AI in cyber security

P.V. Ananda Mohan

Life Fellow IEEE

Bangalore

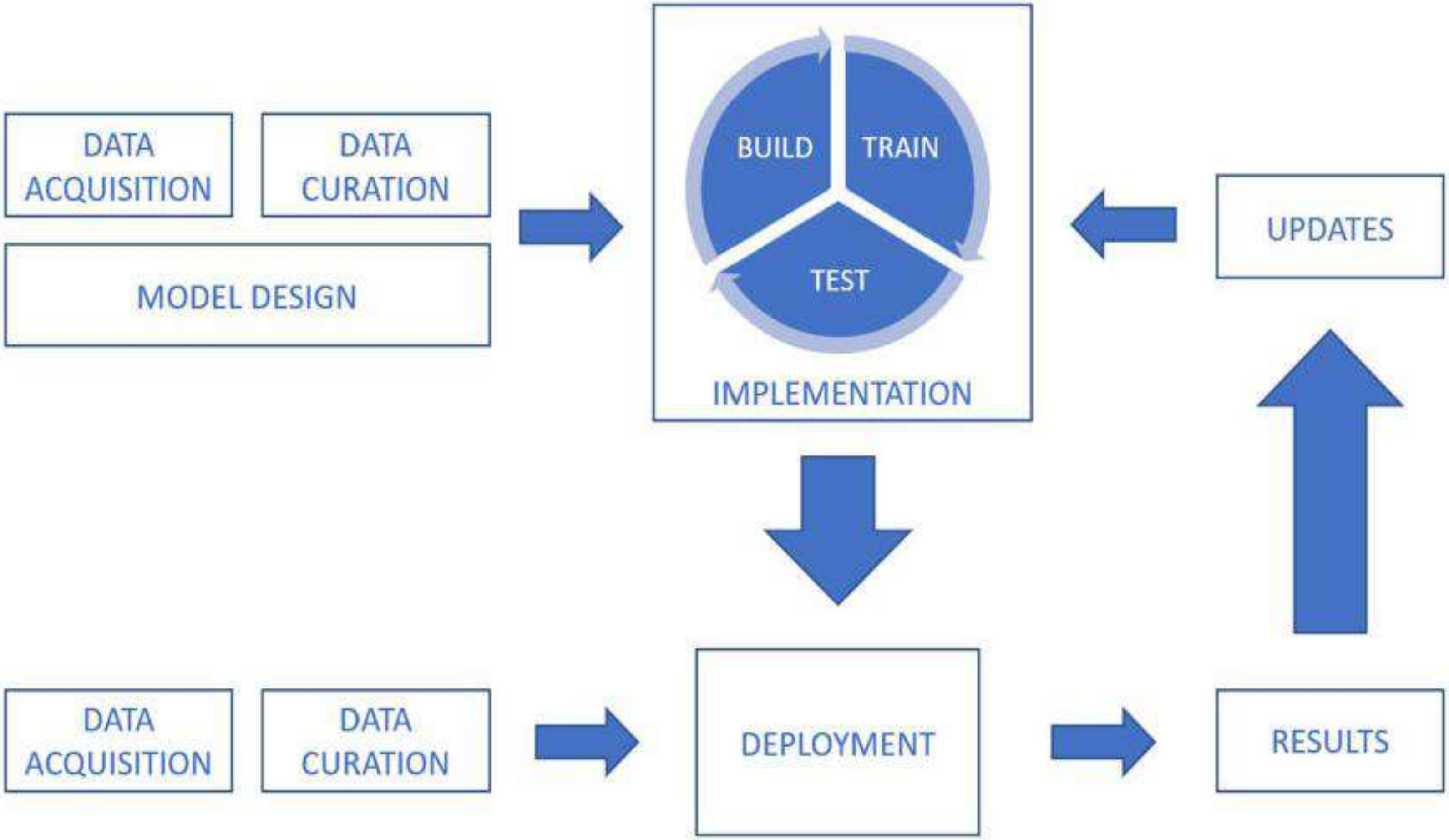
10th February 2022

Artificial Intelligence

- McCullocch and Pitts 1943
- John McCarthy, 1956 *“The science and engineering of making intelligent machines, especially intelligent computer programs”.*
- **Artificial intelligence is the ability of a system to handle representations, both explicit and implicit, and** procedures to perform tasks that would be considered intelligent if performed by a human.

AI terminology

- Machine Learning, Expert System, Deep learning are all subsets today’s AI technology.
- Expert System : It usually comprises a set of pre-configured rules.
- Machine Learning: If it comes from labeled data, it is called supervised learning
- deep learning: It is modern “scalable machine learning” which does not require humans to conduct feature extraction


DATA ACQUISITION
DATA CURATION
MODEL DESIGN
BUILD
TRAIN
TEST
IMPLEMENTATION
UPDATES
DATA ACQUISITION
DATA CURATION
DEPLOYMENT
RESULTS

- Source: ETSI

Learning

- Supervised learning –labelled data
- Semi-Supervised learning – partially labelled data
- Un-supervised learning-unlabelled data
- Reinforced learning – experience based adaptive larning

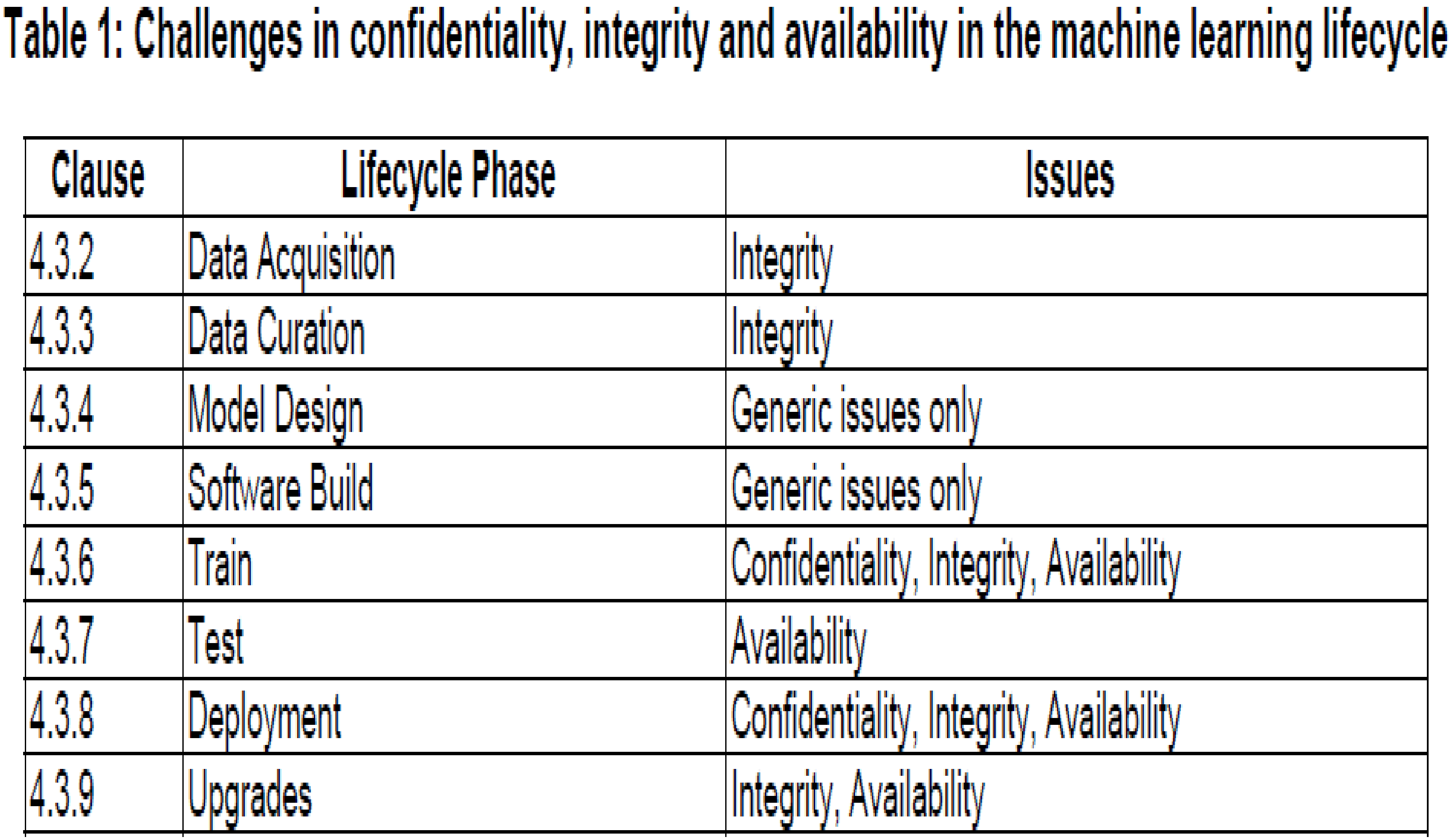
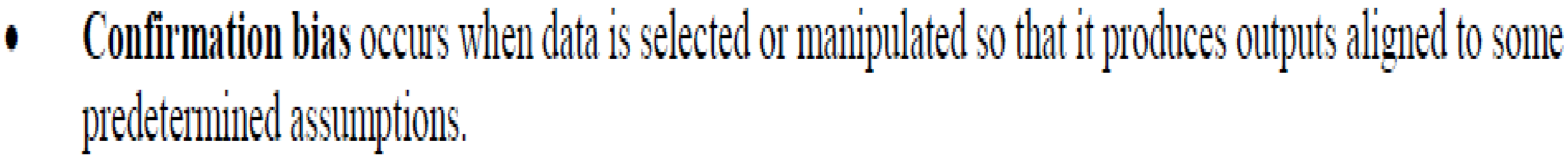


Table 1: Challenges in confidentiality, integrity and availability in the machine learning lifecycle

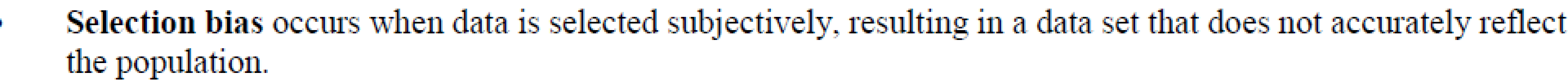
Clause	Lifecycle Phase	Issues
4.3.2	Data Acquisition	Integrity
4.3.3	Data Curation	Integrity
4.3.4	Model Design	Generic issues only
4.3.5	Software Build	Generic issues only
4.3.6	Train	Confidentiality, Integrity, Availability
4.3.7	Test	Availability
4.3.8	Deployment	Confidentiality, Integrity, Availability
4.3.9	Upgrades	Integrity, Availability

- Source ETSI

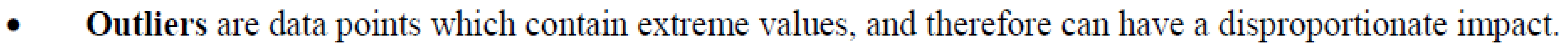
Bias



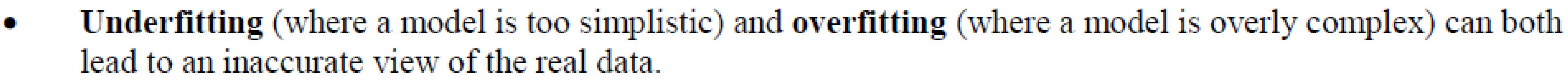
- **Confirmation bias** occurs when data is selected or manipulated so that it produces outputs aligned to some predetermined assumptions.



- **Selection bias** occurs when data is selected subjectively, resulting in a data set that does not accurately reflect the population.



- **Outliers** are data points which contain extreme values, and therefore can have a disproportionate impact.



- **Underfitting** (where a model is too simplistic) and **overfitting** (where a model is overly complex) can both lead to an inaccurate view of the real data.

Attacks

- **Poisoning attacks:**
- Data set poisoning
- Algorithm poisoning
- Model poisoning
- **Input attacks**
- **Backdoor attacks**
- **Reverse Engineering (model extraction attack)**
- **Ad-Blocker attacks** (obfuscating HTML /metadata; look at content to prevent)
- **Malware obfuscation**
- **Deepfakes**
- **Fake conversation**

Goals in Cyber security

- Sifting through the myriad of alerts
  and identifying true from false positives at a certain level of accuracy
- Discovering patterns across signals automatically
- Based on the context, provide actionable recommendations
- Learning from past behavior to recommend relevant insights or suggestions to provide new context

Deep learning proliferation

- Availability of Big data sets
- New GPU based hardware
- Convolutional Neural networks (CNN) = Deep learning

Main Challenges Cyber Security

- attacks are becoming ever more dangerous.
- **Geographically distant IT systems —** Geographic distance makes it more difficult to manually track accidents.
- **Manual threat detection –** can be expensive and time-consuming, leading to more unexpected attacks.
- **Cyber Security's reactive nature —** companies can only fix issues once they've already occurred. It is a huge challenge for security experts to predict threats before they occur.
- **Hackers frequently cover their IP addresses and modify them —** hackers use different programmes such as Virtual Private Networks (VPN), Proxy servers, Tor browsers, and more. These systems aid in keeping hackers anonymous and undetected.
- **global cybersecurity skills shortage**

Possible Benefits of using AI

- Offers instant insights by curating the information of threats from millions of research papers, forums and news reports
- Can identify irregularities in the network by analyzing user actions and studying the patterns.
- Machines can learn based on past behavior to identify new attacks, distinguish malicious signals from benign signals, and can trigger workflows or recommend remediation actions.
- By Automating Repetitive and boring security tasks, Human security resources can also be allocated efficiently with the help of machine learning.

- **Regression**—detects correlations between different datasets and understand how they are related to each other. You can use regression to predict system calls of operating systems, and then identify anomalies by comparing the prediction to an actual call.
- **Clustering**—identifies similarities between datasets and groups them based on their common features. Clustering works directly on new data without considering previous examples.
- **Classification**—Classification algorithms learn from previous observations, and try to apply what they learn to new, unseen data. Classification involves taking artifacts and classifying them under one of several labels. For example, classify a binary file under categories like legitimate software, adware, ransomware, or spyware.

How does AI Enhance Cyber Security?

- **1) Threat detection**
- Traditional safety techniques employ signatures or IoC (Indicators of Compromise) to identify threats.
- AI can increase the detection rate of traditional techniques up to 95 percent. The problem is that you can get multiple false positives. Best solution is to merge conventional approaches and AI. This will lead to a detection rate of 100 % and eliminate false positives.

2) **Vulnerability Management**

- User and Event Behavioural Analytics (UEBA) can analyse user account, endpoint, and server baseline behaviours and identify anomalous behaviours that could signal an unknown zero-day attack.
- Even before vulnerabilities are officially reported and patched, this can help protect organizations.

3) **Data Centres**

- AI can optimise and monitor many essential processes in the data centre such as backup power, cooling filters, power consumption, internal temperatures and use of bandwidth.

4) Network Security

- Traditional network security has two time-consuming aspects: (a) Security policies and (b) network topography.
- **Policies —** With the vast number of networks, the real difficulty is designing and managing the policiesfor big networks.
- **Topography –**
- Using AI can improve network security by studying network traffic patterns and recommending both functional workloads grouping and security policy.

Disadvantages and Weaknesses of Using AI for Cyber Security

- Needs enormous quantities of resources - memory, data, and computing power
- AI systems might be educated by data sets- have several different malware codes, non-malicious codes, and anomaly data sets.
- Hhackers can also use AI themselves to check and refine their malware so that it can actually become AI-proof.

Desirable Features

- Most of AI algorithms are known be a black box and the results are hard to interpret.
- ***Explainable:*** insights should always be supported by data and evidence
- ***Controllable:*** Analysts should have the flexibility to edit what AI suggests in an intuitive way. AI shall be under the security analysts' control.
- ***Adaptive:*** An AI engine must continuously compute and listen to the security analyst's feedback and be able to provide refreshed insights about a unique environment and new data sets.
- ***Alerts:*** In the event that whenever the AI calculations notice uncommon exercises or any conduct that falls outside your standard examples.

Artificial Intelligence and the Attack/Defense Balance
Bruce Schneier

- You can divide Internet security tasks into two sets: what humans do well and what computers do well.
- computers excel at speed, scale, and scope. They can launch attacks in milliseconds and infect millions of computers.
- They can scan computer code to look for particular kinds of vulnerabilities, and data packets to identify particular kinds of attacks.
- Humans, conversely, excel at thinking and reasoning. They can look at the data and distinguish a real attack from a false alarm, understand the attack as it’s happening, and respond to it. They can find new sorts of vulnerabilities in systems. Humans are creative and adaptive, and can understand context.
- Computers—so far, at least—are bad at what humans do well. They’re not creative or adaptive. They don’t understand context. They can behave irrationally because of those things.
- Humans are slow, and get bored at repetitive tasks. They’re terrible at big data analysis. They use cognitive shortcuts, and can only keep a few data points in their head at a time. They can also behave irrationally because of those things.

Paul Kocher- Expert on Side-Channel analysis

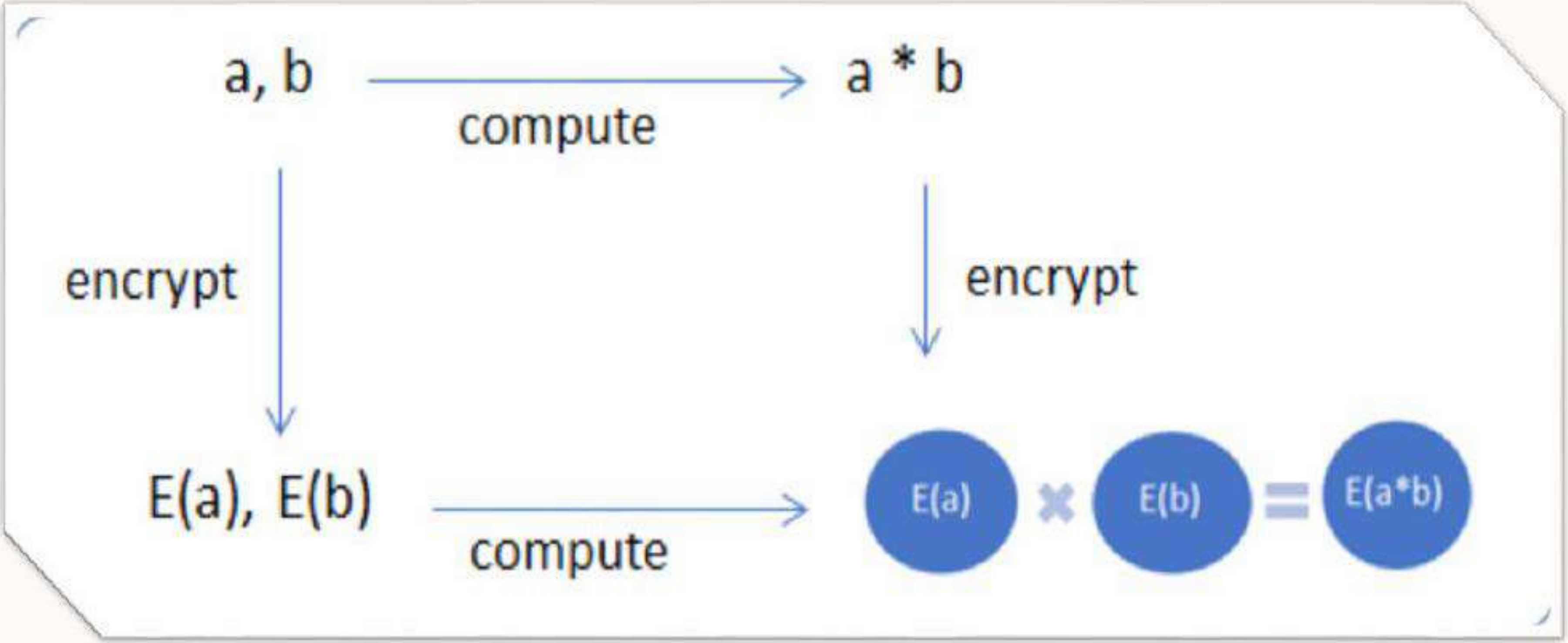
“There’s an open question about whether AI can be taught to understand properties of software-hardware design and tell us useful things about them; for example, whether the design is one that might have certain categories of bugs in it. There’s an open question about how far AI can go there,” he explained. “The current AI applications tend to be ones where you’re optimizing some kind of a search space or you have a relatively straight forward set of problems with very large amounts of training data. Understanding complex logic doesn’t fit very well into that mold [and] there are clearly some advances in AI that are needed for that to happen.”

Current trends

Private AI

- Machine learning on Encrypted Data
- In any system, Data needs to be input to the Algorithms to get predictions.
- May be through cloud through some agent and hence susceptible to be attacked.
- How to protect?
- Encrypt the data at the source.
- Result: You need to perform operations on Encrypted data – homomorphic Encryption!

Homomorphic Encryption



- Adapted from Kristin Lauter, Facebook AI research

Homomorphic Encryption

- Source Encryption adds noise to the computed result.
- Receiver Decryption subtracts the noise and retrieves the information.
- Old examples: RSA for Secure Addition
- Messages of A and B are m1 and m2. Public key e , Private key d, modulus n. Receiver is Y.
- A sends C1 = e^m1 (mod n)
- B sends C2 = e^m2 (mod n)
- Y computes (C1xC2)^d (mod n) = m1+m2.
- ElGamal can do multiplication operation on encrypted data.
- But now you can use ***Ring Learning with Errors (RLWE)***

SEAL

- Microsoft has developed in 2018 **Simple Encrypted Arithmetic library**

Applications in Cloud

- Private storage and computation
- Privatie AI prediction services
- Hosted Private training
- Private set intersection
- Secure collaborative computation
- Password breach detection

CRYPTEN: Secure Multi-Party Computation Meets Machine Learning – Facebook research

- Facilitates training of ML models on private data sets owned by different parties.
- Uses GPUs for computations
- Security against *semi-honest* corruption.
- Applicable for three party setting
- Uses Tensor computation (integer computations)
- Uses Arithmetic and binary secret sharing
- Uses Beaver triples supplied by Trusted Third parties (TTP)- computed offline (pre-processing by user)
- Performs linear functions, nonlinear functions, dot products, matrices, convolutions, sigmoid, softmax, exponentials, comparators
- Uses specialized primitives needed in AI computations

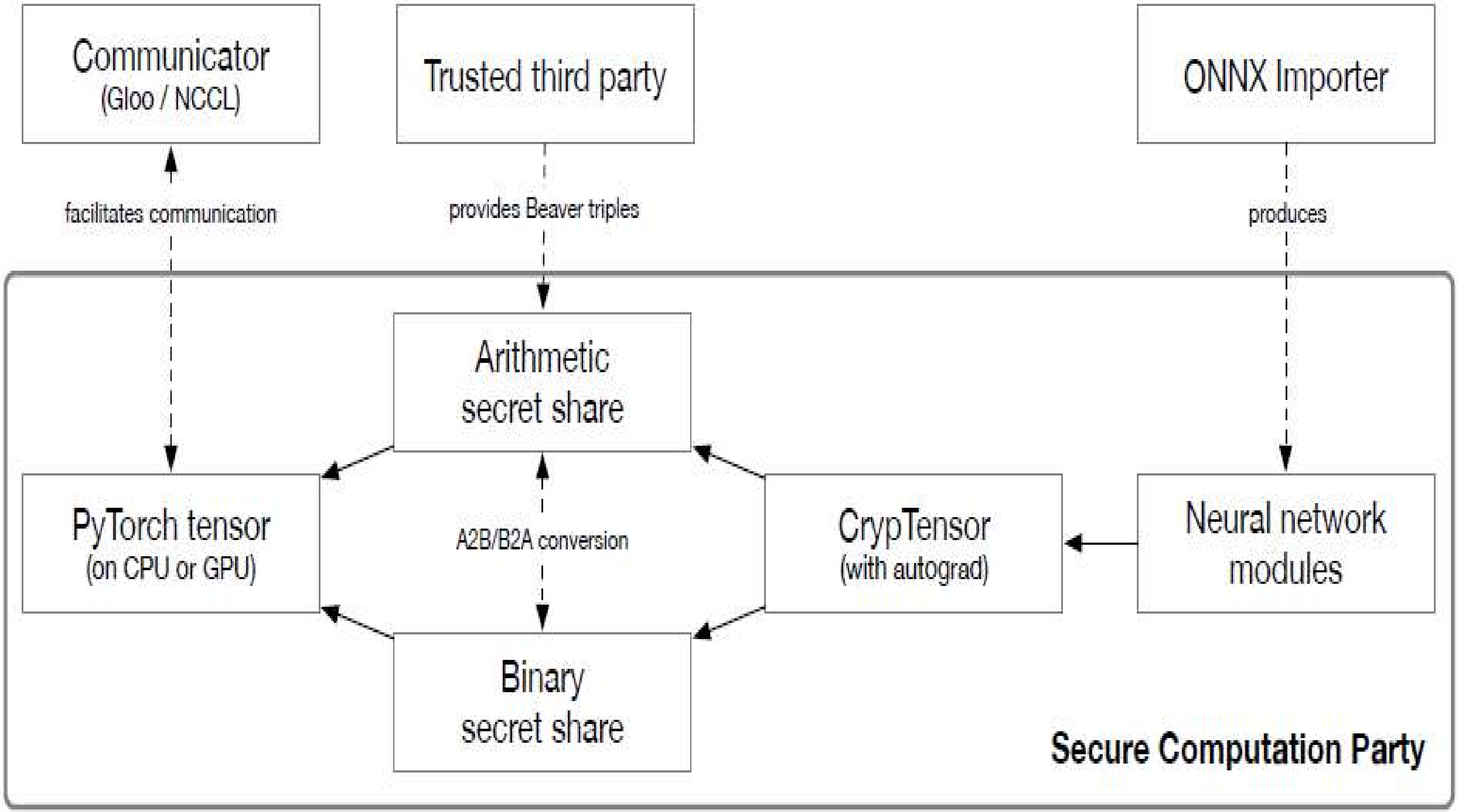
Secret sharing

- **Arithmetic secret sharing:**
- Secret ‘’X’’ is shared among parties. Sum of the shares is X.
- **Binary secret sharing:**
- Secret X in binary form
- Shares are Xor-ed to get X.

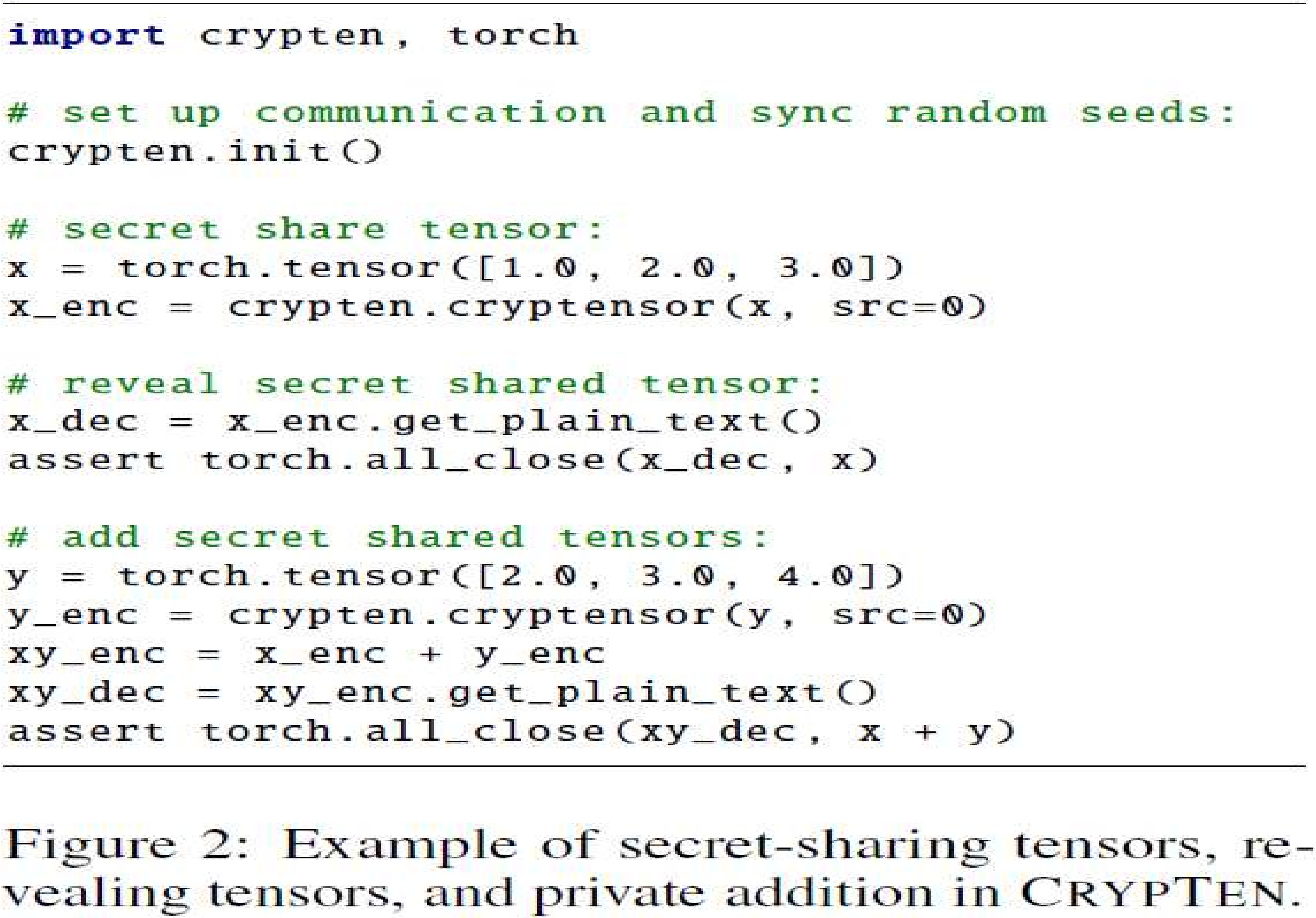
Beaver triple for secure multiplication

- Unique [a],[b], [c] such that $c = a \times b$ supplied by Trusted third party for each computation.
- To multiply x and y , A computes $\epsilon = [x] - [a]$, and B computes $\delta = y - [b]$.
- Now $[x] \times [y] = [c] + \epsilon [b] + [a] \delta + \epsilon \delta$!!!

Donald Beaver, Crypto 1991.



- Adapted from CRYPTEN, Brian Knott et al 2109.00984.pdf arXiv

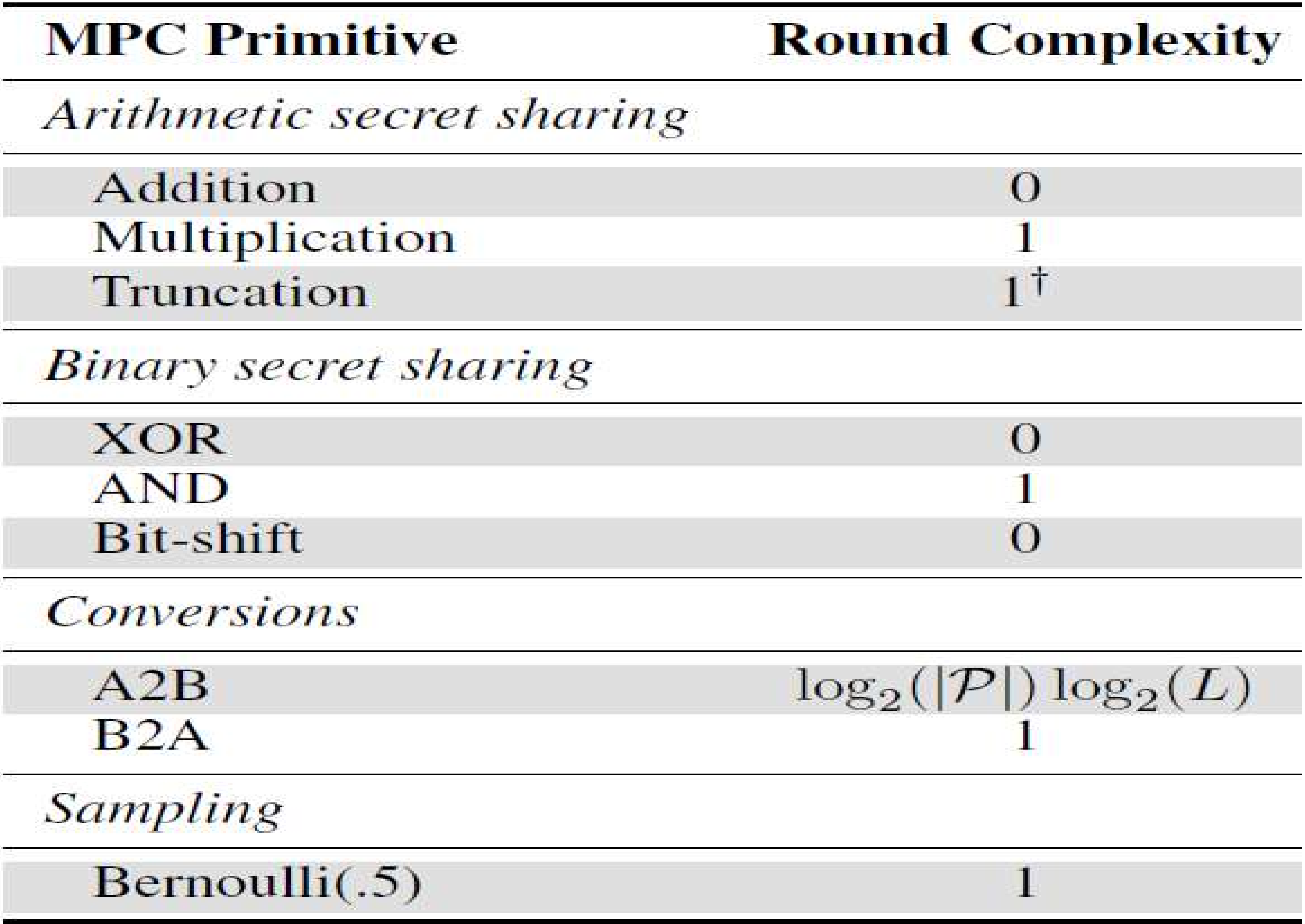


```
import crypten, torch

# set up communication and sync random seeds:
crypten.init()

# secret share tensor:
x = torch.tensor([1.0, 2.0, 3.0])
x_enc = crypten.cryptensor(x, src=0)

# reveal secret shared tensor:
x_dec = x_enc.get_plain_text()
assert torch.all_close(x_dec, x)

# add secret shared tensors:
y = torch.tensor([2.0, 3.0, 4.0])
y_enc = crypten.cryptensor(y, src=0)
xy_enc = x_enc + y_enc
xy_dec = xy_enc.get_plain_text()
assert torch.all_close(xy_dec, x + y)
```

Figure 2: Example of secret-sharing tensors, revealing tensors, and private addition in CRYPTEN.

- Adapted from CRYPTEN, Brian Knott et al 2109.00984.pdf arXiv

Primitives used in CRYPTEN

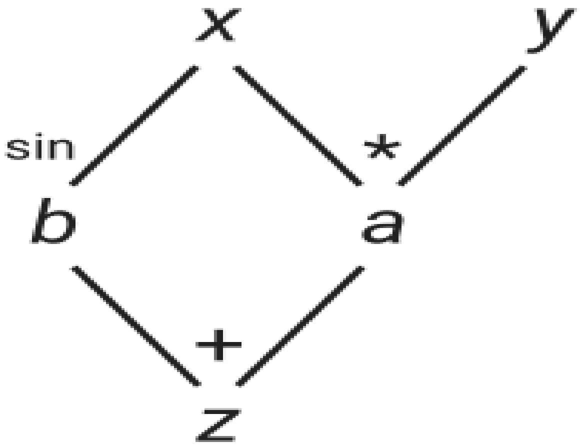


MPC Primitive	Round Complexity
Arithmetic secret sharing	
Addition	0
Multiplication	1
Truncation	$1^{\dagger}$
Binary secret sharing	
XOR	0
AND	1
Bit-shift	0
Conversions	
A2B	$\log_2(\lvert\mathcal{P}\rvert)\log_2(L)$
B2A	1
Sampling	
Bernoulli(.5)	1

Framework	Malicious security	Triple generation	Supports GPUs	Supports training	General purpose[†]
Two parties					
Chameleon [62]	✗	✗	✗	✗	✗
Delphi [49]	✗	✓	✗	✗	✗
EzPC [16]	✗	✓	✗	✗	✗
Gazelle [40]	✗	✓	✗	✗	✗
MiniONN [47]	✗	✓	✗	✗	✗
PySyft [64]	✗	✓	✓	✗	✗
SecureML [51]	✗	✓	✗	✓	✗
XONN [63]	✓	N/A	✗	✗	✗
Three parties					
ABY3 [50]	✗	N/A	✗	✓	✗
Astra [17]	✗	✓	✗	✓	✗
Blaze [59]	✗	✓	✗	✓	✗
CrypTFlow [43]	✗	N/A	✗	✗	✓
CryptGPU[‡] [67]	✗	✗	✓	✓	✓
Falcon [72]	✓	N/A	✗	✓	✓
SecureNN [71]	✗	N/A	✗	✓	✗
Four parties					
FLASH [11]	✓	N/A	✗	✓	✗
Trident [60]	✓	N/A	✗	✓	✗
Arbitrary number of parties					
CrypTen (ours)	✗	✗[§]	✓	✓	✓

Reverse mode automatic differentiation (RMAD)

- Derivative of ReLU is required for computing the gradients
- $z = x \times y + \sin(x)$
- Let us say you want next derivative of z.
- RMAD enables calculation in one step. Derivative also is available.
- Needed for training gradient based neural networks.
- Gradient is available in the last two equations.



$gz = ?$

$gb = gz$

$ga = gz$

$gy = x * ga$

$gx = y * ga + \cos(x) * gb$

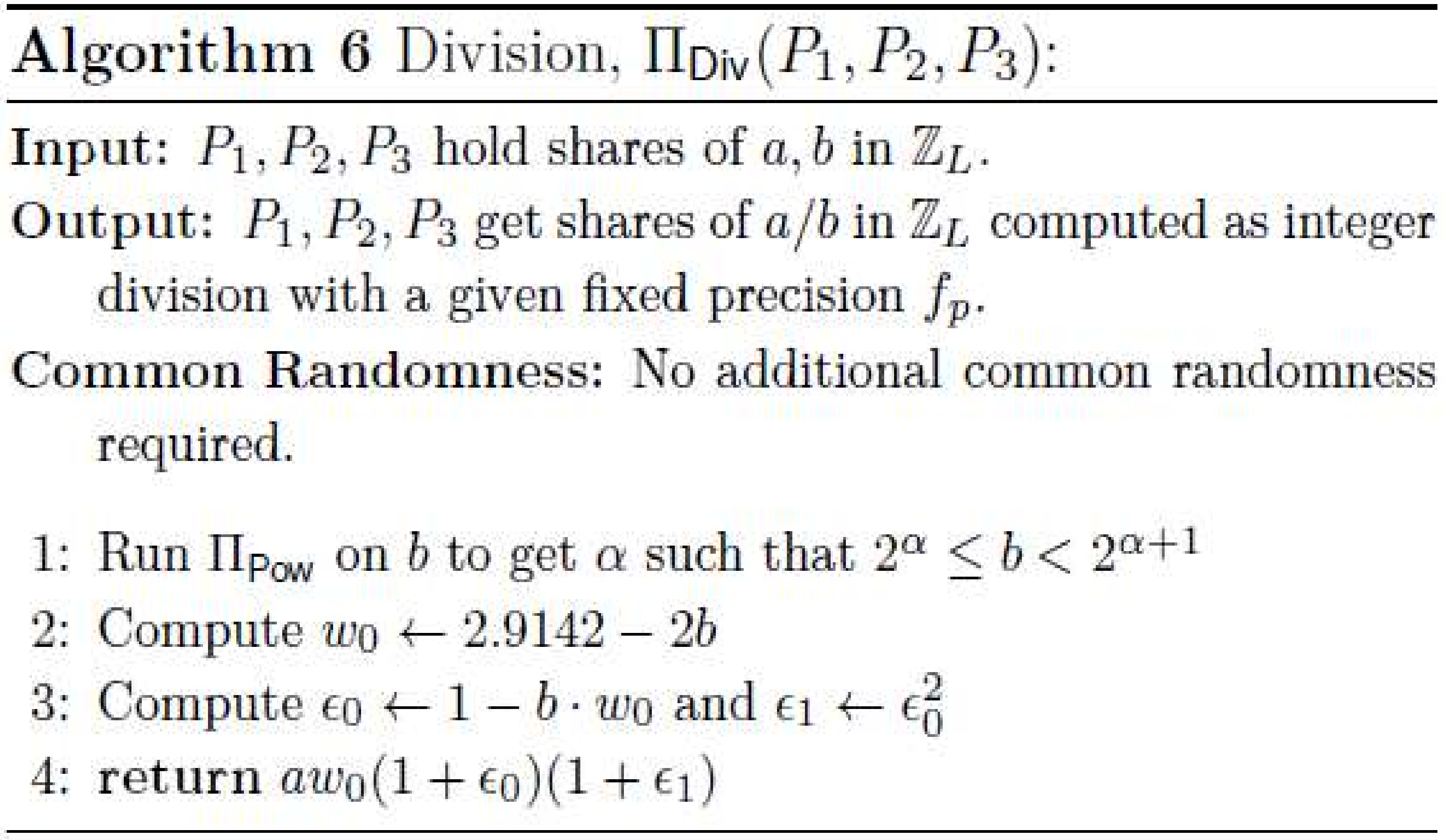
SAI- ETSI Working Group

- ETSI GR-SAI 004 Focuses on Integrity, Confidentiality, availability, ethics, applicability and avoidance of bias, attack vectors

Falcon

- Honest-Majority Maliciously Secure Framework for Private Deep Learning
- Application for image classification of child exploitative imagery
- 2 out of 3 party MPC (atmost one party can be corrupt)
- Batch normalization using mean and variance to improve distribution of inputs in various layers of the deep neural network. Needs computing inverse of a number.
- Semi-honest (sticks to protocols) and malicious (can change the protocols) cases can be handled
- Actors: Data holders-query users- three computing servers
- Wrap function (like carry in conventional arithmetic) is needed in ReLU.
- https://arxiv.org/abs/2004.02229

Division in FALCON to find a/b



Algorithm 6 Division, $\Pi_{\mathsf{Div}}(P_1, P_2, P_3)$:

Input: P_1, P_2, P_3 hold shares of a, b in $\mathbb{Z}_L$.

Output: P_1, P_2, P_3 get shares of a/b in $\mathbb{Z}_L$ computed as integer division with a given fixed precision f_p.

Common Randomness: No additional common randomness required.

```
Run $\Pi_{\mathsf{Pow}}$ on $b$ to get $\alpha$ such that $2^\alpha \leq b < 2^{\alpha+1}$
Compute $w_0 \leftarrow 2.9142 - 2b$
Compute $\epsilon_0 \leftarrow 1 - b \cdot w_0$ and $\epsilon_1 \leftarrow \epsilon_0^2$
**return** $aw_0(1 + \epsilon_0)(1 + \epsilon_1)$
```

- (1/b) is first computed.

Carnegie-Mellon Robust and Secure AI

- National AI engineering initiative Robust against model errors
- Robust against umodeled phenomena
- Underspecification in ML - picking one of the solutions – explainable AI (*XAI*)
- Dataset shift (real world data is different)
- Redundancy
- Adversarial machine learning: *Learn the wrong thing, Do the wrong thing, and reveal the wrong thing*

Conclusion

- Private AI = Earlier work on PPDM (Privacy preserving data mining) + Secure Multi-party computation.
- Extension to more than 3 parties is still to be developed.
- Performance has been reported on several applications.