson, L., Tychsen-Smith, L., Suzuki, H., Overs, L., Kendoul, F., Wagner, G., Palmer, D., Milani, P., O'Brien, M., Jiang, S., Chen, S., and Arkin, R. C. (2022). Heterogeneous ground and air platforms, homogeneous sensing: Team csiro data61's approach to the darpa subterranean challenge. *Field Robotics*, 2:595–636.

Interface Control Document (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/official_docs/blob/main/SubT_Challenge_Finals_ICD.pdf.

Lu, C.-L., Huang, J.-T., Huang, C.-I., Liu, Z.-Y., Hsu, C.-C., Huang, Y.-Y., Huang, S.-C., Chang, P.-K., Ewe, Z. L., Huang, P.-J., Li, P.-L., Wang, B.-H., Yim, L.-S., Huang, S.-W., Bai, M. R., and Wang, H.-C. (2022). A heterogeneous unmanned ground vehicle and blimp robot team for search and rescue using data-driven autonomy and communication-aware navigation. *Field Robotics*, 2:557–594.

Mine Rescue Teams (2022). 30 C.F.R. §49. https://www.ecfr.gov/current/title-30/chapter-I/subchapter-H/part-49.

Montgomery, D. (2022). The Pentagon's $82 million super bowl of robots. Retrieved April 6, 2022, from https://www.washingtonpost.com/magazine/2021/11/10/darpa-robot-competition/.

Ohradzansky, M. T., Rush, E. R., Riley, D. G., Mills, A. B., Ahmad, S., McGuire, S., Biggie, H., Harlow, K., Miles, M. J., Frew, E. W., Heckman, C., and Humbert, J. S. (2022). Multi-agent autonomy: Advancements and challenges in subterranean exploration. *Field Robotics*, 2:1068–1104.

Orekhov, V. L. and Chung, T. H. (2022). The darpa subterranean challenge: A synopsis of the circuits stage. *Field Robotics*, 2:735–747.

Palmer, A. N. (1991). Origin and morphology of limestone caves. *GSA Bulletin*, 103(1):1–21.

Piazza, E., Lima, P. U., and Matteucci, M. (2022). Performance models in robotics with a use case on SLAM. *IEEE Robotics and Automation Letters*, 7(2):4646–4653.

Rogers, J. G., Gregory, J. M., Fink, J., and Stump, E. (2020). Test Your SLAM! The SubT-Tunnel dataset and metric for mapping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 955–961. IEEE.

Scherer, S., Agrawal, V., Best, G., Cao, C., Cujic, K., Darnley, R., DeBortoli, R., Dexheimer, E., Drozd, B., Garg, R., Higgins, I., Keller, J., Kohanbash, D., Nogueira, L., Pradhan, R., Tatum, M., Viswanathan, V. K., Willits, S., Zhao, S., Zhu, H., Abad, D., Angert, T., Armstrong, G., Boirum, R., Dongare, A., Dworman, M., Hu, S., Jaekel, J., Ji, R., Lai, A., Lee, Y. H., Luong, A., Mangelson, J., Maier, J., Picard, J., Pluckter, K., Saba, A., Saroya, M., Scheide, E., Shoemaker-Trejo, N., Spisak, J., Teza, J., Yang, F., Wilson, A., Zhang, H., Choset, H., Kaess, M., Rowe, A., Singh, S., Zhang, J., Hollinger, G. A., and Travers, M. (2022). Resilient and modular subterranean exploration with a team of roving and flying robots. *Field Robotics*, 2:678–734.

SubT Challenge Competition Rules Final Event (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/official_docs/blob/main/SubT_Challenge_Finals_Rules.pdf.

SubT Challenge Virtual Competition: Cloud-hosted Simulation (2020). Retrieved April 6, 2022, from https://youtu.be/n3PdWObDBEw.

SubT Ground Truth Datasets (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/subt_resources.

SubT Reference Datasets Repository (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/tunnel_urban_reference_datasets.

SubT Tech Repo (2018). Retrieved April 6, 2022, from https://subtchallenge.world/openrobotics/fuel/collections/SubT%20Tech%20Repo.

SubT Virtual Testbed Repository (2018). Retrieved April 6, 2022, from https://github.com/osrf/subt.

Test Mapping Server (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/test_mapping_server.

Test Scoring Server (2021). Retrieved April 6, 2022, from https://github.com/subtchallenge/test_scoring_server.

Tranzatto, M., Mascarich, F., Bernreiter, L., Godinho, C., Camurri, M., Khattak, S., Dang, T., Reijgwart, V., Löje, J., Wisth, D., Zimmermann, S., Nguyen, H., Fehr, M., Solanka, L., Buchanan, R., Bjelonic, M., Khedekar, N., Valceschini, M., Jenelten, F., Dharmadhikari, M., Homberger, T., De Petris, P., Wellhausen, L., Kulkarni, M., Miki, T., Hirsch, S., Montenegro, M., Papachristos, C., Tresoldi, F., Carius, J., Valsecchi, G., Lee, J., Meyer, K., Wu, X., Nieto, J., Smith, A., Hutter, M., Siegwart, R., Mueller, M., Fallon, M., and Alexis, K. (2022). Cerberus: Autonomous legged and aerial robotic exploration in the tunnel and urban circuits of the darpa subterranean challenge. *Field Robotics*, 2:274–324.

U.S. Dept. of the Army (2019). *Subterranean Operations (ATP 3-21.51)*. Washington, DC, USA. [Online] https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN19656_ATP%203-21x51%20%20FINAL%20WEB.pdf.

How to cite this article: Orekhov, V. L., Maio, A. C., Daniel, R. P., & Chung, T. H. (2023). Inspiring Field Robotics advances through the design of the DARPA Subterranean Challenge. *Field Robotics, 3*, 560–604.

Publisher's Note: Field Robotics does not accept any legal responsibility for errors, omissions or claims and does not provide any warranty, express or implied, with respect to information published in this article.

9. Appendix

9.1. Competition Documents

The official competition documents listed below are available in the following repository: https://github.com/subtchallenge/official_docs

Competition Rules: The intent of the SubT Challenge Competition Rules document was to provide participants guidance on competition design and scoring objectives to inform their development efforts in preparation for the Final Event. Final rules were released three months before each event.

Operations Guide: This SubT Challenge Operations Guide was intended for Systems Competition teams who qualified to participate in the SubT Challenge Final Event. The purpose of this document was to provide details regarding (1) competition activities, (2) event logistics and operations, and (3) supporting information to ensure a safe and successful event.

Qualification Guide: This document describes the qualification guidelines and submission instructions for the DARPA Subterranean (SubT) Challenge.

Transponder and Emergency (E-stop) Stop Guide: The purpose of this document was to provide teams with specifications and integration guidelines for the DARPA-provided Tier 2 E-Stop receiver in line with the requirements in the rules. The Tier 2 E-Stop helped to ensure that platforms could be brought to a halt so that DARPA personnel could safely enter the course to recover the platforms after the run.

Interface Control Document (ICD): The intent of this document was to convey the overall concept of operations for interaction with the Command Post (Scoring Server and Map/Telemetry Server) during competition and to describe the hardware and software interfaces necessary to successfully interact with the servers.

Artifact Specification Guide: This document provided specifications for the artifacts used in the DARPA Subterranean (SubT) Challenge Final Event. An "artifact" was an object or feature of interest that are commonly found in subterranean environments.

Simulation Model Preparation Guide: This document describes the pathway for innovative simulation assets to be contributed to the SubT Tech Repo and incorporated into the SubT Challenge Virtual Competition events.

9.2. Video Resources

SubTv: Subterranean Challenge-branded content streamed on DARPA's YouTube Channel throughout the Final Event. The content allowed viewers to experience the highlights from each day as one would experience a sports broadcast, with updated scoring, produced video packages, team interviews, and expert commentary. The videos listed below are available in the following Subterranean Challenge playlist: https://www.youtube.com/playlist?list=PL6wMum5UsYvYpbhQALOcbhzXYTt3qnzqA

- Final Event - Virtual Competition Preliminary Round
- Final Event - Day 1 - Introduction to the SubT Challenge
- Final Event - Day 2 - Competition Coverage
- Final Event - Day 3 - Competition Coverage
- Final Event - Day 4 - Prize Round Coverage
- Final Event - Day 4 - Awards Ceremony and SubT Summit

SubT Challenge Playlist: Additional Content produced during the Circuits Stage and in advance of the Final Event is available in the Subterranean Challenge playlist: Subterranean Challenge YouTube Playlist.

Finals Course Videos:

- Finals Course Point Cloud Flythrough - Tunnel Segment
- Finals Course Point Cloud Flythrough - Urban Segment
- Finals Course Point Cloud Flythrough - Cave Segment
- Tunnel Course 360° Flythrough
- Urban Course 360° Flythrough
- Cave Course 360° Flythrough

Highlights and Recaps:

- DARPA Subterranean Challenge - Final Event Wrap-Up
- DARPA Subterranean Challenge - Final Event Compilation
- DARPA Subterranean Challenge - Mapping
- DARPA Subterranean Challenge Final Event - Virtual Competition Preliminary Round

Prize Round Run Videos:

- Team CERBERUS Final Event Full Run
- Team CSIRO Data61 Final Event Full Run
- Team MARBLE Final Event Full Run
- Team Explorer Final Event Full Run
- Team CoSTAR Final Event Full Run
- Team CTU-CRAS-NORLAB Final Event Full Run
- Team Coordinated Robotics Final Event Full Run
- Team Robotika Final Event Full Run