Autonomous Aggregate Data Analytics in Untrusted Cloud

Ganapathy Mani, Denis Ulybyshev, Bharat Bhargava
Jason Kobes*, Puneet Goyal^

CS & CERIAS, Purdue University
***Northrop Grumman Corporation**
^Department of CSE, IIT Ropar, India

AIKE 2018

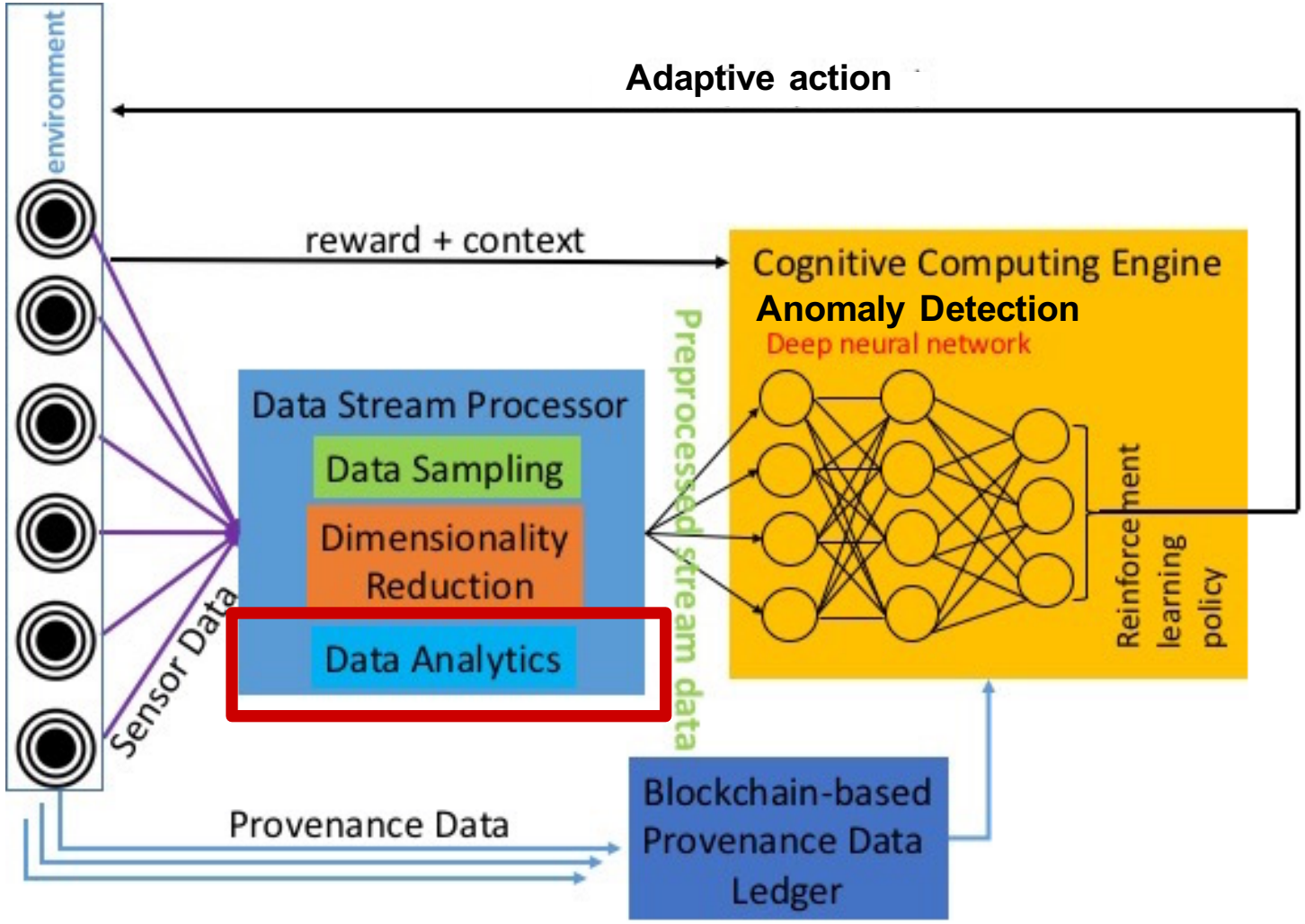
Intelligent Autonomous Systems

- 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 will be on the smart cyber systems.

Comprehensive IAS Architecture



Motivation

- 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

- 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

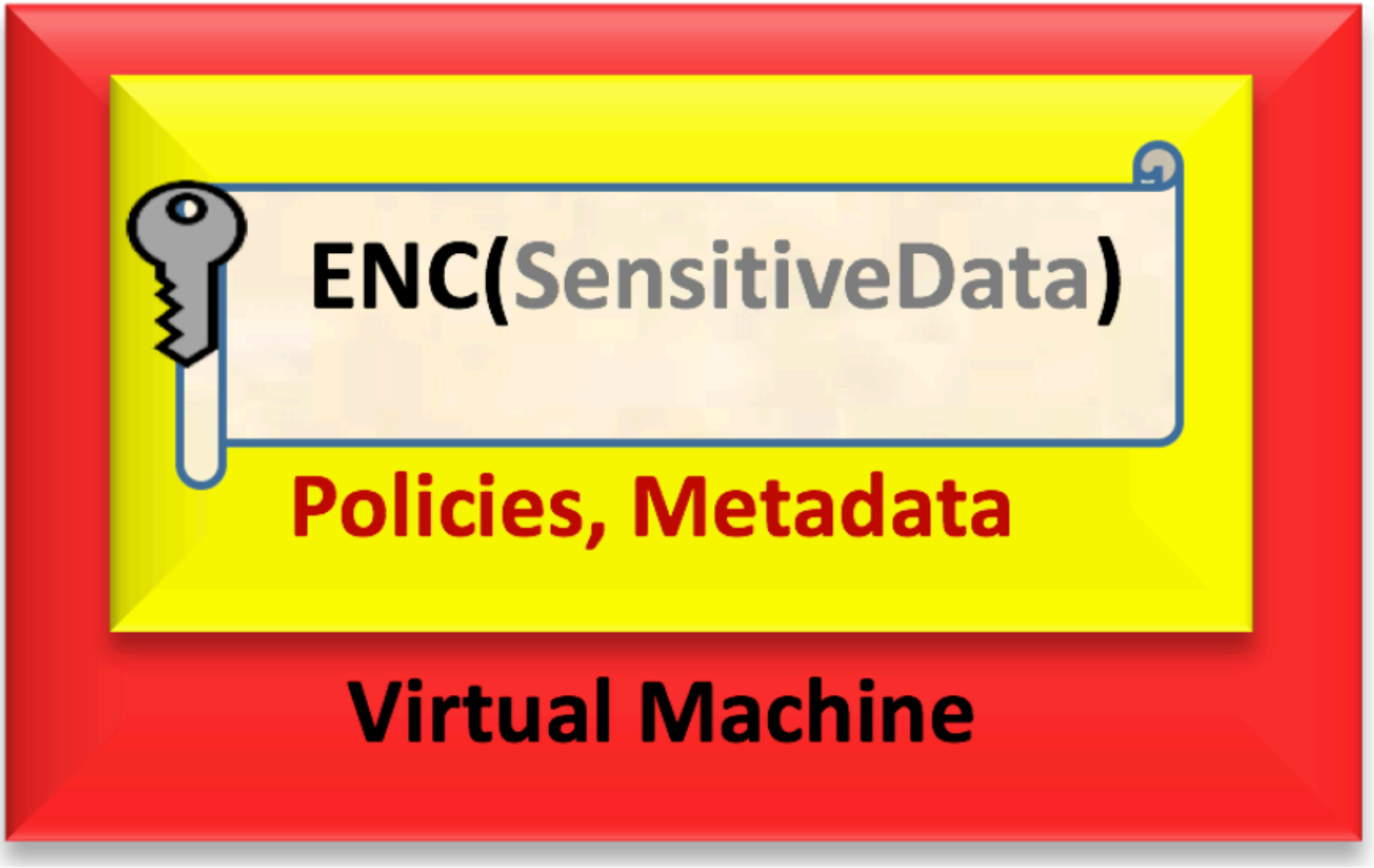
- Instead of checking AB's authentication protocol every time, an AB can obtain a one-time pass to access other ABs data per aggregate query.

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

Active Bundle (AB)

- Active Bundle (AB) is a distributed self-protecting entity with policy enforcement engine.
- Sensitive data is stored in a non relational database in the form of key-value pair. E.g. {*PatientID = "ENC(123456)"}.*
- Authentication of client services is based on digital certificates. The services present their X.509 certificates signed by a trusted Certificate Authority (CA).
- After authentication, policy enforcement engine enforces policies of data access depending upon the service's access level.

Active Bundle (AB)



AB Authentication Protocol - Problem

- Every time a service requests a particular data from active bundle, it has to go through authentication and enforcement policies.

- For each Active Bundle, based on number of policies, the data access time increases.
  - Around 500 msec for 16 policies

- For each Active Bundle, based on security protocols of authentication, the authentication time increases.
  - Around 550 msec for two-way encryption

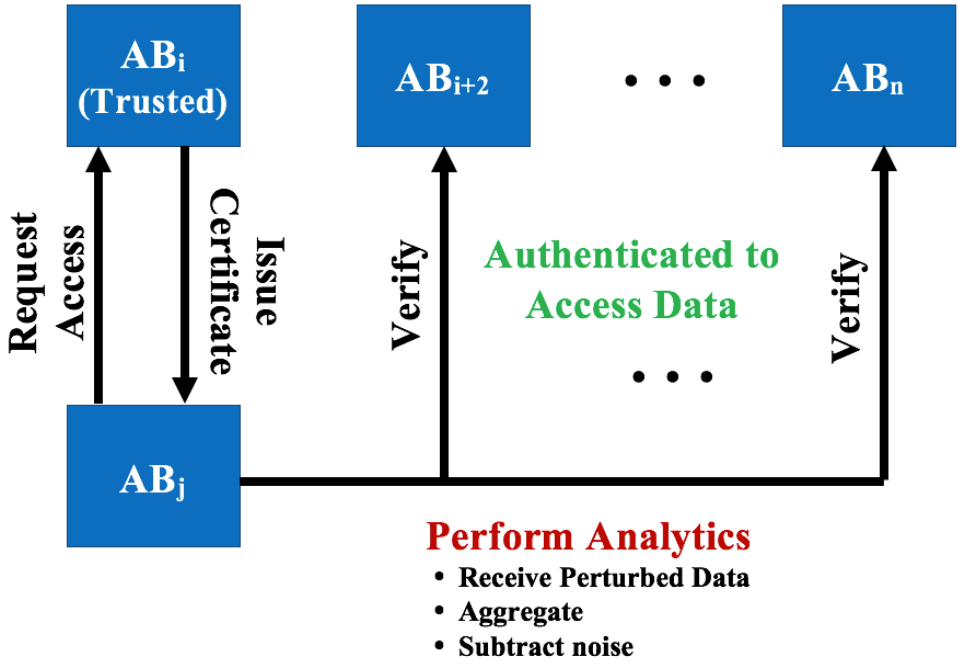
- So the system is not scalable for large databases and data analytics will become enormously time consuming.

One-time AB Authentication Protocol - Solution

- We propose a solution: one-time authentication per aggregated query.
- Here, each autonomous entity such as active bundle can be given a one-time certificate to perform a specific task without going through policies and authentication for each AB.
- One trusted Certificate Authority (CA) can provide the autonomous entity a one-time access pass and restrict the pass to the requested data.
- With this one time authentication, AB can surpass other ABs' policies and authentication, making it faster.

One-time AB Authentication Protocol

- Here, a trusted AB_i provides access certificate to another AB_j.
- AB_j uses the certificate to access other ABs without having to go through policies again.



One-time AB Authentication Protocol

Data: AB_i and AB_j as inputs
Result: Certificate issued/denied/issued with restrictions
if *Type*(AB_i) *is same as* *Type*(AB_j) **then**
  if *Trust*(AB_i) *is greater than* *Trust*(AB_j) **then**
    Generate authentication certificate;
    Issue the certificate to AB_j;
  else
    Generate Certificate with restrictions (only access encrypted data);
  end
else
  Deny the request;
  Report to administrator;
end

Privacy Preserving Data Aggregation

- 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

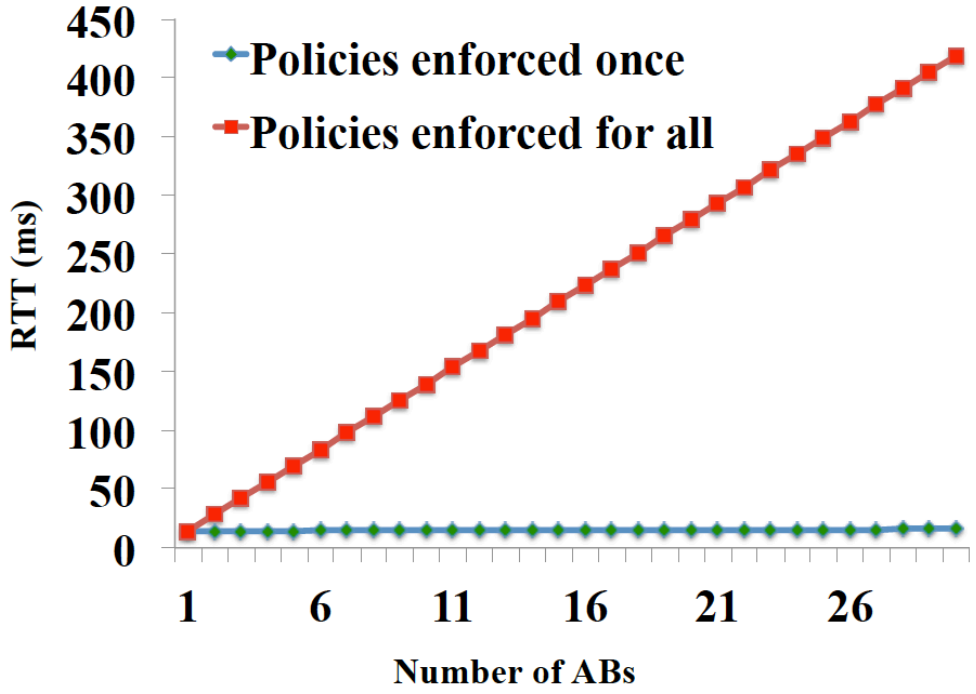
- We measure the latency of data request sent to AB, which is hosted by a local server, located in the same network with the client.

- 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

- 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.



Future Work

- Changing policies on-the-fly is a non-trivial problem in autonomous cyber systems.

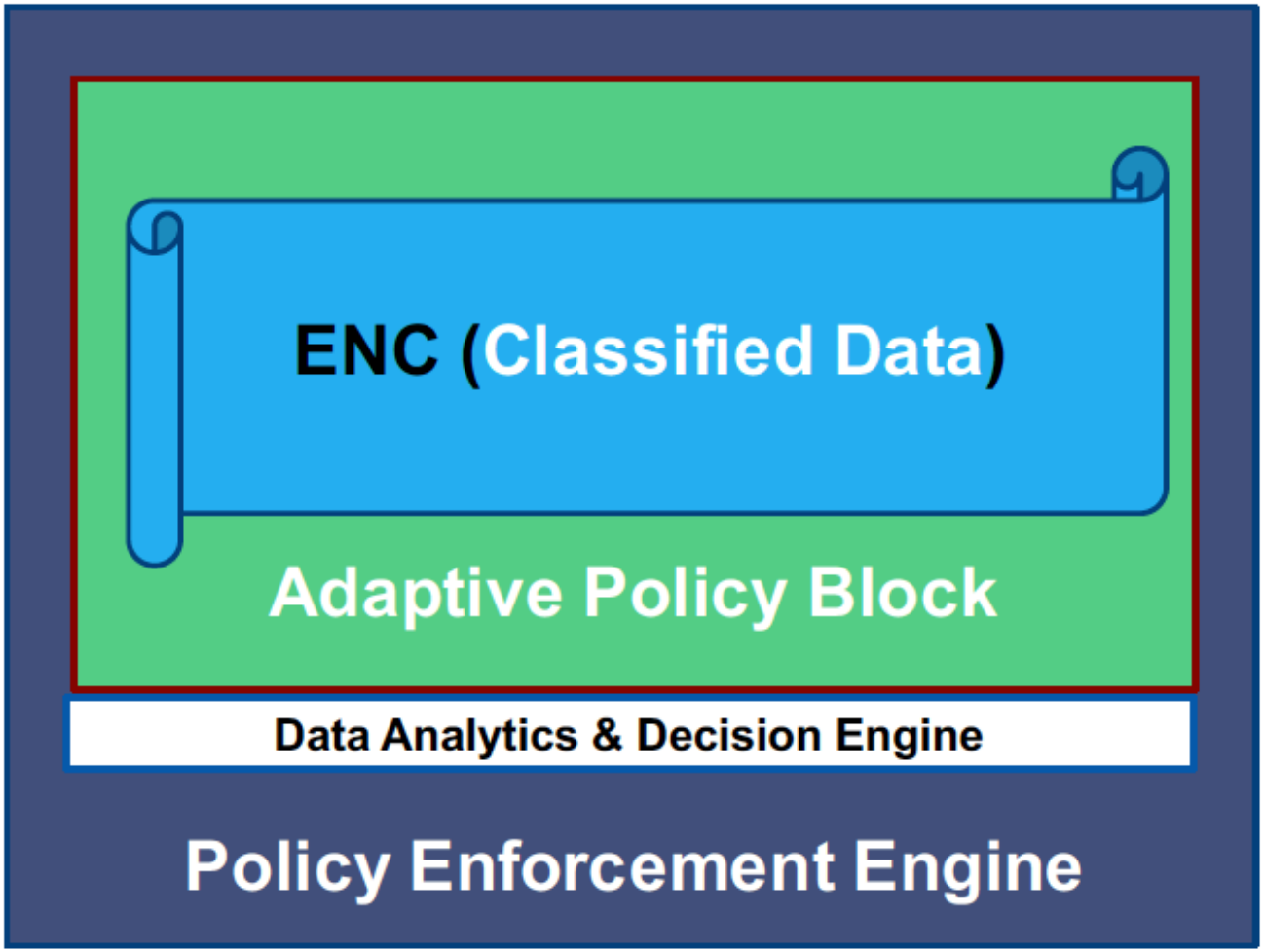
- Autonomous policy changes based on the data analytics can be achieved by introducing an adaptive block with probabilistic rules.

- We plant to implement deep learning methodologies for adapting to new and unknown scenarios, learn from the data, and make probabilistic reasoning to enforce policies.

-

Future Work

- Autonomous policy changes based on the data analytics.



References

L. B. Othmane, *Active bundles for protecting confidentiality of sensitive data throughout their lifecycle*. Western Michigan University, 2010.

L. Lilien and B. Bhargava, "A scheme for privacy-preserving data dissemination," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 36, no. 3, pp. 503–506, 2006.

D. Ulybyshev, B. Bhargava, M. Villarreal-Vasquez, A. O. Alsalem, D. Steiner, L. Li, J. Kobes, H. Halpin, and R. Ranchal, "Privacy-preserving data dissemination in untrusted cloud," in *Cloud Computing (CLOUD), 2017 IEEE 10th International Conference on*. IEEE, 2017, pp. 770–773.

"W3c web cryptography api," 2018. [Online]. Available: https://www.w3.org/TR/WebCryptoAPI/

"Web authentication: an api for accessing scoped credentials," 2018. [Online]. Available: http://www.w3.org/TR/webauthn/

Thank you!!!