Northrop Grumman Cybersecurity Research Consortium (NGCRC)

Intelligent Autonomous Systems based on Data Analytics and Machine Learning

2017-2018 Report

08 November 2018

Bharat Bhargava
Purdue University

Technical Champions: Jason Kobes, Paul Conoval, Jeffrey Ciocco, Will Chambers, Miguel Ochoa, Steve Seaberg, Peter Meloy, Jessica Trombley-Owens, Robert Pike, Brock Bose, Sam Shekar, Roderick Son

Intelligent Autonomous Systems

NORTHROP GRUMMAN

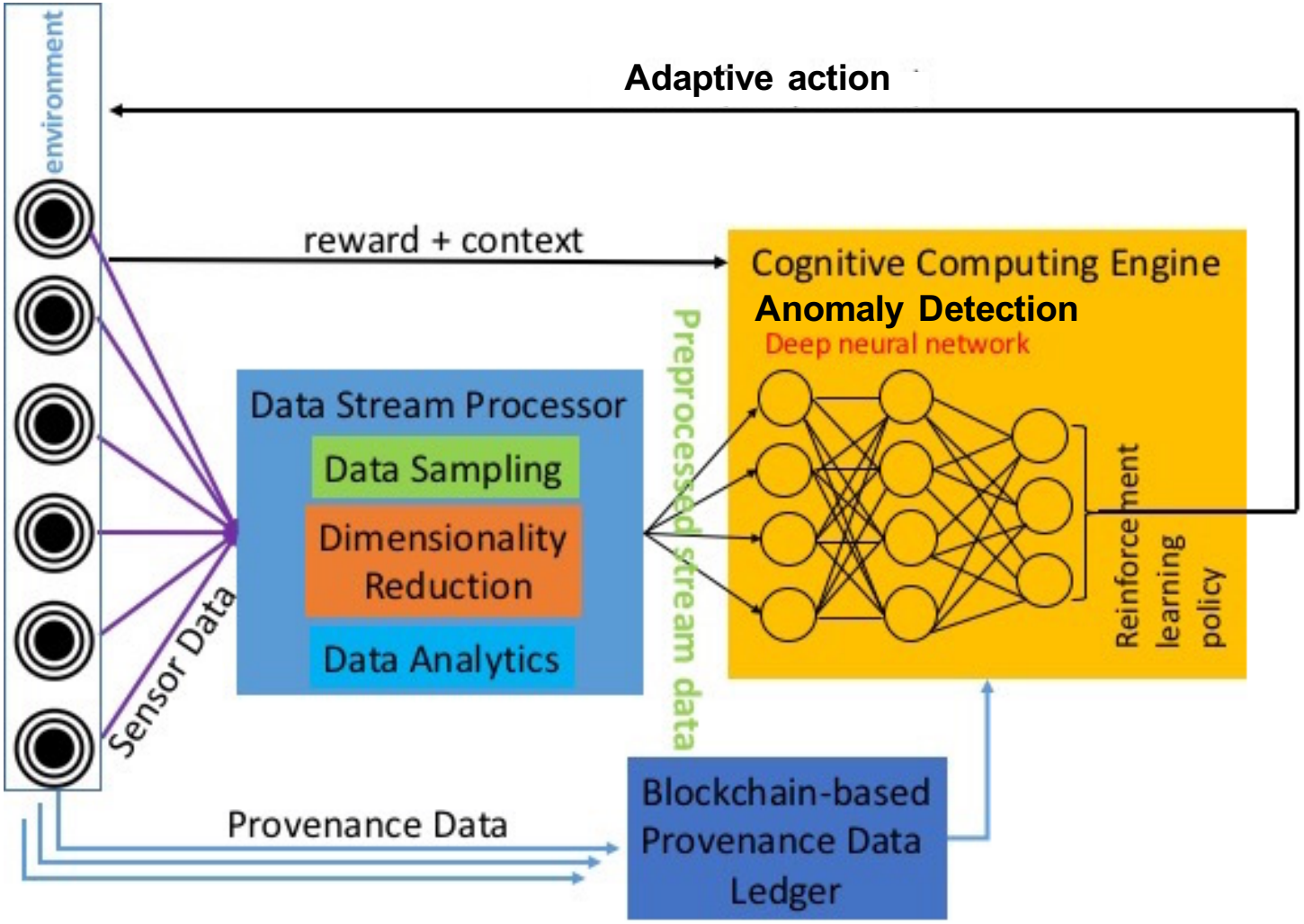
- Autonomous Systems should be
  - Able to perform complex tasks without or with limited ongoing connection to humans.
  - Cognitive enough to act without a human's judgment lapses or execution inadequacies.

- Intelligent Autonomous Systems (IAS) are characterized as highly **Cognitive**, effective in **Knowledge Discovery**, **Reflexive**, and **Trusted**.

- The focus of this research was on the smart cyber systems.

Comprehensive IAS Architecture

NORTHROP GRUMMAN



Accomplished Research Tasks

NORTHROP GRUMMAN

- We completed and demonstrated the tasks that advance the major characteristics of autonomous systems:
  - **Cognitive Autonomy**
    - Developed light-weight ML / deep learning models for system profiling (black/white listing) & anomaly detection.
  - **Knowledge Discovery**
    - Developed fuzzy-based clustering for scalable learning to discover new knowledge & patterns.
    - Developed privacy-preserving aggregated data analytics.
  - **Reflexivity**
    - Implemented combinatorial design-based graceful degradations with Bayesian inference.
  - **Trust**
    - Implemented Blockchain for provenance storage.

Publications

https://www.cs.purdue.edu/homes/bb/#research
https://www.cs.purdue.edu/homes/bb/#colloquia

NORTHROP GRUMMAN

1. G. Mani, B. Bhargava, P. Angin, M. Villarreal-Vasquez, D. Ulybyshev, D. Steiner, J. Kobes. "Machine Learning Models to Enhance the Science of Cognitive Autonomy." In *IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE),* Laguna Hills, 2018
2. G. Mani, B. Bhargava, B. Shivakumar, J. Kobes. "Incremental Learning Through Graceful Degradations in Autonomous Systems." In *IEEE International Conference on Cognitive Computing (ICCC),* San Francisco, 2018.
3. G. Mani, B. Bhargava, J. Kobes. "Scalable Deep Learning Through Fuzzy-based Clustering in Autonomous Systems." In *IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE),* Laguna Hills, 2018
4. D. Ulybyshev, M. Villarreal-Vasquez, B. Bhargava, G. Mani, S. Seaberg, R. Pike, J. Kobes. "Blockhub: Blockchain-Based Software Development System for Untrusted Environments." In *IEEE International Conference on Cloud Computing (CLOUD),* San Francisco, 2018.
5. G. Mani, D. Ulybyshev, B. Bhargava, J. Kobes, P. Goyal. "Autonomous Aggregate Data Analytics in Untrusted Cloud." In *IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE),* Laguna Hills, 2018
6. D. Ulybyshev, B. Bhargava, A. Alsalem. "Secure Data Exchange and Data Leakage Detection in Untrusted Cloud." In *International Conference on Applications of Computing and Communication Technologies (ICACCT).* Zurich, 2018 Springer.

Abstract

NORTHROP GRUMMAN

“Machine Learning Models to Enhance the Science of Cognitive Autonomy.”

Examples of IAS include software systems that are capable of automatic reconfiguration, autonomous vehicles, network of sensors with reconfigurable sensory platforms, and an unmanned aerial vehicle (UAV) respecting privacy by deciding to turn off its camera when pointing inside a private residence. Research contributes to build systems that can monitor their environment and interactions, learn their capability as well limitations, and adapt to meet the mission objectives with limited or no human intervention. The systems is fail-safe and allows for graceful degradations while continuing to meet the mission objectives.

In this paper, we propose new methodologies and workflows, and develop approaches that can advance the science of autonomy in smart systems through enhancements in real- time control, auto-reconfigurability, monitoring, adaptability, and trust.

Abstract

NORTHROP GRUMMAN

"Incremental Learning Through Graceful Degradations in Autonomous Systems."

Intelligent Autonomous Systems (IAS) are highly cognitive, reflexive, multitasking, trustworthy (secure and ethical), and rich in knowledge discovery. IAS are deployed in dynamic environments and connected with numerous devices of different types, and receive large sets of diverse data. They receive new types of raw data that was not present in either training or testing data sets thus they are unknown to the learning models. In a dynamic environment, these unknown data objects cannot be ignored as anomalies. Hence the learning models should provide incremental guarantees to IAS for learning and adapting in the presence of unknown data. The model should support progressive enhancements when the environment behaves as expected or graceful degradations when it does not. In the case of graceful degradations, there are two alternatives for IAS: (1) weaken the acceptance test of data object (operating at a lower capacity) or (2) replace primary system with a replica or an alternate system that can pass the acceptance test.

In this paper, we provide a combinatorial design-MACROF configuration-built with balanced incomplete block design to support graceful degradations in IAS and aid them to adapt in dynamic environments. The architecture provides stable and robust degradations in unpredictable operating environments with limited number of replicas. Since the replicas receive frequent updates from primary systems, they can take over primary system's functionality immediately after an adverse event. We proposed a Bayesian learning model to dynamically change the frequency of updates. Our experimental results show that MACROF configuration provides an efficient replication scheme to support graceful degradations in autonomous systems.

Abstract

NORTHROP GRUMMAN

"Scalable Deep Learning Through Fuzzy-based Clustering in Autonomous Systems

Autonomous cyber systems continuously receive large streams of diverse data from numerous entities operating and interacting in their environment. It is imperative that the learning models in autonomous systems to scale up to process the new and unknown data items. Scalable learning is a method to achieve maximum classification without rejecting any unknown data item that were not present in the training or testing datasets as anomalies.

In this paper, we present Bitwise Fuzzy-based Clustering (BFC) technique through errorcorrecting codes to address the problem. Through BFC, we can approximate the classes of multidimensional features of data items by reversing standard forward error-correction coding. Approximating classes problems generally arise in autonomous systems that are processing fuzzily cataloged data items. These data items can be classified by applying binary vectors to their corresponding features (1: feature is present or 0: feature is absent) to obtain message words. These codewords are used as cluster centers. In BFC technique, binary vectors of 23 bits are mapped into codewords (labels or indices) of 12 bits. Two different 23-bit binary vectors with the Hamming distance of 2 will have a few common labels. This setting enables the clustering of neighboring 23-bit binary vectors with at most 2-bit variation (mismatch) from a given input. BFC technique has 223 codeword space, which makes it ideal for scalability in clustering of millions of categories and their associated features. With reasonable redundancy, the clustering can be accomplished in O (N) time.

Abstract

NORTHROP GRUMMAN

“Blockhub: Blockchain-Based Software Development System for Untrusted Environments.”

To ensure integrity, trust, immutability and authenticity of software and information (cyber data, user data and attack event data) in a collaborative environment, research is needed for cross-domain data communication, global software collaboration, sharing, access auditing and accountability. Blockchain technology can significantly automate the software export auditing and tracking processes. It allows to track and control what data or software components are shared between entities across multiple security domains.

The blockchain-based solution relies on role-based and attribute-based access control and prevents unauthorized data accesses. It guarantees integrity of provenance data on who updated what software module and when. Furthermore, this solution detects data leakages, made behind the scene by authorized blockchain network participants, to unauthorized entities. The approach is used for data forensics/provenance, when the identity of those entities who have accessed/ updated/ transferred the sensitive cyber data or sensitive software is determined. All the transactions in the global collaborative software development environment are recorded in the blockchain public ledger and can be verified any time in the future. Transactions can not be repudiated by invokers. We propose modified transaction validation procedure to improve performance and to protect permissioned IBM Hyperledger-based blockchains from DoS attacks, caused by bursts of invalid transactions.

Abstract

"Autonomous Aggregate Data Analytics in Untrusted Cloud

Intelligent Autonomous Systems (IAS) are highly reflexive and very cognizant about their limitations and capabilities, interactions with neighboring entities, as well as the interactions with its operational environment. IAS should be able to conduct data analytics and update policies based on those analytics. These tasks should be performed autonomously i.e. with limited or no human intervention.

In this paper, we introduce advanced aggregate analytics over untrusted cloud and autonomous policy updates as a result of those analytics. We used Active Bundle (AB), a distributed self- protecting entity, wrapped with policy enforcement engine as our implementation service. We proposed an algorithm that can enable individual ABs to grant or limit permissions to their AB peers and provide them with access to anonymized data to conduct analytics autonomously. When these processes take place, ABs do not need to rely on policy enforcement engine every time, which increases scalability. This workflow also creates an AB environment that is decentralized, privacy- preserving, and autonomous.

Abstract

NORTHROP GRUMMAN

"Secure Data Exchange and Data Leakage Detection in Untrusted Cloud."

In service-oriented architecture, services can communicate and share data amongst themselves. It is necessary to provide role-based access control for data. In addition, data leakages made by authorized insiders to unauthorized services should be detected and reported back to the data owner.

In this paper, we proposed a solution that uses role- and attribute-based access control for data exchange among services, including services hosted by untrusted environments. This approach provides data leakage prevention and detection for multiple leakage scenarios. We proposed a damage assessment model for data leakages. The implemented prototype supports a privacy-preserving exchange of Electronic Health Records that can be hosted by untrusted cloud providers, as well as detecting leakages made by insiders.

Collaboration with NGC

NORTHROP GRUMMAN

- Thanks to Paul Conoval, Jason Kobes, and Steve Seaberg for their extended discussions for setting definitive goals for autonomous systems research. Steve had discussions with ONR for deploying some ideas for consistency of information among vessels.

- IAS implementations have application to the NGC IRADs
  - Adaptive Real-Time Detection and Examination Network IRAD
  - Automated Mission Planning for Autonomous Systems IRAD
  - Enterprise Information Management System and Analytics IRAD
  - Information Analytics IRAD
  - Rapid Autonomy Prototype Implementation & Demonstration IRAD
  - Reliability Analysis Data System IRAD
  - Smart Autonomy IRAD

Implementation of Components of IAS

NORTHROP GRUMMAN

- **Cognitive Autonomy & Knowledge Discovery:**
  - Monitors and records system's activities (Data provenance and sequence of system calls)
  - Conducts privacy-preserving aggregated analytics on provenance data.
  - Utilizes Deep learning based anomaly detection by analyzing sequence of system calls.
- **Reflexivity:**
  - Adaptive actions are performed through graceful degradations without disrupting the ongoing critical processes by incremental learning.
- **Trust:**
  - Uses blockchain to store provenance data for trust.

Implementation and Deliverables

NORTHROP GRUMMAN

- **Reflexivity prototype for combinatorial replica scheme:**
  Source code: Node.js implementation, Bayesian model, simulation software developed for combinatorial design, and Data used for simulation. Demo Link: https://goo.gl/M4rXCN

  The prototype is built with FAYE framework (https://faye.jcoglan.com/node.html) with Node.js.

  Replica updates are done through a combinatorial design simulator (https://goo.gl/pgVHdk).
- **Deep Learning based anomaly detection prototype:** In progress.
- **Blockhub prototype for secure blockchain-based data distribution:**
  Source code: https://github.com/Denis-Ulybysh/Waxedprune2018
  codebase is taken from open-source "Marbles" project https://github.com/IBM-Blockchain/marbles/tree/v4.0
- **Documentation:** Demo video and User manual for running the prototype.

Demo Description

NORTHROP GRUMMAN

- There are three components that are demonstrated.
- **Demo 1 (Cognitive Autonomy/Knowledge Discovery):**
  - System is monitored and its interactions with client services are recorded as provenance data.
  - Privacy-preserving aggregated data analytics are performed on the provenance data.
  - Sensitive data is perturbed with random noise and the noise is removed at the end to obtain aggregated result, protecting the privacy of individual entities.
  - A Deep Learning based anomaly detection is implemented to protect against code-hijacking attacks.

Demo Description

NORTHROP GRUMMAN

- **Demo 2 (Reflexivity):**
  - Under anomalous operating contexts or attacks, the replicas in the replacement scheme based on Combinatorial balanced designs take over the processing from primary module.
  - Replicas are updated with system states periodically (Update interval is determined through Bayesian inference of system's operating context).
  - Unused replicas are used for other processes simultaneously, which makes the system faster and fault-tolerant.

Demo Description

NORTHROP GRUMMAN

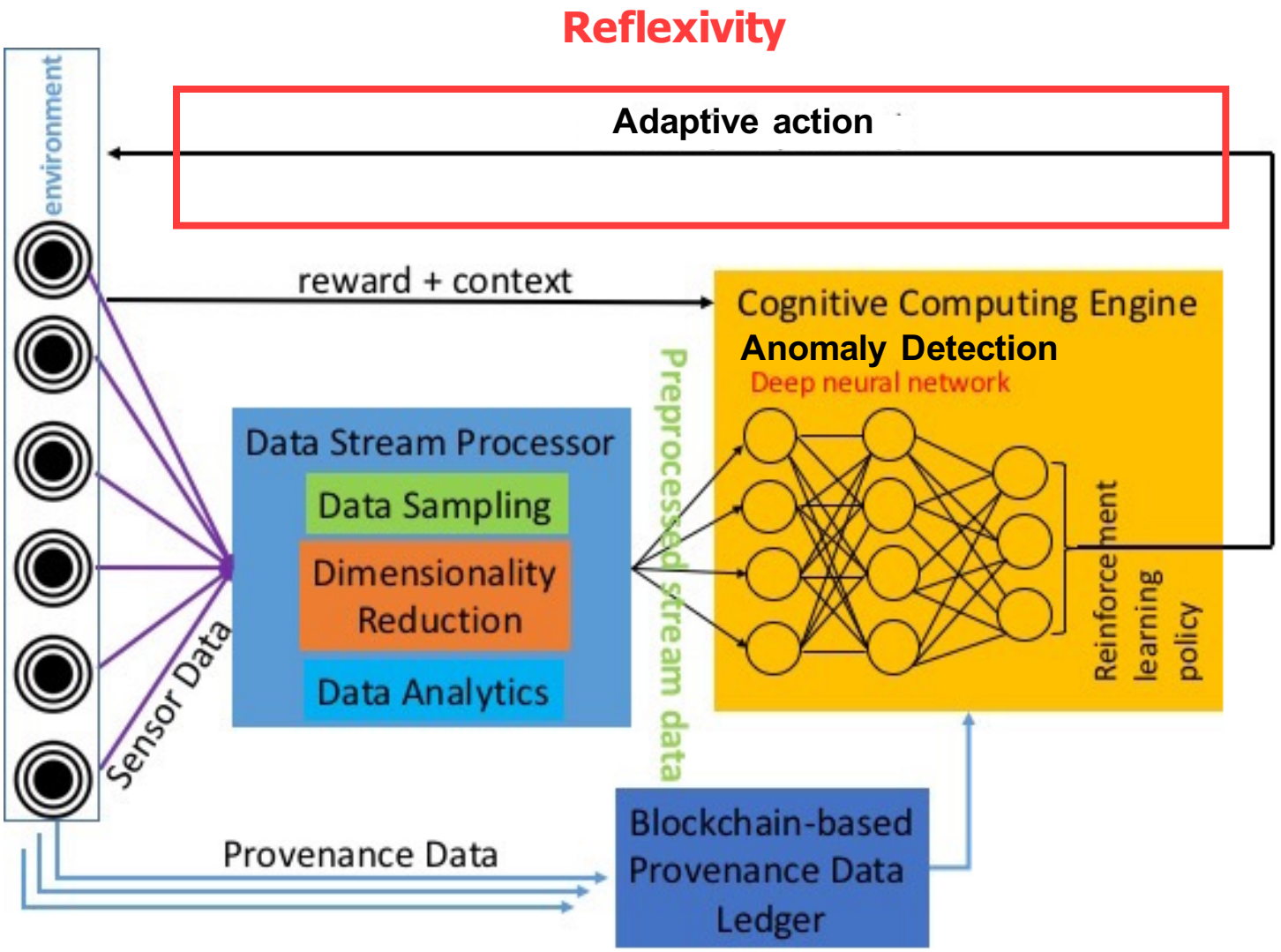
- **Demo 3 (Trust):**
  - A scheme that guarantees the integrity of provenance data is implemented.
  - Capability to verify every transaction in IAS.

Reflexivity

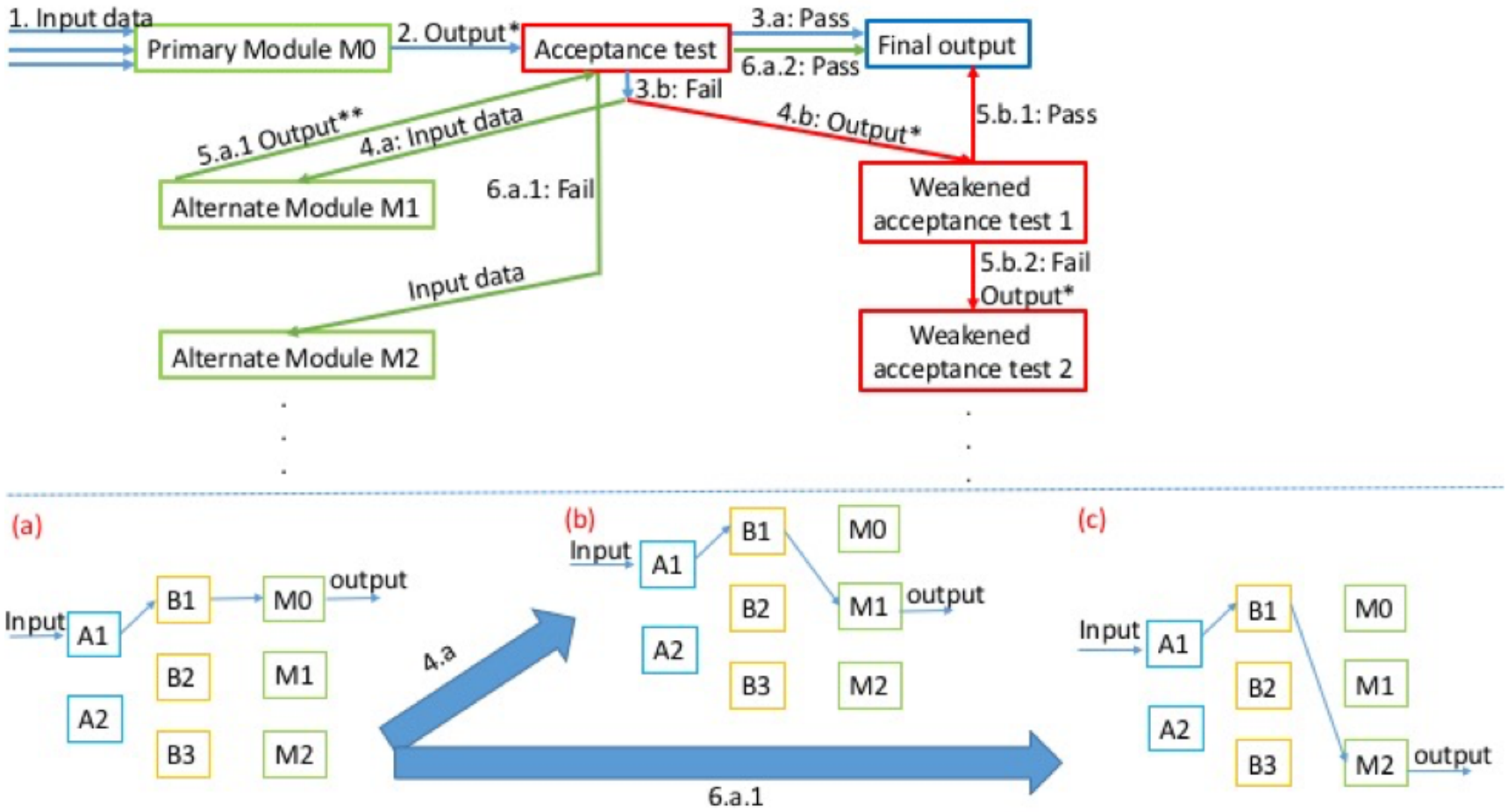
A Solution Based on Graceful Degradation

Comprehensive Architecture of IAS

NORTHROP GRUMMAN



Generic Model of Dynamic Adaptation



Problem Statement

NORTHROP GRUMMAN

Given a smart cyber system operating in a distributed computing environment, it should be able to:

1. Replace anomalous/underperforming modules

2. Swiftly adapt to changes in context

3. Achieve continuous availability even under attacks and failures.

Graceful Degradations: Combinatorial Replica Replacement Scheme

NORTHROP GRUMMAN

Replica replacement by Combinatorial Balanced-blocks:

- N systems (S1…S7) are split into M subset blocks (DAB1…DAB7) of size R (3 : S1, S5, S7). Each system appears in C blocks (3 out of M). Each system pair appears in Δ blocks (only 1). We implemented (N, M, R, C, Δ) = (7, 7, 3, 3, 1). Example on next slide.

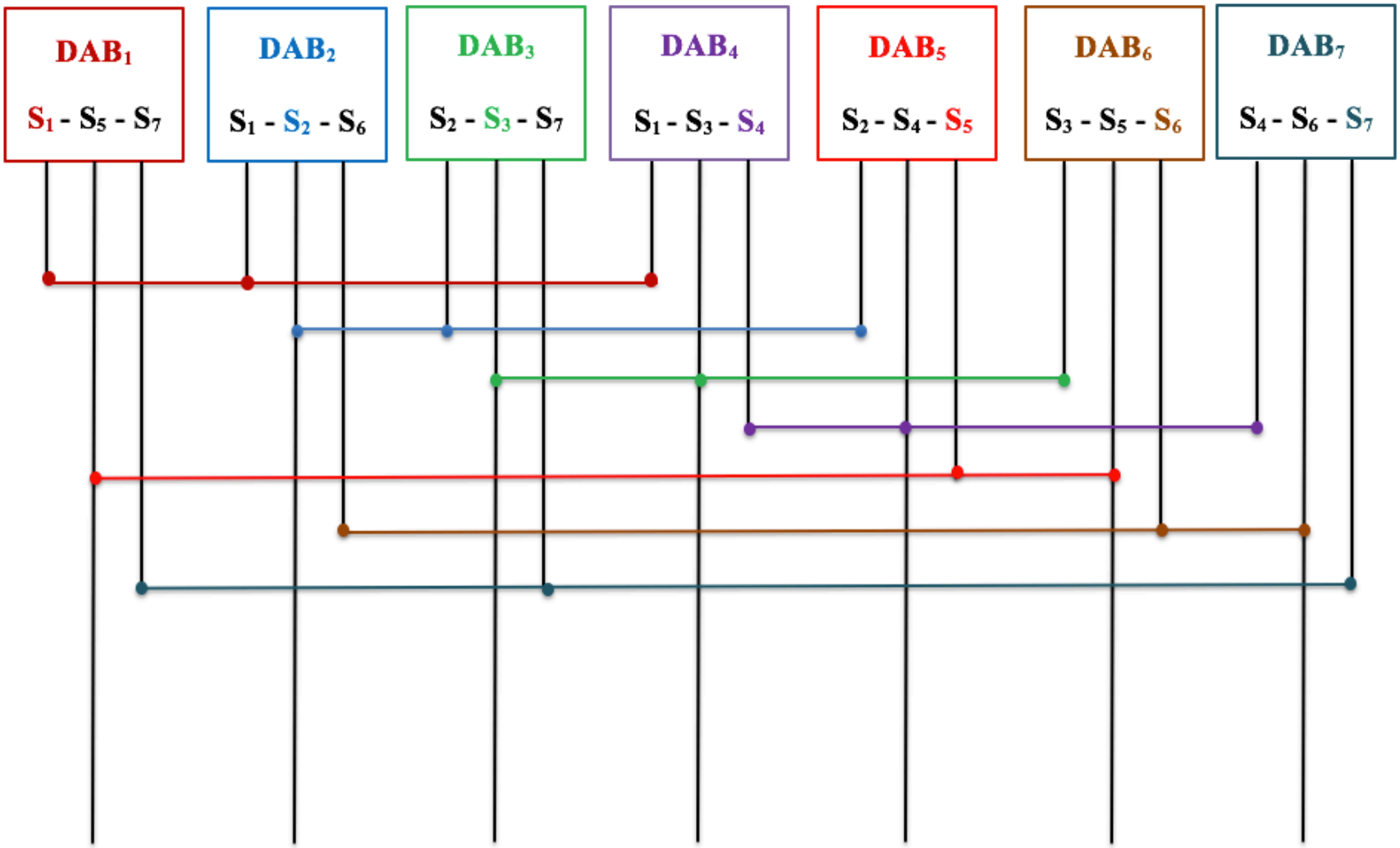
- Each distributed block contains a subset of systems and their replicas that are mathematically distributed and connected, providing balanced resource usage.

- The replicas periodically receive updates from their primary modules. Update interval is set based on Bayesian inference.

- Replicas can be used to perform other tasks in parallel while primary module is functioning properly.

(7, 7, 3, 3, 1)-configuration
DAB: Distributed Autonomous Block

NORTHROP GRUMMAN



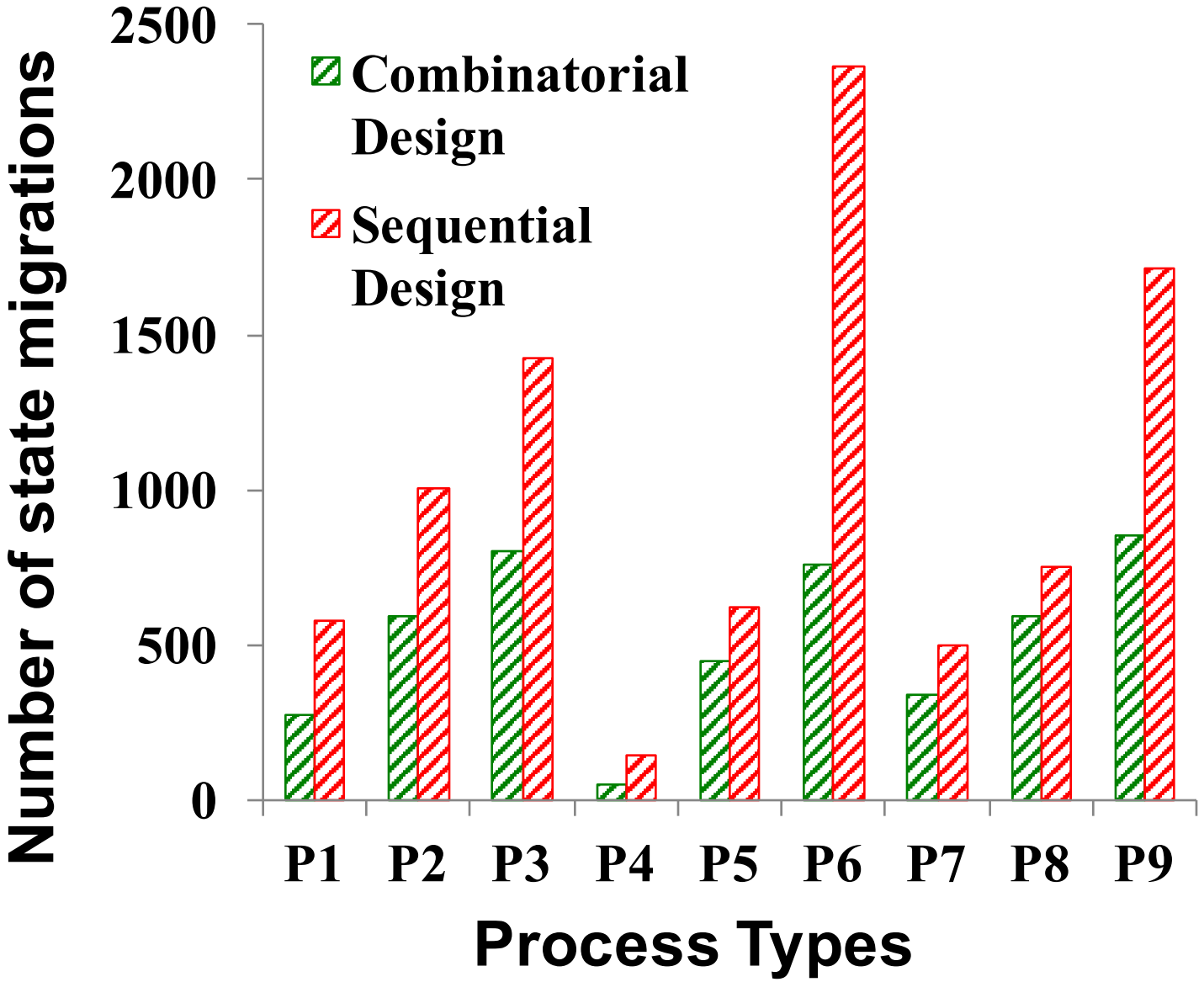
Measurements for Various Process Completions

NORTHROP GRUMMAN

Process Type	Process Name	Speed Up Due to Combinatorial Replica Scheme (Compared to regular sequential design)
P1	FIBSEARCH	1.3
P2	DOUBLE MULT	1.4
P3	FIBB	1.5
P4	SEARCH	1.8
P5	COPY	1.8
P6	SCALAR	2
P7	SUM	2.1
P8	PRINT	3
P9	MOVEMENT	3.1

Measurements for Various Process Completions

NORTHROP GRUMMAN

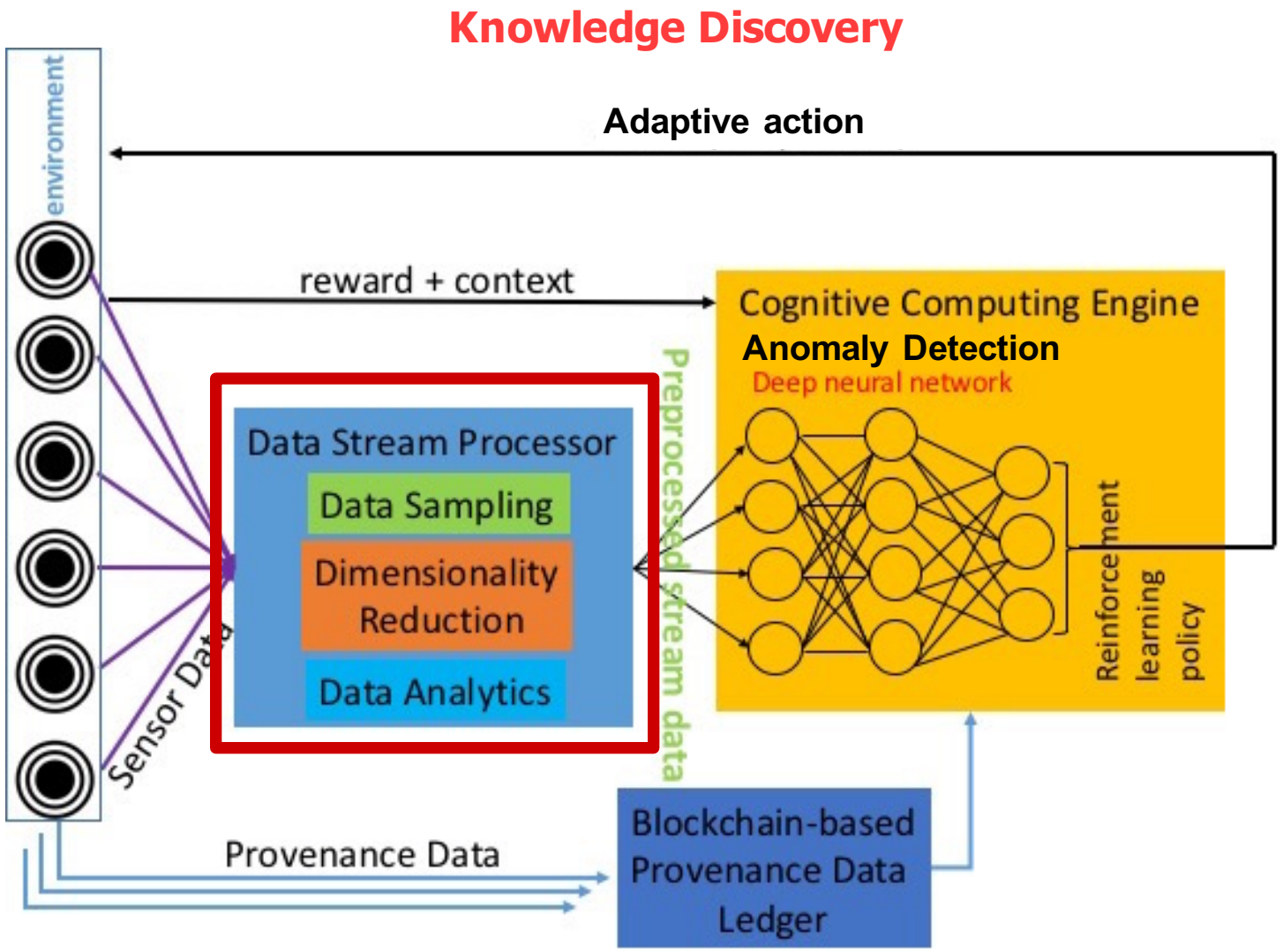


Knowledge Discovery

Scalable Learning & Clustering

Comprehensive Architecture of IAS

NORTHROP GRUMMAN



Problem Statement

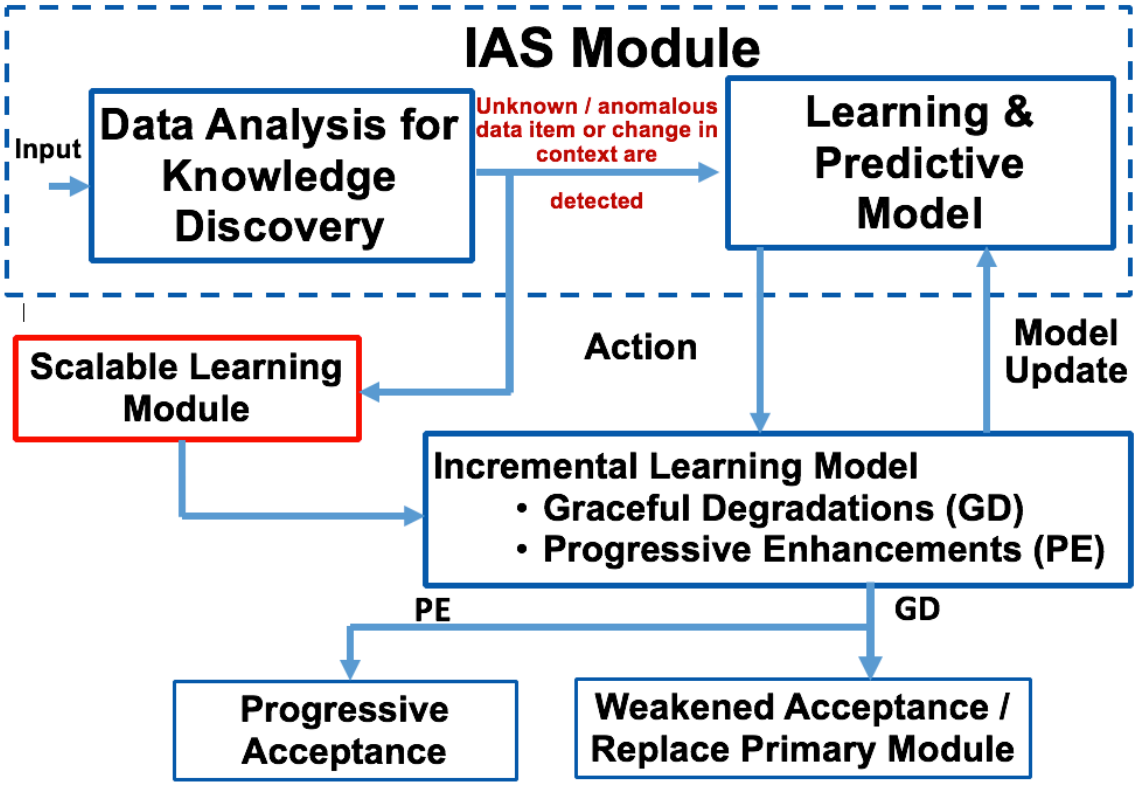
NORTHROP GRUMMAN

- Autonomous systems receive continuous streams of diverse data from numerous sources.
- Disregarding new and unknown data or broadly classifying them into few categories would cause an inadequate learning environment.
- IAS should be trained to work with
  - Meta-data, limited data, incomplete data, and unknown (new) data
  - Dynamic, unpredictable, and adversarial environment

Scalable Learning

NORTHROP GRUMMAN

- It's a method to achieve maximum classification without rejecting any unknown data item that was not present in the training or testing datasets as anomalies.



Bitwise Fuzzy-based Clustering (BFC)

NORTHROP GRUMMAN

- BFC is implemented through Perfect Error-correcting codes or Golay codes.
- Error Correcting Codes (ECC) are used for controlling errors in data (any information that could be represented in bits 0/1).
- When there is an error in data, the error correcting codes can approximately match the distorted data to the original.
  - For example, take the message (m) = 000. Consider 1 bit distortions of m: 100, 010, 001.
  - All three distortions are 1 hamming distance (it takes 1 bit flip to get to 000) away from 000. So they can be easily corrected.

Why Error Correcting Codes?

NORTHROP GRUMMAN

- BFC creates clusters based on fuzzily (approximately) matched data items similar to error correction.
  - For example, take the message (m) = 000 as a data item. m's 1 bit distortions (100, 010, and 001) will be clustered into one.

- 0/1 bits are used to label binary features.
  - Assume that 0 – Absent and 1 – Present. Based on number of features of a data item, we can create a binary classification.
  - For example, data item D has 3 features. Presence or absence of each feature creates a code word, say, 101.
  - Code word such as 101 will be a **label** for that data item.
  - Using ECC provides scalability (2^n combinations of clusters) and fault tolerance (distorted labels can be clustered correctly).

Clustering by BFC

NORTHROP GRUMMAN

- A binary vector template is created. Based on the presence and absence of it, 0 or 1 is encoded.

- Consider an image data item:

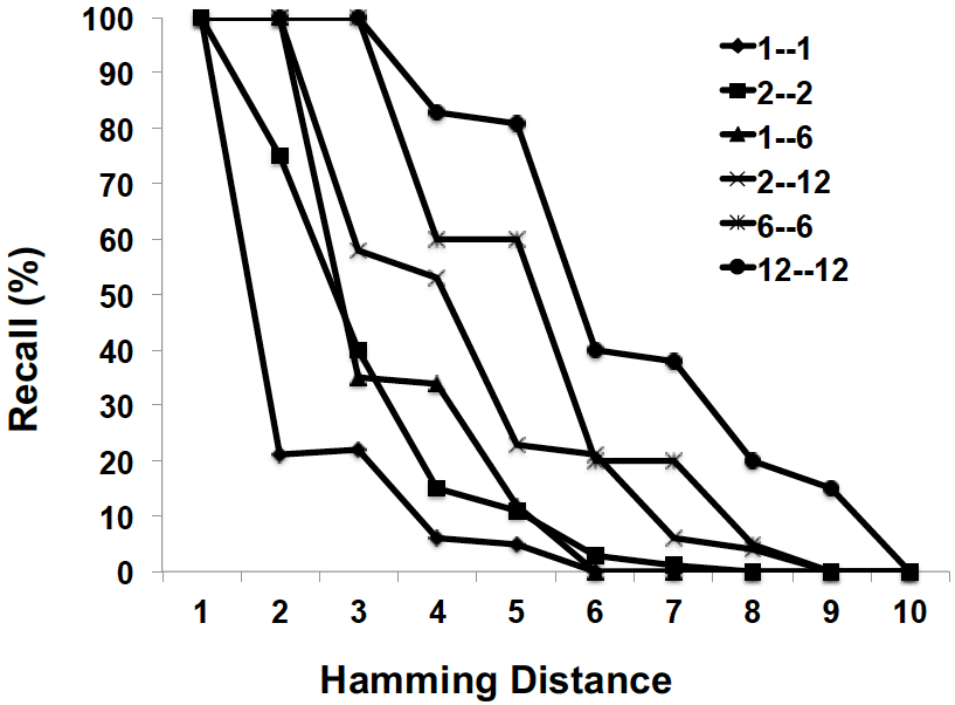
Features	YES/NO
F1: is it red?	0
F2: is it female?	1
.	.
.	.
.	.
F22: is it tall?	0
F23: is it animal?	0

- BFC follows Golay error-correction code for labeling which produces 2^{23} unique labels and items can be clustered properly even with 3-4 features were misidentified.

Evaluation – Recall

NORTHROP GRUMMAN

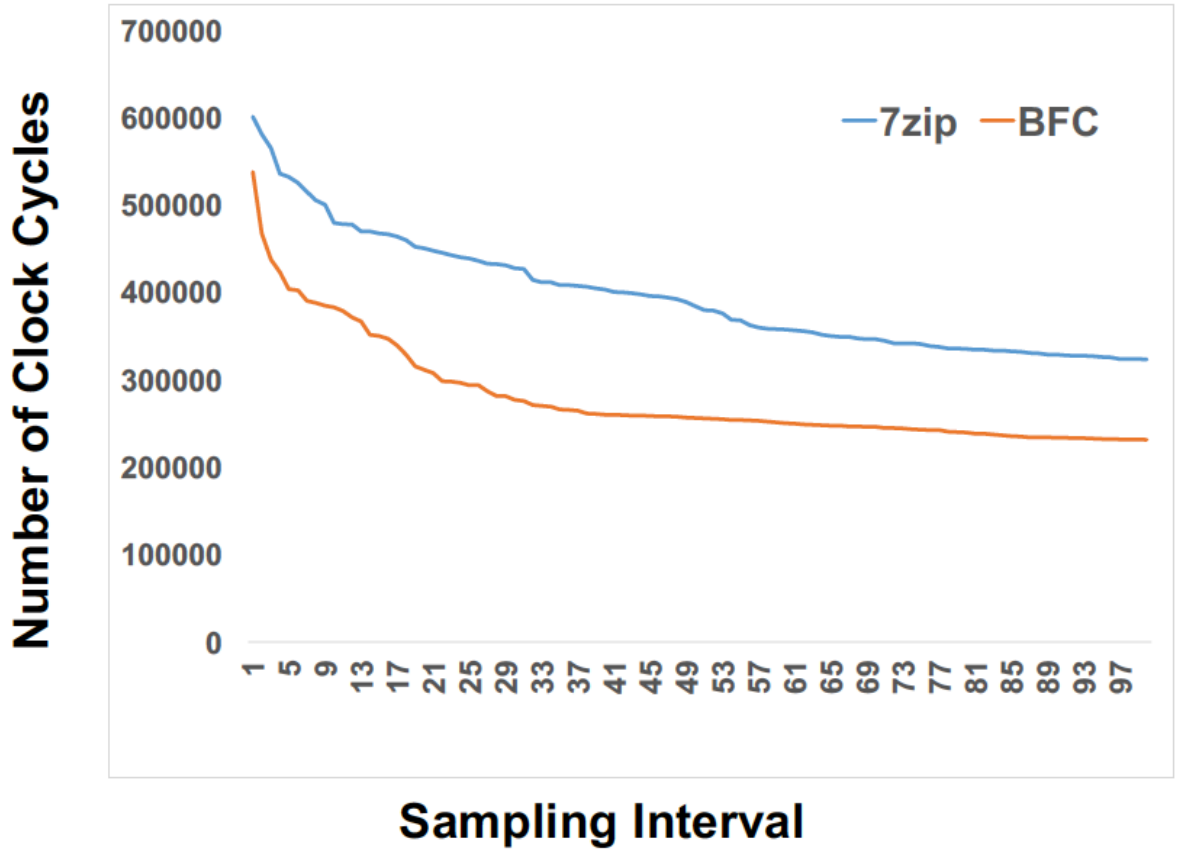
- Recall = (relevant items ∩ retrieved items) / relevant items
- For higher hamming distances, the recall probabilities are small. 1—1 means direct matching with 1 hamming distance cluster with the same hamming distance.



Evaluation – CPU Performance

NORTHROP GRUMMAN

- Number of clock cycles for encoding. We used process thread API to collect data, sampled every 1000k instructions.



Evaluation – Time Complexity

NORTHROP GRUMMAN

- Number of clock cycles for encoding. We used process thread API to collect data, sampled every 1000k instructions.

Clustering Algorithms	Time Complexities
k-means	$\mathcal{O}(nkd)$
Hierarchical Clustering	$\mathcal{O}(n^2)$
Clustering using REpresentatives (CURE) [21]	$\mathcal{O}(n^2\ log\ n)$
ROCK [22]	$\mathcal{O}(min(n^2, nm_m m_a))$
CLICK [23]	$\mathcal{O}(n\ log\ n)$
BFC	$\mathcal{O}(n)$

Autonomous Aggregate Data Analytics in Untrusted Cloud – Motivation

NORTHROP GRUMMAN

- Autonomous systems operating in distributed environment have to collectively learn from one another.
- It is important to maintain the privacy of individual entities generating data and humans interacting with them.
- Autonomous systems should be able to
  - Learn from restricted information
  - Preserve privacy while collectively learning about the distributed environment.

Autonomous Aggregate Data Analytics in Untrusted Cloud – Motivation

NORTHROP GRUMMAN

- Autonomous systems operating in distributed environment have to collectively learn from one another.
- It is important to maintain the privacy of individual entities generating data and humans interacting with them.
- Autonomous systems should be able to
  - Learn from restricted information
  - Preserve privacy while collectively learning about the distributed environment.

Privacy Preserving Autonomous Data Aggregation

NORTHROP GRUMMAN

- Using Active Bundle (AB), a distributed self-protecting entity with policy enforcement engine, we implement
  - One-time access certificate used to query other ABs
  - Privacy preserving aggregation analytics on numerical data

- Instead of checking AB's authentication protocol every time, an AB can obtain a one-time pass to access other ABs data per aggregate query.

- Numerical data is perturbed for the analytics and at the end the perturbation is removed.

Privacy Preserving Autonomous Data Aggregation

NORTHROP GRUMMAN

- After passing the authentication and policies enforced by AB's policy enforcement engine, aggregate data analytics can be performed.
- AB's provenance data is used for aggregated analytics such as *Count, Average, etc.* on qualified attributes.
- These aggregate analytics guarantee privacy of individual ABs. Consider an aggregation,
  - AB_1's age attribute is perturbed: "Age (a) " + "Random Perturbation (R)" ➔ $2AB_1(a + r = a_n) + 2AB_2(a + a_n = a_{n1}) + \ldots$
  - Final average = $(a_{nn} - R) / count(2AB)$

Evaluation

NORTHROP GRUMMAN

- We measure the latency of data request sent to AB, which is hosted by a local server, located in the same network with the client.

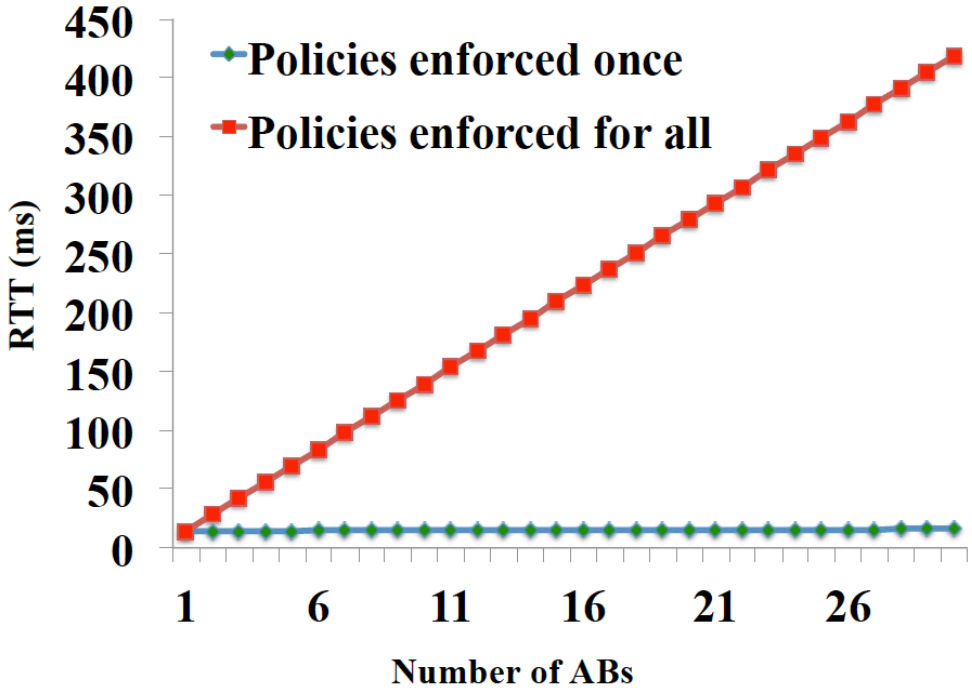
- As a latency parameter, we record Round-Trip Time (RTT) for the data request processing at the server side (Note: we do not consider network delays in this experiment).

- ApacheBench v2.3 is used to calculate RTT measurements. We run 50 requests in a row and compute RTT average.

Evaluation

NORTHROP GRUMMAN

- Our initial work shows that the policies enforced for each AB access raise the access time exponentially where as a simple python simulation of file access (one time authentication example) stays almost constant for multiple entities.

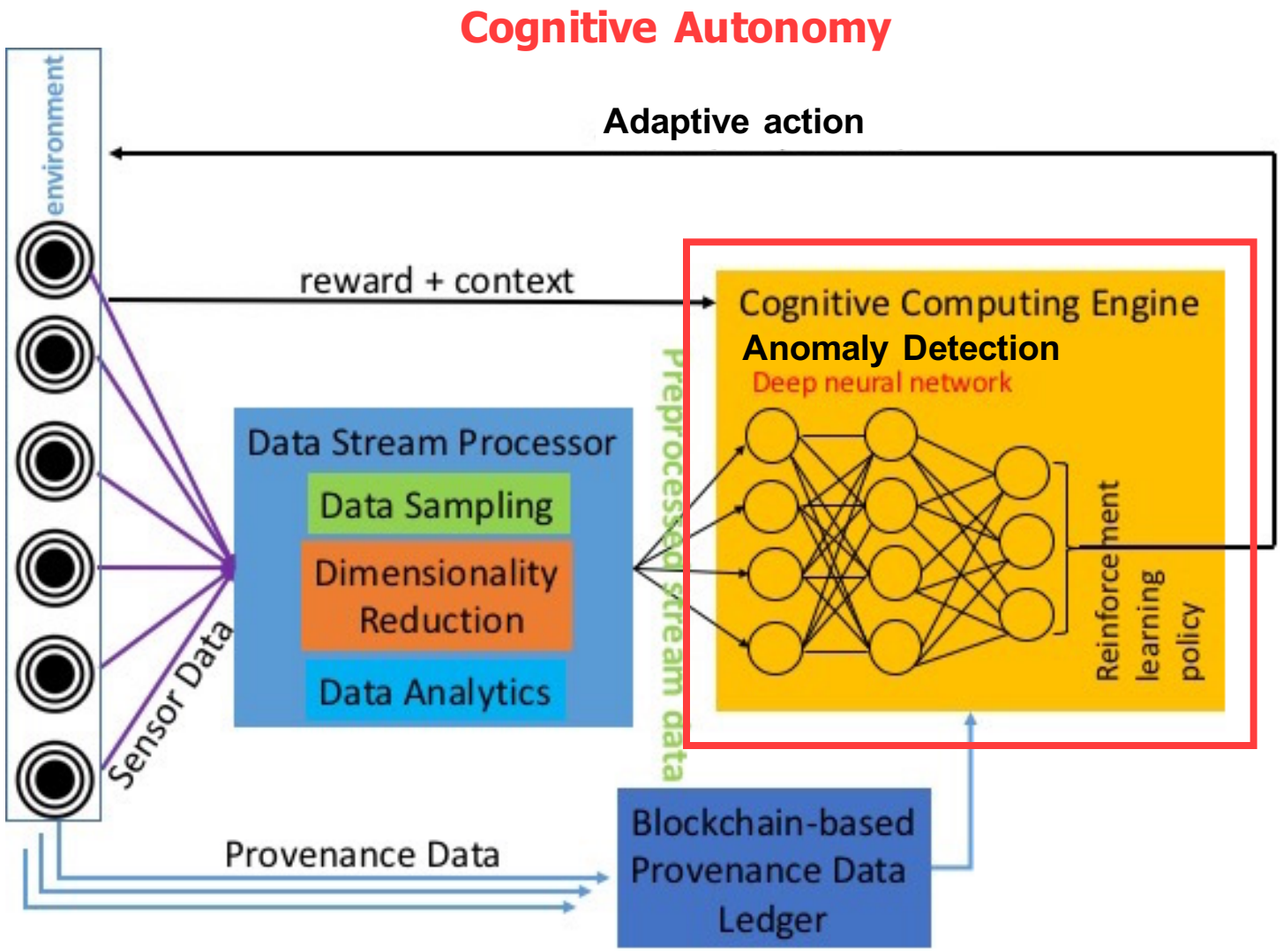


Cognitive Autonomy

Anomaly Detection

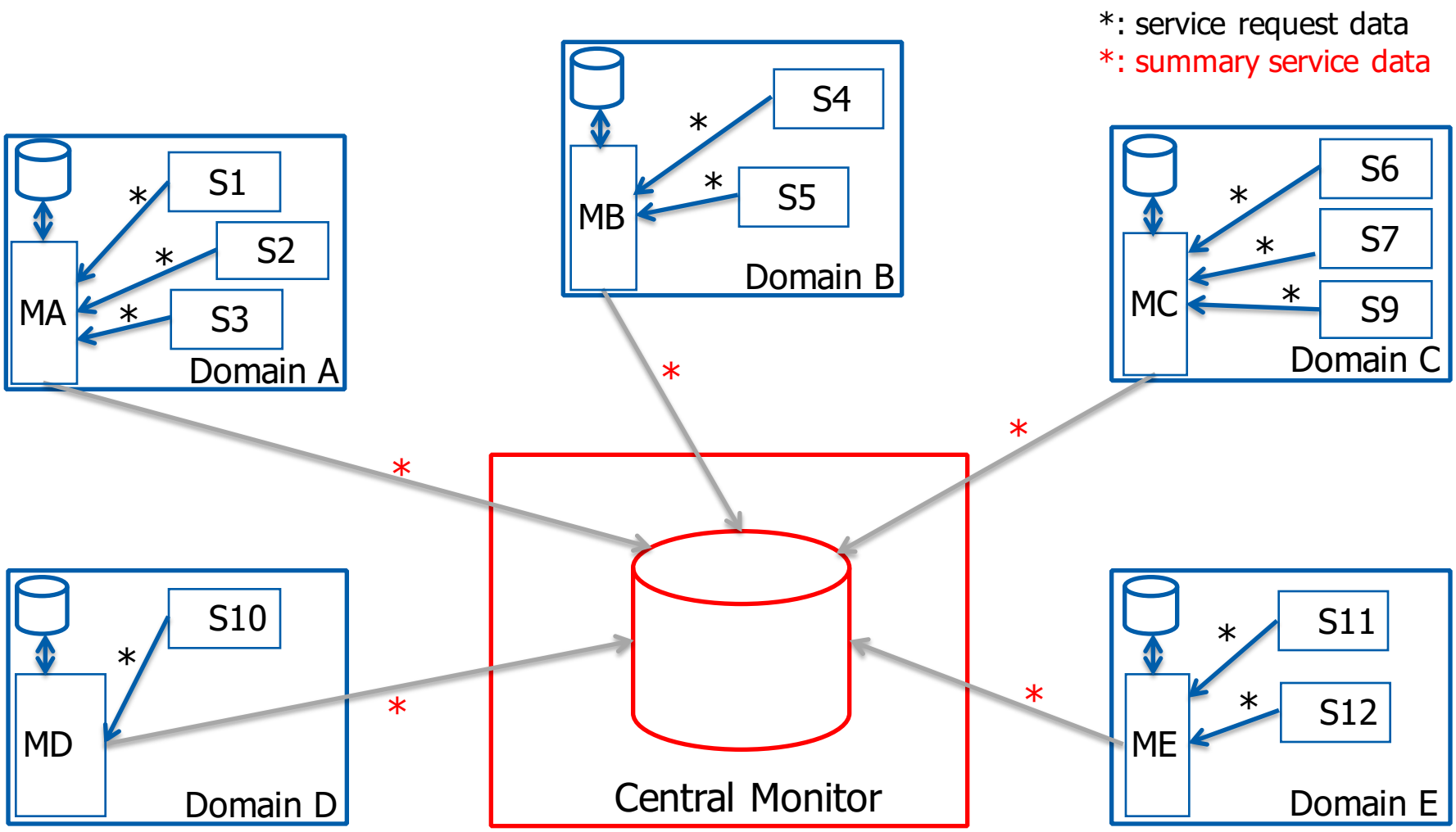
Comprehensive Architecture of IAS

NORTHROP GRUMMAN



Distributed Service Monitoring / Anomaly Detection System

NORTHROP GRUMMAN



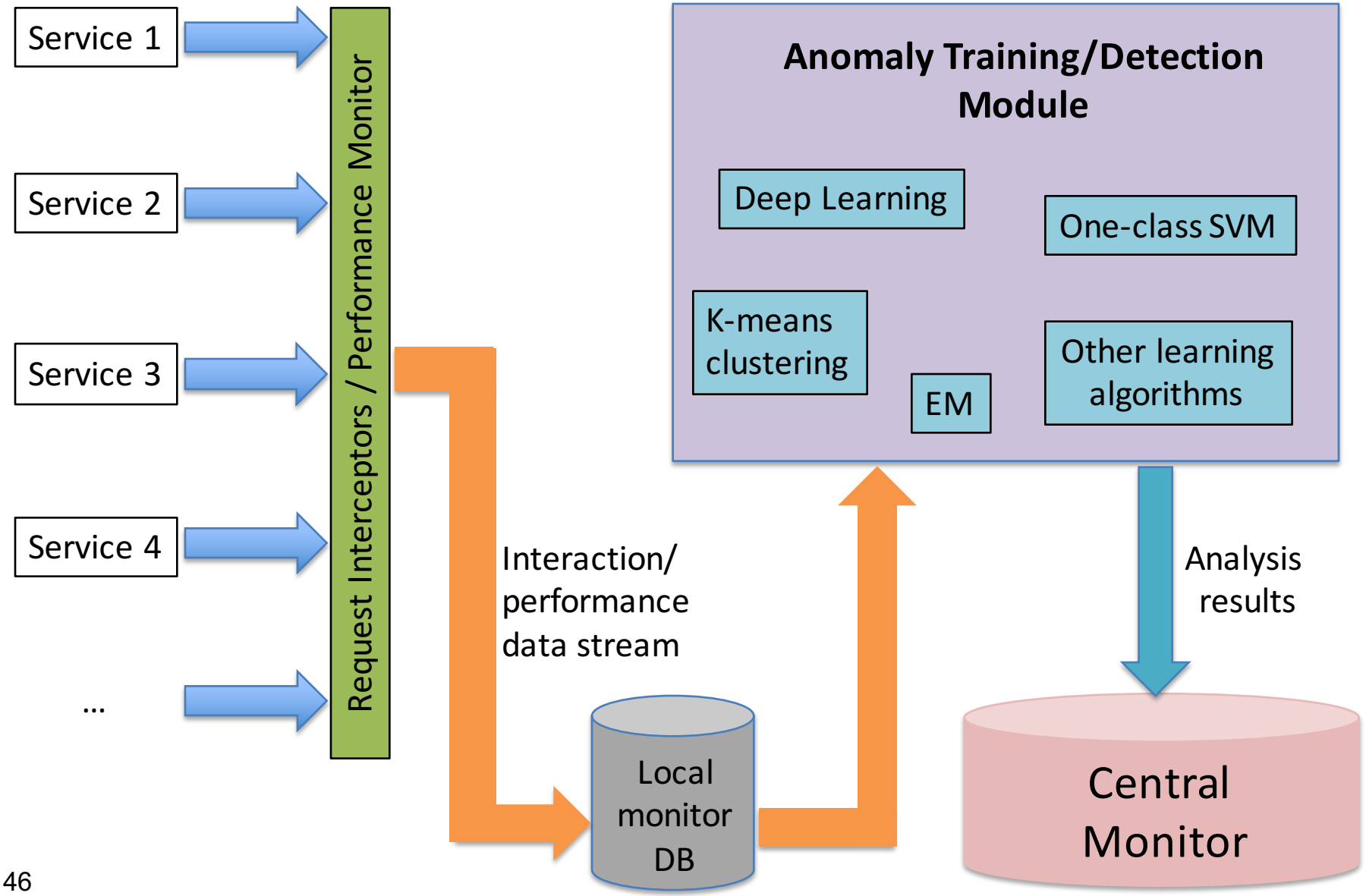
Distributed Service Monitoring / Anomaly Detection System

NORTHROP GRUMMAN

- Distributed service monitoring allows for the **collection**, **analysis** and **reaction** to dynamic cyber events across all domains involved, and **prevents propagation of threats** within or outside the domain of the anomalous service by taking proactive measures (service isolation, replication).
- The data (service requests, service performance data etc.) gathered by the monitor M_x of each service domain x is stored in the monitoring database of the domain.
- Service monitoring is distributed across domains, with one monitor for each domain. Each monitor is responsible for reporting the health status of the services in its own domain to the central monitor.
- Service monitor of each domain mines the data stored in its database to detect anomalies with services in the domain and takes measures accordingly (re-deployment, backup service creation).
- Service monitor of each domain sends summary health status data of services to the central monitor, which is utilized for dynamic service composition.

Distributed Service Monitoring / Anomaly Detection System

NORTHROP GRUMMAN



Distributed Service Monitoring / Anomaly Detection System: Technology Overview

NORTHROP GRUMMAN

- A novel distributed monitoring tool to:
  - Audit and detect service behavior and performance changes[1]
  - Gather service trust data and share them securely in various domains
  - Dynamically reconfigure service orchestrations based on security context and QoS requirements[2]
  - It will also include a deep learning based approach resilient to *poisoning or causative attacks* as models can be re-trained in an unsupervised manner by observing the real events of the systems (i.e. attacker cannot generate malicious examples to alter the model).
  - The deep learning approach can detect variable-length anomalous sequences that span for long periods of time.

[1]B. Bhargava, P. Angin, R. Ranchal, S. Lingayat. "A Distributed Monitoring and Reconfiguration Approach for Adaptive Network Computing." DNCMS in conjunction with SRDS 2015 (**Best paper award**).
[2]M. Azarmi. 2"End-to-End Security in Service-Oriented Architecture," PhD Thesis, Purdue University, April 2016.

Distributed Service Monitoring / Anomaly Detection System: Benefits of Technology

NORTHROP GRUMMAN

- System modules for *service anomaly detection*, *service performance monitoring* and *trust management* can be easily integrated into NGC cybersecurity software.
- The **modular architecture** and use of **standard software** in the monitoring framework allows for **easy plugin** to any system.
- The different components of the adaptability framework can be extended and integrated with various IRADs and demonstrated at TechFest:
  - Behavior-based Analytics IRAD
  - Cognitive Autonomy Engine (CAE) IRAD
  - Infromation Analytics IRAD
  - ADEN Information Operations IRAD

Unsupervised Learning: Training Parameters

NORTHROP GRUMMAN

- No class labels needed for training.
- Training data: input vectors without any corresponding target values
- Unsupervised learning with security and performance parameters used to find clusters
- Values outside clusters will be detected as outliers and help detection of anomalies.

Model training parameters used:

Parameter	Cloud services	Cloud data services
Number of requests/sec	X	
Bytes downloaded/sec	X	X
Bytes uploaded/sec	X	X
Total error rate	X	
CPU utilization	X	
Memory utilization	X	
Number of authentication failures	X	
Number of connections	X	
Number of connection failures	X	
Number of disk reads/writes	X	X
Network latency	X	
Service response time	X	
Disk space usage	X	X
Throughput	X	
Number of database connections	X	X
Service/cluster health status	X	

Unsupervised Learning: Algorithms

NORTHROP GRUMMAN

- Two unsupervised learning algorithms implemented.
- Training performed offline, under normal system operation. Classification time negligible.
- Different learning algorithms pluggable into system.

K-Means Clustering:

- Training:

  Input: Matrix $V_{d \times t}$ of service performance record

  d: number of performance parameters

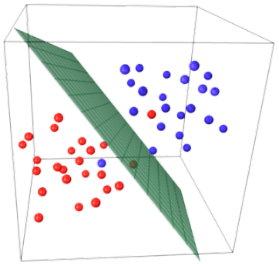
  t: number of time points observed

  Cluster each set of performance parameter values using K-means algorithm

- Testing (system operation):

  for each service interaction log:
  measure distance of performance parameter values to each cluster, assign time point to closest cluster
  if latest interaction does not belong to any cluster
  raise anomaly signal
  end_if
  end_for

One-class SVM (Support Vector Machines):



* Figure from http://stackoverflow.com/questions/9480605/

- Learns decision function for novelty detection
- Decision hyper-plane boundary based on normal runtime conditions
- Multiple features of services enter the model for classification

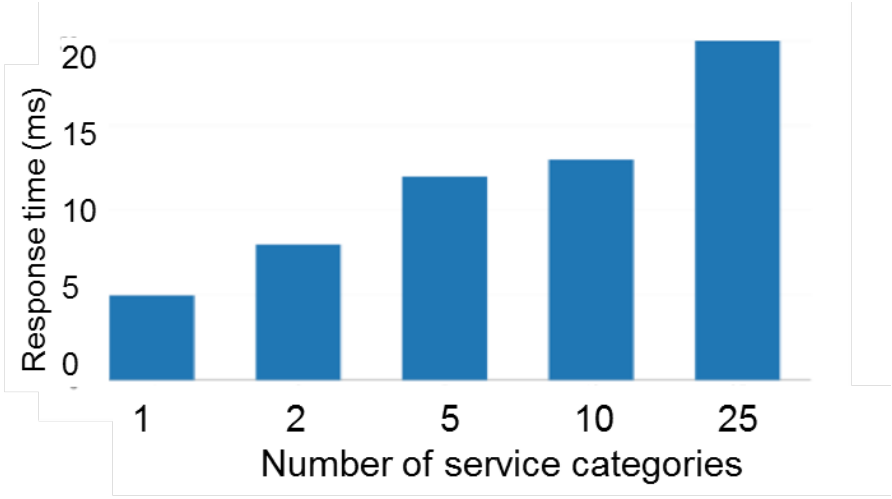
Dynamic Service Composition Experiments

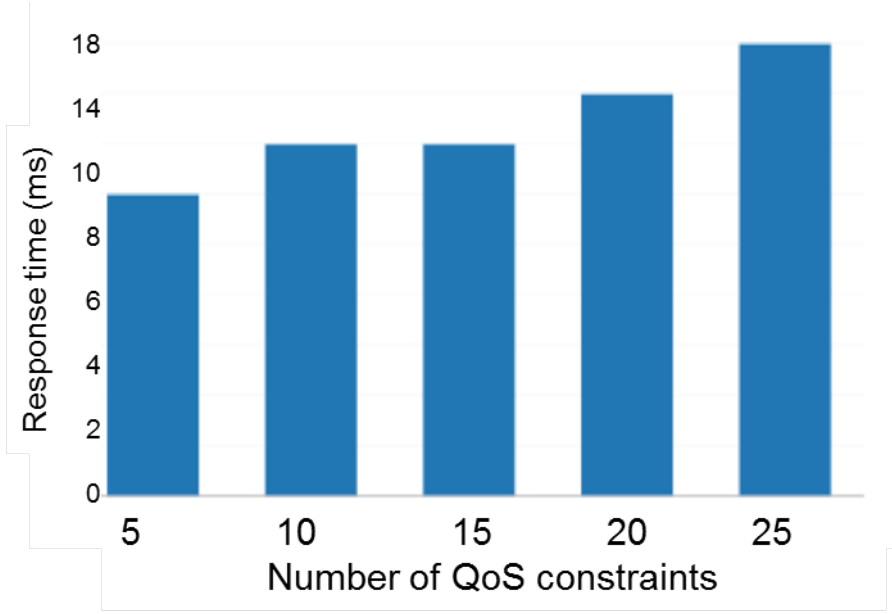
NORTHROP GRUMMAN

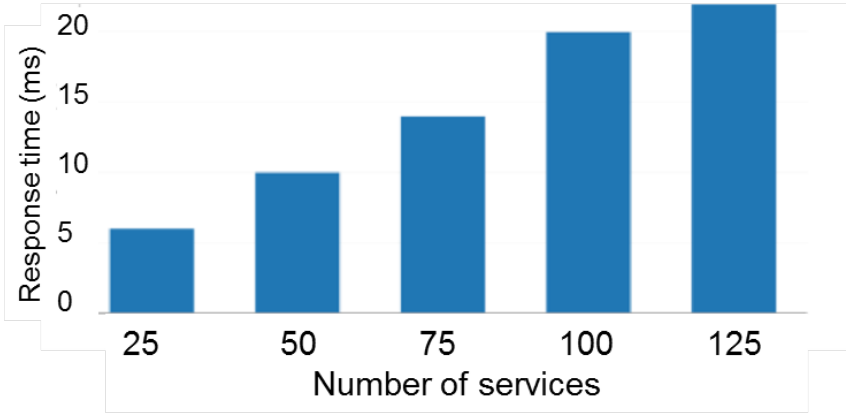
Experiment settings:

Service instance	t2.small (1 vCPU, 2GB memory, and EBS storage)
PE and TM instances	t2.small (1 vCPU, 2GB memory, and EBS storage)
Client instance	t2.micro (1 vCPU, 1GB memory, and EBS storage)
Operating system	Amazon Linux 2015.03 64-bit OS
Geographical region	US-w2 (Oregon region)

- Overhead evalution for three cases:
  - Different number of service categories in composition
  - Different number of services to choose from for each category
  - Different number of QoS contraints
- Composition time not affected significantly by the number of QoS constraints. Number of services and service categories have more visible effect, with still reasonable overhead.







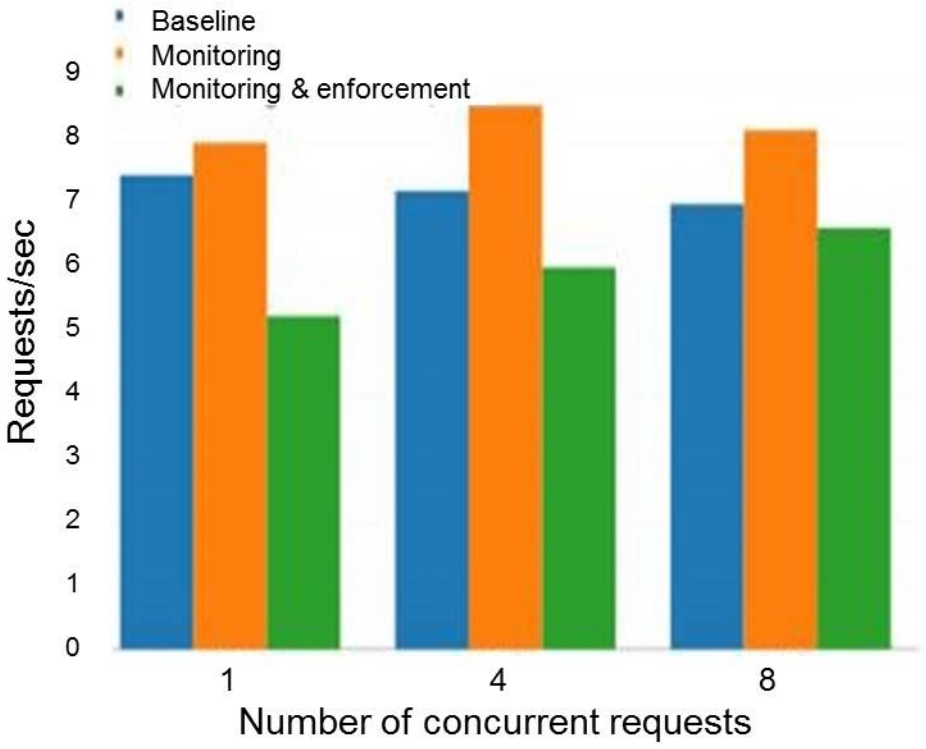
Overhead of Service Monitoring / Request Interception

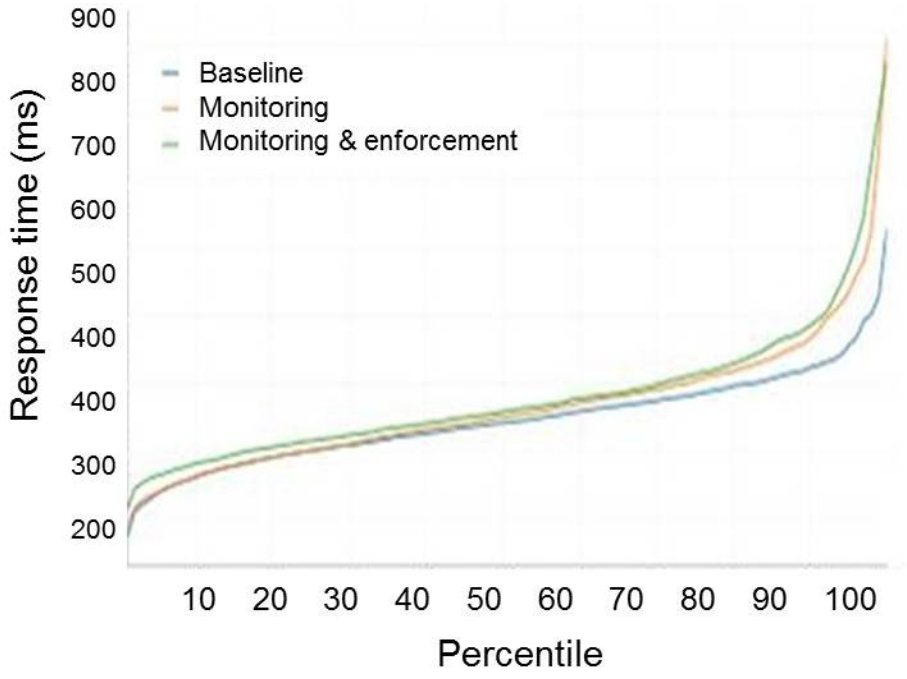
NORTHROP GRUMMAN

- Performance evaluation of service domain in terms of throughput and service response time
- Negligible overhead incurred by monitoring

Experiment settings:

Service instance	t2.small (1 vCPU, 2GB memory, and EBS storage)
PE and TM instances	t2.small (1 vCPU, 2GB memory, and EBS storage)
Client instance	t2.micro (1 vCPU, 1GB memory, and EBS storage)
Operating system	Amazon Linux 2015.03 64-bit OS
Geographical region	US-w2 (Oregon region)



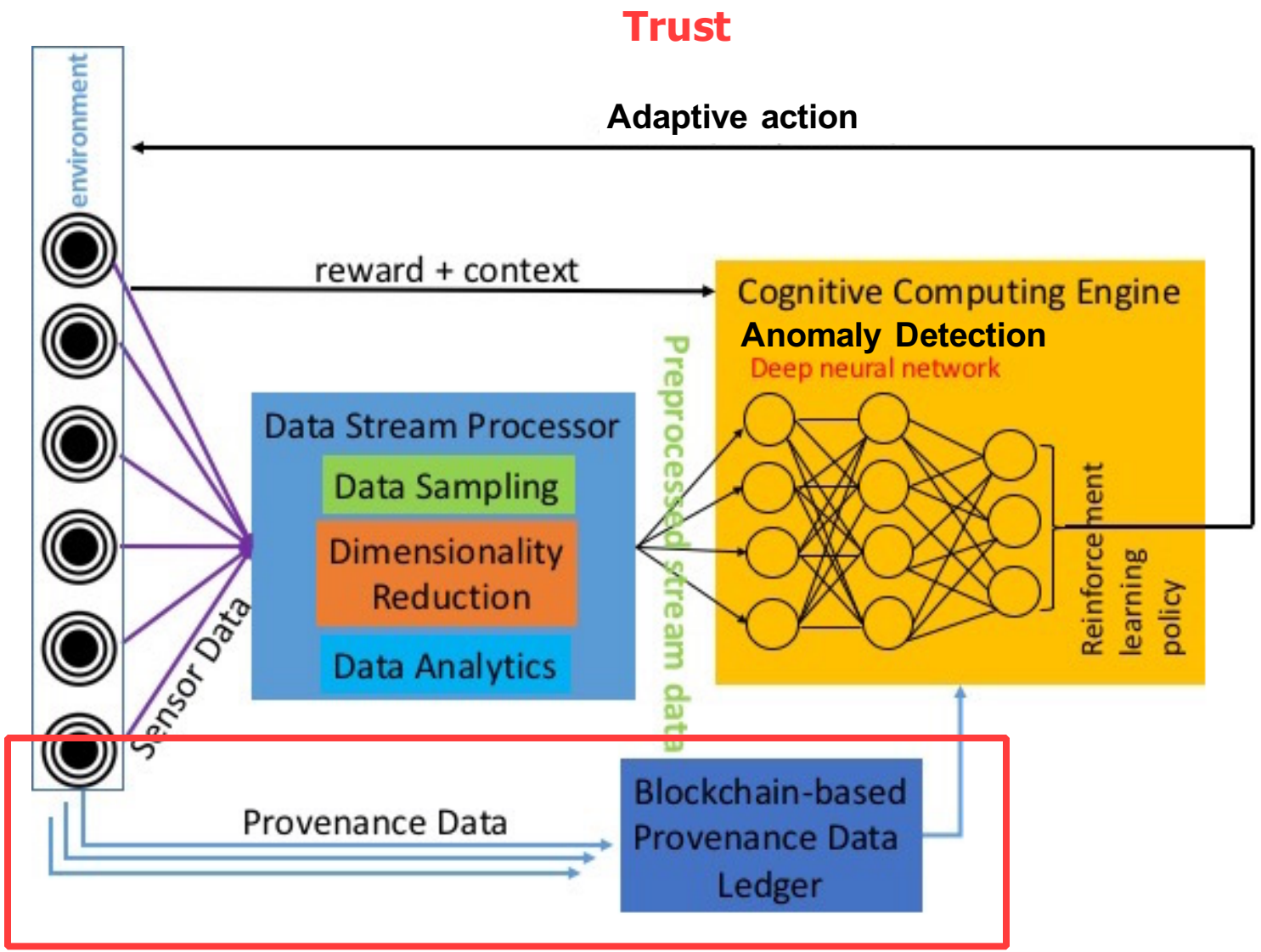


Trust

Blockhub: Blockchain-Based Solution for Secure Distribution of Software and Data in IAS

Comprehensive Architecture of IAS

NORTHROP GRUMMAN



Problem Statement

NORTHROP GRUMMAN

- Provide trust (integrity, confidentiality, verifiability) to provenance data in IAS
  - Interactions between services are logged
  - Log records can not be corrupted
- Provide trust for network participants in IAS
  - Ensure data confidentiality
  - Ensure data integrity
- Provide privacy-preserving data exchange in IAS

Blockchain Technology Deployment

NORTHROP GRUMMAN

- Need for fine-grained role- and attribute-based access control with data leakage detection capabilities

Solution: Integrate WAXEDPRUNE project into blockchain-based framework: every data/software request is registered in blockchain network before reaching the Software Bundle

- Performance

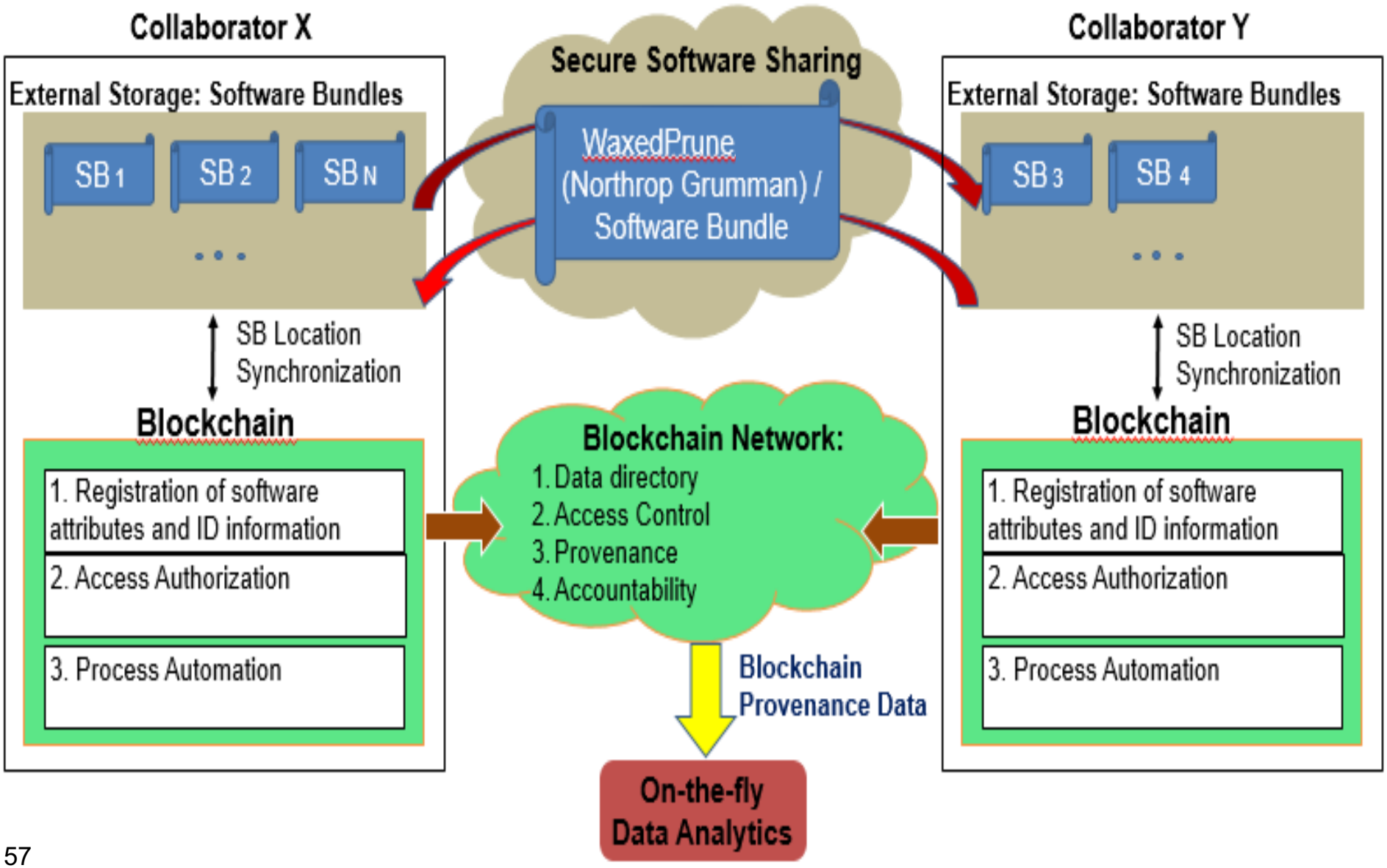
(a) Transaction latency on IBM Hyperledger Fabric blockchain platform (ver. 1.0.x) is about 6 seconds*

(b) Transaction verification takes long when chain has many blocks

Solution: Depth-robust graphs (in collaboration with Prof. Jeremiah Blocki, Purdue) to store blockchain for faster transaction verification no need to verify all the links in the chain

Blockhub: blockchain-platform for IAS

NORTHROP GRUMMAN



Future Work

NORTHROP GRUMMAN

- Building MTD-style defense mechanism using graceful degradations.
- Enhancing update times with context-aware Hidden Markov Models (HMM).
- User profiling and network intrusion detection through deep learning methodologies.
- Failure recovery for blockchain mechanism in mobile environments and quantification of performance parameters.

Thank you!!!