UAS Service Supplier Network Performance

Results and Analysis from Flight Testing Multiple USS Providers in NASA's TCL4 Demonstration

Joseph L. Rios
Irene Smith
Daniel Mulfinger
Brandon Cook
NASA Ames Research Center, Moffett Field, California

Priya Venkatesan
Vijay Baskaran
Madhavrao Lakshminarasimha
Abraham Ishihara
Daniel Liddell
Qiang Li
SGT Inc., Moffett Field, California

Punam Verma
Universities Space Research Association, Moffett Field, California

David Smith
Sherry Jurcak
Logyx LLC, Moffett Field, California

Hemil Modi
Science and Technology Corp., Moffett Field, California

January 2020

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.
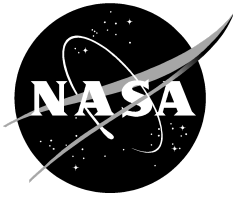
For more information about the NASA STI program, see the following:

Access the NASA STI program home page at http://www.sti.nasa.gov

E-mail your question to help@sti.nasa.gov

Phone the NASA STI Information Desk at 757-864-9658

Write to:
NASA STI Information Desk
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

UAS Service Supplier Network Performance

Results and Analysis from Flight Testing Multiple USS Providers in NASA's TCL4 Demonstration

Joseph L. Rios
Irene S. Smith
Daniel G. Mulfinger
Brandon M. Cook
NASA Ames Research Center, Moffett Field, California

Priya Venkatesan
Vijayakumar Baskaran
Madhavrao Lakshminarasimha
Abraham K. Ishihara
Daniel L. Liddell
Qiang Li
SGT Inc., Moffett Field, California

Punam Verma
Universities Space Research Association, Moffett Field, California

David R. Smith
Sherry Jurcak
Logyx LLC, Moffett Field, California

Hemil Modi
Science and Technology Corp., Moffett Field, California

Acknowledgments

This report is available in electronic form at

https://www.sti.nasa.gov

Overview

This report summarizes the performance of Unmanned Aircraft System (UAS) Service Suppliers (USS) in the Technical Capability Level 4 (TCL4) flight test performed by NASA and its partners in support of the UAS Traffic Management (UTM) concept. TCL4 is the final in a series of TCL demonstrations of a traffic management system for small UAS (sUAS). All demonstrations have been executed in collaboration with industry partners.

The [FAA 2018] UTM Concept of Operations document describes UTM as:

> ...the manner in which the FAA will support operations for predominantly sUAS operating in low altitude airspace. UTM utilizes industry's ability to supply services under FAA's regulatory authority where these services do not currently exist. It is a community-based traffic management system, where the Operators are responsible for the coordination, execution, and management of operations, with rules of the road established by FAA. UTM is designed to support the demand and expectations for a broad spectrum of operations with ever-increasing complexity and risk.

UTM should be considered a collection of services rather than a monolithic application. The following section describes the architecture at a high level. For further insight, the FAA's concept document [FAA 2018] or NASA's earlier concept publication [NASA 2016] should be consulted.

UTM Operational Architecture

The proposed system for managing small UAS at low altitude is much different than traditional air traffic management systems for conventional commercial airspace operations in the US. Within UTM, the operations are managed collaboratively by a collection of USSs. Each small UAS operation is managed by a USS to provide appropriate operational data to other USSs and to the operator. These data exchanges are in support of the Concept of Operations for UTM and defined by a set of Application Programming Interfaces and the UAS Service Supplier Specification [Rios 2019b]. The Air Navigation Service Provider (ANSP) still maintains authority over the airspace, but certain services are delegated to the USSs so that the ANSP does not directly manage this class of operations in nominal cases. This allows for scalability of the system and leverages industry's ability to innovate and provide services. There are other requirements that may be levied on the operator and the vehicles within UTM, but this paper focuses on the data exchange infrastructure of UTM.
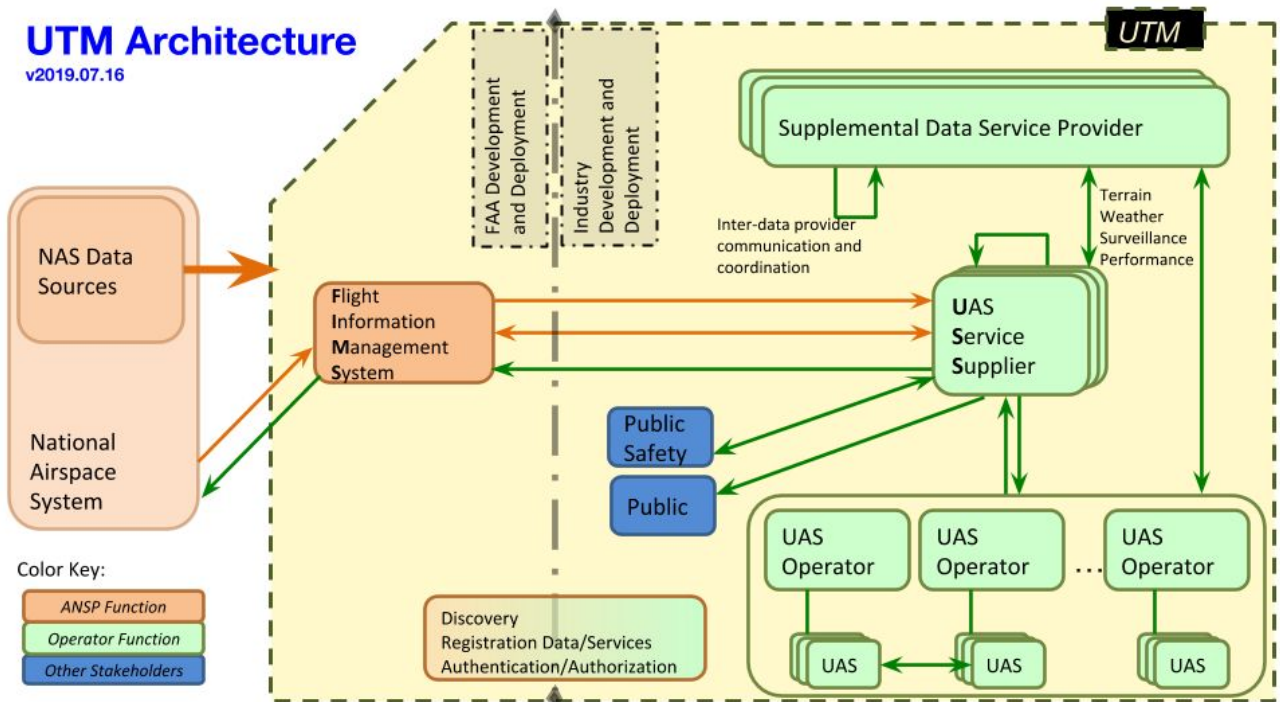
Figure 1: UTM Application Programming Interfaces

NASA UTM Data Collection Architecture

The UTM data collection architecture is the result of lessons learned from previous TCLs as well as industry best-practices as applied to UTM testing. There are two primary flows of data into NASA from the stakeholders involved in testing. First there is the data exchanged as part of the operational concept. These would include data such as operation plans, positions, and other UTM messages. All of these data were collected live with a designated USS Data Collector (UDC) that interacts with the other USSs as if it were a USS. The other flow of data originates from NASA's Data Management Plan (DMP) for TCL4, and includes data that are used for research purposes and are not necessarily part of the UTM operational concept. These data would include elements such as latency measurements, vehicle logs, weather data, and others. These data were submitted to the separate DMP system during or after testing.

Both flows of data end up in a Universal Data Store (UDS). The UDS within this architecture provides a single source of truth for all TCL4 data, implementing a key architectural pattern in data collection and analysis. UDS is used for all visualization and analysis of TCL4 testing data. Figure 2 summarizes this data collection architecture.
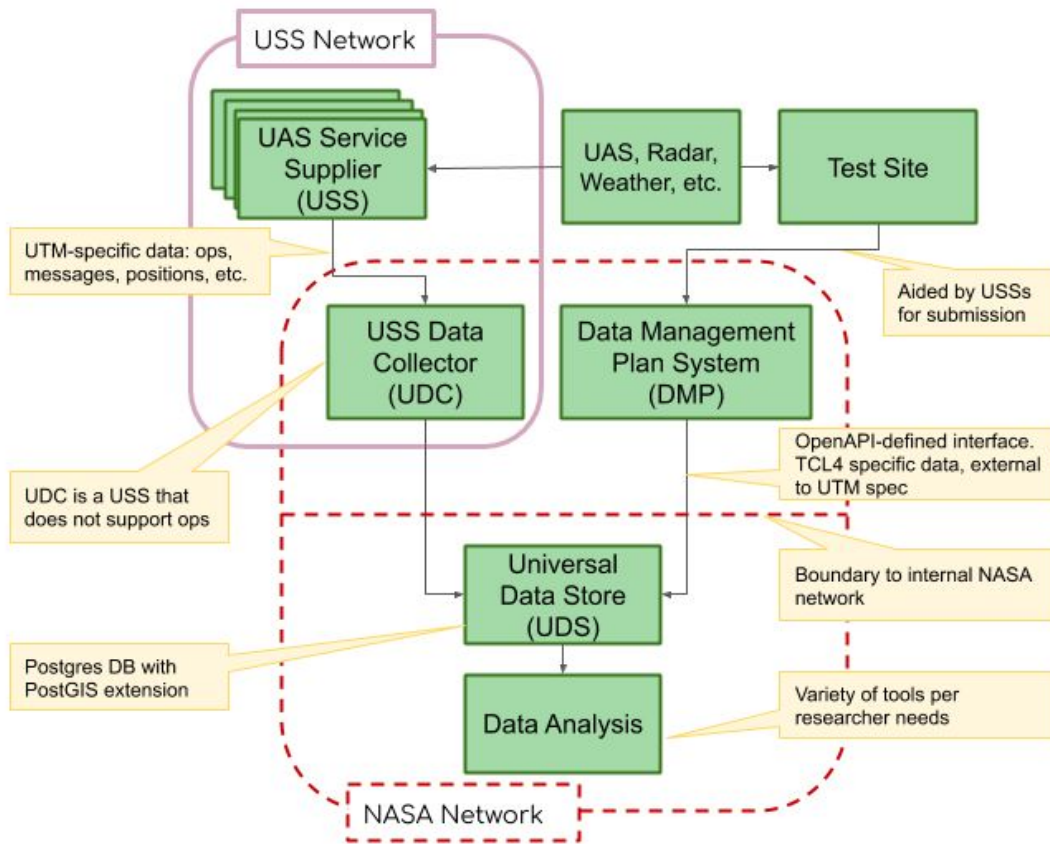
Figure 2. TCL4 data management architecture.

TCL4 Flight Test Summary

The overall planning and execution of TCL4 will be described in a future NASA document. A brief summary is provided here for context within the rest of the document.

The TCL4 Demonstration was the final in a series of demonstrations by the NASA UTM Project. Each demonstration from TCL1 to TCL4 increased in the number of system capabilities and the complexity of the operational environment. Overall, the UTM Project sought to demonstrate in concert with the FAA and industry how UTM should work and to investigate areas in need of further research. UTM is such a paradigm shift in how the airspace will be managed that often the correct questions to ask were not easily known in advance. The TCL Demonstrations helped shape the architecture and concept of UTM and aided in the formulation of the requirements and gaps.

TCL4 was executed at two of the FAA-designated UAS Test Sites: Nevada and Texas, managed by the Nevada Institute for Autonomous Systems and Lone Star UAS Center of Excellence & Innovation, respectively. Both sites were provided the same statement of work (SOW) and each developed a series of test plans to meet the SOW. These plans were finalized in coordination with the NASA UTM Project.

The test sites developed a team of partners to execute the flight testing. These included USS providers, UAS equipment manufacturers, weather service providers, cellular service providers, radar providers, operators, public safety agencies, and others. The SOW described five detailed scenarios with specific characteristics and test events that were considered the primary requirements of the flight test.

A key set of partners were the USS implementers. This group represents key stakeholders in the future of the UTM System. Given the primacy of the USSs in UTM, NASA developed a series of collaborative simulations and checkout exercises that each USS needed to complete successfully in order to be allowed to participate in the TCL4 flight testing. Some of this process is described in [Rios 2018] and [Rios 2018b]. This process also helped NASA inform potential future requirements related to USS Checkout procedures for a future operational UTM System [Smith 2019].

In the months leading up to the flight testing, the test sites worked with NASA to execute tabletop exercises, simulations, and shakeout flights to prepare for the actual field tests. At each of these stages, the data collection approach was solidified to support the analysis that NASA needed to accomplish.

Measures of Performance

The NASA UTM Project defined twenty Measures of Performance (MOPs) to aid in quantification of the UTM System as designed and tested in TCL4. Each of these MOPs will be reported internally to NASA. Additionally, NASA will report on most of these MOPs in various venues in the future. Three of the MOPs will be detailed in this report.

The approach to defining MOPs is driven by NASA's "NASA Systems Engineering Processes and Requirements" [NASA 2013]. Requirements for a system are supported by Measures of Effectiveness (MOEs) which are qualitative in nature. These MOEs are supported by MOPs which are quantitative in nature. Formal definitions for requirements, MOEs, and MOPs are provided below. These relationships are illustrated in Figure 1.

Figure 3. Relationships of requirements, MOEs, and MOPs.

The UTM System has several high-level requirements. The formal definition of a requirement used within the UTM Project is defined as:

The agreed upon need, capability, capacity, or demand for personnel, equipment, facilities, or other resources or services by specified quantities for specific periods of time or at a specified time expressed as a "shall" statement. Acceptable form for a requirement statement is individually clear, correct, feasible to obtain, unambiguous in meaning, and can be validated at the level of the system structure at which stated. In pairs of requirement statements or as a set, collectively, they are not redundant, are adequately related with respect to terms used, and are not in conflict with one another.

Each MOP is connected to a defined Measure of Effectiveness (MOE). An MOE is defined by NASA as:

A measure by which a stakeholder's expectations will be judged in assessing satisfaction with products or systems produced and delivered in accordance with the associated technical effort. An MOE is deemed to be critical to not only the acceptability

of the product by the stakeholder but also critical to operational/mission usage. An MOE is typically qualitative in nature or not able to be used directly as a "design-to" requirement.

A MOP is defined by NASA [NASA 2013] as:

A quantitative measure that, when met by the design solution, will help ensure that an MOE for a product or system will be satisfied. MOPs are given special attention during design to ensure that the MOEs with which they are associated are met. There are generally two or more measures of performance for each MOE.

Requirement UTM-REQ-4

The UTM System SHALL provide common situational awareness for stakeholders related to sUAS operations.

This requirement, like the four others in the set of UTM's first five requirements (not detailed here), is driven from the first Concept of Operations document published by NASA [NASA 2016], which in turn was driven by stakeholder engagement across operators, regulators, researchers, pilots, and others.

Measure of Effectiveness UTM-MOE-4

UTM allows for common situational awareness of the airspace and operations within it to support sUAS operations.

Summary of Measures of Performance

The following table is a summary of the three MOPs reported in this document. Further details provided in the subsequent subsections.

Table 1. Summary of MOPs reported in this document.

MOP ID	Title	Description	Minimum Success	Target Success
UTM-MOP-11	Successful response rate by USSs to unauthorized data exchanges	The NASA USS submits intermittent requests to other USSs during TCL4 with bad access tokens.	> 90% appropriate response rate by USSs to access tokens	100% appropriate response rate by USSs to access tokens
UTM-MOP-13	USS latency within USS Network	Calculate the average latency of USS exchanges, broken down by various categories.	< 1000 ms average latency for data exchanges	< 400 ms average latency for data exchanges
UTM-MOP-16	Successful High Density Operations	This MOP measures the number of (live and simulated) aircraft per defined 0.2 nmi^2 of UTM operations.	> 10 aircraft (at least 3 live operations) airborne and managed by UTM for > 20 continuous minutes within 0.2 nmi^2	> 15 aircraft (at least 3 live operations) airborne and managed by UTM for > 20 continuous minutes within 0.2 nmi^2

Measure of Performance 11: Unauthorized Data Access

UTM-MOP-11 is entitled "Successful response rate by USSs to unauthorized data exchanges." The high-level description is documented as "Using [NASA USS], submit intermittent requests to other USSs during TCL4 with bad access tokens."

This MOP aids in building confidence that data exchanges are appropriately authorized within the context of UTM. Thus, the MOP supports UTM-MOE-4 by ensuring that the exchanged data are valid in building up common situational awareness. Recall that these data exchanges between USSs describe operations, positions, and other operational data, which are the basis for a common operating picture.

UTM-MOP-11 was conducted using a test driver leveraging the NASA USS (NUSS) using the USS-API. NUSS was authenticated using the Flight Information Management System (FIMS) Authorization roles and authorizations. The UTM-MOP-11 tests were automatically executed daily, at random times during the live Nevada and Texas Flight Tests. Each day there were six different tests run against each USS, with the overall pass rate recorded for that USS. USSs were manually and directly notified of failures through an instant messaging service.

Tests

The six tests are described in the table below. Each one tests a different way that operational data could be manipulated if the appropriate checks are not performed by the receiving USS. These include trying to send messages about another USS's operation to other USSs or a USS submitting spoofed position reports for an operation that it does not manage. The USS-under-test (UUT) must perform cross-checks between the access token and certain values of the data model and if they do not match the UUT must reject this message as forbidden. As detailed in [Rios 2019], the cross-checks protect against several attacks including token re-use and USS spoofing. The collection of the six tests will be referred to as the "test suite" and are summarized in Table 2. The general data and action flow for each test is described in Figure 4. The flow requires the test driver to setup initial valid data with the UUT and then attempt some illegitimate access or modification of those initial data, with the expectation that illegitimate access would be appropriately blocked by the UUT to pass the test.

Table 2. Summary of the six tests (a.k.a the "test suite") executed each run for UTM-MOP-11

Test #	Title	Description
1	Bad USS name in position report	Given an existing operation supported by a USS, provide a position report with a valid token from that supporting USS. The position report, however, has the uss_name field that is not for that USS.
2	Spoof position report	Given a valid operation supported by USS A, USS B submits a position report for that operation using an access token for USS B.
3	Spoof operation plan	USS A submits an operation with a valid token for USS A, however the uss_name in the operation plan is not USS A.
4	Hijack operation management	Given a valid operation supported by USS A, USS B attempts to submit an update to that operation with USS B as the uss_name and a valid token for USS B.
5	Bad USS name in UTM Message	USS A submits a UTM Message related to an operation managed by USS A, however the uss_name in the UTM Message is not USS A.
6	Spoof UTM Message	Given a valid operation supported by USS A, USS B submits a UTM Message related to that operation using an access token for USS B.
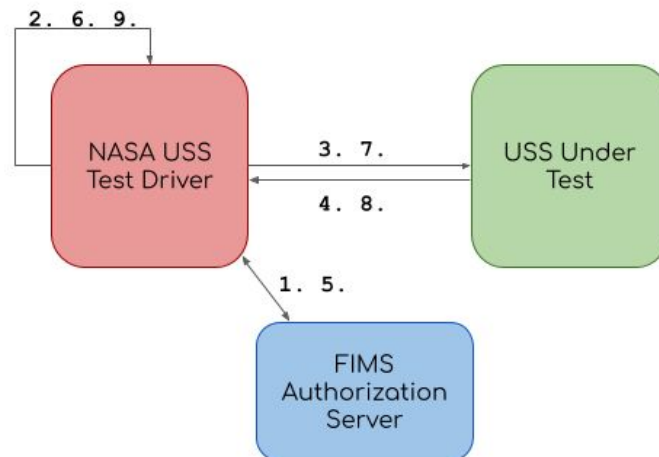

```
1. Get access_token as USS A
2. Create data payload consistent with token
3. Send request, expect HTTP status code 200
4. Send HTTP response
5. Get access_token as USS B
6. Create data payload inconsistent with token
7. Send bad request, expect HTTP status code 403.
8. Send HTTP response.
9. If status code == 403, pass; else fail
```

**Figure 4. General flow for a single test for UTM-MOP-11.**

Results

Figure 5 shows that of the 162 test suite runs, 144 times the UUT passed all six tests in the test suite. The next largest cluster of results shows that the UUT failed the entire test suite, but that only occurred thirteen times. These failures were likely caused by deployment regressions. The process of developing a web service application includes developing the source code and building the artifact, then deploying the artifact, typically to cloud. Network boundary services at the layer directly exposed to the Internet handle routing, load balancing and various security operational controls. Typically these network boundary services are maintained by separate network personnel, not by the application developers. During TCL4 testing, problems in this layer contributed to regressions.
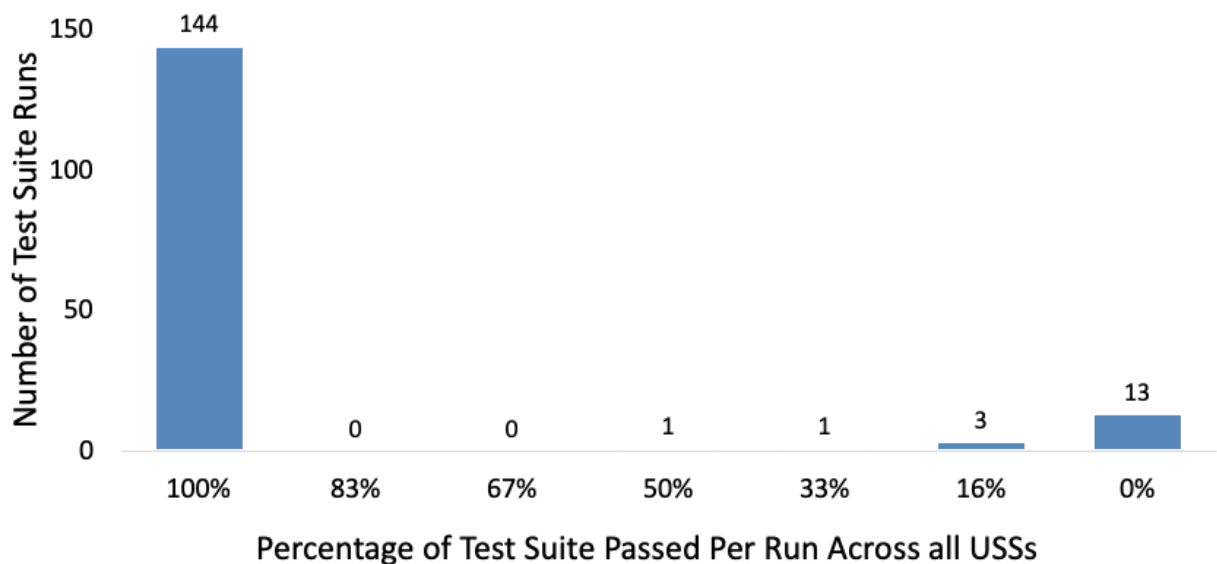
Figure 5: Performance of USSs each time the test suite was run.

Figure 6 shows UTM-MOP-11 results for each USS during flight testing. Each USS was tested once for each operational flight day between June 17, 2019 and August 8, 2019 for a total of 497 data points. The failures of "USS D" occurred during flight testing at Texas after "USS D" was completely successful during flight testing in Nevada. The regression was not completely explained and there was insufficient time to make the fix to USS D system thus repeated failures were detected. Failures in USS I were erratic and could have been caused by a problem in the network boundary layer such as a "sticky cache." USS I could have corrected the problem had it been notified of the failure, however at the time of testing, the test framework did not have automated, per-USS test report delivery. NASA's automation solution was implemented in another framework but not MOPS-11 framework.
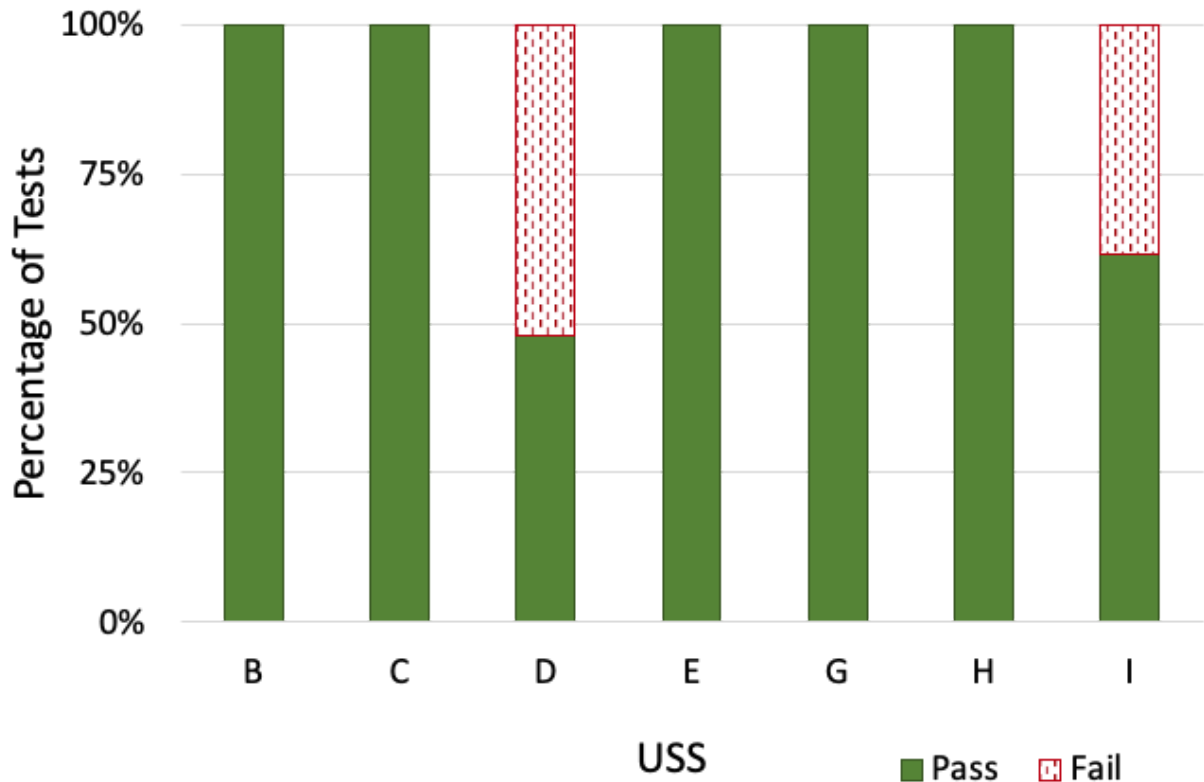
Figure 6: Results of testing authorization implementation of each USS during flight testing.

UTM-MOP-11 established a minimum success bar of 90%. On a per-USS view, 5 of the 7 USSs met the minimum success criterion. Aggregating all test results for all flight days across all USSs, the result was a 89.70% pass rate, failing to achieve the overall minimum success criteria.

$$\frac{\texttt{num\_illegal\_access\_attempts} - \texttt{num\_allowed\_illegal\_access\_attempts}}{\texttt{num\_illegal\_access\_attempts}}$$

Figure 7: Formula for percentage of successful tests.

Discussion

There are several key takeaway messages from the data collection and analysis related to UTM-MOP-11. This section will list and provide discussion on each.

All or Nothing Results

The results tended to be "all or nothing," either completely succeeded or failed. When NASA notified a USS of its failure status, those USSs were provided full insight to the failures from NASA's perspective. Most of the USSs that participated in Nevada also participated in Texas, therefore the USSs as a group were able to reduce overall regressions over the course of TCL4. Despite a large gap between the full success (144 instances) on each test suite run and the full failure (13 instances), those two groups represent nearly all of the results, excluding just 5 of the overall results. This result points to failures in this security implementation being wider than just a single coding error on a particular aspect of the service. As noted above, this

14

likely had more to do with whether infrastructure failures at various layers of a USS's architecture were automatically discovered and reported. This is typical in industry and government collaboration efforts where the management of the network and the applications may involve completely separate personnel with completely different operating objectives.

Test Automation

The TCL4 Demonstration helped NASA further define efficient methods for testing USSs. For the most complete assessment of USS test data, ideally, the testing process would be fully automated, including scheduling, executing, capturing results, writing generous log data, and automatically pushing the results to the UUT. Note that because NASA need to push the report package securely, the automatic reporting required coordinating credentials, and therefore not all tests were instrumented with automatic reporting. Some delays in USSs addressing their failures and regressions resulted in delayed notification of failures.

As noted above, if the USS under test fails to perform cross-checks between the security token and the data payload, a security threat is no longer mitigated. During a live test flight which uses prescribed data and scenarios, almost all of the data exchanges are expected to return OK and none are expected to return "Forbidden." This illustrates the benefit of running separate automated tests concurrently with the flight tests, as these provide quality of service data, in addition to the functional and performance data provided by the flight tests.

These observations point to the need to monitor and automatically report failures to the USS to protect against regressions in the dynamic deployment processes of modern software. To support this objective, considerations for requirements related to software updates would also be reasonable.

Bad Apples

Having one or two stakeholders fail to meet security criteria for the system will have an impact on the overall security of the system. This illustrates the need for compliance by all USSs in an operational system, else cracks in the implementation will diminish the overall trust and security of the system.

Ordering of Checks

The USS Specification requires validation checks but does not prescribe in what order they are implemented. Depending on the order the checks were implemented by the UUT, more than one response is received by NASA. For example, if the USS implemented a time format validation check to occur before the token 'sub' claim check, the response error code would be "Invalid Data" (HTTP Response Code 400). However, if the order of these checks are reversed the response would be something like "Message USS name does not match the token sub" (HTTP Response Code 401). During our testing process we discussed this condition with our partners and in some cases adjusted our tests to be more flexible.

For an operational system, the ordering of checks needs to be formally defined to ensure interoperability. TCL 4 provides valuable lessons learned on how to prescribe ordering. Both the network technology stack and the data elements under inspection need to be considered. Future USS interfaces will likely contain different network connection technologies such as streaming data endpoints. For example, websockets are based on a stateful connection, where authentication and authorization is done once at the connection step, in contrast to RESTFul which authorizes each data exchange. Authorization checks require inspection of several data elements. Some checks inspect only the security context, others require cross-checking between the security context and the data payload and finally some checks require

cross-checking of data from a prior data exchange that is likely stored in the application's persistence service.   Checks requiring persisted data would need to be performed last.

Measure of Performance 13: USS Latency

UTM-MOP-13 is entitled "USS latency within USS Network." The high-level description is documented as "Calculate the average latency of USS exchanges, broken down by various categories." The UTM Project established a minimum success criteria of an average latency of < 1000ms and a target success of < 400ms for various data exchanges.

This MOP provides insight into the timeliness of data exchanged within the USS Network. Thus, it directly supports UTM-MOE-4 by ensuring that the exchanged data are timely and actionable by the USSs and the other stakeholders those USSs support.

All USSs were required to log the latency to complete the primary USS-USS data exchanges. These exchanges were broken into the following categories and logged as such by the USSs:

- Discovery: Data exchanges with the underlying USS Discovery system
- UTM Messages: Messages regarding Operations
- Positions: Reports of Positions from USS to USS
- Operations: Operation plans shared from USS to USS
- Negotiation Messages: Messages related to strategic deconfliction of Operations

The following subsections examine various aspects of these USS Exchange messages, culminating in a report on UTM-MOP-13.

Data Model

Each USS logged data exchanges with other USSs and the underlying USS Discovery system. The UTM System does not impose requirements on the structure of those logs. However, for NASA TCL4 testing, a unified model for reporting information about exchanges between USSs was required. For each exchange of Operation, Position, UTM Message, Negotiation Message, or Discovery data, each USS was required to submit an instance of the USSExchange data model. This model was part of the overall Data Management Plan for the participating stakeholders and was documented (amongst other ways) as an API published on GitHub[1].

The model collects elements such as the source USS, the target USS, the reporting USS (which must be either the source or target USS), HTTP response, data type exchanged, time initiated, time completed, and other related elements.

Data Scrubbing

There were a total of 1,149,926 USS Exchange data samples collected from USSs during TCL4 operational flight tests. Note that there were additional samples (over 2 million total) collected during testing and shakeout events, which are not included in this analysis. Some of the samples are specifically excluded from the analysis below. These exclusions were due to three separate sources of error:

1. Outlier latency
2. Target system latency

---

[1] UTM API Documentation on GitHub, as of October 2019.

3. Invalid data submission

The first condition, outlier latencies, was applied to allow for reasonable averages to be computed. As a reasonable cutoff, all data exchanges taking over 5 minutes were excluded; 33 of the 1,149,926 (< 0.003%) samples fell into this category. Typically these samples were from the automated servers logging exchanges with downed or unavailable servers, so the ultimate HTTP status code would be of the 50X-type (server error). The HTTP timeout for USS exchanges was not specified in TCL4 testing, so USSs used various values for waiting for responses from other servers.

The second condition, target system latency, was applied to address two concerns. First, there were few samples (25,330) relative to the overall data set of more than 1.1 million (~2.2% of all samples), which raised the possibility of selection or reporting bias in the data. Second, the latency for a target server is embedded in the latency report for a source server. This is because the latency of a request from a source server as reported in the USS Exchange model is the sum of transport latency to the target server, processing time of the target server, and the transport latency back to the source server. Thus, the target latencies would generally be lower than the source latencies and would actually be "double counted" if reporting was complete and consistent.

The third and final condition involved corrupted data submitted by a particular USS. The error was introduced through the serialization and deserialization process of timestamps. In that process, some information was lost due to mis-implementation of the API specification. There were 30,850 samples, or 2.7% of all samples, in this category. With over 95% of all data still available for analysis covering 6 other USSs involved in TCL4 flight testing, it was determined to be reasonable to ignore these samples rather than manually correcting the data.

Some of the scrubbed data occurred in both the second and third conditions, i.e. both submitted as a target USS and from the mis-submitting USS. Thus overall, there were 49,633 of the 1,149,926 samples excluded, or ~4.3% of the collected data. The breakdown of excluded data are illustrated in Figure 5 as a Venn diagram.
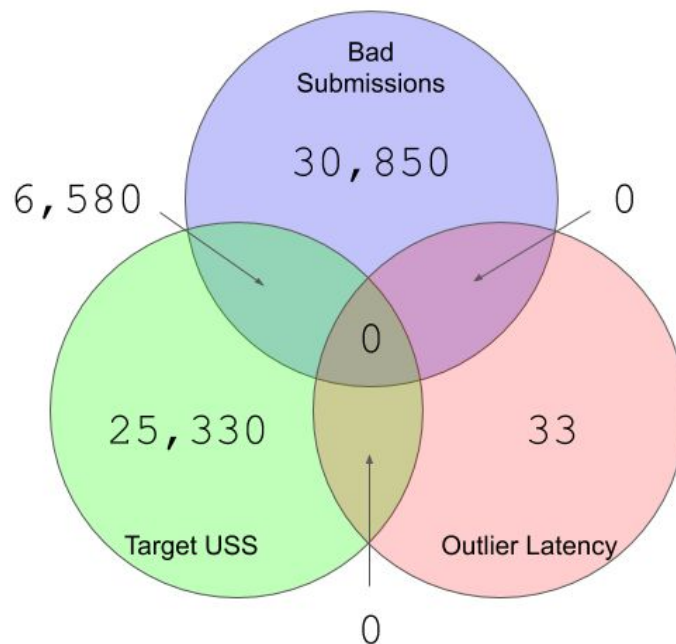
Figure 8. Data samples specifically excluded from analysis due to known conditions producing errors that affected ~4.3% of samples.

Simulated Versus Live Operations

The data analysis includes USS exchanges related to both live and simulated operations during TCL4. From the perspective of USS to USS data exchanges, there is no difference between a simulated operation and a live operation, thus the latencies are equally valid. This is due to the fact that all USSs are remotely-served systems, likely all cloud-based. The measured latencies are between these systems, and it does not matter what triggers the data exchange (i.e. a live or simulated event). For other operational data analysis, it would be reasonable to separate the classes of operations. However, there were more simulated operations than live operations in TCL4; one potential impact to analysis would be if a USS that handled more simulated operations had skewed latency values for any reason.   This situation is addressed below to some degree.

Message Count and Latency

Due to the type of data and the requirements for data exchange with the TCL4 UTM System, there was wide variation in the quantity and latency of data exchanges of each type, which are illustrated in Figure 9. A takeaway from this graph, which will be discussed in more depth in the Discussion section below, is that the message type with the highest latency (Negotiations) was used much less frequently than the most common data exchange (Positions), which has a relatively low latency. Note that the latencies presented here are as measured by the requesting system, thus the measurements represent total latency from request to response from the requestor's perspective.
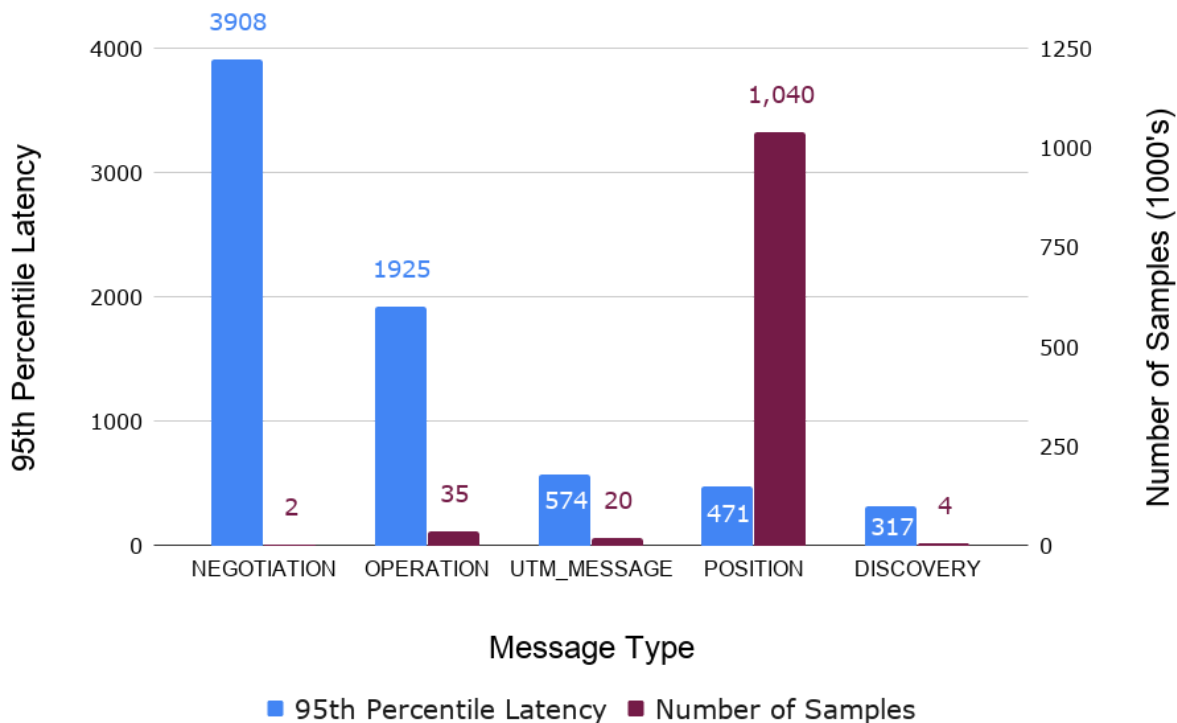

Figure 9. 95th Percentile latency values and number of samples by message type.

Per USS Performance

The samples were dominated by a single USS that diligently recorded and reported its data calls to other USSs. Together with the USS that submitted bad data, another USS was excluded since it submitted data as a target USS exclusively, which met the exclusion criteria as discussed above. Thus, the data analysis is based on samples from 4 distinct USSs. To ensure that the large portion of data submitted by a single USS did not overly skew the data, the median latencies for each of the 4 USSs was compared (Table 3).

Table 3. Samples and median latency per USS.

	USS 1	USS 2	USS 3	USS 4
Samples	220,720	23,460	844,061	12,052
Median Latency (ms)	82	115	83	104

Since the two dominantly-represented USSs have similar median latency values, the analysis was performed with the assumption that there was not an obvious bias due to the large sample set from a single USS. A deeper analysis may provide further insight, but is not presented here.

UTM-MOP-13

As mentioned above, this MOP is defined as "Calculate the average latency of USS exchanges, broken down by various categories." The UTM Project established a minimum success criteria of an average latency of < 1000ms and a target success of < 400ms for various data exchanges. The results are presented in Figure 10.
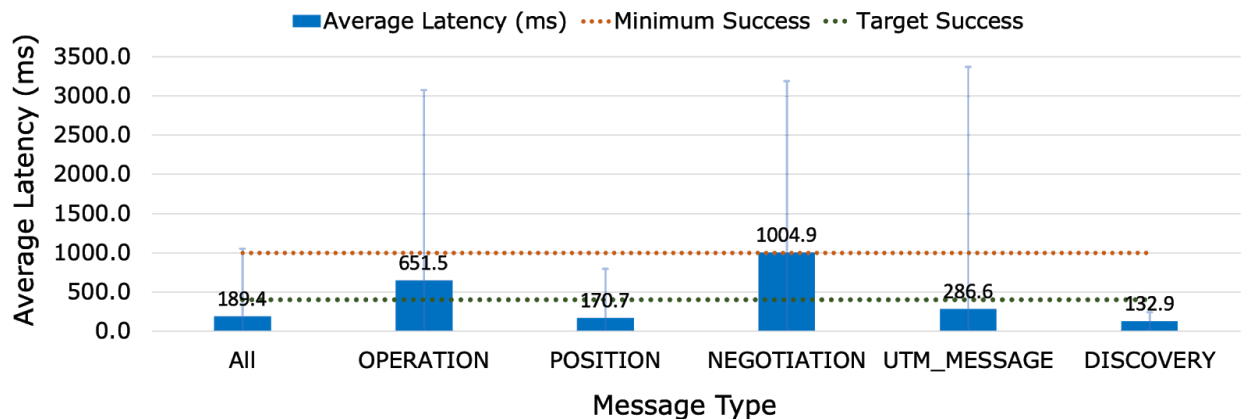
Figure 10. UTM-MOP-13 results for average latency for USS data exchanges.

Over the full set of messages, this MOP was met, with an overall average latency of 189.4 ms. When broken down by message type, the minimum success criteria were met for all but the Negotiation message type. Variability was high amongst most message types (over 2 seconds of standard deviation for Operation, Negotiation, and UTM_Message exchanges), with tighter standard deviation on Position and Discovery messages (623 ms and 110 ms, respectively).

Another view of the data is provided via the cumulative distribution of all samples, for all message types, illustrated in Figure 11. This plot shows that 95% of all data exchanges occur with 531ms or less latency. As with the average values presented in Figure 10, the 95th percentile latency changes with message type. To illustrate, note Figures 12 and 13 showing

the cumulative distributions for Position and Negotiation messages, respectively. The 95th percentile for Position messages is 471 ms, while it is 3,908 ms for Negotiation messages.

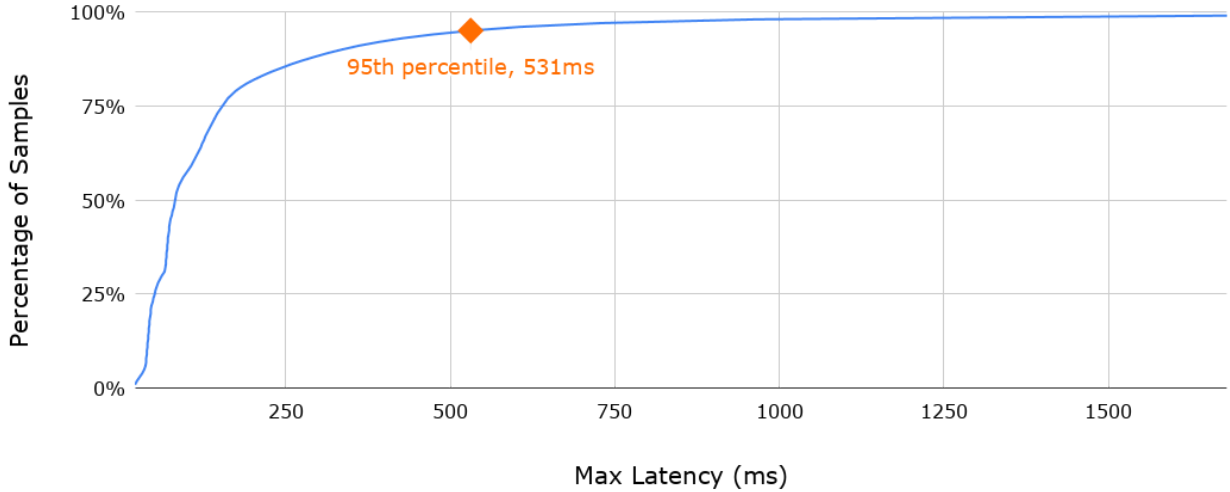USS-USS Communications Latency Cumulative Distribution

Figure 11. UTM-MOP-13 results for average latency for USS data exchanges.

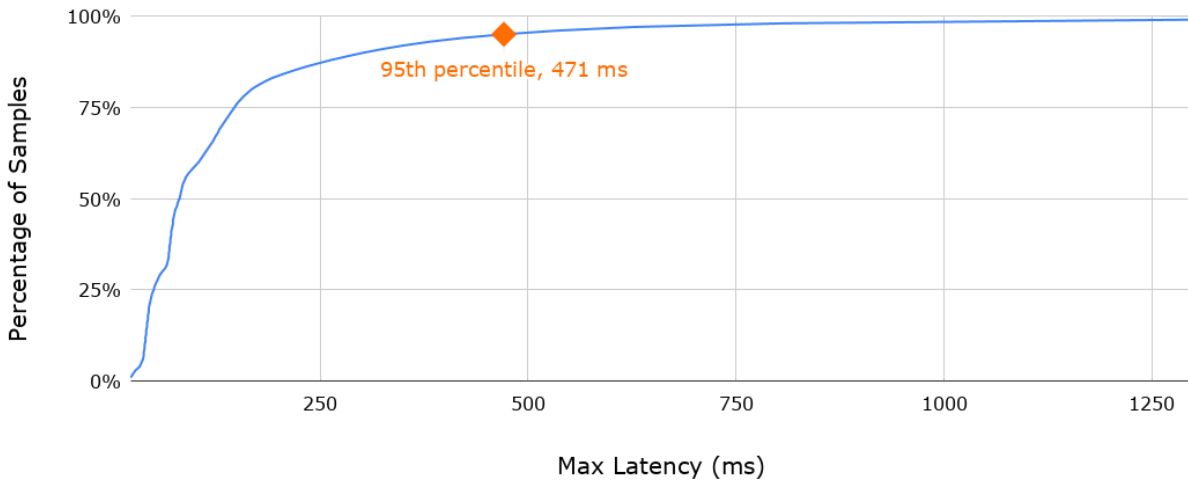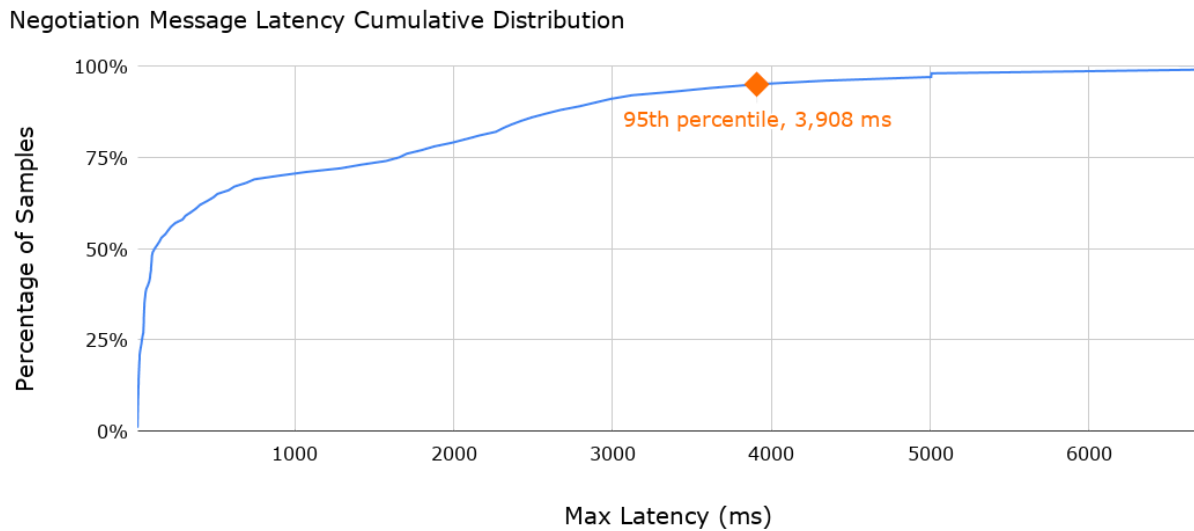Position Message Latency Cumulative Distribution

Figure 12. Cumulative distribution of sampled latencies for Position messages.

Negotiation Message Latency Cumulative Distribution

Figure 13. Cumulative distribution of sampled latencies for Negotiation messages.

Discussion

There are several key takeaway messages from the data collection and analysis related to UTM-MOP-13. This section will list and provide discussion on each.

Varied Performance Per Message Type

There was wide latency variation across message types. This finding makes it difficult to state a blanket "good" latency for data exchanges between USSs. The implication is that any future performance specifications for USS communications needs to take into consideration the type of data being exchanged when establishing performance requirements. Otherwise, a lowest common denominator approach will place inappropriate performance requirements on potentially low-latency exchanges. The extra latency for certain message types derives from the potential processing that a USS must perform on the message to provide a response. For example, for a negotiation message, there may be several latent or expensive computations that must be performed to generate the appropriate response. On the other hand, a position message requires virtually no processing by the receiving USS before an appropriate response can be provided, thus the wide variation in observed latency by the source USS.

Minimal Operational Impact Due to Latency

A subjective, qualitative observation from TCL4 flight testing is that there was little impact to operations due to the latency of data exchanges between USSs. There was negative impact observed due to downed servers, inappropriately high http-request rates, or misimplemented specifications, but not purely to latency of data exchanges. The implication is that USS-USS communications will not likely be a bottleneck to the implementation of an operational UTM System. It is also useful to note that the implementations of USSs for TCL4 were not necessarily hardened or tuned for true operational use, thus there is likely further room for reasonable performance enhancement where needed.

Streaming Interface

In the course of UTM development, NASA and partners often discussed implementing Position reporting use cases over a streaming endpoint such as websockets. During TCL2 and

-3 partner testing, NASA evaluated the position reporting use cases and decided to bypass streaming implementations until UTM concepts and the USS Specification matured. This strategy of using one simple, uniform interface technology (REST using JSON-over-HTTP) helped NASA and their partners maintain continuous progress in developing UTM. Thus all participants were able to focus on data models and protocols instead of additional development considerations and deployment issues. The work presented here should function as a performance benchmark upon which further design decisions can be implemented and measured.

Viable Initial Model for Communication Performance

The data for this analysis were collected using a specified model for TCL4. This model may also prove useful in an operational UTM System. There will likely be reporting requirements for USSs to the regulator and/or ANSP which may be related to incident investigations, service-level monitoring, or other aspects of the UTM System. The standardization of the data formatting for issues related to USS-USS communications will be important for rapid analysis, consistent requirements, and compliance reporting across USSs. The USSExchange model should be considered a benchmark from which the UTM industry stakeholders and regulators can tune to operational requirements.

Measure of Performance 16: High-Density Operations

UTM-MOP-16 is entitled "Successful High-Density Operations." The high-level description is documented as "This MOP measures the number of (live and simulated) aircraft per defined 0.2 nmi^2 of UTM operations." The original minimum success criteria as underspecified as "> 10." For this analysis, the minimum success criteria was more precisely defined as "> 10 aircraft (at least 3 live operations) airborne and managed by UTM within an area of 0.2 nmi^2." Similarly the original target success criteria was further detailed to become: "> 15 aircraft (at least 3 live operations) airborne and managed by UTM within an area of 0.2 nmi^2."

In some cases, the requirement as written in the Statement of Work to the Test Site was more stringent in order to push the limits on subsystem capabilities. For example, in one scenario, NASA defined a requirement for > 90% of the operational time to be spent in a high-density condition of > 10 aircraft per 0.2 nmi^2, which could amount to nearly an hour spent in that condition, depending on scenario execution time.

Density Calculation

The parameter of 0.2 nmi^2 was derived qualitatively based on range size, subject-matter expert input, and ease of translation between units. Many of the underlying calculations are performed in meters and then converted to feet or nmi. The value of 0.2 nmi^2 is roughly equivalent to 720 m^2 or a circle with radius 406m. For all calculations in this paper, a circle with 406m radius was used to calculate density.

Given that the operations are dynamic, it is not feasible to determine a static "center point" for calculating density. It is possible to pre-define a grid of the desired dimensions, but operations could be densely packed separated by a static grid line, reducing the actual density calculation, since operations could be split into different static grid cells. Even if NASA attempted to use the planned operational locations to determine a reasonable center point, the calculations would be affected by last-minute operational decisions to ground certain operations or move their operating areas. Therefore, NASA implemented the following data-driven approach to determining a center from which to calculate density.

After collecting all current positions for all active operations at any given time, a density center was calculated using a geometric median. The geometric median is the point that minimizes the sum of the distances to the points in the collection. This is preferred for this use case over a center of mass calculation as outliers can inordinately affect the center of mass calculation, whereas the geometric median finds the largest grouping more often. Also, the geometric median is shown to be robust to errors within the source data, still providing a good estimate of the median when up to half the source positions are corrupted[2]. This approach could be affected by an evenly divided pair of equally sized groups of points, but this is unlikely and was not experienced in TCL4. NASA leveraged a Geographic Information System (GIS) capable relational database (Postgres with the PostGIS extension), which has an implementation of geometric median. The ST_GeometricMedian function in PostGIS implements an approach described in [Weiszfeld] (with an English translation available by [Plastria]). Pseudocode is provided in Figure 14, with the efficient implementation being provided by Postgres and PostGIS, thus the actual geometric calculations are not necessarily "for loops" as presented in the pseudocode, but the concept is consistent.

---

[2] The data for this calculation shows no signs of corruption, but given some natural uncertainties in position measurement, the calculation robustness is a nice feature.

```
for each time t in scenario s

     for each operation, o, in scenario s
          if o is active
                add most recent position of o to positions collection, P
          fi
     done for each operation

     calculate geometric median, med, of P

     for each position, pos, in P
          if pos is within 406m of med
                add pos to positions collection, Q
          fi
     done for each position

     return size of Q as current density at time t

end for
```

Figure 14. Pseudocode for calculation of the density values.

Data Model

The primary data model for the calculation of this MOP was the Position model. An instance of this model is provided by each USS for each of their operations at a rate of 1Hz. This is considered an "operational" data model as it is expected to be used (in its current form or some related form) for a future operational system. These position reports were received by a NASA-deployed testing artifact within the TCL4 environment called the USS Data Collector or UDC. UDC implemented the USS API and, thus, interacted with other USSs during the test as if it were another USS. However its primary purpose was to collect USS-USS communication data as it occurred. UDC performed other functions related to data validation, data visualization, and related functions that will not be detailed here.

Results

Each test site had several runs of the five scenarios developed by NASA. Each run at each test site was given an identifier by test site, scenario, and run. The goal of UTM-MOP-16 was not to maximize density of operations for every run, as each scenario had different requirements. However, most good[3] runs reached the minimum density and many runs met the target density value. Figure 15 shows the maximum achieved density from several runs of

---

[3] The "goodness" of a run was qualitative in nature as determined by the NASA UTM flight test team. This qualitative assessment was supported by live data and reports from the field. A "not good" run may have experienced technical difficulty with multiple live vehicles or USSs, for example. This would be visible in the collected data in the form of missing operations.

multiple scenarios from both test sites.
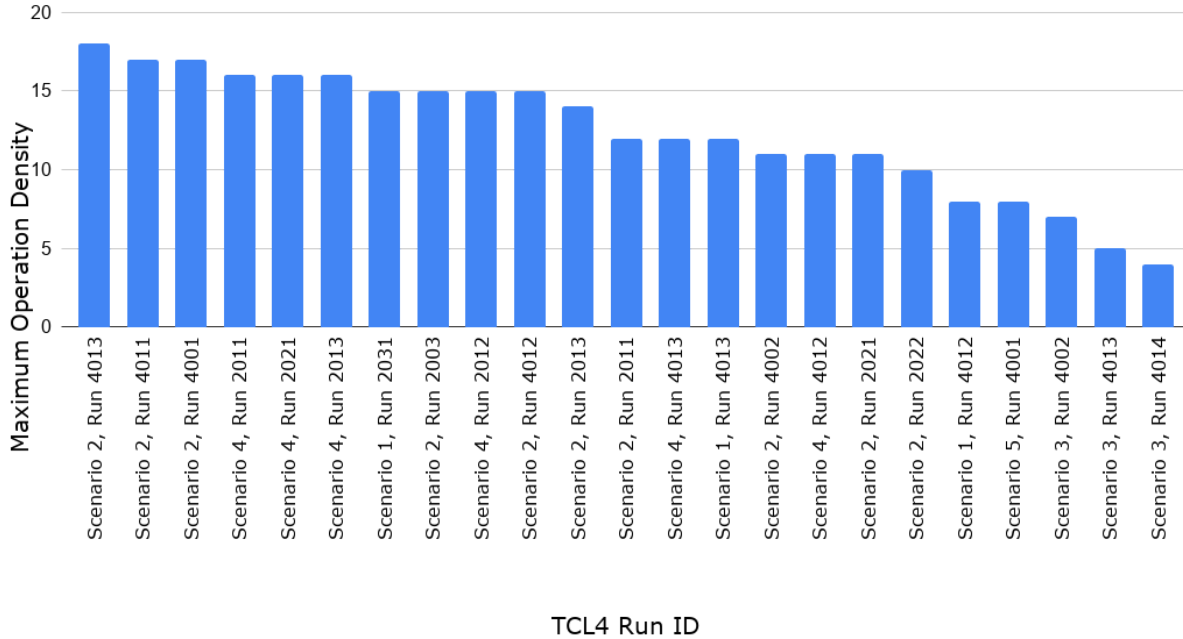
Figure 15. Maximum density across a sampling of scenarios, both test sites, and many runs.

For Run 2021 of Scenario 4 (S4R2021) from Nevada flight testing, further analysis is illuminating. Scenario 4 required the simulation of a large-scale loss of communications. For example, if most or all operators relied upon the cellular network and there was a catastrophic loss of this service for some reason, then all operations would need to execute contingencies and attempt to land. In this run, there were at most two live operations active at any time. The large number of simulated operations are the primary reason that this run was able to maintain a consistent level of density, with a summary provided in Table 4. However this run did not reach the threshold of three for live operation count.
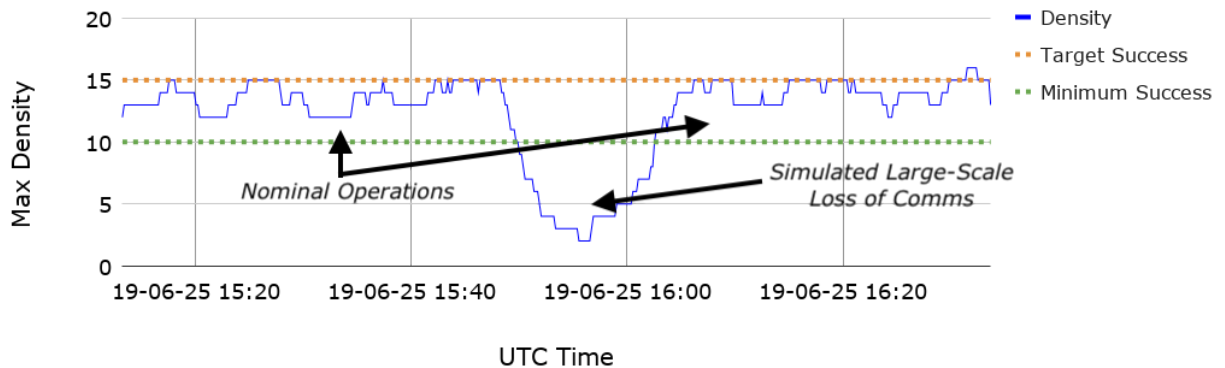
Figure 16. S4R2021 Density of operations with simulated large-scale loss of communications.

Figure 17[4]. S4R2021 at 16:31 with 14 simulated and 2 live operations (airplane symbol) within 406m of density center (black marker, obscured) in downtown Reno.

Table 4. S4R2021 and UTM-MOP-16 target goals for density.

	Density >= 10	Density >= 15
% Operational Time	83.30%	26.19%
Total Mins	67.3	21.2

For Run 4013 of Scenario 4 (S4R4013) in Corpus Christi, Texas, the minimum success density was achieved for a short period. In this case, there were five live operations involved at the time of peak density (12 operations within $0.2nm^2$). Figure 19 shows one of the peaks in density (at 16:00 UTC). There were additional simulated operations occurring at that time in downtown Corpus Christi, indicated by the orange markers in Figure 19. Those operations are not included in the density count due to their distance from the calculated median point of all of the operations, indicated with the black marker in Figure 19.

---

[4] Underlying map data is © OpenStreetMap contributors. https://www.openstreetmap.org/copyright.
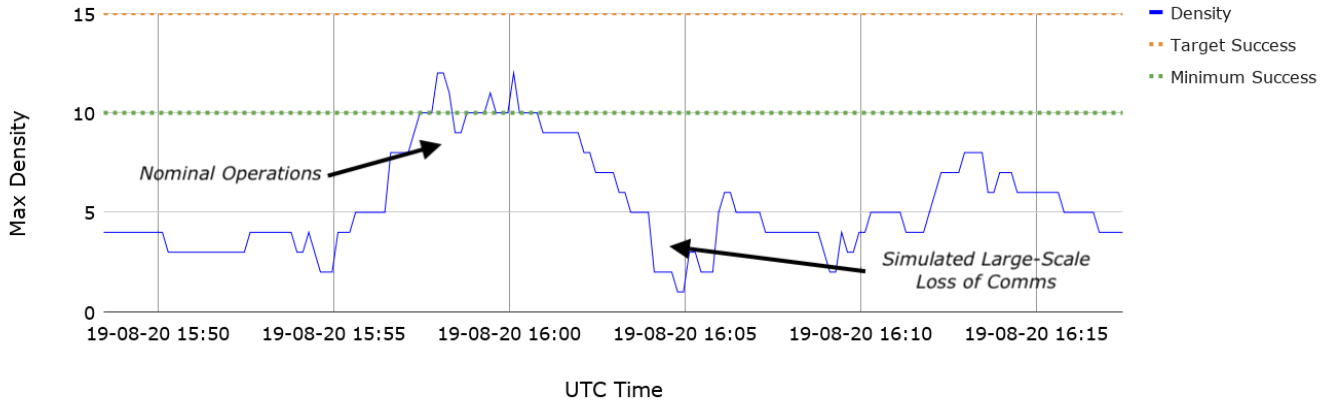
Figure 18. S4R2021 Density of operations with simulated large-scale loss of communications.

Figure 19. S4R4013 at 16:00 with 7 simulated and 5 live operations (airplane symbol) within 406m of density center (black marker) in Corpus Christi. Other simulated operations at that time away from the density center indicated in orange.

Table 5. S4R4013 and UTM-MOP-16 target goals for density.

	Density >= 10	Density >= 15
% Operational Time	10.80%	0.00%
Total Mins	3.2	0.0

Discussion

There are several key takeaway messages from the data collection and analysis related to UTM-MOP-16. This section will list and provide discussion on each.

Viable Architecture
Overall, this measure of performance indicates that a reasonable volume of small UAS traffic can be handled by the proposed UTM architecture, which meets the needs of many identified UTM use cases. There are still open questions about scalability that were not directly addressed, for example having this density of operations in several locations in a given city would require appropriately scaled infrastructure. However, it will be some time before much higher densities from those tested in TCL4 will become a safe reality. Thus the scalability issues can be addressed in the meantime with no major architectural blockers for reaching reasonable densities. There are also still protocol issues related to negotiation and emergency procedures, but, again, the architecture is amenable to scaling operations.

Operational Orchestration
Target densities were often missed in a given run due to orchestrating multiple stakeholders, platforms, and software systems. While this outcome impacted running a flight test as planned, it did not directly provide data on the ability of an operational system to handle diverse stakeholders in a UTM environment.

Overall Discussion

The analysis of the three MOPs presented in this document resulted in a number of insights, summarized in the following table.

Table 6. Summary of lessons and insights from USS network performance testing.

MOP	Lesson/Insight	Summary
UTM-MOP-11	All or Nothing Results	For certain requirements and tests, there are likely interdependencies introduced via the deployment process of the software. This may cause test failures to occur in sets.
UTM-MOP-11	Testing Automation	Efficient, continuous testing approaches for USSs are possible and likely necessary for an operational UTM System.
UTM-MOP-11	Bad Apples	With respect to security, a single, poorly-implemented USS will diminish trust and security in the overall system. Compliance and compliance monitoring of security implementations will be of high importance in an operational UTM System.
UTM-MOP-11	Ordering of Checks	USSs checked various security and message requirements in different orders, thus impacting interoperability. Ordering of required checks needs formalization for an operational system.
UTM-MOP-13	Varied Performance Per Message Type	Performance and QoS requirements likely need to be tailored for each type of data exchange.
UTM-MOP-13	Minimal Operational Impact Due to Latency	In general, there were not any observable impacts to active operations due to USS data exchanges.
UTM-MOP-13	Streaming Interface	In UTM TCL4, a design decision in favor of development velocity over data exchange efficiency was made early in development. It is appropriate to use the results of this flight test to reexamine the types of communication interfaces between USSs to potentially include streaming data.
UTM-MOP-13	Viable Initial Model for Communication Performance	The approach to collecting USS communication performance data may be relevant to logging and auditing requirements in an operational UTM System.
UTM-MOP-16	Viable Architecture	Testing indicates that a reasonable volume of small UAS traffic can be handled by the proposed UTM architecture.
UTM-MOP-16	Operational Orchestration	Target densities were often missed in a given run due to orchestrating multiple stakeholders, platforms, and software systems. While this has an impact on running a flight test as planned, it does not directly comment on the ability of an operational system to handle diverse stakeholders in a UTM environment.

The MOPs for USS Network performance provide insight into how a future operational UTM System may operate in a secure and responsive manner to handle a volume of traffic reasonable for the near future. With further work on scalability and the design of protocols, the architecture is likely amenable to denser and more complex operational densities.

The results of the NASA UTM Project are being transferred to the FAA via a Research Transition Team. Some findings are also being transferred to industry via standards bodies and publications such as this one. NASA plans to continue research on UTM. This includes graduating concepts and architectures to other aviation domains such as Urban Air Mobility (UAM), high altitude operations (over 60,000 ft), and space traffic management. Thus, the results from UTM testing will likely have broad impact on future aerospace applications.

References

[FAA 2018] FAA NextGen Office, "Unmanned Aircraft Systems (UAS) Traffic Management (UTM) Concept of Operations", v1.0, May 2018.

[Plastria 2009] Plastria, F., "On the point for which the sum of the distances to n given points is minimum," translation of Weiszfeld's "Sur le point pour lequel la somme des distances de n points donnés est minimum," Annals of Operations Research, March 2009, Volume 167, Issue 1, pp 7–41.

[NASA 2013] NASA, NASA Systems Engineering Processes and Requirements, NPR 7123.1B, April 2013.

[NASA 2016] Kopardekar, P., Rios, J., Prevot, T., Johnson, M., Jung, J., Robinson III, J., "Unmanned Aircraft System Traffic Management (UTM) Concept of Operations", AIAA Aviation Forum 2016, June 2016.

[Rios 2018] Rios, J., Smith I., Venkatesen P., Smith, D., Baskaran, V., Jurcak, S., Strauss, R., Iyer, S., Verma, P., "UTM UAS Service Supplier Development: Sprint 1 Toward Technical Capability Level 4", NASA Technical Memorandum, NASA/TM-2018-220024, November 2018.

[Rios 2018b] Rios, J., Smith I., Venkatesen P., Smith, D., Baskaran, V., Jurcak, S., Iyer, S., Verma, P., "UTM UAS Service Supplier Development: Sprint 2 Toward Technical Capability Level 4", NASA Technical Memorandum, NASA/TM-2018-220050, December 2018.

[Rios 2019] Rios, J., Smith I., Venkatesen P., "UTM Authentication and Authorization Framework", NASA Technical Memorandum, NASA/TM-2019-220364, October 2019.

[Rios 2019b] Rios, J., et al., "UAS Service Supplier Specification," NASA Technical Memorandum, NASA/TM-2019-220376, October 2019.

[Smith 2019] Smith, I., et al., "USS Checkout," NASA Technical Memorandum, to be published.

[Weiszfeld 1937] Weiszfeld, E., "Sur le point pour lequel la Somme des distances de n points donnés est minimum," Tohoku Mathematical Journal, 1937, Volume 43, Pages 355-386,