Outsourcing Secured Machine Learning (ML)-as-a-Service for Causal Impact Analytics in the Multi-Tenant Public Cloud

Prof. (Dr.) Yuh-Jong Hu
hu@cs.nccu.edu.tw

Emerging Network Technology (ENT) Lab.
Department of Computer Science
National Chengchi University, Taipei, Taiwan

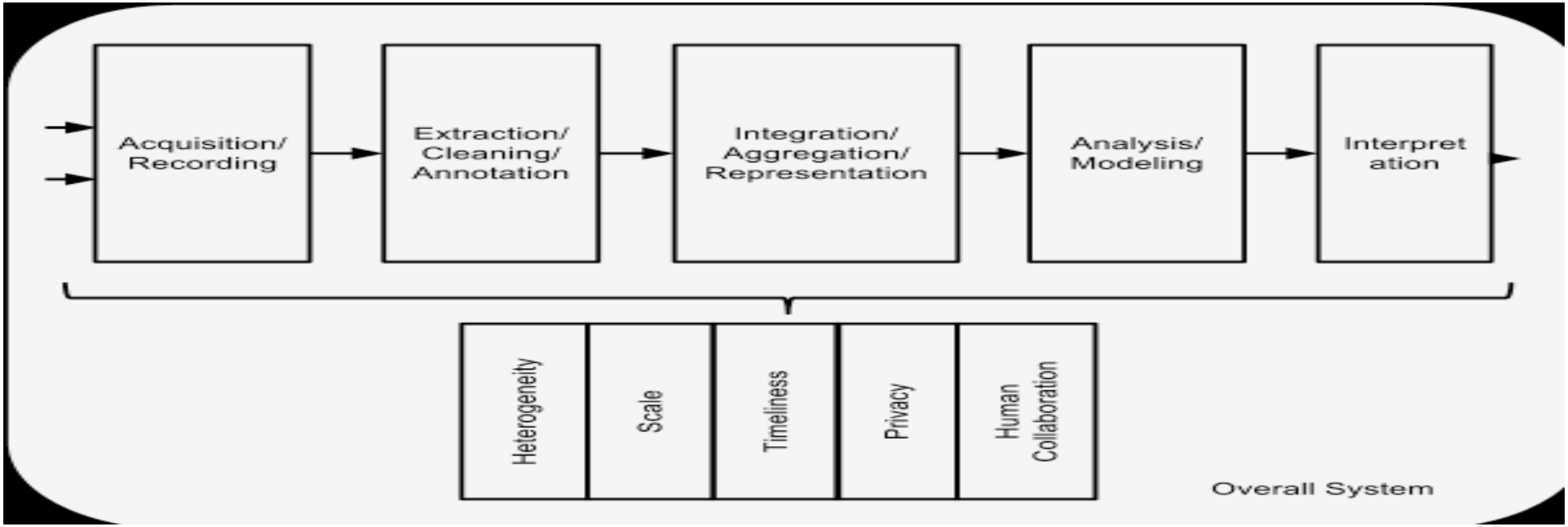
IEEE 2nd Int. Conf. TEL-NET-2017, Noida, India


Outline


Motivations

1. Big data analytics applies statistical inference to discover features' correlation, prediction, and through causal inference to discover cause-effect relationship.
2. Exploit the implications of high-dimensional statistical and causal inference on big data analytics.
3. Public cloud platform can provide multi-tenant software leased services for different stakeholder.
4. Outsource various software-as-a-service (SaaS) in the public cloud: security-as-a-service, machine learning-as-a-service, and data broker-as-a-service.


A Simple Big Data Analytics Pipeline



–Labrindis, A. and H. V. Jagadish, Challenges and Opportunities with Big Data, Proc. of the VLDB Endowment, 5(12), 2012.

Outline

Research Challenges

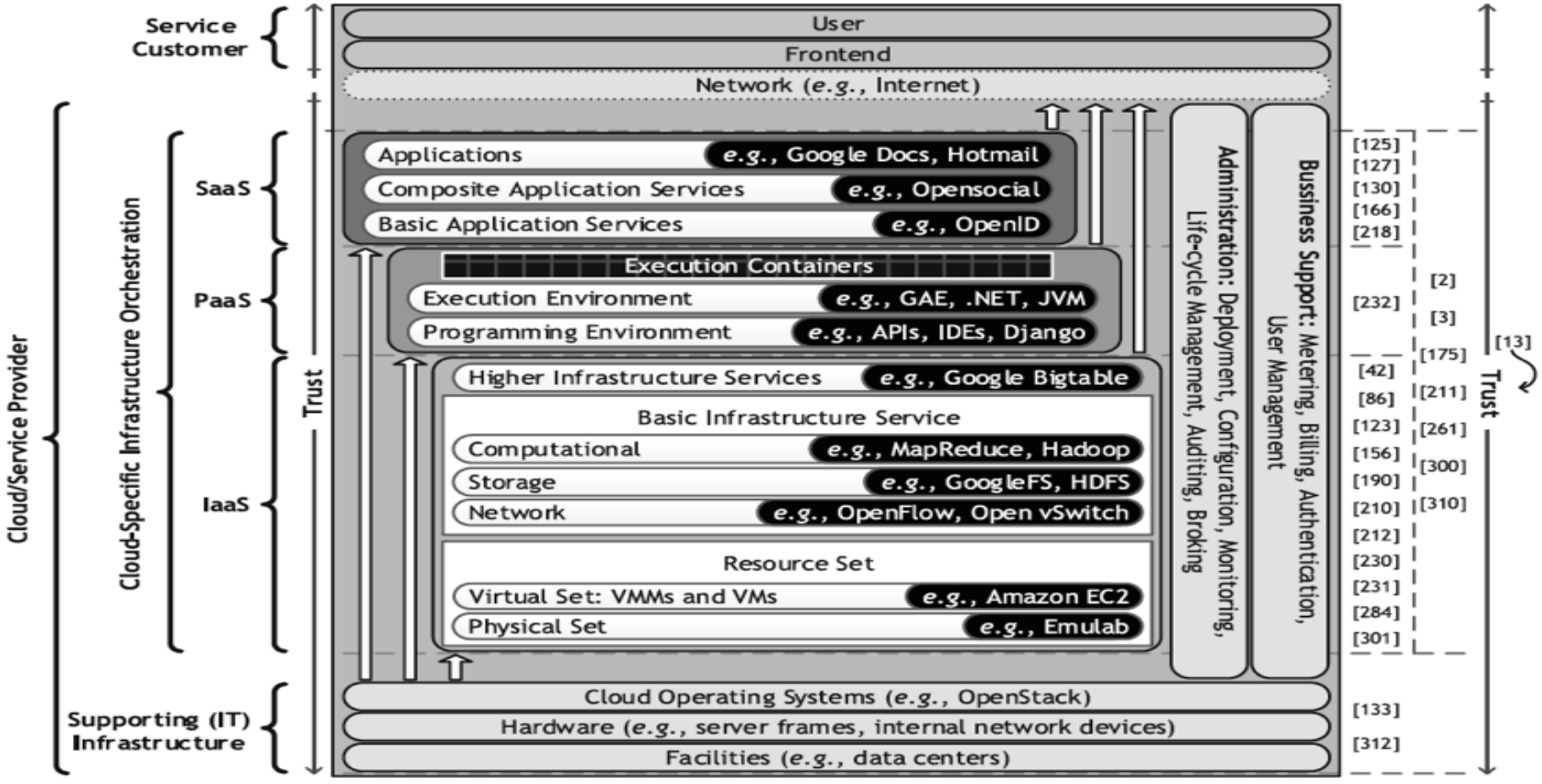
- We are aiming at the following research issues in the public multi-tenant cloud:
  1. Is it possible to outsource various big data analytics services?
  2. How to ensure security and privacy of outsourced big data analytics services?
  3. How to provide (automated) secured machine learning service to protect data privacy and discover features' correlation, prediction, and cause-effect from the datasets?
  4. How to balance data protection and data utility while enforcing secured machine learning?

Current Status

- We have achieved the following goals In the multi-tenant public cloud:
  1. Exploit up-to-date research status of secured machine learning.
  2. Propose secured causal inference concept for big data analytics.
  3. Preliminary results about automated machine learning on causal impact analytics for a policy evaluation.
  4. Propose and implement a scenario of automated causal inference.

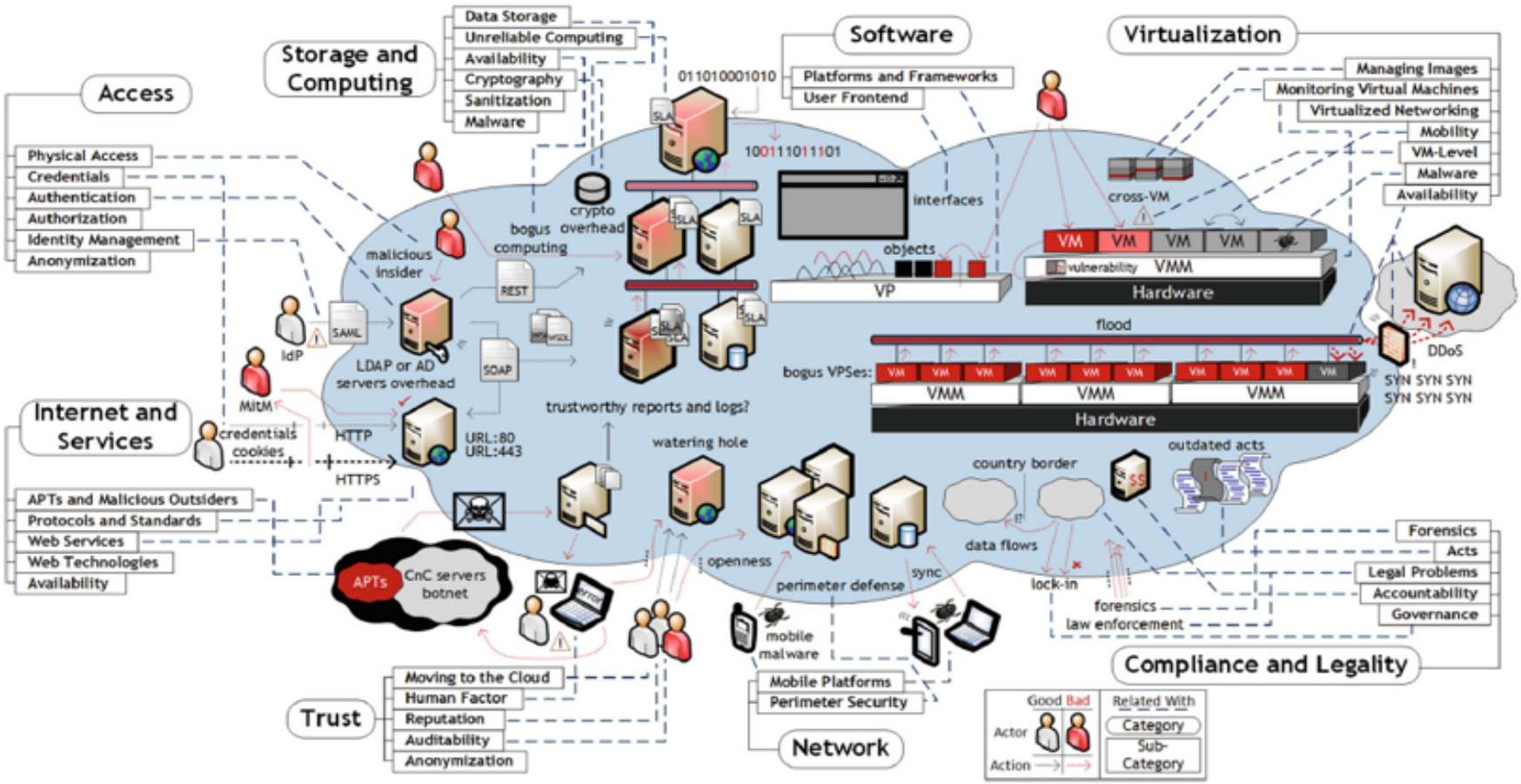

Outline

Cloud Service Delivery Models



–Fernades, A. B. D., et al., Security issues in cloud environments: a survey. International Journal Information Security, 13, pp. 113-170, Springer, 2014.

Security Issues in the Public Cloud



–Fernades, A. B. D., et al., Security issues in cloud environments: a survey. International Journal Information Security, 13, pp. 113-170, Springer, 2014.

Outline

What is Secured Machine Learning?

- A machine learning algorithm finds the regularity of data as a function (or model) through bottom-up data analytics services.
- The goal of machine learning is to search an optimal model through training, validation, and testing phases of data analytics services.
- Software agents can be created for cloud provider, data broker, data owner, modeler to initiate various SaaS services.
- Secured machine learning models and datasets can be achieved as:
  1. A data broker agent provides *DB-as-a-Service* and *Sec-as-a-Service* without disclosing personal private information from the datasets.
  2. A machine learning modeler's agent provides *ML-as-a-Service* without disclosing its propriety model and optimal hyper-parameters.

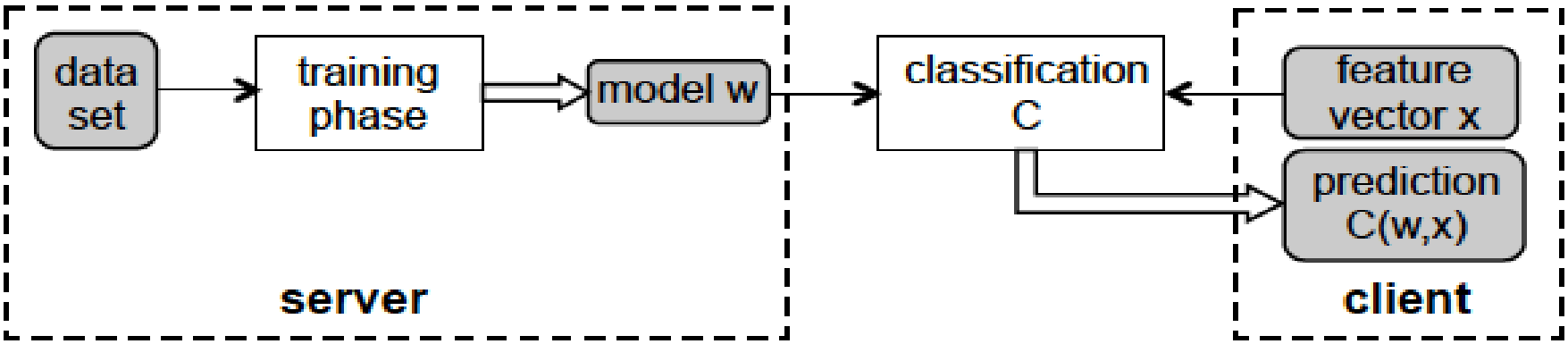

What is Secured Machine Learning?

- A machine learning algorithm finds the regularity of data as a function (or model) through bottom-up data analytics services.
- The goal of machine learning is to search an optimal model through training, validation, and testing phases of data analytics services.
- Software agents can be created for cloud provider, data broker, data owner, modeler to initiate various SaaS services.
- Secured machine learning models and datasets can be achieved as:
  1. A data broker agent provides *DB-as-a-Service* and *Sec-as-a-Service* without disclosing personal private information from the datasets.
  2. A machine learning modeler's agent provides *ML-as-a-Service* without disclosing its propriety model and optimal hyper-parameters.


What is Secured Machine Learning?

- A machine learning algorithm finds the regularity of data as a function (or model) through bottom-up data analytics services.
- The goal of machine learning is to search an optimal model through training, validation, and testing phases of data analytics services.
- Software agents can be created for cloud provider, data broker, data owner, modeler to initiate various SaaS services.
- Secured machine learning models and datasets can be achieved as:
  1. A data broker agent provides *DB-as-a-Service* and *Sec-as-a-Service* without disclosing personal private information from the datasets.
  2. A machine learning modeler's agent provides *ML-as-a-Service* without disclosing its propriety model and optimal hyper-parameters.


Secured Machine Learning



–Bost, R., Machine learning classification over encrypted data. NDSS'15, Feb. 2015.

Secured Machine Learning Algorithms

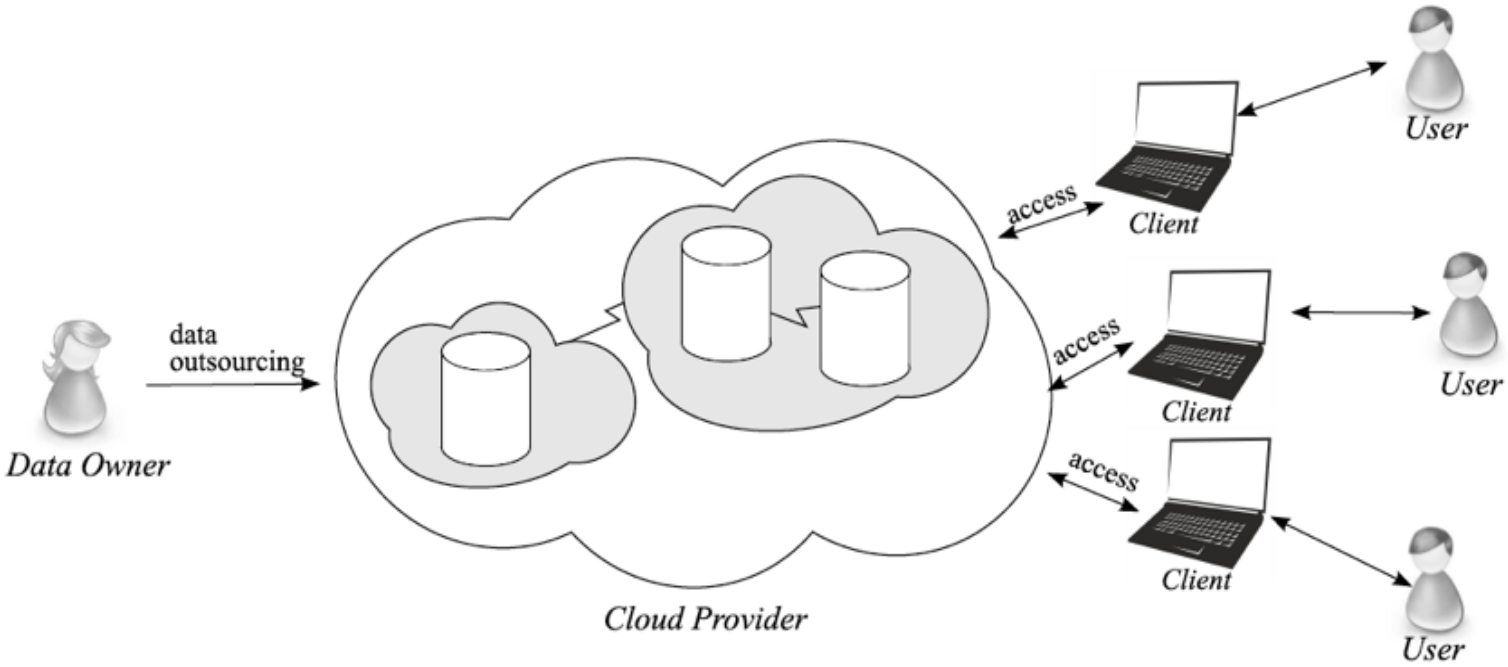
- Perceptron: hyperplane decision
- Least Squares: hyperplane decision
- SVM: hyperplane decision
- Naive Bayes vs. Logistic Regression
- Decision Trees, Ensemble Trees

–Bost, R., Machine learning classification over encrypted data. NDSS'15, Feb. 2015.


Core Technologies of Secured Machine Learning

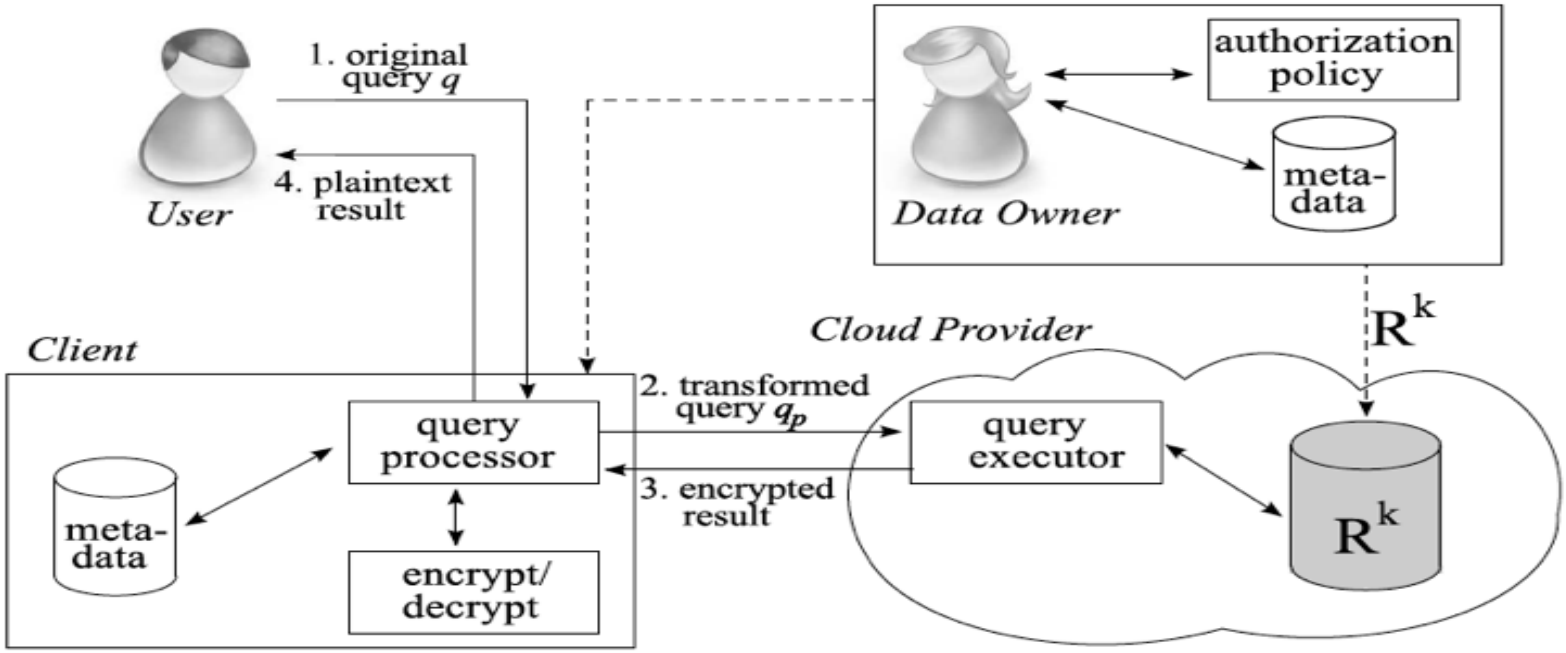
- Machine learning algorithms with various arithmetic operations - add, multiplication, inner dot products, comparison, branching, and argMax.
- Simply anonymizing data to de-identify Personally Identifiable Information (PII) is unsafe - so sanitizing data through Differential Privacy (DP) is an emerging research trend.
- Fully Homomorphic Encryption (FHE) - getting practical through Somewhat Homomorphic Encryption (SwHE) development.
- CryptDB - protecting confidentiality with encrypted query processing over encrypted data without disclosing query expression and sensitivity data.


Protecting and Managing Data in the Cloud



–Vimercati, di S. De C., et al., Practical techniques building on encryption for protecting and managing data in the cloud. The New Codebrakers, Vol. 9100, LNCS, Springer, 2015.

Protecting and Managing Data in the Cloud (Conti.)



–Vimercati, di S. De C., et al., Practical techniques building on encryption for protecting and managing data in the cloud. The New Codebrakers, Vol. 9100, LNCS, Springer, 2015.

Outsourcing Intelligent Intrusion Detection System (IDS)

A Scenario of Outsourcing IDS with Data Protection

An TwnECC ECommerce (EC) Inc. intrusion detection system (IDS) is outsourced to another SecInt Corp. for security data analytics. On one hand, TwnECC's user surfing website log records are encrypted or added noisy before forwarding to the SecInt Corp.'s IDS for analytics to ensure log records' security and users' privacy.

A Scenario of Outsourcing IDS with Model Protection

On the other hand, the SecInt's machine learning optimal model with its hyperparameter for IDS analytics services are also protected to avoid disclosure its propriety machine learning model. How do you provide secured machine learning through the outsourcing intelligent IDS to balance between data utility and model/data protection in the multi-tenant public cloud?

Outsourcing Intelligent Intrusion Detection System (IDS)

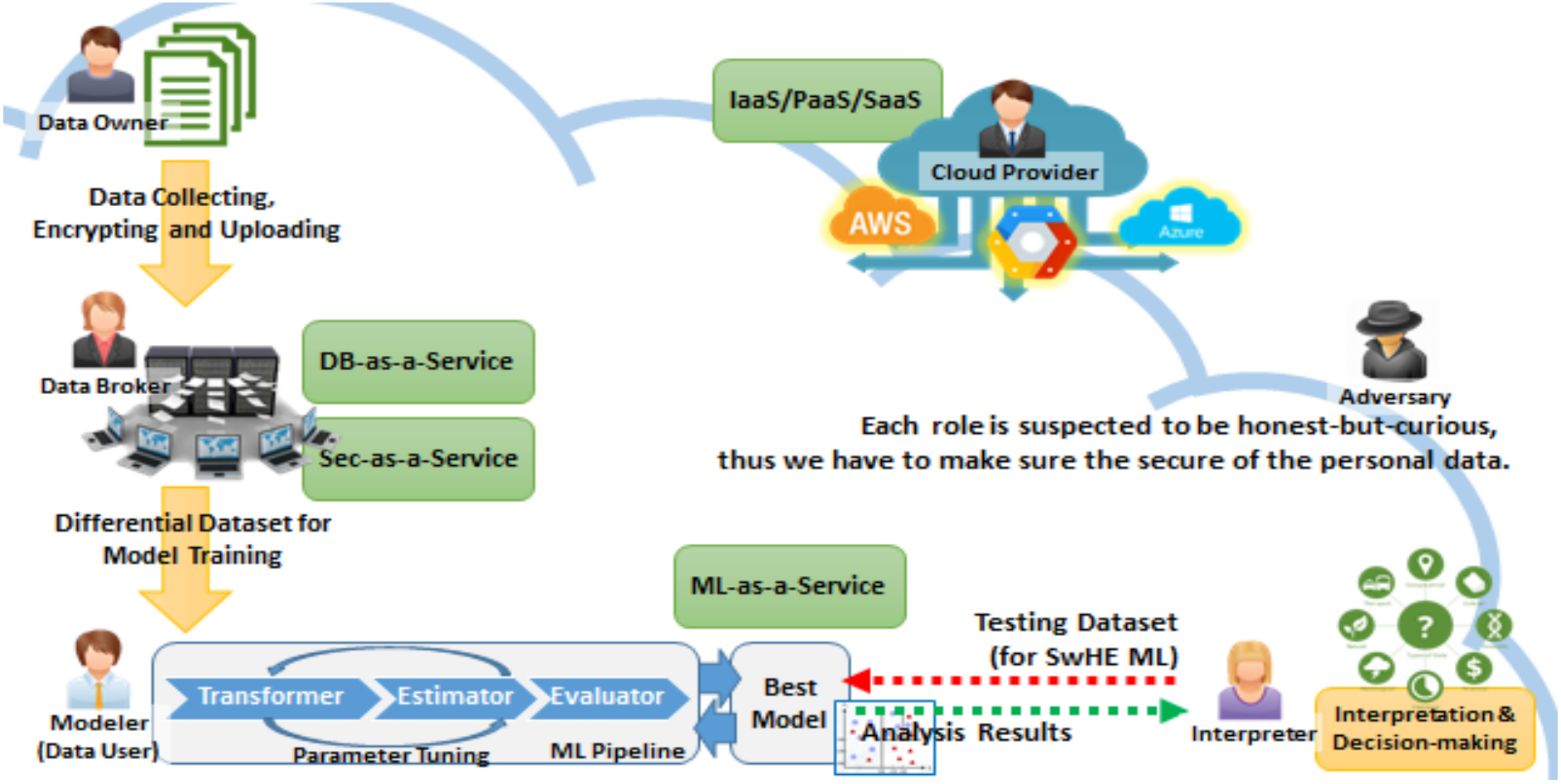
A Scenario of Outsourcing IDS with Data Protection

An TwnECC ECommerce (EC) Inc. intrusion detection system (IDS) is outsourced to another SecInt Corp. for security data analytics. On one hand, TwnECC's user surfing website log records are encrypted or added noisy before forwarding to the SecInt Corp.'s IDS for analytics to ensure log records' security and users' privacy.

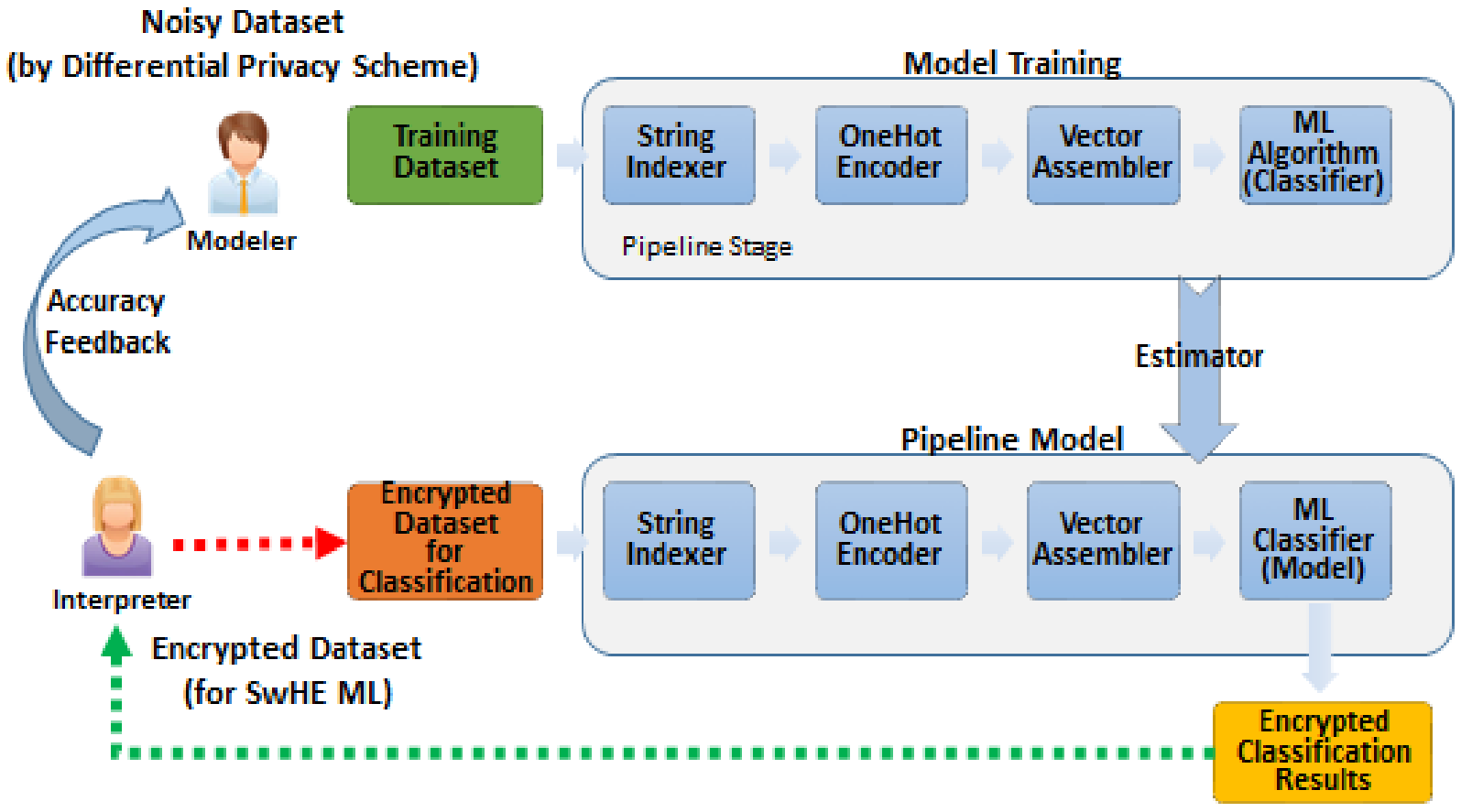
A Scenario of Outsourcing IDS with Model Protection

On the other hand, the SecInt's machine learning optimal model with its hyperparameter for IDS analytics services are also protected to avoid disclosure its propriety machine learning model. How do you provide secured machine learning through the outsourcing intelligent IDS to balance between data utility and model/data protection in the multi-tenant public cloud?

Outsourcing Secured Machine Learning



Outsourcing Secured Machine Learning (Conti.)



Fully Homomorphic Encryption (FHE)

Public Key Encryption for Fully Homomorphic Encryption

- Encryption maps a message $m \in M$ from message space to an element of cypher text space $c \in C$, public key k_p and secret key k_s with $Enc(k_p, \cdot)$ and $Dec(k_s, \cdot)$ which encrypt and decrypt messages respectively.
- If encryption and decryption are functions, then $m = Dec(k_s, Enc(k_p, m)), \forall m \in M$.
- A scheme is said to be homomorphic for some operators $\circ \in S_M$ acting in message space, there are corresponding operators $\diamond \in S_C$ acting in the cipher text space satisfying the property:
$$Dec(k_s, Enc(k_p, m_1) \diamond Enc(k_p, m_2)) = M_1 \circ M_2, \forall m_1, m_2 \in M$$
- A scheme is *fully* homomorphic if it is homomorphic for both addition and manipulation operators.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.

Applying FHE for Machine Learning Algorithms

- Some limitations of FHE:
  1. Machine learning algorithms are beyond the computational reach of existing homomorphic schemes.
  2. Substantial inflation in the size of encrypted messages.
  3. No homomorphic schemes capable of natively supporting divisions, only addition and multiplication.
  4. The noise tends to accumulate when there is randomness injected into the cipher text in the encryption schemes,
- One mitigating proposal is to initially encrypt message using a non-homomorphic, size efficient encryption algorithms, such as AES, and to encrypt the AES decryption key with a homomorphic scheme.
- *Bootstrap* a cipher text is an operation which resets the noise to that of a freshly encrypted message.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.

Applying FHE for Machine Learning Algorithms

- Some limitations of FHE:
  1. Machine learning algorithms are beyond the computational reach of existing homomorphic schemes.
  2. Substantial inflation in the size of encrypted messages.
  3. No homomorphic schemes capable of natively supporting divisions, only addition and multiplication.
  4. The noise tends to accumulate when there is randomness injected into the cipher text in the encryption schemes,
- One mitigating proposal is to initially encrypt message using a non-homomorphic, size efficient encryption algorithms, such as AES, and to encrypt the AES decryption key with a homomorphic scheme.
- *Bootstrap* a cipher text is an operation which resets the noise to that of a freshly encrypted message.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.

Applying FHE for Machine Learning Algorithms

- Some limitations of FHE:
  1. Machine learning algorithms are beyond the computational reach of existing homomorphic schemes.
  2. Substantial inflation in the size of encrypted messages.
  3. No homomorphic schemes capable of natively supporting divisions, only addition and multiplication.
  4. The noise tends to accumulate when there is randomness injected into the cipher text in the encryption schemes,
- One mitigating proposal is to initially encrypt message using a non-homomorphic, size efficient encryption algorithms, such as AES, and to encrypt the AES decryption key with a homomorphic scheme.
- *Bootstrap* a cipher text is an operation which resets the noise to that of a freshly encrypted message.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.

Applying FHE for Machine Learning Algorithms (Conti.)

- Two approaches work with the constraints imposed by homomorphic encryption for machine learning algorithms are:
  1. Use existing algorithms amendable to homomorphic encryption so that all stages of model fitting and prediction can be computed encrypted.
  2. Invoking secured two-(multi-)party protocol so that each participant can preserve its secrecy but still can cooperate on the machine learning algorithm computation.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.


Applying FHE for Machine Learning Algorithms (Conti.)

- Two approaches work with the constraints imposed by homomorphic encryption for machine learning algorithms are:
  1. Use existing algorithms amendable to homomorphic encryption so that all stages of model fitting and prediction can be computed encrypted.
  2. Invoking secured two-(multi-)party protocol so that each participant can preserve its secrecy but still can cooperate on the machine learning algorithm computation.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.


Applying FHE for Machine Learning Algorithms (Conti.)

- Two approaches work with the constraints imposed by homomorphic encryption for machine learning algorithms are:
  1. Use existing algorithms amendable to homomorphic encryption so that all stages of model fitting and prediction can be computed encrypted.
  2. Invoking secured two-(multi-)party protocol so that each participant can preserve its secrecy but still can cooperate on the machine learning algorithm computation.

–Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.

What do you mean Differential Privacy (DP)?

Differential Privacy (DP)

For all neighboring datasets D_1, D_2 that differ in one person's value for any set S, if $M = (\epsilon, \delta)$-differentially private randomized algorithm, then:

$$Pr(M(D_1) \in S) \leq e^{\epsilon} Pr(M(D_2) \in S) + \delta$$

Pure differential privacy: $\delta = 0$

$$Pr(M(D_1) \in S) \leq (1 + \epsilon) Pr(M(D_2) \in S)$$

- The outcome of any analysis is essentially equally likely, independent of whether any individual joins, or refrains from joining, the dataset.
- Privacy risk (or loss) is measured by a parameter (ϵ, δ).
- Small (ϵ, δ) guarantee more privacy, however, choosing ϵ for privacy budget (or loss) is a challenge.

–Dwork, C., A firm foundation for private data analysis. CACM, 54(1), 2011.


What do you mean Differential Privacy (DP)?

Differential Privacy (DP)

For all neighboring datasets D_1, D_2 that differ in one person's value for any set S, if $M = (\epsilon, \delta)$-differentially private randomized algorithm, then:

$$Pr(M(D_1) \in S) \leq e^{\epsilon} Pr(M(D_2) \in S) + \delta$$

Pure differential privacy: $\delta = 0$

$$Pr(M(D_1) \in S) \leq (1 + \epsilon) Pr(M(D_2) \in S)$$

- The outcome of any analysis is essentially equally likely, independent of whether any individual joins, or refrains from joining, the dataset.
- Privacy risk (or loss) is measured by a parameter (ϵ, δ).
- Small (ϵ, δ) guarantee more privacy, however, choosing ϵ for privacy budget (or loss) is a challenge.

–Dwork, C., A firm foundation for private data analysis. CACM, 54(1), 2011.

What do you mean Differential Privacy (DP)? (Conti.)

Global Sensitivity of DP

The global sensitivity, Δ_M describing how much a randomized machine learning algorithm M changes when D changes,

$$\Delta_M := Max_{D,D' \subseteq \chi s.t. d_H(D,D') \leq 1} |M(D) - M(D'|$$

Output $M(D) + \omega$, where ω is a noise variable drawn from $Lap(0, \Delta_{M/\epsilon})$

Local Sensitivity of DP

The local sensitivity, Δ_M describing how much a randomized machine learning algorithm M changes when a specific D changes,

$$\Delta(D)_M := Max_{D' \subseteq \chi s.t. d_H(D,D') \leq 1} |M(D) - M(D'|)$$

Output $M(D) + \omega$, where ω is a noise variable drawn from $Lap(0, \Delta_{M/\epsilon})$

–Dwork, C., A firm foundation for private data analysis. CACM, 54(1), 2011.

What do you mean Differential Privacy (DP)? (Conti.)

Global Sensitivity of DP

The global sensitivity, Δ_M describing how much a randomized machine learning algorithm M changes when D changes,

$$\Delta_M := Max_{D,D' \subseteq \chi s.t. d_H(D,D') \leq 1} |M(D) - M(D'|$$

Output $M(D) + \omega$, where ω is a noise variable drawn from $Lap(0, \Delta_{M/\epsilon})$

Local Sensitivity of DP

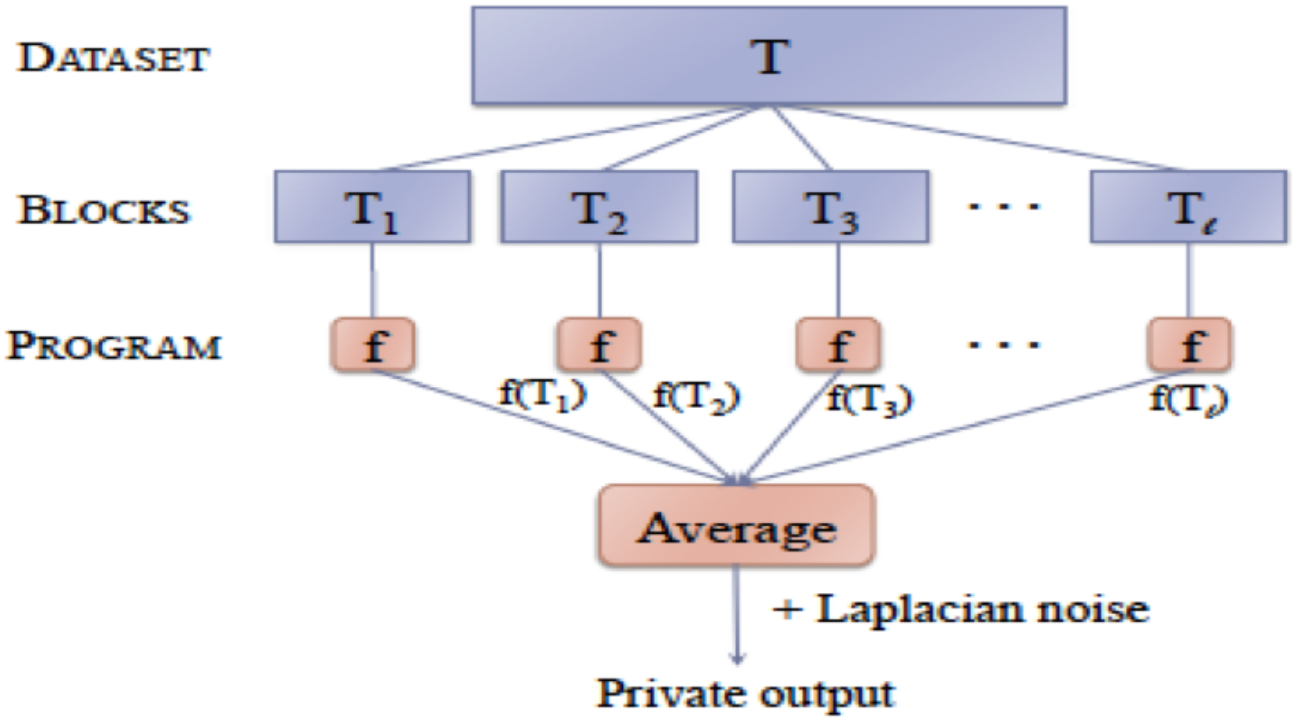
The local sensitivity, Δ_M describing how much a randomized machine learning algorithm M changes when a specific D changes,

$$\Delta(D)_M := Max_{D' \subseteq \chi s.t. d_H(D,D') \leq 1} |M(D) - M(D'|)$$

Output $M(D) + \omega$, where ω is a noise variable drawn from $Lap(0, \Delta_{M/\epsilon})$

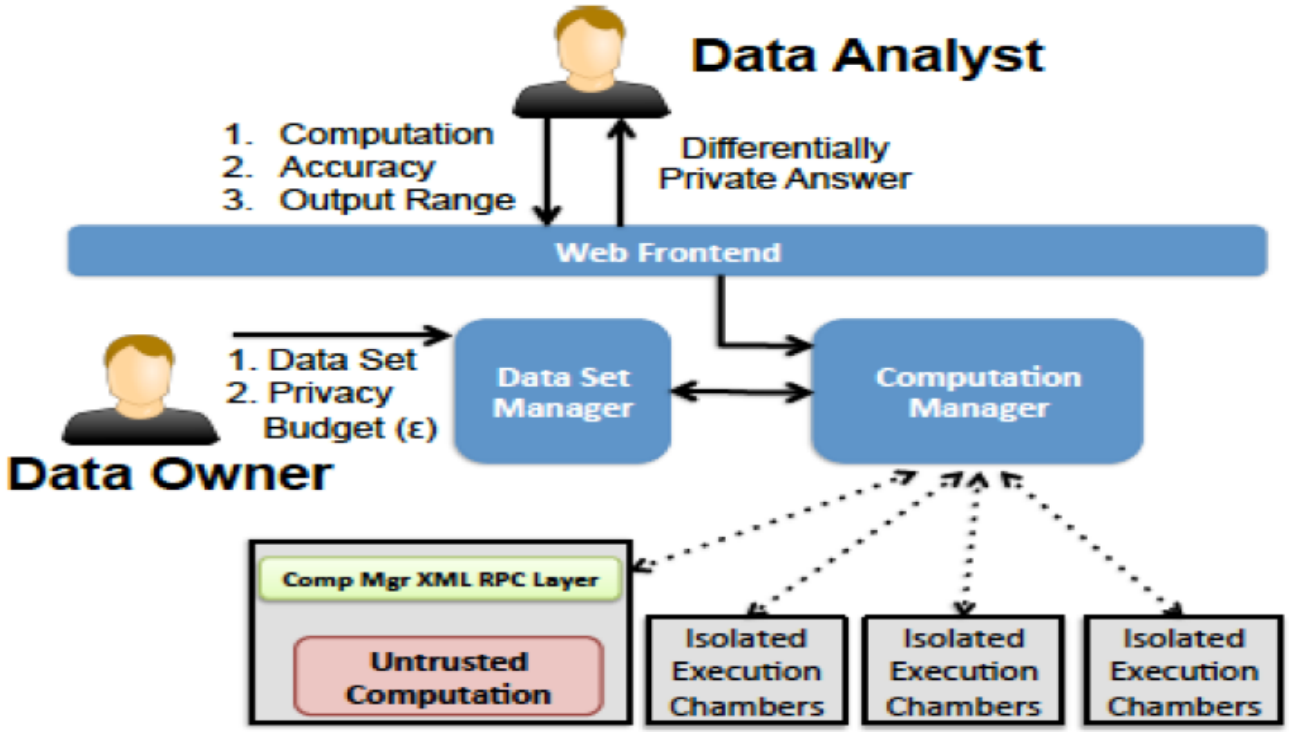
–Dwork, C., A firm foundation for private data analysis. CACM, 54(1), 2011.

Privacy Preserving Data Analysis System: GUPT



–Mohan, P., et al., GUPT: privacy preserving data analysis made easy. SIMOD'12, 2012.

Privacy Preserving Data Analysis System: GUPT (Conti.)



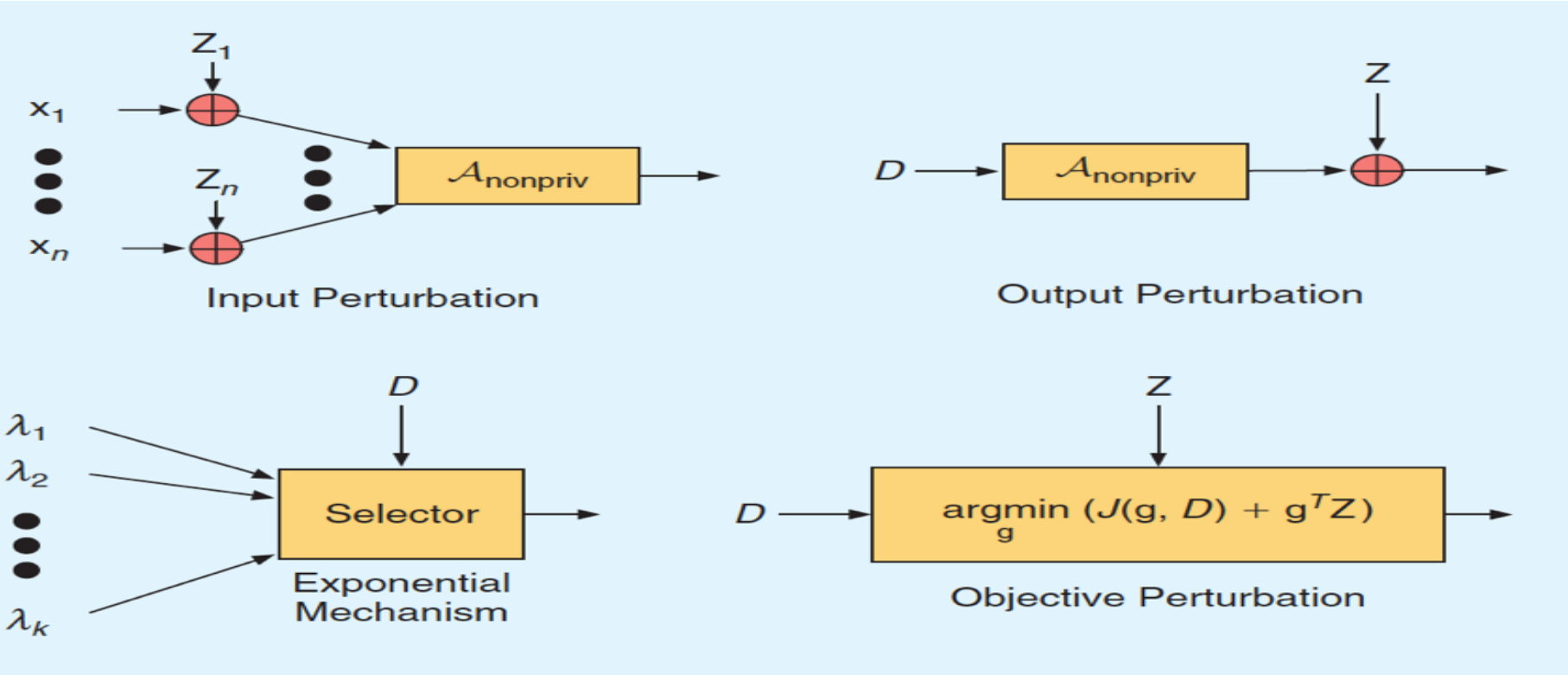
–Mohan, P., et al., GUPT: privacy preserving data analysis made easy. SIMOD'12, 2012.

Applying DP for Machine Learning Algorithms

- Input perturbation:
  Add noise to the input before running a machine learning algorithm.
- Output perturbation:
  Run a machine learning algorithm, then add noise to the output.
- Objective function perturbation:
  Randomize the internals of machine learning algorithm.

–Sarwate, D. A. and K. Chaudhuri, Signal processing and machine learning with differential privacy: algorithms and challenges for continuous data. IEEE Signal Processing Magazine. Sep. 2013.


Applying DP for Machine Learning Algorithms (Conti.)



–Sarwate, D. A. and K. Chaudhuri, Signal processing and machine learning with differential privacy: algorithms and challenges for continuous data. IEEE Signal Processing Magazine. Sep. 2013.

Outline

Causal Inference

- Machine learning algorithms are used for classification to discover correlation and prediction, but not for causal inference.
- Why bother for using causal inference when we already have conventional statistical inference?
- Because simple statistical inference cannot be directly applied for causal impact analytics.
- Three approaches for structural causal model (SCM) construction:
  1. Potential Outcomes
  2. Structured Equation Model
  3. Causal Graph

Causal Inference (Conti.)

- Given a causal structure is *known*, a total effect can be computed by using several techniques, such as covariate adjustment, inverse probability weighting, or instrumental variables.
- However, when causal structure is *unknown* we should apply causal structure learning methods, such as constrain-based, score-based, and additive noise model (ANM) methods;
- There is the R-package `pcalg` that can perform the causal structure discovery methods.
- Apply manipulation (or intervention) to endogenous features to see what has a total effect on the target feature.


Private Causal Inference

Private Causal Inference

Two random variables X, Y with joint density $p(x, y)$ are said to 'satisfy an Additive Noise Model (ANM)' $X \to Y$, if there exists a nonlinear function $f : R \to R$ and a random noise variable N_Y, independent from X, i.e., $X \perp\!\!\!\perp N_Y$, such that:

$$Y = f(X) + N_Y$$

$P_{X,Y}$ for this ANM $X \to Y$ to be different from the one induced by the ANM $Y \to X$.

- ANM is proposed for inferring whether $X \to Y$ or $Y \to X$ for two variables, where X and Y do not simultaneously cause each other.
- Given a (non-)linear function of the cause plus independent noise.
- Data can be discrete or continuous.

–Kusner, J. M., et al., Private Causal Inference. arXiv:1512.05469v2, 20 Aug 2016.

Private Causal Inference

Private Causal Inference

Two random variables X, Y with joint density $p(x, y)$ are said to 'satisfy an Additive Noise Model (ANM)' $X \to Y$, if there exists a nonlinear function $f : R \to R$ and a random noise variable N_Y, independent from X, i.e., $X \perp\!\!\!\perp N_Y$, such that:

$$Y = f(X) + N_Y$$

$P_{X,Y}$ for this ANM $X \to Y$ to be different from the one induced by the ANM $Y \to X$.

- ANM is proposed for inferring whether $X \to Y$ or $Y \to X$ for two variables, where X and Y do not simultaneously cause each other.
- Given a (non-)linear function of the cause plus independent noise.
- Data can be discrete or continuous.

–Kusner, J. M., et al., Private Causal Inference. arXiv:1512.05469v2, 20 Aug 2016.


Private Causal Inference (Conti.)

Algorithm 1 ANM Causal Inference [23]

1: **Input:** train/test data $\{x_i, y_i\}_{i=1}^{n}$, $\{x'_i, y'_i\}_{i=1}^{m}$
2: Regress on training data, to yield $\hat{f}, \hat{g}$, such that:
3: $\hat{f}(x_i) \approx y_i, \quad \hat{g}(y_i) \approx x_i, \quad \forall i$
4: Compute residuals on test data:
5: $\mathbf{r}'_Y := \mathbf{y}' - \hat{f}(\mathbf{x}'), \quad \mathbf{r}'_X := \mathbf{x}' - \hat{g}(\mathbf{y}')$
6: Calculate dependence scores:
7: $s_{X \to Y} := s(\mathbf{x}', \mathbf{r}'_Y), \quad s_{Y \to X} := s(\mathbf{y}', \mathbf{r}'_X)$
8: **Return:** $s_{X \to Y}$, $s_{X \to Y}$, and D, where
9: $$\mathrm{D} = \begin{cases} X \to Y & \text{if } s_{X \to Y} < s_{Y \to X} \\ Y \to X & \text{if } s_{X \to Y} > s_{Y \to X} \end{cases}$$

–Kusner, J. M., et al., Private Causal Inference. arXiv:1512.05469v2, 20 Aug 2016.


Challenges of Applying DP for Secured Causal Inference

- Existing secured machine learning techniques might not be useful to deal with secured causal inference. Why?
- Only observational datasets might not rich enough for a causal model search and inference. Moreover, additional research challenges are:
  1. How to apply different privacy techniques of secured machine learning for causal graph with multiple random variables?
  2. How to preserve personal privacy in the causal structure discovery and derive the intended cause-effect with observational datasets only?
  3. How to decide which features are added noise to protect sensitivity data and also ensure model secrecy in the causal inference processing?

Challenges of Applying DP for Secured Causal Inference

- Existing secured machine learning techniques might not be useful to deal with secured causal inference. Why?
- Only observational datasets might not rich enough for a causal model search and inference. Moreover, additional research challenges are:
  1. How to apply different privacy techniques of secured machine learning for causal graph with multiple random variables?
  2. How to preserve personal privacy in the causal structure discovery and derive the intended cause-effect with observational datasets only?
  3. How to decide which features are added noise to protect sensitivity data and also ensure model secrecy in the causal inference processing?


Challenges of Applying DP for Secured Causal Inference

- Existing secured machine learning techniques might not be useful to deal with secured causal inference. Why?
- Only observational datasets might not rich enough for a causal model search and inference. Moreover, additional research challenges are:
  1. How to apply different privacy techniques of secured machine learning for causal graph with multiple random variables?
  2. How to preserve personal privacy in the causal structure discovery and derive the intended cause-effect with observational datasets only?
  3. How to decide which features are added noise to protect sensitivity data and also ensure model secrecy in the causal inference processing?

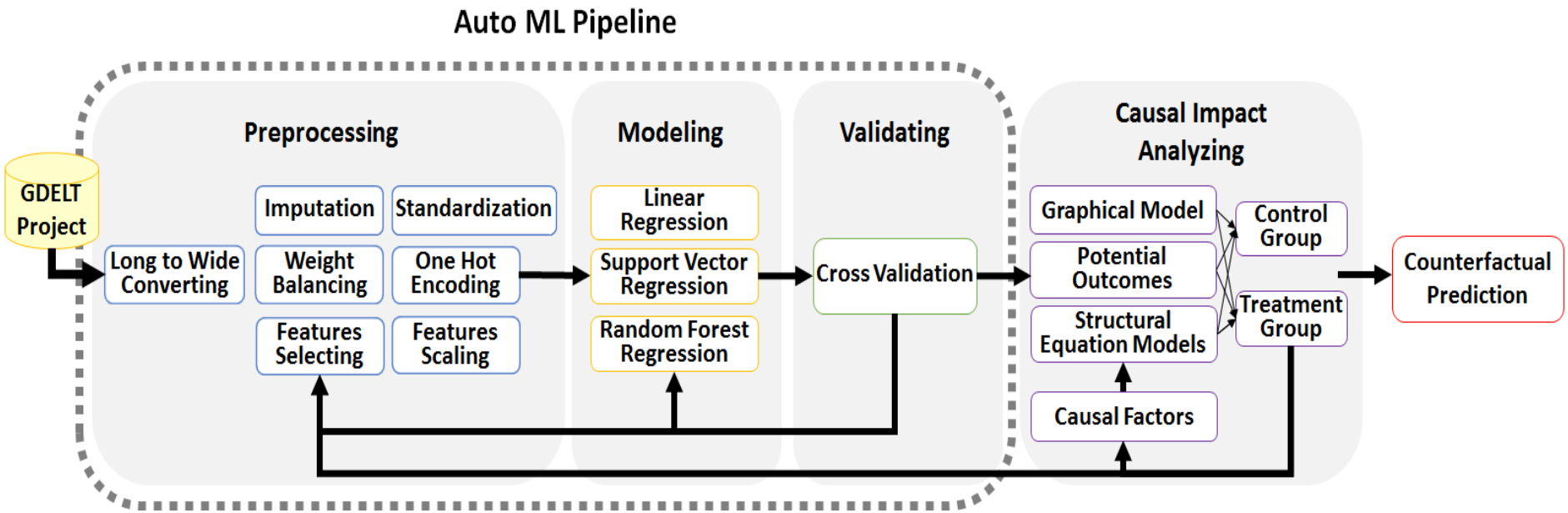
Challenges of Applying DP for Secured Causal Inference

- Existing secured machine learning techniques might not be useful to deal with secured causal inference. Why?
- Only observational datasets might not rich enough for a causal model search and inference. Moreover, additional research challenges are:
  1. How to apply different privacy techniques of secured machine learning for causal graph with multiple random variables?
  2. How to preserve personal privacy in the causal structure discovery and derive the intended cause-effect with observational datasets only?
  3. How to decide which features are added noise to protect sensitivity data and also ensure model secrecy in the causal inference processing?

Challenges of Applying DP for Secured Causal Inference

- Existing secured machine learning techniques might not be useful to deal with secured causal inference. Why?
- Only observational datasets might not rich enough for a causal model search and inference. Moreover, additional research challenges are:
  1. How to apply different privacy techniques of secured machine learning for causal graph with multiple random variables?
  2. How to preserve personal privacy in the causal structure discovery and derive the intended cause-effect with observational datasets only?
  3. How to decide which features are added noise to protect sensitivity data and also ensure model secrecy in the causal inference processing?


AutoML Pipeline for Causal-Impact Analytics



–Challenges of Automated Machine Learning on Causal Impact Analytics for Policy Evaluation, TEL-NET-2017

Outline

Summary

- We are dealing with the research issues of outsourcing secure machine learning for causal impact analytics in the multi-tenant public cloud:
  1. First we ensure the data protection principles, then provide a secured machine learning model while outsourcing data management and analytics modeling SaaS services.
  2. Fully (or somewhat) homomorphic encryption and differential privacy are two parallel techniques for secured machine learning on sensitive datasets and private model.
  3. We intend to merge differential privacy for data/model protection in the *offline* machine learning phase, and homomorphic encryption for data/model protection in the *online* machine learning phase.
  4. Balancing data protection and data utility is plausible on outsourcing various SaaS services for data analytics, protection and modeling.

Summary

- We are dealing with the research issues of outsourcing secure machine learning for causal impact analytics in the multi-tenant public cloud:
  1. First we ensure the data protection principles, then provide a secured machine learning model while outsourcing data management and analytics modeling SaaS services.
  2. Fully (or somewhat) homomorphic encryption and differential privacy are two parallel techniques for secured machine learning on sensitive datasets and private model.
  3. We intend to merge differential privacy for data/model protection in the *offline* machine learning phase, and homomorphic encryption for data/model protection in the *online* machine learning phase.
  4. Balancing data protection and data utility is plausible on outsourcing various SaaS services for data analytics, protection and modeling.

Summary

- We are dealing with the research issues of outsourcing secure machine learning for causal impact analytics in the multi-tenant public cloud:
  1. First we ensure the data protection principles, then provide a secured machine learning model while outsourcing data management and analytics modeling SaaS services.
  2. Fully (or somewhat) homomorphic encryption and differential privacy are two parallel techniques for secured machine learning on sensitive datasets and private model.
  3. We intend to merge differential privacy for data/model protection in the *offline* machine learning phase, and homomorphic encryption for data/model protection in the *online* machine learning phase.
  4. Balancing data protection and data utility is plausible on outsourcing various SaaS services for data analytics, protection and modeling.

Summary

- We are dealing with the research issues of outsourcing secure machine learning for causal impact analytics in the multi-tenant public cloud:
  1. First we ensure the data protection principles, then provide a secured machine learning model while outsourcing data management and analytics modeling SaaS services.
  2. Fully (or somewhat) homomorphic encryption and differential privacy are two parallel techniques for secured machine learning on sensitive datasets and private model.
  3. We intend to merge differential privacy for data/model protection in the *offline* machine learning phase, and homomorphic encryption for data/model protection in the *online* machine learning phase.
  4. Balancing data protection and data utility is plausible on outsourcing various SaaS services for data analytics, protection and modeling.

Summary

- We are dealing with the research issues of outsourcing secure machine learning for causal impact analytics in the multi-tenant public cloud:
  1. First we ensure the data protection principles, then provide a secured machine learning model while outsourcing data management and analytics modeling SaaS services.
  2. Fully (or somewhat) homomorphic encryption and differential privacy are two parallel techniques for secured machine learning on sensitive datasets and private model.
  3. We intend to merge differential privacy for data/model protection in the *offline* machine learning phase, and homomorphic encryption for data/model protection in the *online* machine learning phase.
  4. Balancing data protection and data utility is plausible on outsourcing various SaaS services for data analytics, protection and modeling.

Future Work

- Possible future work are:
  1. Private causal impact analytics in the multi-tenant public cloud is an important research field to be further studied.
  2. A lot of work is needed for future research and development.
  3. We call for International Academic Research Cooperation.


Future Work

- Possible future work are:
  1. Private causal impact analytics in the multi-tenant public cloud is an important research field to be further studied.
  2. A lot of work is needed for future research and development.
  3. We call for International Academic Research Cooperation.


Future Work

- Possible future work are:
  1. Private causal impact analytics in the multi-tenant public cloud is an important research field to be further studied.
  2. A lot of work is needed for future research and development.
  3. We call for International Academic Research Cooperation.


Future Work

- Possible future work are:
  1. Private causal impact analytics in the multi-tenant public cloud is an important research field to be further studied.
  2. A lot of work is needed for future research and development.
  3. We call for International Academic Research Cooperation.


Outline


References I

1. Aslett, J. M., L., et al., Encrypted statistical machine learning: new privacy reserving methods. arXiv preprint arXiv:1508.06845v1, 27 Aug 2015.
2. Aslett, J. M., L., et al., A review of homomorphic encryption and software tools for encrypted statistical machine learning. arXiv preprint arXiv:1508.06574v1, 26 Aug 2015.
3. Ateniese, G., et al., Hacking smart machines with smarter ones: how to extract meaningful data from machine learning classifiers. Int. Journal of Security and Networks, 10(3), 2015.
4. Barreno, M., et al., The security of machine learning. Machine Learning, 81(2), Springer, pp. 121-148, 2010.
5. Ben-David, A., et al., FairlayMP : a system for secure multi-party computation. CCS?08, 2008.
6. Bos, W. J., et al., Private predictive analysis on encrypted medical data. Microsoft Tech Report, 200652, 2013.
7. Bost, R., Machine learning classification over encrypted data. NDSS'15, Feb. 2015.
8. Chaudhuri, K., et al., Differentially private empirical risk minimization. Journal of Machine Learning Research. 12, 2011.
9. Ciriani, V., et al., Microdata protection. Secure Data Management in Decentralized Systems, Springer, pp. 291-322, 2007.


References II

10. Dwork, C., A firm foundation for private data analysis. CACM, 54(1), 2011.
11. Fan, J. and F. Vercauteren, Somewhat practically fully homomorphic encryption. ICAR Cryptology ePrint archive, 2012.
12. Fernades, A. B. D., et al., Security issues in cloud environments: a survey. International Journal Information Security,13, pp. 113-170, Springer, 2014.
13. Fox, A. and D. Patterson, Engineering Software as a Service: An Agile Approach Using Cloud Computing. Strawberry Canyon, 2014.
14. Gentry, G., Fully homomorphic encryption using ideal lattices. STOC'09, 2009.
15. Gentry, C., Computing arbitrary functions on encrypted data. CACM, 55(3), 2010.
16. Gentry, C., et al., Homomorphic evaluation of the AES circuit (updated implementation). IACR Cryptography ePrint Archive, 2015.
17. Graepel, T., et al., ML confidential: machine learning on encrypted data. Information Security and Cryptology ? ICISC, LNCS, Springer, 2012.
18. Henecka, W., et al., TASTY: Tool for Automating Secure Two-partY computations. CCS?10, pp. 451-462, 2010.
19. Ji, Z., et al., Differential privacy and machine learning: a survey and review. ArXiv preprint arXiv:1412.7584v1, 24 Dec. 2014.


References III

20. Kusner, J. M., et al., Private Causal Inference. arXiv:1512.05469v2, 20 Aug 2016.
21. Labrindis, A. and H. V. Jagadish, Challenges and opportunities with big data. Proc. of the VLDB Endowment, 5(12), 2012.
22. Laskov, P. and R. Lippmann, Machine learning in adversarial environments. Machine Learning, 81(2), Springer, pp. 115-119, 2010.
23. Lauter, K., et al., Can homomorphic encryption be practical? Proc. of the 3rd ACM workshop on cloud computing security, ACM, pp113-124, 2011.
24. Mohan, P., et al., GUPT: privacy preserving data analysis made easy. SIMOD'12, 2012.
25. Papernot, N., et al., SoK: towards the science of security and privacy in machine learning. arXiv preprint arXiv:1611.03814v1, 11 Nov. 2016.
26. Pawar, S. P., et al., Security-as-a-Service in multi-cloud and federated cloud environments. IFIPTM 2015, IFIP AICT 454, pp. 251-261, 2015.
27. Pearce, M., S. Zeadally, and R. Hunt. Virtualization: issues, security threats, and solutions. ACM Computing Surveys, 45(2), Feb. 2013.
28. Popa, Ad. R., et al., CryptDB: protecting confidentiality with encrypted query processing. SOSP'11, ACM, 2011.


References IV

29. Rivest, R. L. et al., On data banks and privacy homomorphisms. Foundations of Secure Computing, 4(11), pp. 169-180, 1978.
30. Samarati, P. and S. De C. di Vimercati, Data protection in outsourcing scenarios: issues and directions. ASIACCS'10, 2010.
31. Samarati, P. and S. De C. di Viemrcati, Cloud security: issues and concerns. Encyclopedia on Cloud Computing, Wiley, 2016.
32. Sarwate, D. A. and K. Chaudhuri, Signal processing and machine learning with differential privacy: algorithms and challenges for continuous data. IEEE Signal Processing Magazine. Sep. 2013.
33. Sen, J., Homomorphic encryption: theory & application. The Theory and Practice of Cryptography and Network Security Protocols and Technologies. InTech, 2013.
34. Thornton, C., et al., Auto-WEKA- combined selection and hyperparameter optimization of classification algorithms. arXiv preprint arXiv:1208.3719v2, 6 Mar. 2013.
35. Tramer, F., et al., Stealing machine learning models via prediction APIs. arXiv preprint arXiv:1609.02943, 3 Oct. 2016.
36. Vimercati, di S. De C., et al., Practical techniques building on encryption for protecting and managing data in the cloud. The New Codebrakers, Vol. 9100, LNCS, Springer, 2015.
37. Zhang, F., et al., CloudVisor: retrofitting protection of virtual machines in multi-tenant cloud with nested virtualization. SOSP'11, ACM, 2011.