In partnership with NVIDIA

Booz | Allen | Hamilton

DETECTING THE UNKNOWN: USING UNSUPERVISED BEHAVIOR MODELS TO EXPOSE MALICIOUS NETWORK ACTIVITY

Aaron Sant-Miller
Senior Lead Data Scientist, Booz Allen Hamilton

MARCH 2019

CONSULTING | ANALYTICS | DIGITAL SOLUTIONS | ENGINEERING | CYBER

Cyber-attacks are becoming more creative and have a much higher impact on daily life.

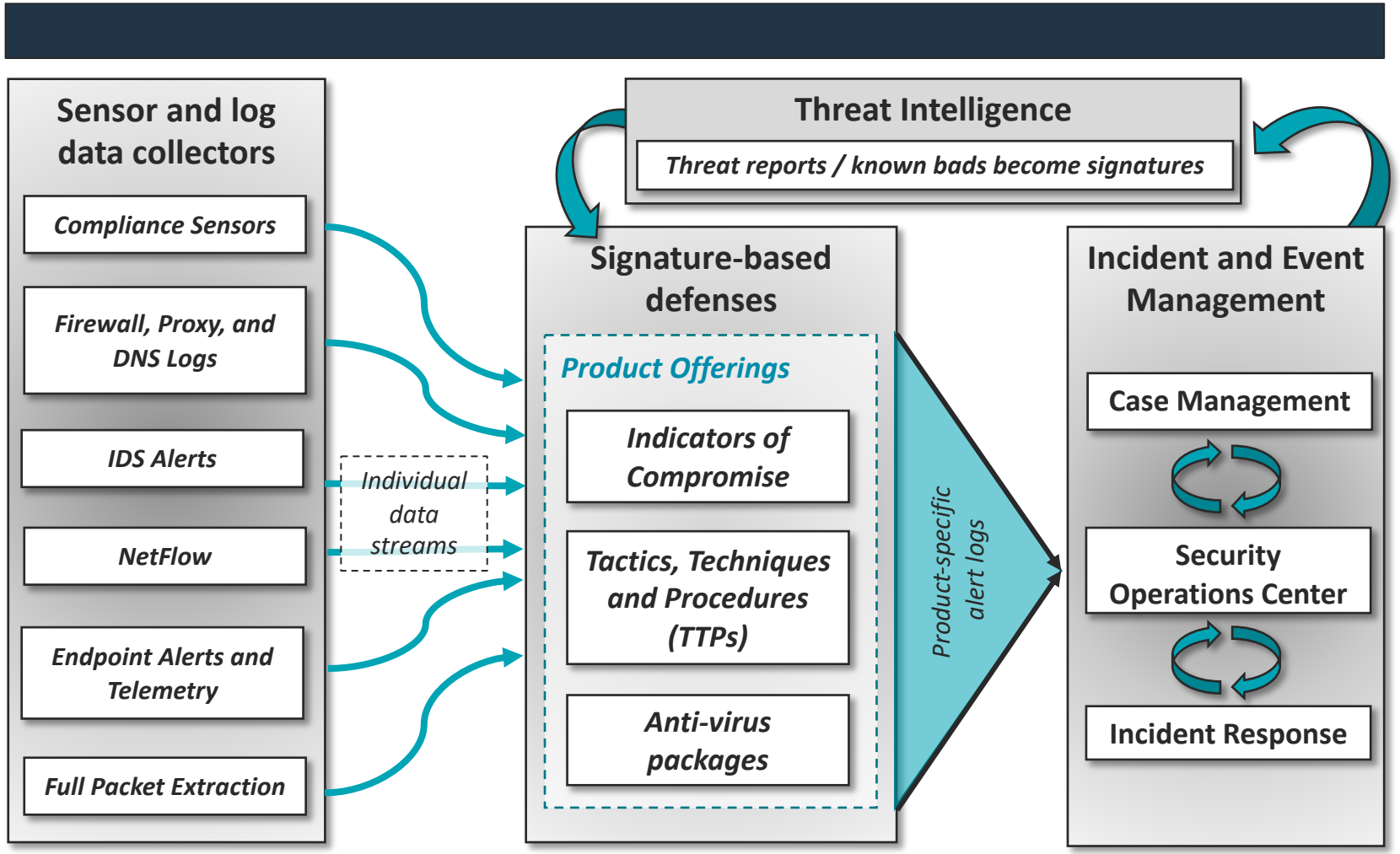
- *Rapidly expanding attack surface*
- *Inundation of cyber tools*
- *Attacks are more sophisticated*
- *Cyber talent shortage*

The average intrusion is detected

200 days

after the fact

Traditional defenses lack real-time adaptability

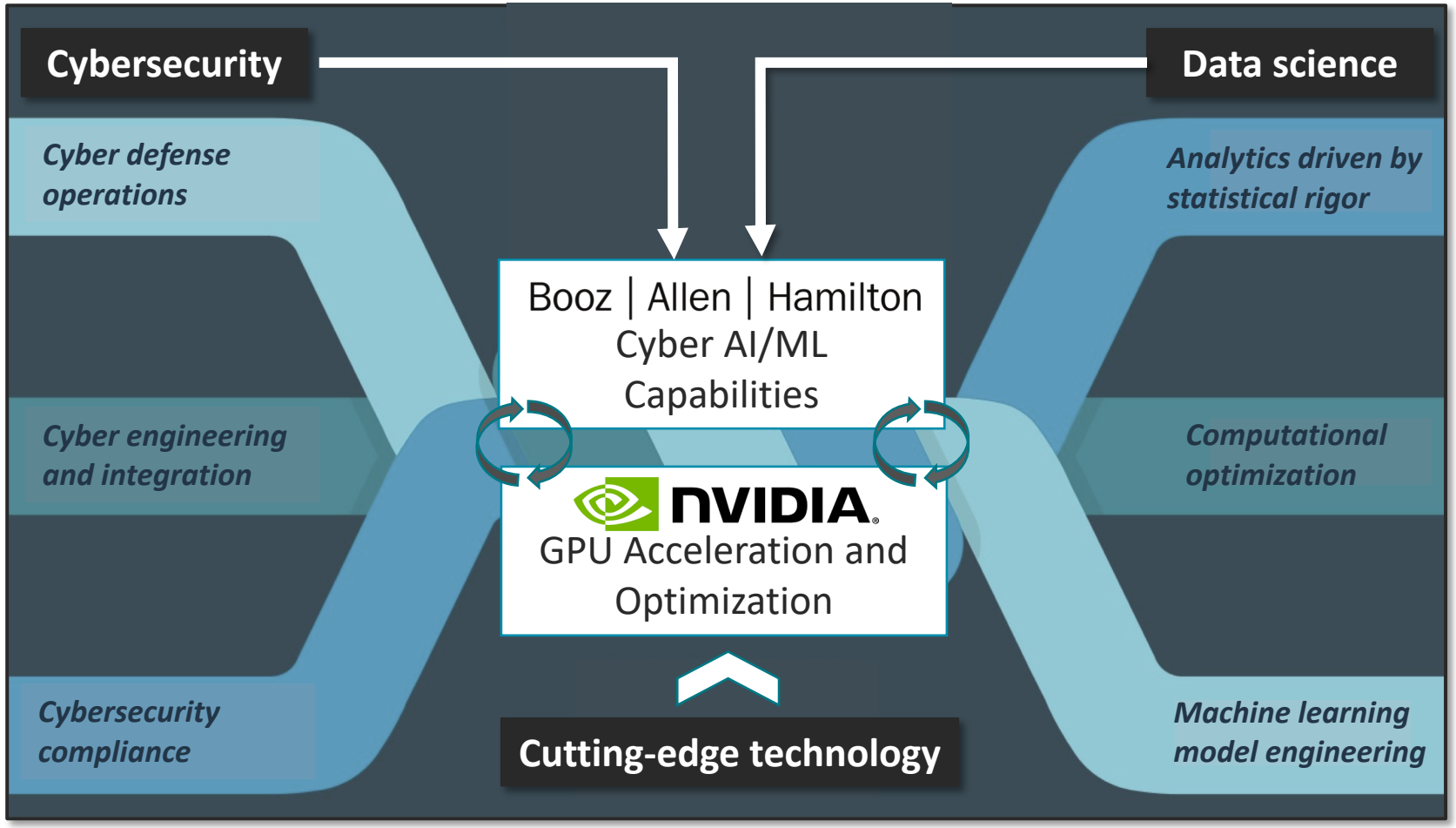


Characteristics

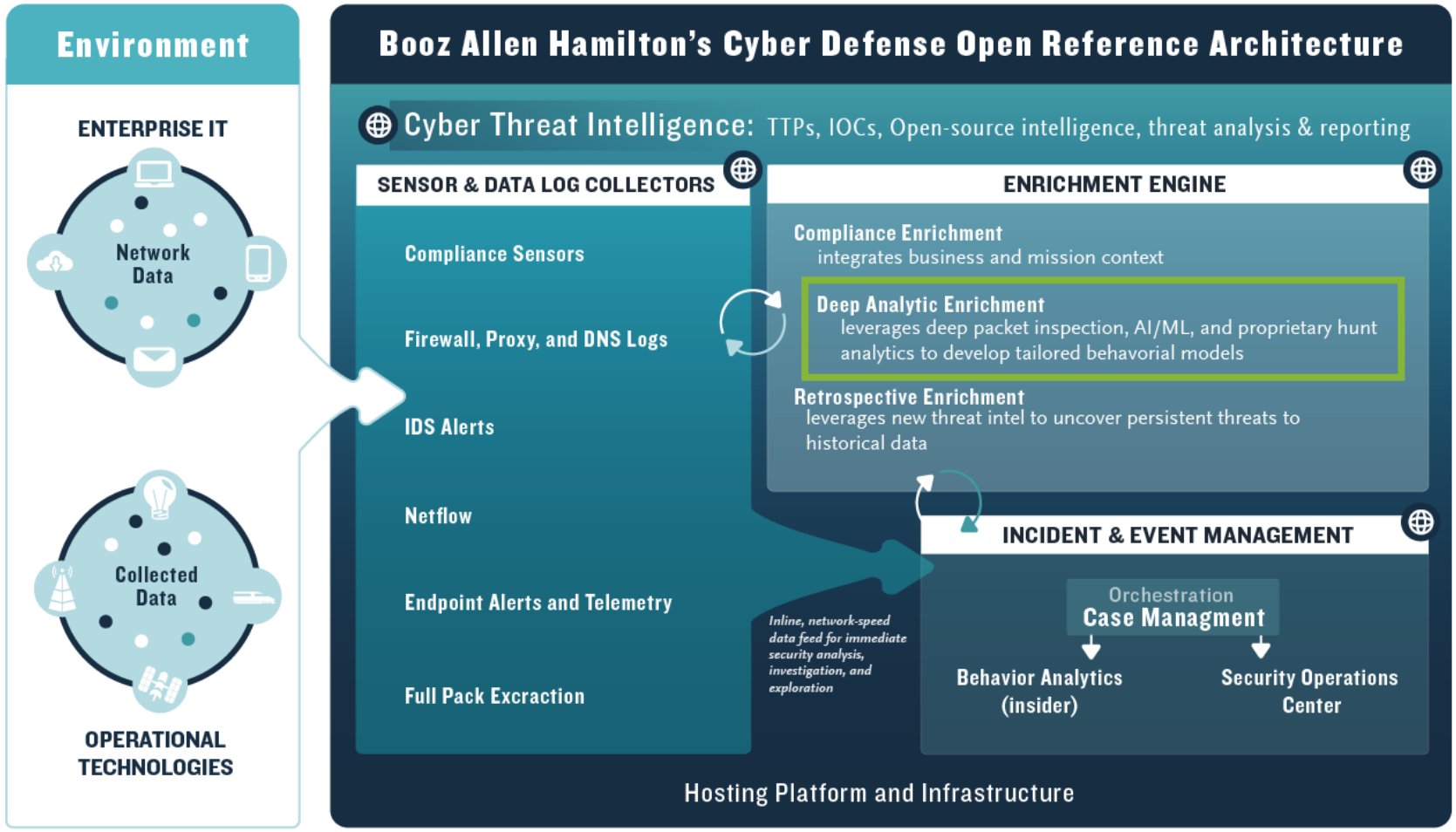
- Inherently Reactive
- Defined by Rules
- Built by Products
- Lacks Customization
- Human-dependent

Modernization requires collaboration and innovation

Booz Allen and NVIDIA fuse capability offerings across domains to maximize solution impact



Optimal solutions are both adaptive and additive



Characteristics

- Inherently Adaptive
- AI Accelerates Analysts
- Enrichment is Additive
- Network Tailored
- Network Speed Alerts

NVIDIA's RAPIDS aims to accelerate everything on GPUs

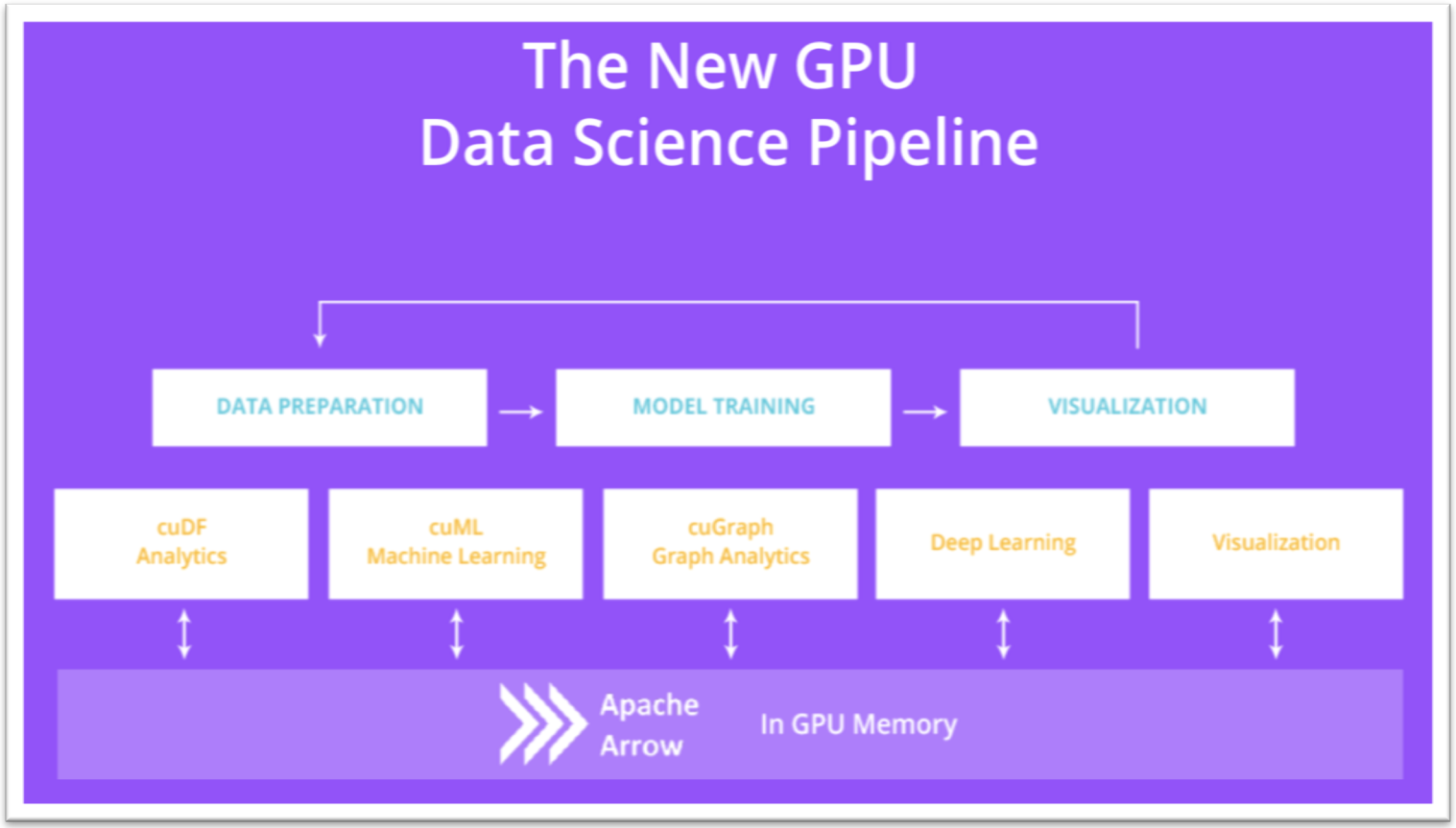
Why use GPUs?

- ***A CPU is an Executive:*** composed of a few cores with lots of cache memory – they can do anything and everything
- ***A GPU is a Laborer:*** GPUs composed of hundreds of cores that can handle thousands of threads simultaneously, optimized for performing the same operation over and over

What is RAPIDS?

- Suite of open-source, end-to-end data science tools
- Built on CUDA
- Pandas-like API for data cleaning and transformation
- Scikit-learn-like API for ML
- A unifying framework for GPU data science

RAPIDS is powerful, easy to use, and a great fit for the cyber use case



Hassle free integration requiring minor code edits

Reduced ML workflow and model compute times

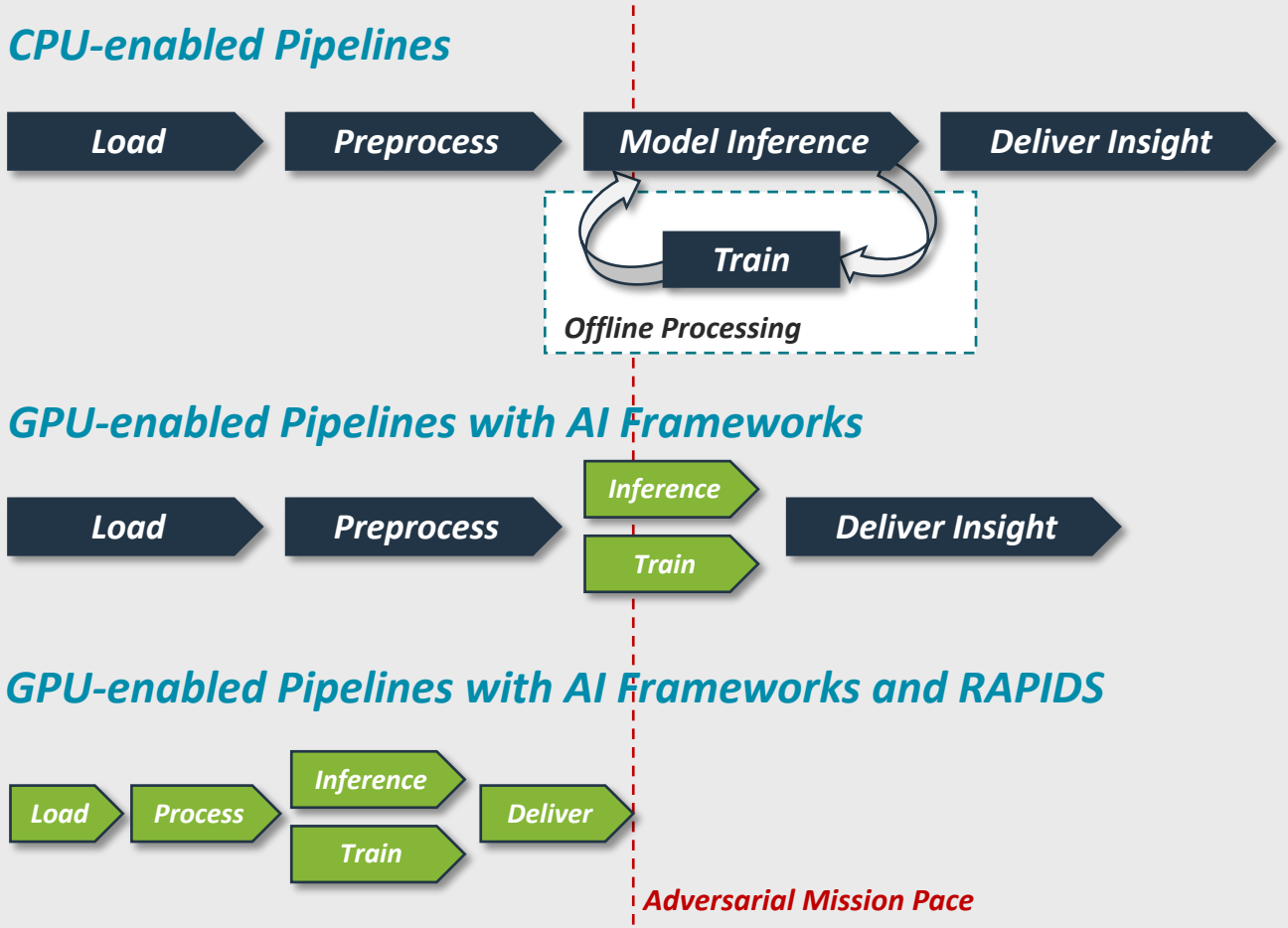
Open-source, quickly evolving and improving

Improves performance for select ML implementations

In cybersecurity,
speed is paramount.

Data moves at high velocity, and every second in delays in alerting and detection is more time for adversaries to cause more damage.

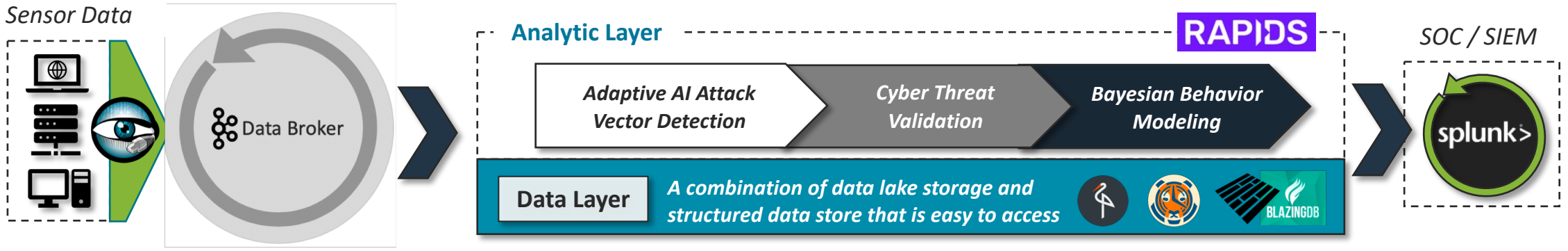
AI Pipeline Speed Up is Different than AI Speed Up



With end-to-end GPU processing, our AI defense can keep up

Booz Allen's Cyber Precog: AI-enabled enrichment

Cyber Precog: *Combining network speed alerting with adaptive, endpoint behavioral learning for use case oriented defenses*



Cyber Precog is:

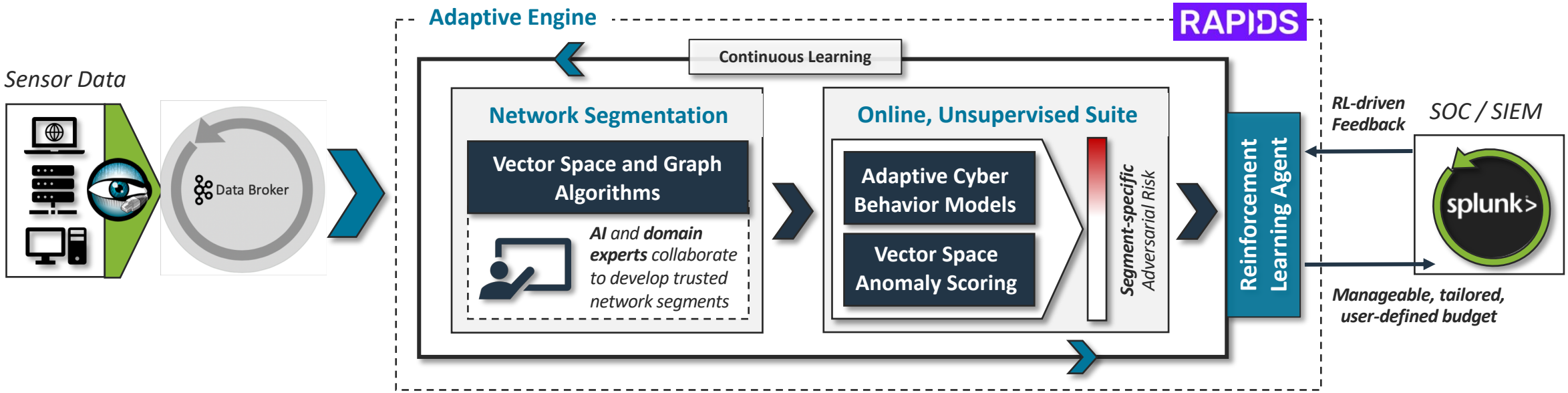
- ***Fast***: *End-to-end, fully GPU and RAPIDS accelerated*
- ***Flexible:*** *Agnostic of any particular tool or architecture*
- ***Highly effective:*** *Designed to mitigate false positives*
- ***Cyber-oriented:*** *All analytic outputs mirror cyber mission*

But… There are some limitations:

The current method assumes use cases can be architected at the adversarial pace, as adversaries identify new attack vectors. ***Cyber Precog needs an adaptive and efficient anomaly detection engine that can detect new attack vectors as they are implemented.***

Intelligent identification of truly adversarial anomalies

An unsupervised engine *that can decrease the alert burden, learn over time, and differentiate adversarial anomalies from random human behavior*



Defining Characteristics:

- ***Segmentation structures learner feedback, improving impacts of reinforcement learning agent***
- ***Anomaly scoring is done with network segment and cyber context, minimizing false positives***
- ***User feedback encoded in reinforcement learning algorithm differentiates adversarial anomalies***
- ***Engine is fully network tailored and remains assumption agnostic, without pretraining bias***

In partnership with NVIDIA

Booz | Allen | Hamilton

Case Study: Impacts of RL on a GPU-enabled ensemble

Background

- **Network flows:** over 80 features, CICFLOWMETER
- **Simulated attacks**
- **One week of traffic**

Canadian Institute of Cybersecurity *released over 50GB of test data from a week of simulating attacks in 2017,* ***the CICIDS17 dataset***

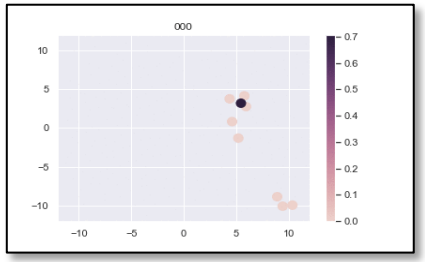
An ensemble of models

K-Means in RAPIDS

K-Means *normalizes the feature space and builds clusters to optimize cluster distance and tightness in vector space*

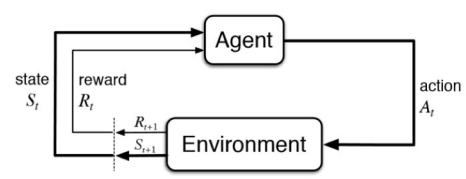
CLOF in CuPy / FAISS

Composite Nearest Neighbor Local Outlier Factor scores anomalousness within retrospective windows using distance measures and neighborhood algorithms



Reinforcement Learning Agent

Implements a contextual bandit variant of a fitted Q-iteration algorithm

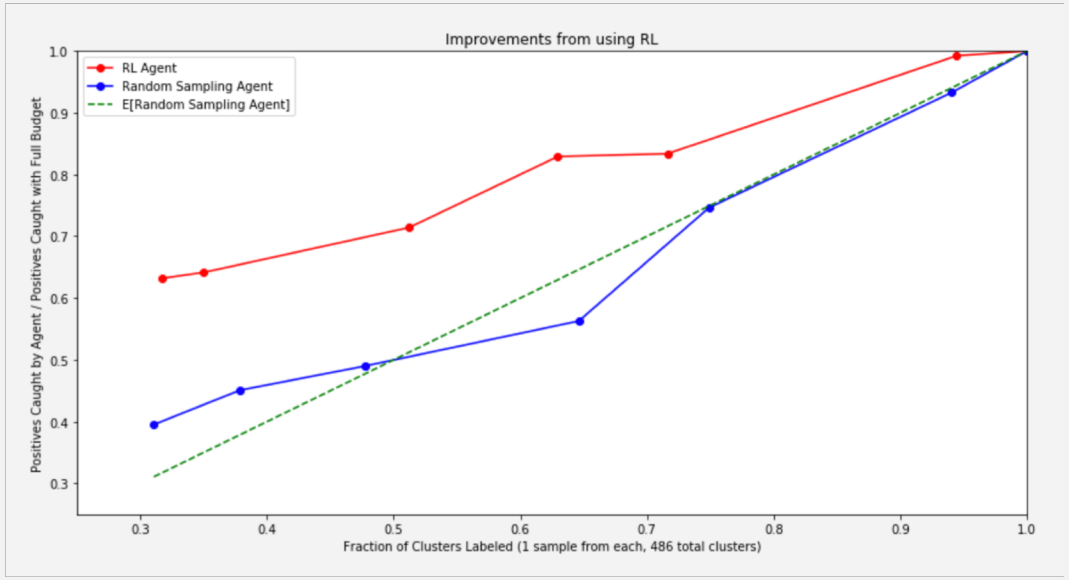


Actions are assumed to not effect scenario state, but method integrates modeling of long term action effects

Allows the user to optimize the tradeoff of detection and exploration

Allows for a principled, but dynamic, balance of exploitation and exploration (e.g., near-term and long-term benefits)

Implemented fully in RAPIDS, entirely on GPUs, and deployed using XGBoost



Implementation Performance

96% alert accuracy (4% false positive rate)

3x improvement on recall with RL

Deployment Functionality

Inference at over 5 GB/s on single node single GPU

Single node, 4 GPU rates jump up over 12 GB/s

Future Directions

- *Organization scale test and parameter optimization*
- *Embed with NVIDIA GPU parsers and graph analytics*
- *Deploy on network for validation*

Thank you!

Special thanks to the development team!

Andre Nguyen is a machine learning researcher at Booz Allen Hamilton working primarily on adversarial defense, cyber, and synthetic data research. Previously, Andre led development on Booz Allen's algorithmic warfare strategic investment, architected cloud data platforms, and delivered for clients in the pharmacovigilance space. Andre graduated from Harvard with a Bachelor's in Applied Mathematics and Computer Science and is an Amazon Web Services certified cloud solutions architect.

Sarah Olson is a Data Scientist and machine learning engineer at Booz Allen Hamilton, with current focus areas in cyber security, climate science, analytic tool development, machine learning parallelization, and natural language processing. Sarah brings deep experience in NVIDIA's RAPIDS platform to the team Sarah graduated from the University of Notre Dame with a Bachelor's in computer science and a minor in philosophy.

Will Badart is a Machine Learning Engineer at Booz Allen Hamilton who researches AI-driven cyber defenses and designs the larger systems which deliver them, with a focus on GPU-enabled solution design and deployment. In past lives, he was a software engineer at Facebook, full-stack developer at Booz Allen, and freelance web developer. He has a degree in Computer Science from the University of Notre Dame.

Jesse Shanahan is a data scientist at Booz Allen working primarily on cyber anomaly detection and cyber risk modeling. Jesse is also focusing on AI ethics and developing effective AI for humanitarian aid. Previously, Jesse worked as a researcher, studying supermassive black holes. She did her graduate studies in Astrophysics at Wesleyan University and undergraduate in North African Linguistics as an Echols Scholar at the University of Virginia.

As well as all of our NVIDIA collaborators on the RAPIDS team

In partnership with

NVIDIA

Booz | Allen | Hamilton