National Aeronautics and Space Administration

Next-Gen Space Safeguards

ML-Based Protocol Anomaly Detection for Securing Space-to-Ground Data Links

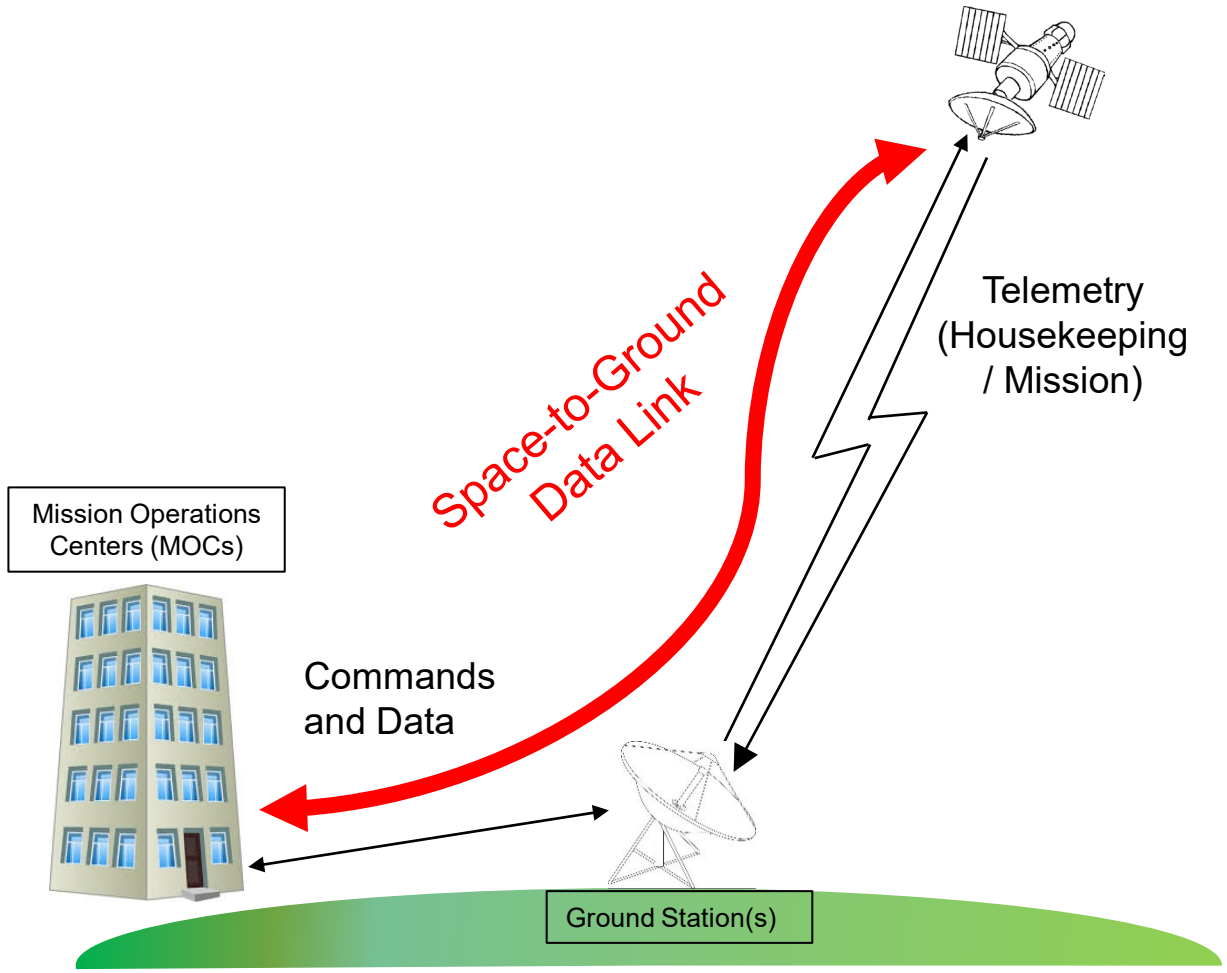
William Stanton and Shannon Bull, NASA GSFC

GSAW 2024
Track 2: Intelligent Systems
Tuesday, Feb. 27th, 2024
Los Angeles, California

Agenda

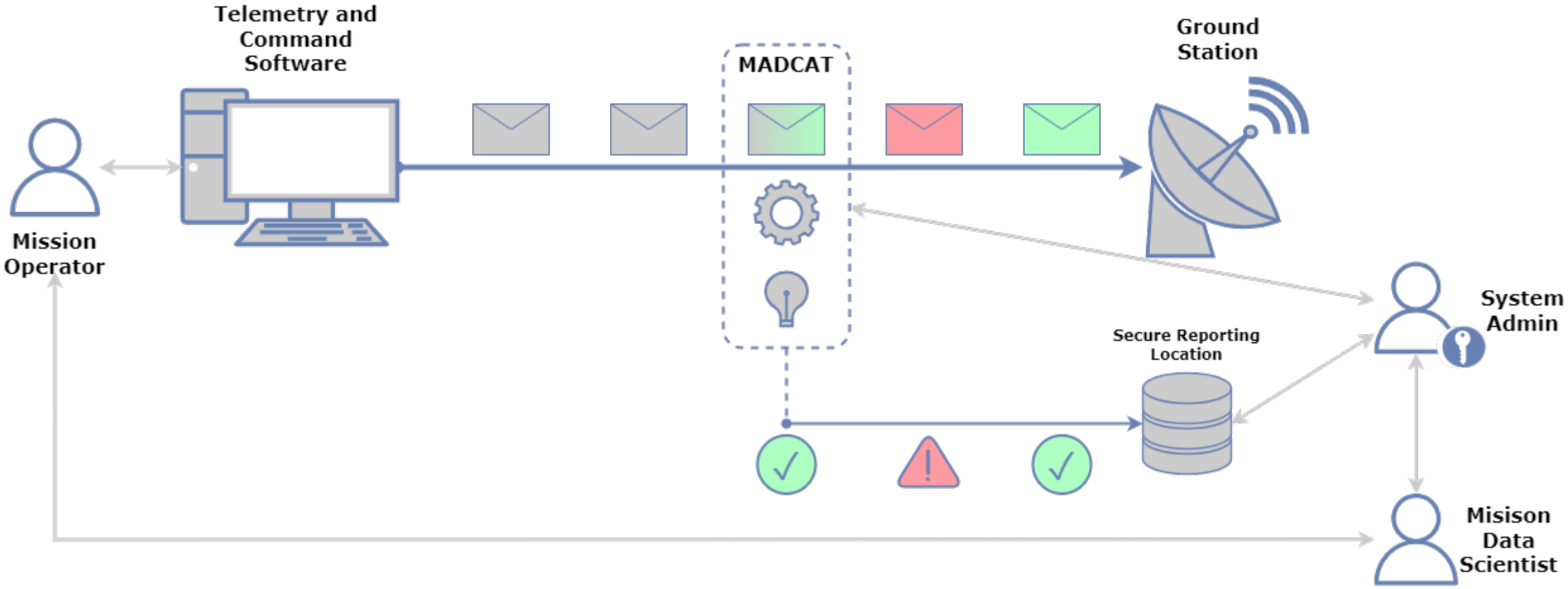
- Context
- MADCAT Concept of Operations
- Anomalous Message Categories
- Anomaly Detection Algorithms
- MADCAT Architecture
- Results to Date
- Lessons Learned and Best Practices
- Future Work

Context



- The threat landscape for protecting the Space-to-Ground link is evolving
- Recent White House Executive Orders (e.g. 14028) and Policy Directives (SPD-5) call for Federal agencies to improve cybersecurity for critical assets including space systems
- Missions continue to be susceptible to human error
- Missions use protocols such as CCSDS SPP, CFDP and OMG C2MS to enable communications between operational systems
- Typical practice for Protocol Verification and Anomaly Detection is limited to what is predefined
- Machine Learning is a growing field and technologies are becoming easier to use/adopt
- **Message Anomaly Detection for Command and Telemetry Systems (MADCAT) is a new technology in development at NASA GSFC that leverages Machine Learning to perform protocol anomaly detection**

Concept of Operations



Anomalous Message Categories

- **Message**: Broad term for a discrete unit of communication. Can include packets, sessions, telemetry, frames, etc.

- Standard messages can be classified as both **valid** and **nominal**.

- MADCAT will identify messages that fall into any of the following **anomalous** message categories:

    1. Anomalous Context
    2. Anomalous Content
    3. Anomalous Volume & Rate
    4. Anomalous Format
    5. Anomalous Communicants
    6. Anomalous Temporal

Anomaly Detection Algorithms (1)

Leverage both **Signature-based** and **Machine Learning** algorithms to detect anomalous messages.

Signature-based algorithms: based on pre-stated rules, identify anomalous messages

1. Kaitai
2. Sliding Window
3. Term Frequency Inverse Document Frequency (TFIDF)

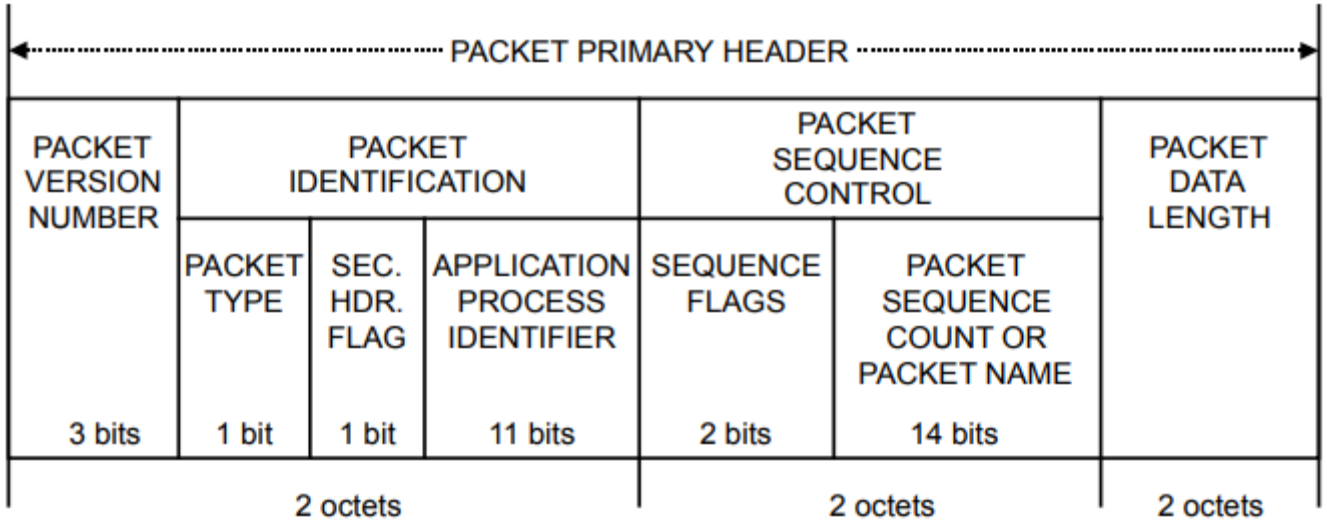
Machine Learning algorithms: use ML to identify patterns and detect anomalies we can’t or don't program for

1. One Support Vector Machine (One SVM)
2. Hierarchical Density-based Clustering (HDBSCAN)
3. Isolation Forest

Generating useful anomaly detection algorithms requires **knowledge of message anomalies** and/or **realistic mission data**

Anomaly Detection Algorithms (2)

Kaitai protocol verification plugin (Signature Based)



Source: CCSDS SPP Blue Book

Catch malformed packets and anomalous format messages (i.e. non CCSDS SPP compliant messages)

Anomaly Detection Algorithms (3)

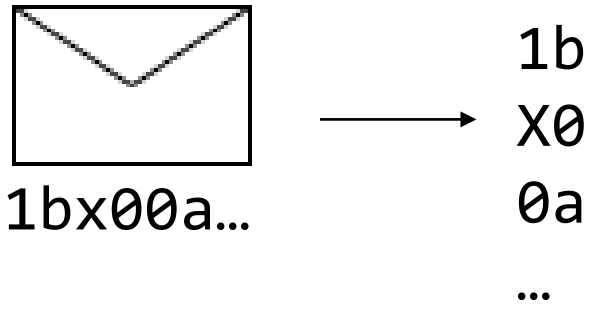
Sliding window plugin (Signature Based)

User Description:
Looking at the application ID and user data field
Y: 10 messages
X: 4 occurrences
13, 17, 20, 20, 23, 15, 13, 17, 20, 20, 20, 15, 23, 17, 18, 17, 17, 17, 17, 13, 17, 18, 17, ...

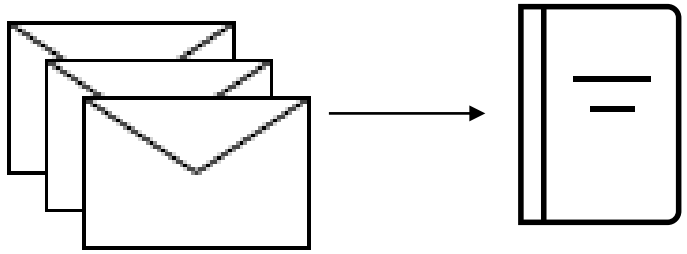
Easy to implement and understand; flags anomalous volume and rate sequences

Anomaly Detection Algorithms (4)

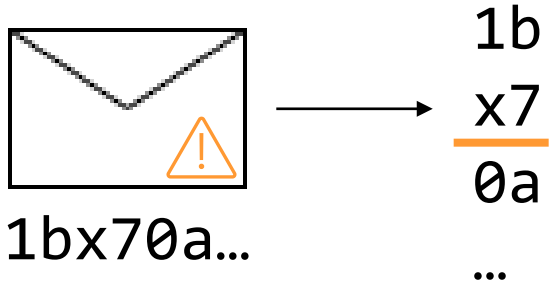
Term Frequency Inverse Document Frequency (Signature Based)



Parse messages and treat pieces as words



Using many messages, generate a dictionary of known words. Assign each word a frequency score.

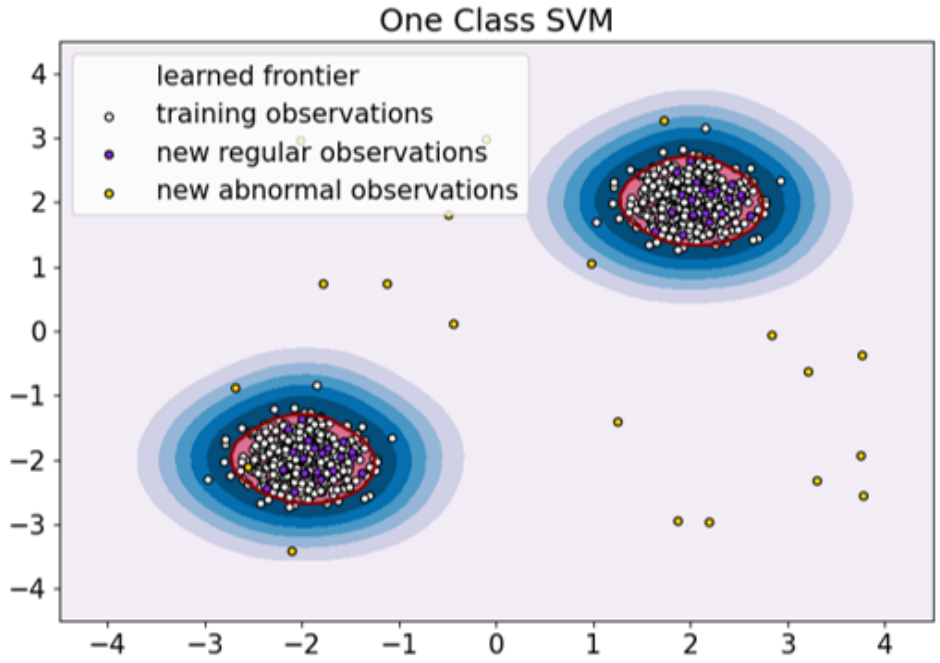


Parse new messages. Flag messages containing ‘rare’ words as anomalous.

Use learned patterns in messages to identify when a rare message is being sent

Anomaly Detection Algorithms (5)

One Support Vector Machine Plugin (ML)

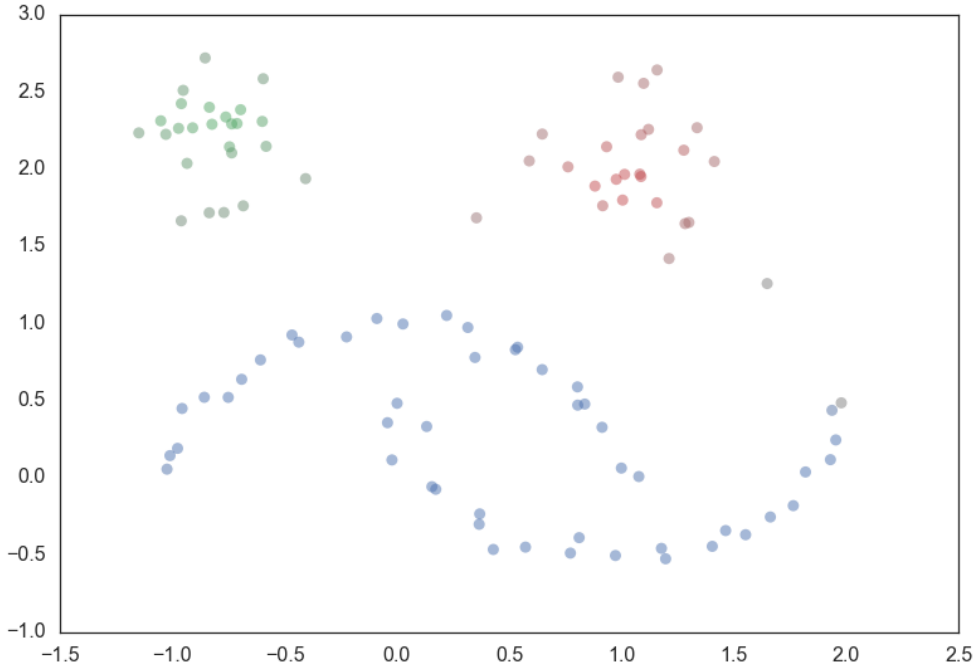


Source: scikit-learn OneSVM. This image does not contain any data or results from our algorithm training.

OneSVM learns specified features of a training dataset to flag anomalous values in test dataset

Anomaly Detection Algorithms (6)

Hierarchical Density-based Clustering Plugin (ML)

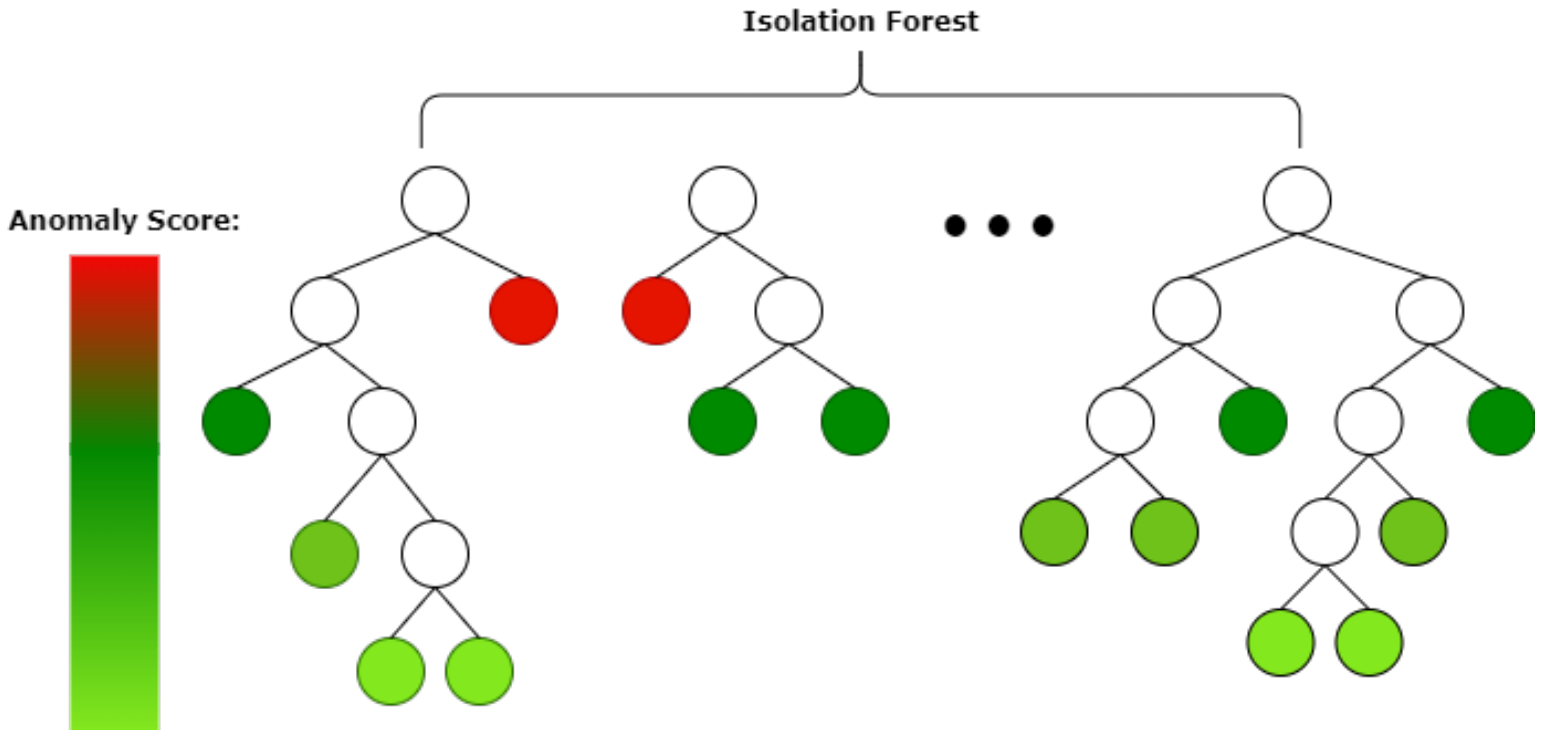


Source: scikit-learn HDBSCAN. This image does not contain any data or results from our algorithm training.

HDBSCAN generates clusters from nominal training data and will attempt to categorize new test data in known clusters

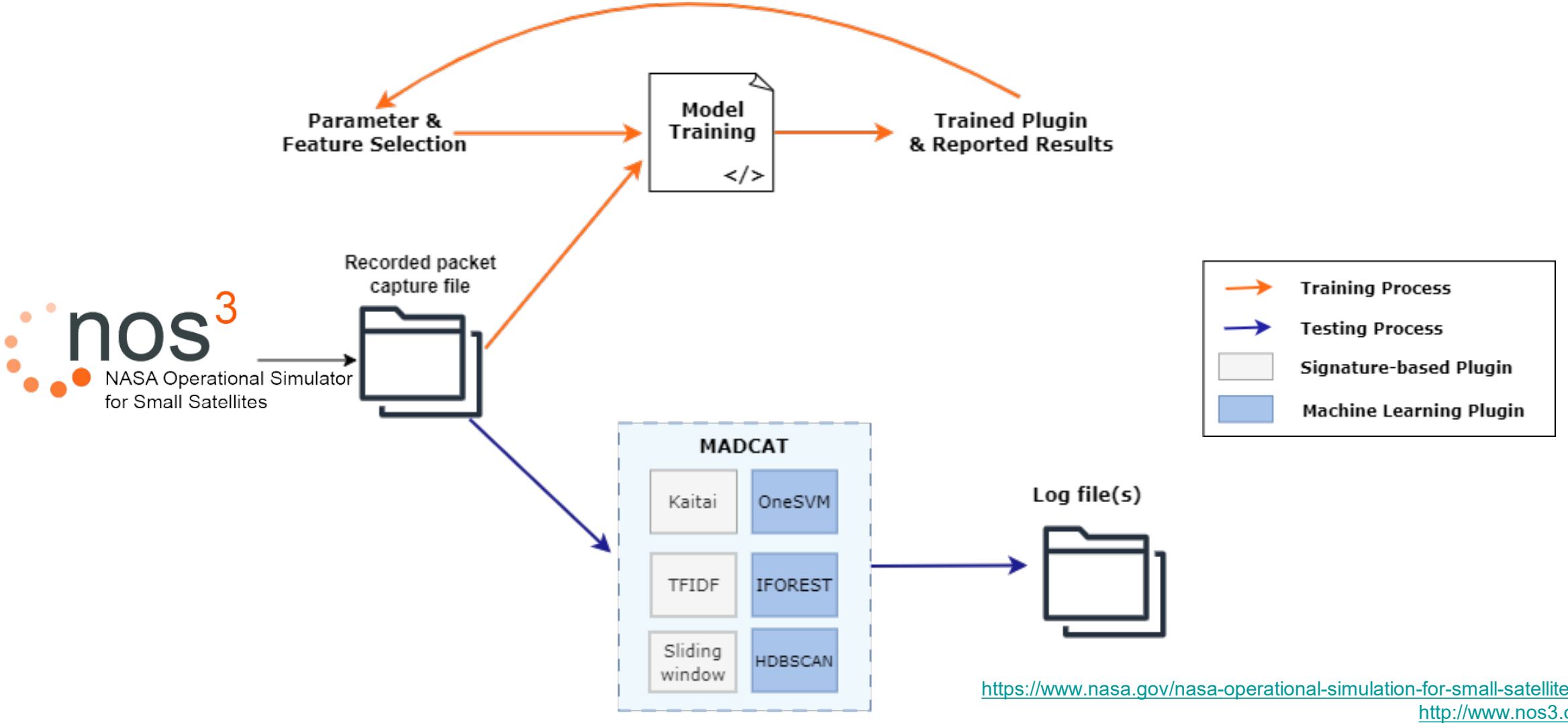
Anomaly Detection Algorithms (7)

Isolation Forest Plugin (ML)



Based on principle that anomalies are observations that are few and different. Anomalous messages are easy to isolate in binary trees

MADCAT Current Architecture



https://www.nasa.gov/nasa-operational-simulation-for-small-satellites/
http://www.nos3.org

Findings to Date

Algorithm	Anomaly class detected	Implemented	Preliminary performance and/or strengths	Observed weaknesses
Kaitai	Format	✓	Quick processing, parses messages for other plugins, adaptable for other messages	Can only check detect validity of message
Sliding Window	Volume and Rate	✓	Quick processing, intuitive, easy to implement	Does not consider timing of messages
TFIDF	Content	✗	Easily expanded with other NLP techniques, new way of looking at messages	Requires large amounts of nominal training data, retraining needed
OneSVM	Content	✓	Low memory usage, effective on high dimensional data	Complicated hyperparameters to tune, requires some anomalous data in training dataset
HDBSCAN	Content	✓	Intuitive, produces meaningful visual results, good performance with command messages	Struggles to cluster headers of telemetry messages
Isolation Forest	Content	✗	Low memory usage, quick processing	Requires some anomalous data in training dataset

Lessons Learned and Best Practices

Simple models should be used to improve explainability and reduce complexity

ML models require large amounts of data for training

Training datasets must be realistic and comprehensive

Maintain collaborative relaitonship with mission users/data providors to understand anomalies

Models must be retrained over time as data and needs change

Keep up with the latest secure AI news and guidelines

Future Work

- Integrate MADCAT into a Ground System testbed and test the message protocol anomaly detection capability in a real-time scenario
- Continue to research, implement and test anomaly detection algorithms for commanding and telemetry message protocols
  - Continue to acquire sets of mission data to support this effort
- Extend this technology for use on additional protocols such as C2MS
- Continue to engage with the potential user community to encourage and support the use of this technology in their ground system solutions

Questions? Comments?

Thank you ☺

Backup

Term	Definition
Message	A broad term for a discrete unit of communication. Can include packets, sessions, telemetry, frames, etc.
Valid Message	Messages that are properly formatted and contain legal and approved commands for a given mission operation and are sent from an approved source to an approved destination.
Nominal Message	Messages that are considered by both the ground and space segments of a mission to be ordinary or expected at a given point during mission operations.
Anomalous Message	Messages that through any one or more categories of: context, content, volume & rate, format, communicants, temporal, or other unlisted qualifications, could be considered by a mission operation to not be nominal messages

Backup

Anomaly Class	Definition	Example
Anomalous Context Message	Messages that are Valid but when evaluated in relation to Valid messages sent prior to the anomalous message or given context of the status of the ground or space segments of the mission, would not be considered to be nominal.	A Valid message commanding a spacecraft to deploy science system antennas after the science system antennas have already been deployed.
Anomalous Content Message	Messages that are not Valid messages due to invalid commands or invalid command parameters for the given mission.	An invalid message issuing a command to set the value of a science instrument to 11 when the science instrument has a maximum setting of 10
Anomalous Volume and Rate Messages	Messages that are Valid or invalid that are transmitted in such a high volume or at a considerable rate that would not be considered to be nominal.	A series of 2000 Valid messages sent over a period of 20 seconds to a spacecraft that typically receives fewer than 100 messages per day.
Anomalous Format Message	Messages that are not Valid messages due to the format of any of the layers of the message containing missing fields, improperly formatted values, or being ordered in a way not prescribed by established standards.	An invalid message containing an IPv6 address in an IPv4 address location
Anomalous Communicants Message	Messages sent to/from unanticipated communicants, ports, or devices, or to/from communicants, ports, or devices that would not be considered to be nominal.	A Valid message sent from the backup ground station while the primary ground station is sending messages
Anomalous Temporal Message	Messages that are Valid but are sent at a time of day that would not be considered to be nominal.	A Valid message that is typically sent at 1100 GMT that is sent at 1900 GMT.

Backup

Acronym	Definition
C2MS	Command and Control Message Specification
CCSDS	Consultative Committee for Space Data Systems
CFDP	CCSDS File Delivery Protocol
HDBSCAN	Hierarchical Density-based Clustering
MADCAT	Message Anomaly Detection for Command and Telemetry Systems
NOS^3	NASA Operational Simulator for Small Satellites
OMG	Object Management Group
OneSVM	One Support Vector Machine
SPP	Space Packet Protocol
TFIDF	Term Frequency Inverse Document Frequency

References

- *Scikit-learn: Machine Learning in Python*, Pedregosa et al., Journal of Machine Learning Research (JMLR) 12, pp. 2825-2830, 2011.
- McInnes L, Healy J. *Accelerated Hierarchical Density Based Clustering* In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, pp 33-42. 2017