Research to Support Decision-Driven Data Collection
17th ICCRTS Conference
George Mason University
21 June 2012

David Jakubek
Deputy Director, Information Systems
Office of the Assistant Secretary of Defense for Research and Engineering

Heather McCreary
Research Analyst
ITT Exelis, Advanced Information Solutions

Distribution Statement A. Approved for
public release; distribution is unlimited.

Decision Driven Data Collection

Rather than attempting to “handle” increasing amounts of information … Big Data Problem

Imagine ways to reduce the volume needed to achieve the desired effect … Solve a different problem

“Purpose driven collection” … Performed today in some areas.

Apply these practices in a military environment as Decision Driven Data Collection.

Distribution Statement A. Approved for public release; distribution is unlimited.

Motivation

- Years of video footage without the people and time available.
- Sensors will improve due to technology opportunity.
- Use of sensors and platforms will expand due to availability and need.
- Recent trends towards being reactive.
- Manpower and time are critical constraints for military operations.

Distribution Statement A. Approved for public release; distribution is unlimited.

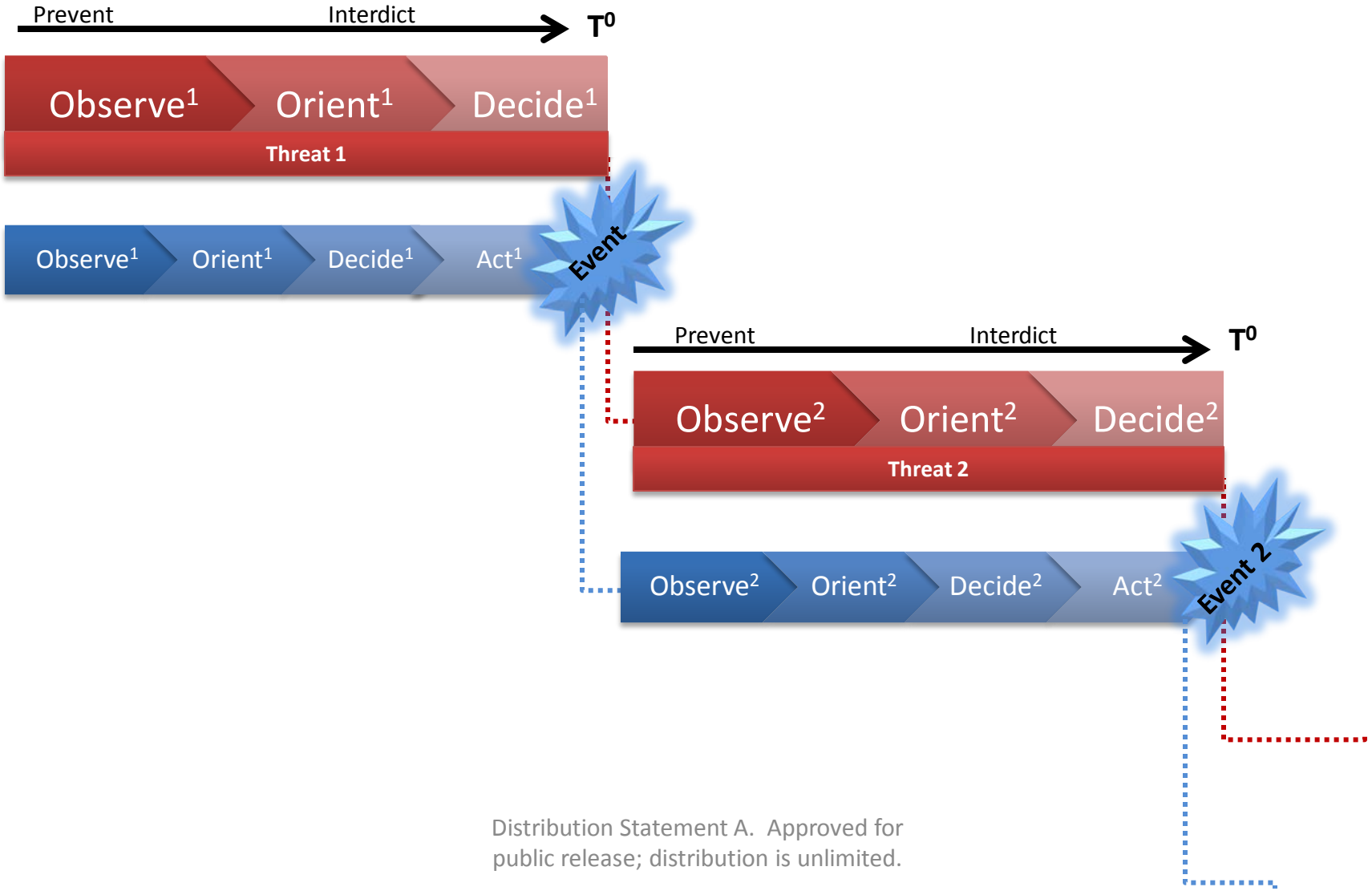
Intelligence and Command and Control

- Intelligence
  - "[…] the range of activities—from targeting through information collection to analysis and dissemination—conducted in secret, and aimed at maintaining or enhancing relative security by providing forewarning of threats or potential threats in a manner that allows for the timely implementation of a preventive policy or strategy, including, where deemed desirable, covert activities"*
- Command and Control
  - Observe-Orient-Decide-Act cycle as defined by USAF Col. John Boyd.
  - Recurring process typically described as a decision and action loop.
  - In some contexts the goal is to accelerated this process to get inside an opponents decision cycle.

* Gill, P., Phythian, M. (2004). *Issues in the Theorization of Intelligence*. Paper presented at the International Studies Association conference in Montreal, March 2004.

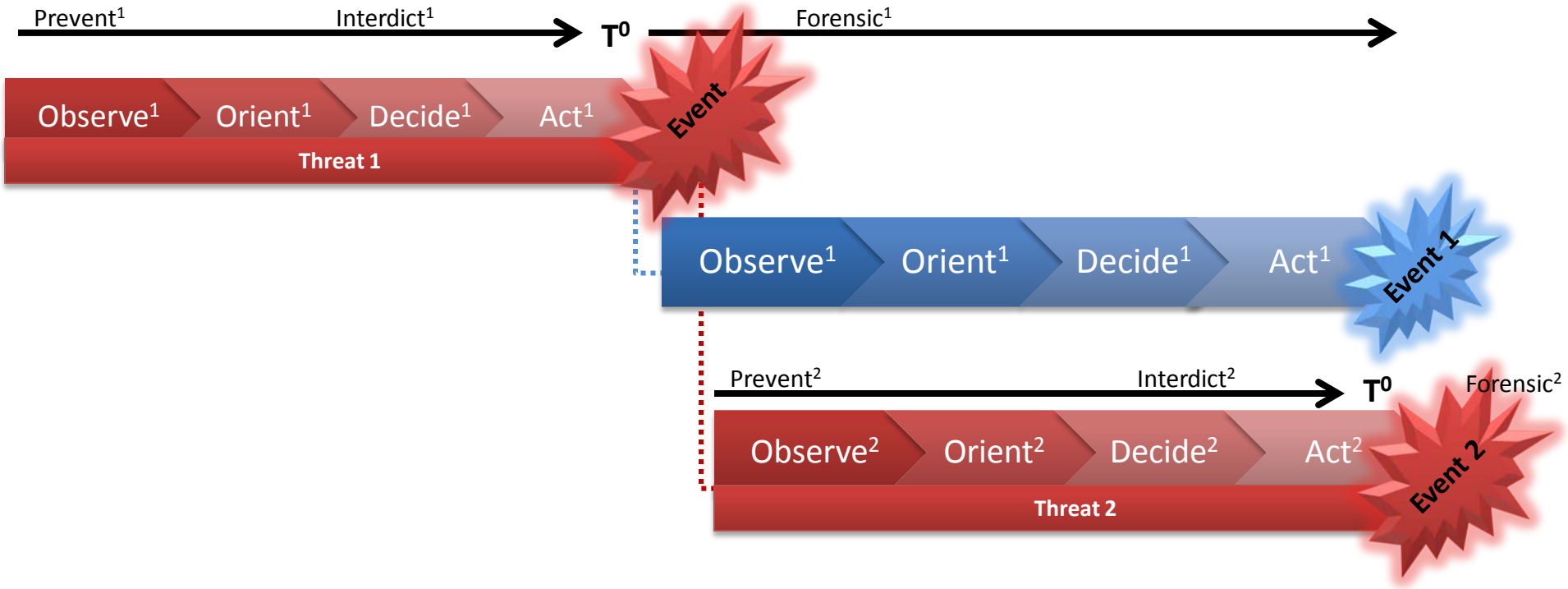
Distribution Statement A. Approved for public release; distribution is unlimited.

Ideal Intelligence and C2 Process



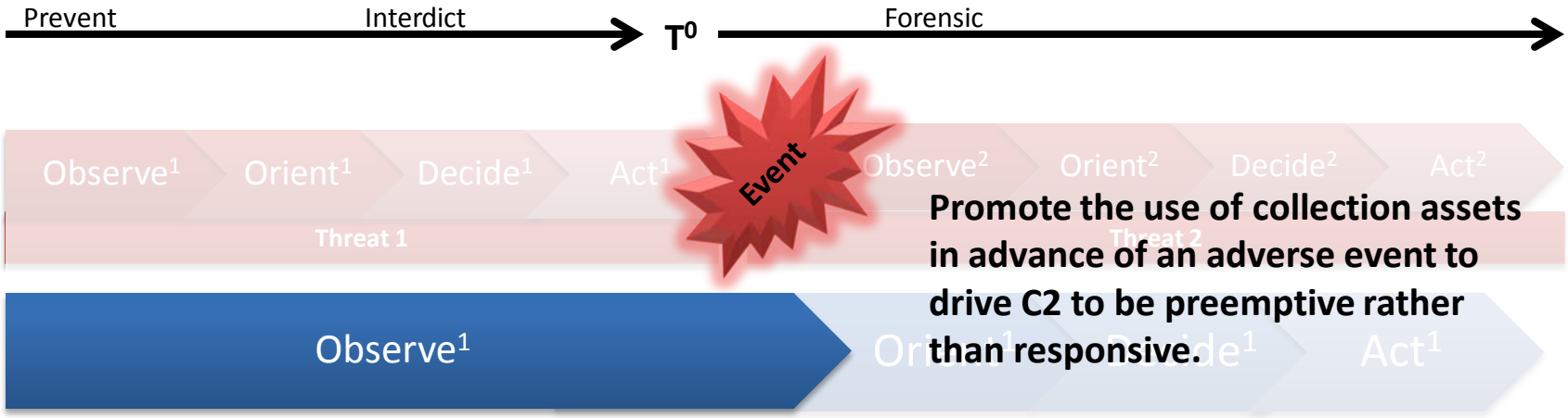
Distribution Statement A. Approved for public release; distribution is unlimited.

Reactive Response



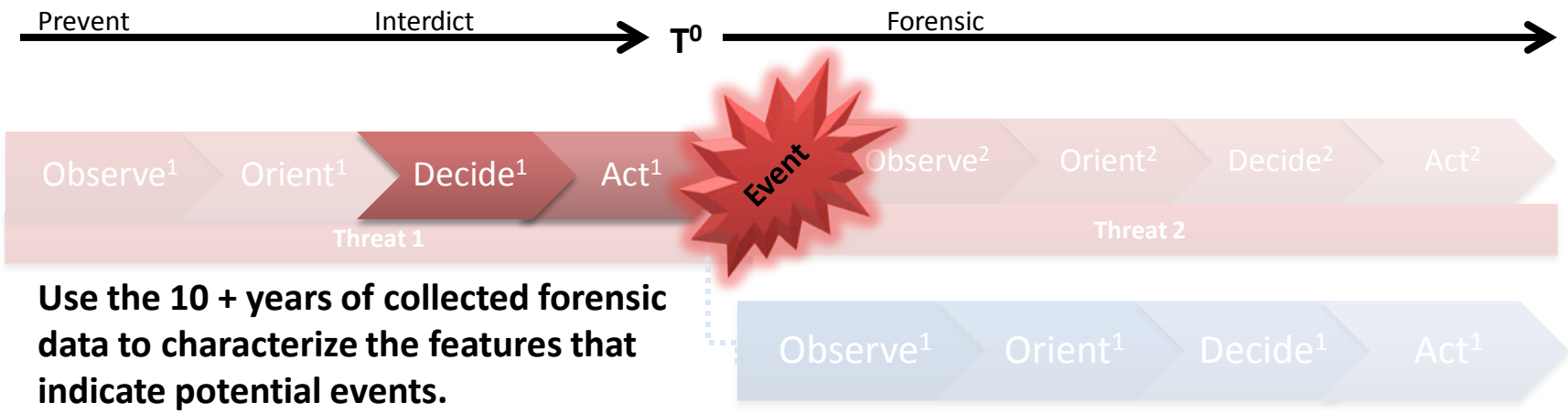
Distribution Statement A. Approved for public release; distribution is unlimited.

Decision Driven Data Collection Step 1



Distribution Statement A. Approved for public release; distribution is unlimited.

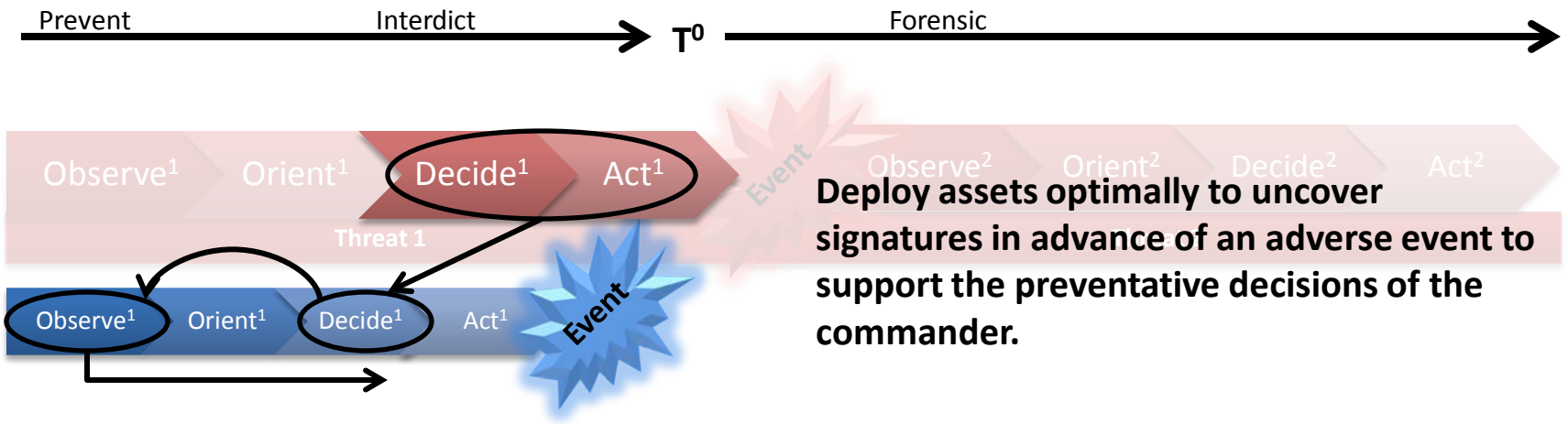
Decision Driven Data Collection Step 2



Use the 10 + years of collected forensic data to characterize the features that indicate potential events.

Distribution Statement A. Approved for public release; distribution is unlimited.

Decision Driven Data Collection Step 3



Deploy assets optimally to uncover signatures in advance of an adverse event to support the preventative decisions of the commander.

Distribution Statement A. Approved for public release; distribution is unlimited.

Near Term Activity (1-3 years)

- Business process analysis is the critical activity.
- Fixed constraints
  - Number of people (analysts, operators, etc.)
  - Level of training
  - Amount a data generated
  - Delays in transferring from collection asset to point of analysis
  - Addition of other contextual information
- Potential results from a business process analysis
  - Define a new required end state. For example, predictive analysis for interdicting as opposed to post event forensic analysis.
  - Determine how resources constraints listed above are then allocated to achieve this new desired end state.
  - Transform existing processes into a specific set of tasks that are outcome oriented and focus on using a people and time more effectively.
  - Develop the observables and define signatures for detection prior to adverse events.
  - Develop measures of effectiveness of the desired end to end of effect.

Distribution Statement A. Approved for public release; distribution is unlimited.

Mid-term Activity (3-5 years)

- Build off of the near term activity with additional research to
  - Use the large amount of data already available and collected
  - Identify the measurable observables that can be collected by existing sensors
  - Discovery and characterization of high-value data to support the identification of “activities” leading up to events
- Challenge in the past has been balancing optimized performance with resources available.
- Move more towards prevention and interdiction as opposed to reaction.
- Today’s opportunity
  - Data availability to identify new signatures
  - Sensor performance improvements to detect new signatures
- Research can investigate how these signatures can be defined to provide future solutions as well as an assessment methodology and metrics to determine how they provide the desired mission effect.
  - Research disciplines within physical, cultural and behavioral science can be used to uncover observable signatures.
  - Goal is to define automated exploitation capabilities and change data collection in new and novel ways to support planning of alternative courses of action that more accurately reflect the definition of intelligence.

Distribution Statement A. Approved for public release; distribution is unlimited.

Far Term Activities (>5 years) (1of2)

- The current trend in research is to accommodate large data volumes through widening the pipes, employing preprocessing, and compression techniques. These methods may reduce the amount of time it takes data to get from the sensor to the analyst, however they do not reduce the data volume encountered by analysis resources.
- Fundamental research in computer science is required to investigate:
  - New techniques that allow for selective reduction in collection, transmission, processing, etc. should be explored as a means close the data volume and analysis resource gap.
  - New techniques to address pedigree and uncertainty.

Distribution Statement A. Approved for public release; distribution is unlimited.

Far Term Activities (>5 years)(2of2)

- Fundamental research in mathematics and statistics is required.
- Research in abstraction, summarization, and visualization of complex data to support warfighter/subject matter "novice" use in the pursuit of scientific discovery, agile decision making, mission responsiveness, and predictive intelligence capability.
  - Graph theory and semantic web are used to improve context without excessive translation into outdated computer architecture models that may lend to improved representations.
- Innovative ways to overcome traditional sources of bias
  - Sensor bias, for example, is impacted by the collection method which can lead to artificial impressions of network and activity shifts in systems over time.
  - Sampling bias, as another example, can lead to an imbalance of coverage creating gaps overtime leading to risks of "undiscoverable nodes".
  - Examine issues with the reuse of existing data sources
  - While statistics is an established area, the use of statistical methods for certain domain areas remain weak and the small number of research statisticians within the DoD focus on more concrete applications such as UGV and robotics.
- Metrics to define the limitations and parameters of data are needed to convey context and ensure that the integrity of data is communicated to the user.

Distribution Statement A. Approved for public release; distribution is unlimited.

Summary

- Big data or large data is perceived as both a problem and an opportunity.
- Current Defense Science and Technology “Data to Decision” priority seeks to reduce the cycle time and manpower required to conduct analysis in support of military missions.
- Decision Driven Data Collection offers different perspective on the issue from what would be typically considered big data or large data.
  - Continue to exploit the full capability of improved sensor performance. Typical large data problem approach.
  - Additionally, focus manpower and time to achieve a desired effect. This is different from a large data problem approach and not limited to post event processing and forensic analysis.
  - Small changes can be instituted in the near term.
  - Emerging research provides an even greater opportunity to explore improved capability and processes.

Distribution Statement A. Approved for public release; distribution is unlimited.