Fault Management Technical Performance Measures (TPMs)

Dr. Stephen B. Johnson
sjohns22@uccs.edu and stephen.b.johnson@nasa.gov

NASA MSFC EV43 Integrated System Health Management and Automation Branch
&
University of Colorado at Colorado Springs - Jacobs ESTS

10 April 2012

Fault Management Theory

- **The operational subset of System Health Management—protection/preservation of system functionality**

- **A set of "meta-control loops" that aim to restore the system to a state that is controllable by nominal (passive and/or active) control systems**
  - Usually the regular (passive or active) control system has been compromised because (for active control) its sensors, processing, or actuators are compromised, or (for passive control) the design margins have eroded to zero or negative.

- **Each loop: failure detection, fault isolation, decision, and response**
  - Variants include different detection types (anomalies or degradations), prognostics, failure identification, and different response types (recovery, goal change, operational failure prevention).

- **The newly-controllable state might or might not be to the system's original goals.**
  - If original goals sustained, then we have failure recovery.
    - Example: computer voting, redundancy management
  - If original goals not sustained, then we have a goal change, usually to some subset or degraded version of the original system goals.
    - Example: vehicle safing or crew abort

- **Control theory applies: state estimation and control = failure detection/isolation and failure response decision/execution**

- **Systems theory applies: system boundaries and recursion**

Overall FM TPMs

- **Purpose of FM is to protect / preserve system functionality, so the overall metric(s) for FM must be against system functionality**

- **In general, there are 3 basic measures:**
  - Reliability (loss of mission---LOM)
  - Availability (mission performance when needed, as often as needed)
  - Safety (loss of crew---LOC, sometimes also loss of critical infrastructure)
    - (Loss of Vehicle (LOV) is also a common top measure)
  - Overall---dependability = RAS (Reliability, Availability, Safety)

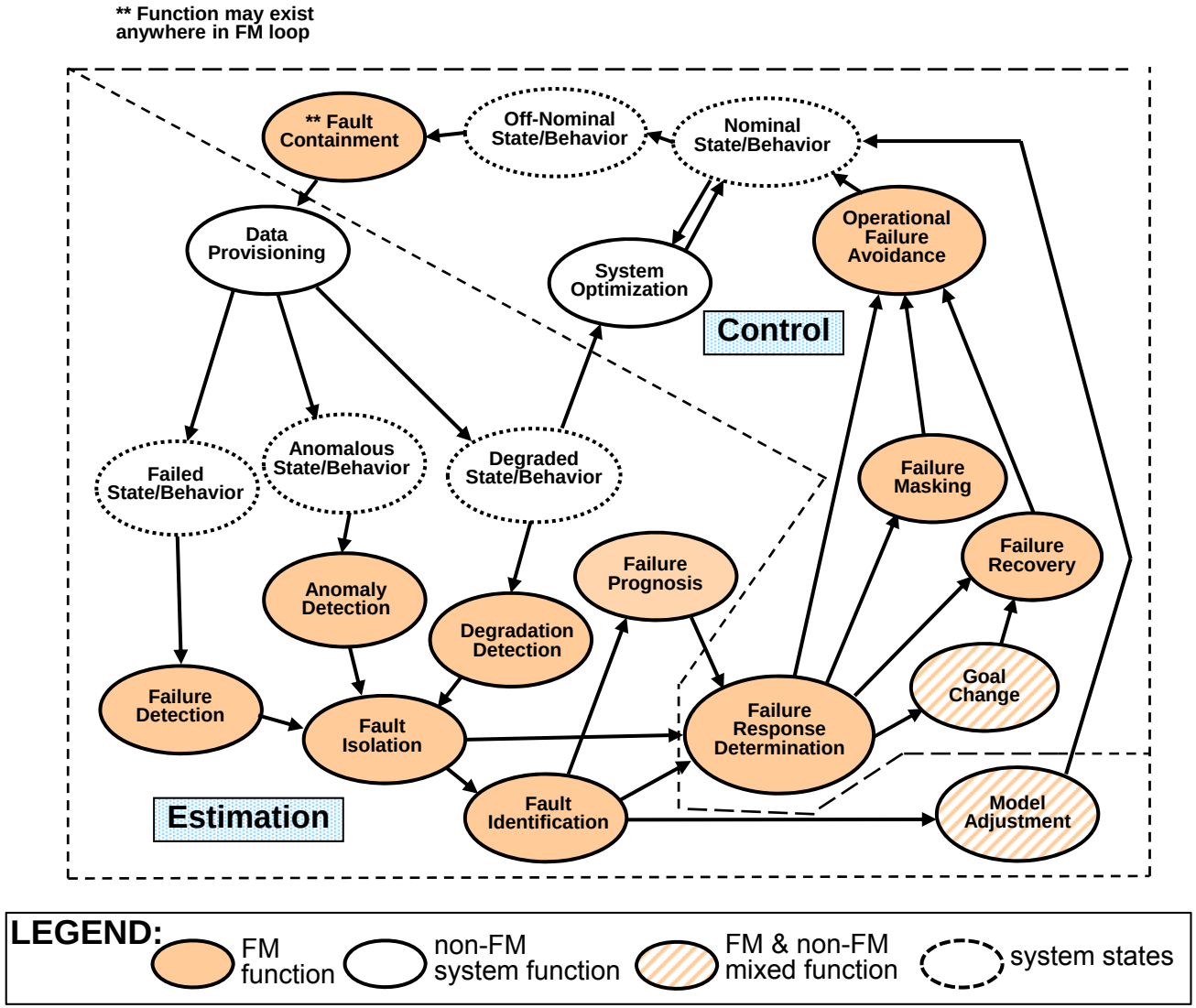
- **The metrics are inherently probabilistic**
  - Estimating uncertainties, not just the mean, is important

- **Within NASA, these top metrics are generally the responsibility of S&MA (Safety & Mission Assurance)**

- **FM's role is to provide the effectiveness measures of the FM design, which contribute to the overall S&MA metrics of LOC, LOM, availability, LOV, etc.**
  - Any engineering organization must understand the effectiveness of their designs

- **This means FM is responsible for calculating the probabilistic effectiveness of each FM Control Loop to protect the mission, crew, and availability**
  - The results passed along to S&MA, which integrates into the overall RAS numbers

Fault Management Functions as State Estimation and Control



Updated since Johnson & Day 2010

Failure Detection Coverage

- **Usually assessed "from the bottom up" using the failure modes**

- **However, "which" failure modes?**
  - NASA S&MA process: Crit 1, 1R, 2, 3, select which criticality is relevant
  - Selection of Criticality level depends on a prior analysis of the worst case EFFECTS of the failure modes, to determine how they affect system functionality, typically through FMEA.
  - This S&MA FMEA process is generally not formal, and can have mistakes… what process was used to determine what the worst case effect might be?
  - In some situations, the worst case may not happen, do you count it in the statistics?
  - Or, use some other selection method, such as all failure modes in a subsystem or box, or all Crit 1 or 2 failure modes in a box, etc.

- **Simplest Approach – non-probabilistic sum**
  - Sum up the failure modes selected as above, and those potentially detected by the failure detection mechanism are divided by the total: $C = \Sigma F_{det} / \Sigma F_{tot}$

- **More complex approach – probability-weighted sum of selected failure modes, divided by probability-weighted sum of all failure modes**

- **Alternative, and perhaps more insightful method**
  - What really matters is the probabilistically weighted coverage of those failures that can compromise a certain function.
  - This is implicit in the Criticality selection of which failure modes to include in the coverage estimate, but only against safety or mission-related functions "in general".

Degradation and Anomaly Detection Coverage

- **Degradation detection coverage**
  - In essence, yellow-line or degradation trending detection coverage
  - Inherently similar methods to failure detection coverage, under the assumption that degradations are
    - Lower-level failures that cannot be observed directly with existing measurements/observations, or
    - Lower-level functions that we observe, but are failures at lower levels not compromising the higher-level function of interest (human spaceflight C&W showing loss of redundancy is a good example!!), or
    - Erosion of margins that may in the future lead to failures that are predicted in FMEAs.

- **Anomaly detection coverage**
  - Anomaly = unexpected performance of intended function
  - By definition, being unexpected, they are not predictable, and use of pre-predicted failure modes in an assessment is inherently incomplete.
  - Any TPM for anomaly detection coverage is likely to use a top-down method against intended functionality.
  - A research topic

Failure Detection:
False Positive (FP) / False Negative (FN)

- **FP/FN methodology derived from radar operational performance analysis – Receiver-Operator Characteristic (ROC) Curves.**

- **True Positive (TP) / True Negative (TN) / FP / FN**
  - False Positive = identifying a failure when no actual failure exists
  - False Negative = not identifying a failure when an actual failure exists
  - True Positive = identifying a failure when an actual failure exists
  - True Negative = not identifying a failure when no actual failure exists

- **Every FM detection (failure/degradation/anomaly) inserted into the system creates the possibility of "false alarm" (FP), and could fail to detect the real problem (FN).**
  - Example: SLS-MPCV--- every "abort trigger" decreases Loss of Crew (TP), but increases Loss of Mission (LOM) (FP); FN treated as if no detection exists

- **Equations must account for the physics of the behavior being monitored, and the failures of the hardware/software/human components of the detection.**
  - Can lead to quite complex equations, and dependencies on engineering, operations, and S&MA for data

- **Also---time to detect will be needed for response purposes.**

Fault Isolation and Identification

- **Fault Isolation = Determining the possible locations of a hypothesized failure or anomaly cause, to a defined level of granularity. *(note this differs from the EE usage)***
- **Fault Identification = Determining the possible causes of a failure or anomaly.**
- **Both essentially the same in methodology; isolation determines the location of the cause, and identification determines the cause.**
  - Example: There are 3 components, each with 5 failure modes. A failure has been detected. Isolation determines which of the 3 components the cause resides in, and identification identifies one of the 5 failure modes inside that component as the cause.
- **As a state estimation function, ROC methodology applies (FP / FN / TP / TN), and time to isolate and identify will be needed for response metrics**
- **Typically, the metrics must deal with "ambiguity groups"**
  - Example: In the example above, isolation may only determine that the cause is in components A and B, but not C. Identification may then identify that of the 10 failure modes in components A and B, 7 of the 10 failure modes are possible. In each case, the 2 components and the 7 failure modes are considered "ambiguity groups" in which you cannot distinguish between locations or failure modes, respectively.
- **Metrics useful to determine if a cause really is in a string of redundant components, if sensors are sufficient to identify causes (root cause analysis), locate a component for repair, etc.**

Prognostics

- **Prognostics = predicting the time at which a component will no longer perform its intended function.**
- **Primary output of prognostics is "Remaining Useful Life" (RUL), which is the time from the present to the time the component fails (no longer performs intended function).**
- **The purpose of prognostics and RUL is to provide information that guides proper response, which consists usually of**
  - Repair
  - Replacement
  - Retirement
  - Other mitigations to extend RUL
- **RUL uses TPMs for accuracy, precision, and convergence.**
  - Accuracy = closeness of predictive estimate to actual value
  - Precision = variability of prediction
  - Convergence = quantification of how accuracy and precision measures improve as RUL decreases (improvement over time before failure occurs)
- **TPMs use sophisticated mathematical, physical, statistical methods.**
- **Do ROC techniques apply?? (TP/TN/FP/FN)**

Model Adjustment

- **Adjustments/modifications to the models on which expectations of system behavior are based**
- **Typical activity after root cause analysis of an anomaly**
  - Determine what really happened, change your models to account for the new knowledge, so if it happens in the future it will not be classified as an anomaly.
- **I know of no research papers on this subject/function or its TPMs.**
- **Research needed**

Failure Response Determination (Decision Function)

- **Failure Response Determination = Selecting actions to mitigate a current or future failure.**
- **Ultimate metric is the probability that the correct action(s) was/is/will be selected.**
- **This is a research topic---I know of no existing metrics or research for this FM function.**

Response Effectiveness

- **Several possible responses**
  - Goal change
  - Failure recovery
  - Failure masking
  - Operational Failure Avoidance
- **All responses must operate more quickly than the failure effects that they are mitigating / responding to**
  - Therefore a race condition analysis of failure effects versus failure responses is required.
  - The race includes latencies for detection, diagnostics (isolation/identification), decision, and response, versus failure effect propagations to the “Critical Failure Effect”.
- **Responses may also have design faults, or if performed by humans, human faults**
  - Most likely cause of response failure, aside from losing the race condition, is interactions between failure responses, and between responses and other system control activities
- **Overall metric is “probability of successful response” prior to the failure of the function that the FM Control Loop is protecting.**

Example: Distributed / Parallel Monitor-Response versus Single-Response Centralized Architectures

- **Trade studies of FM architectures generally determine degrees of centralization versus distribution of FM functions, and allowing of single responses at a time versus allowing parallel responses.**
  - There are other things… this is an example of a common trade.
- **What aspects of the TPMs can be used to quantitatively assess the architectures?**
- **Assume detection and isolation are the same, and the specific responses.**
- **What differs in this case are primarily the effectiveness and interaction of responses.**
- **Distributed, parallel M-R architectures are fast, but can create more response interactions.**
- **Centralized, sequential architectures are slower, but eliminate most response interactions.**
- **Divide the response metrics into a "race condition effectiveness" piece and a "response interaction" piece, and determine which is more effective against the system's RAS requirements.**
  - If speed is essential, then parallel responses are needed to win the race condition.
  - If speed is not essential, it is safer to limit yourself to sequential responses to avoid interaction failures.

Less Rigorous Alternatives

- **What if your project does not have the resources to quantitatively address all of these metrics?**
  - Most projects don't, and use qualitative requirements such as single failure tolerance.
- **Should perform top-down assessment of coverage against system functions and top-level requirements.**
  - This is not a huge effort; it is crucial to determine if you have mechanisms to detect and respond to all critical system objectives.
  - Single failure tolerance is against the critical functions!!!
- **Perform non-probability weighted assessments of coverage.**
  - Consider historical data for similar systems to determine which components and functions are more likely to fail… put efforts there.
- **FP/FN not normally addressed for detection or diagnostics.**
  - Should at least consider whether some detections may be less effective, and focus efforts to determine if these are appropriate
  - For systems with slow on-board responses (slow Time to Criticality), it may be unimportant to isolate to "the" string than to ensure that responses cover all possibilities.
  - Diagnostic effectiveness quite important when operational maintenance and repair is significant.
  - Observability for root cause analysis a critical design feature; it is almost never addressed.
- **Response effectiveness: this is a very important issue for all systems, can be assessed with physical estimates and reasoning.**

Conclusion

- **FM TPMs are ultimately tied to S&MA analyses of reliability, availability, and safety.**
- **Analyze the effectiveness of each loop, which is the summed probabilities of each of the loop component functions---detection, diagnostics (isolation/identification), decision, and response.**
  - Determine that each loop provides enough benefit to be worthwhile to counteract its false positive negative effects (false alarms) and cost.
- **Effectiveness of each loop feeds the overall estimates of RAS.**
- **Some metrics well-known:**
  - failure detection, diagnostics, prognostics
- **Some are not well known or understood---research is needed.**
  - Decision, response (interaction & race condition), anomaly detection, model adjustment
- **Good tools for these metrics mostly don't exist.**
  - Some exceptions—diagnostics is perhaps in the best shape.
- **Implication: requirements for FM should be written in forms amenable to the TPMs and associated processes (i.e. single failure tolerance) for each system.**