# The Bell-LaPadula (BLP) Enterprise Security Architecture Model vs Inference Attacks

Dominic Ayamga
*School of Electrical and Data Engineering*
*University of Technology Sydney*
Sydney, Australia
dominic.ayamga@student.uts.edu.au or 0000-0001-8264-3189

Priyadarsi Nanda
*School of Electrical and Data Engineering*
*University of Technology Sydney*
Sydney, Australia
Priyadarsi.Nanda@uts.edu.au or 0000-0002-5748-155X

Manoranjan Mohanty
*School of Mathematical and Physical Sciences*
*University of Technology Sydney*
Sydney, Australia
Manoranjan.Mohanty@uts.edu.au or 0000-0002-0258-4586

***Abstract*—Protecting information flow, data and assets is paramount to every establishment. Therefore, enterprise security architecture design is essential in achieving this protection as it directly implements enterprise security policies. Existing research revealed that researchers have made little effort to investigate inference security challenges to enterprise security architecture design and to assess how the existing security architecture models fare against inference attacks. It was also discovered that existing security architecture models are too old and susceptible to inference attacks. Hence, this research explores a novel solution for designing effective enterprise security architecture and addressing inference attacks.**

***Index Terms*—Information flow, Security architecture, Inference attacks, impact on design behaviour, frameworks, models**

## I. Introduction

Enterprise architecture describes the overarching forms and functionalities of systems in enterprises, including stakeholders, frameworks, standards, and rules for distinct architecture designs. [13] Security architecture describes the comprehensive structure and recurring procedures that put in place policies, standards, risk management decisions, and planned support that facilitate the development and operations of employees. [37]

The root security of any established enterprise and its resources is its security policies, and the execution of these security policies determines its resilience to attacks from malicious actors. [2] Access control is a security mechanism that restricts and regulates access to certain areas of an organisation or a system using physical and logical means. [47] Inference attack refers to the ability to arrive at an individual's identity from a particular dataset by aggregating data that does not explicitly point to the individual. [17] [36]

### A. Types of inference attacks

The most common types are membership, property, poison, and attribute inference attacks. A membership inference attack (MIA) occurs when a foe tries to rebuild the dataset used in training the model [10] [25] [32] [40], taking advantage of the shortcomings in decision boundaries in susceptible algorithms resulting from the over-parameterisation and complexity. [11] [26] [49] Attribute inference attack is when the adversary exploits a dataset contained in the train model by manipulating the data to deduce data that he/she had not gotten the privilege of knowing. [20] [21] [29] [50] Property inference attack occurs when the adversary makes deductive conclusions or predictions from aggregated data obtained. [32] [16] [34] Poisoning inference attack is a form of MIA that occurs when a foe poisons certain portions of the training dataset to behave in a pattern of interest to the adversary. [32]

## II. OBJECTIVES

- To assess how the BLP model fare against inference attacks.
- To assess the academic research depth of inference attacks against the defensive capabilities of the BLP model through rigorous literature review.
- To perform simulations to select the best ML algorithm for the proposed improved BLP model.

## III. SIGNIFICANCE

i. To advocate for further research to improve existing enterprise security architecture models to defend against inference attacks

ii. To create awareness of the vulnerabilities of the BLP model to inference attacks

## IV. METHOD AND RESULTS

1) The methods employed include rigorous literature review and simulation.

2) Results from the literature review indicate the BLP model is vulnerable to inference attacks. The results of the simulation are found in tables I, II, and III below under the simulation results and conclusion section.

## V. BACKGROUND

A methodical delineation of a master plan that presents a model design development, criteria, execution processes and procedures, and security management processes, such as policy formulations, education, and training, denotes a security framework. [47] A security model is the theoretical and logical representation of the security policies of an enterprise or a system and the implementation processes and procedures while considering the enterprise's shortcomings. [27] Due to the varied nature of each institution's business objectives, security policy formulation cannot have a holistic procedure. Hence, the security framework is influenced by the business objectives. The research, however, only concentrates on the performance of the Bell-LaPdula (BLP) model against inference attacks and how that impacts enterprise security architecture design.

### A. Question

Is the BLP enterprise security architecture model able to mitigate inference attacks?

### B. The BLP Security Architecture Model

The Bell-LaPadula (BLP) Model, advanced in 1976 by David Bell and Leonard LaPadula, is rooted in the confidentiality of information, data, and assets. [4] [33] Though it was not designed to handle data integrity, the model achieves some integrity aspects if fully implemented well, as the lack of unauthorised access leads to fewer data corruption. Data availability is the aspect of the model that is negatively hampered when fully implemented and strictly adhered to. The BLP operates based on data or object classification levels (unclassified (U), confidential (C), secret (S), and top secret (Ts)) and security clearance levels of subjects interacting with data or objects. Moreover, it operates on the following set of rules.

*1) Simple security:* No read-up. [4] [33] A subject can only write up to objects with higher authority clearance than has been sanctioned but cannot read its content. [4] [33] In this property, subjects in the same security clearance level have read, write ( and also, execute though not categorically stated) access to objects under the security classification they are cleared to have access to (note: the model did not say anything or emphasis on the need-to-know principle), as well as being able to have read access to objects under the security classification lower to their security clearance level (note: a weak point to engage in inference attack (inference attacks possible).

Let $S = collection\,of\,subjects$
$O = objects\,of\,the\,system$
$P = privileges(read\,r, write\,a, read\,and\,write\ w,$
$and\,empty/null\,e)$
$M = the\,privileges\,regulation\,matrices$
$f = collection\,of\,three(3) - tuples\,\{f_s, f_o, f_c\}$, where $f_s$ denotes each subject's maximum security clearance level, $f_c$ denotes the current security clearance level, $f_o$ denotes each object security level. A state $(s, o, p) \in S{\times}O{\times}P$ fulfills the simple security property relative to the function $f$ provided one of these is true:
$i.\ p = e\,or\,p = a, \Longrightarrow$ for every subject $s$ of $S$ with security clearance and right $p$ of $P$ to the object $o$ of $O$ with which the subject is cleared, must have right $p$ equal null or empty($e$) (meaning not cleared to read) or $p$ equal write a (meaning cleared to write) or the security level $s$ must dominate $o$. [33]
$ii.\ p = r\,or\,p = w$, and $f_s(s)dom f_0(O) \Longrightarrow$ for subject $s$ a member ($\in$) of S with security clearance and right $p \in P$ to read $r$ an object $o \in O$ or read and write $w$ on the object $o \in O$, the relative security function $f_s$ of the subject $s$ should overshadow the relative security function $f_o$ of the object $o$. [33]

*2) *Security property:* A security property that allows subjects based on specific security clearance to have write (note: read and execute as well, though not categorically stated) privilege to objects within the same authorisation category to which they have been cleared as well as write privilege to objects with authorisation categories above the categories they have been cleared but cannot write to objects with authorisation categories below them. [4] [33] Put simply, no write-down. This prevents confidential data leakage to lower authority categorisation level(s). This leakage can be in the form of a subject with a secret authority category having read privilege to an object with a secret category and then making an unclassified copy of the secret document,

thereby making subjects with classified (lower than secret) authority category have read privilege to the otherwise secret object. [35] (Note: A property inference attack is possible).

A state $(b, m, f, h)$ fulfills $*$ property iff, for each $s \in S$, these are fulfilled:

(1). $b(s : a) \neq \emptyset \implies [\forall o \in b(s : a)\,[f_o(o)\,dom\,f_c(s)]] \implies$ for all subject $s$ a subset of $b$ with security clearance to write on an object $o$, the relative function of the object $f_o(O)$ must dominate the relative security function of the subject $f_c(s)$.

(2). $b(s : w) \neq \emptyset \implies [\forall o \in b(s : w)\,[f_o(o) = f_c(s)]] \implies$ for all subject $s$ a subset of $b$ with security clearance to read and write on an object $o$, the relative security function of the object $f_o(o)$ must equal the relative security function of the subject $f_c(s)$.

(3). $b(s : r) \neq \emptyset \implies [\forall o \in b(s : r)[f_c(s)dom f_o(O)]] \implies$ for all subject $s$ a subset of $b$ with security clearance to read on an object $o$, the relative security function of the subject $f_c(s)$ must dominate the relative security function of the object $f_o(O)$. [33] Fulfillment of the $*-$property is attained when every stage is met. If the subset $S^1$ satisfies the $*-$property, then, the $*-$property is satisfied relative to $S^1 \subseteq S$. [33]

Strong $*$property, categorically permits read-write privileges to objects with the same sanctioned category to which subjects are given security clearance to have access (weakness: inference attacks possible). [33] The set $(b, m, f, h)$ abide by the discretionary property, $iff$, for every member of the triple $(s, o, p) \in b, p \in [s, o] \implies$for every subject $s \in S$, and object $o \in O$ with rights $p \in P$ all being member $b$ in the four $(4) - tuple$ where $b \in P$ ( $s \times o \times p$ ) indicates which subject has rights to which object and categorises the rights. $m \in M$ being the authorisation restriction matrix of the present state. [33]

The Bell-LaPadula model was formulated to ensure secrets remain secret. [23] The BLP model presents a good environment to achieve success against inference attack possibilities, unlike the other existing models, but the existence of some inherent weaknesses in the model makes inference attacks possible. With the BLP model, data/information flows upwards.

Strength

$i$. Adhere to strong confidentiality principles/policies.

$ii$. Creates separation of duties.

$iii$. Have a well-defined reporting format of information flow.

Vulnerabilities

The simple security property vs inference security attack. A subject Ayamga with secret security clearance to read objects with security classification secret (also has write and execute privileges, though not categorically stated since the need-to-know principle was not indicated by the BLP model), and read privileges to objects with security classification level below secret classification, can perform membership inference attacks on all objects within the secret classification level as well as property inference attack by aggregating data within the secret and below the secret clearance levels thereby being able to logically conclude on the data that will be available to the top secret object classification level, a level which Ayamga has not been given security clearance to read and can use the data gained through the inference attack to deductively conclude decisions of persons with security clearance to objects with top-secret security classifications.

The * security property vs inference security attack. Ayamga has write (also has read and execute privileges, though not explicitly stated) access to objects with security classification (secret) to which he has been cleared, as well as write access to objects with security classification (top secret) to which he has not been cleared to read. Ayamga does not have write access to the objects with a security classification below the secret classification.

Ayamga can make membership inference attacks on objects with the same security classification to which he has been cleared through data manipulation and also can perform property inference attacks on objects with security classification to which he has been cleared as well as objects with security classification (top secret) which he has not been cleared to have read access to by performing data aggregation and thereby being able to predict data that will be available to the top secret security classification objects. This enables him to draw conclusions and predict the next moves of subjects cleared to have access to objects with top-secret classifications. The strong * security property vs inference attacks. Ayamga has read-write (and execute, though not categorically stated) privileges to objects within the same authorisation category (secret) and can perform both membership inference attacks through data manipulation and property inference attacks through data aggregation on objects.

### C. The Relationship between Set Theory, Information Theory and Inference Attacks

Given the universal set $\mu$, set $A \in \mu$, set $B \in \mu$, set $C \in \mu$, and $n(A \cup B \cup C) \in n(\mu)$.
$n(A \cup B \cup C) = n(A) + n(b) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$
Let $n(a)$ be the number of elements found in set $A$ only, $n(b)$ be the number of elements found in set $B$ only, and $n(c)$ is the number of elements found in set $C$ only. The focus here is on $n(A \cup B \cup C), n(A \cap B), n(A \cap C), n(B \cap C), n(A \cap B \cap C), a, b, c.$ From set theory,
$n(a) = n(A) - n(A \cap B) - n(A \cap C)$
$n(b) = n(B) - n(A \cap B) - n(B \cap C)$
$n(c) - n(C) - n(A \cap C) - n(B \cap C)$ Relating the set theory to information theory: Let $i = 1, 2, \ldots, n, x_i \in A, m_i \in B \, and \, r_i \in C$
$a = A(x_i \mid m_i; r_i) =$ the uncertainty of set $A$ left after knowing sets $B$ and $C$,
$b = B(m_i \mid x_i; r_i) =$ uncertainty of set $B$ left after knowing sets $A$ and $C$,
$c = C(r_i \mid x_i; m_i) =$ uncertainty of set $C$ left after knowing sets $A$ and $B$,
$n(A \cap B) only = I(x_i; m_i \mid r_i) =$ the interaction of $A$ and $B$ given $C$,
$n(A \cap C) only = I(x_i; r_i \mid m_i) =$ the interaction of $A$ and $C$ given $B$,
$n(B \cap C) only = I(r_i; m_i \mid x_i) =$ the interaction of $B$ and $C$ given $A$, and
$n(A \cap B \cap C) = I(A, B, C)$ =the interaction between $A$,$B$,and $C$. Let $n(A \cup B \cup C) = I(x_i, m_i, r_i)$ =the main information (Actual prediction target of the attacker). Note: The amount of information shared between $A$ and $B$ is impacted by $C$, the amount of information shared between $A$ and $C$ is impacted by $B$, and so is the information shared between $B$ and $C$ impacted by $A$. [1] According to Claude Shannon [1] [43], the conditional interaction of $A$ and $B$, given as $I(n(A \cap B))$, the conditional interaction of $A$ and $C$ given as $I((A \cap C))$ and conditional interaction of $B$ and $C$ given as $I(n(B \cap C))$ are respectively theorised as,
$n(A \cap B) = I(x_i; m_i \mid r_i) = E_r((x_i; m_i) \mid r_i) = \sum_{r_i \in C} P_(r_i)(r_i) \sum_{m_i \in B} \sum_{x_i \in A} P_{x_i, m_i \mid r_i}(x_i, m_i \mid r_i) \log \frac{P_{x_i, m_i \mid r_i}(x_i, m_i \mid r_i)}{P_{x_i \mid r_i}(x_i \mid r_i) P_{m_i \mid r_i}(m_i \mid r_i)}$

$n(A \cap C) = I(x_i; r_i \mid m_i) = E_m(I(x_i; r_i) \mid m_i) = \sum_{m_i \in B} P_{m_i}(m_i) \sum_{r_i \in C} \sum_{x_i \in A} P_{x_i, r_i \mid m_i}(x_i, r_i \mid m_i) \log \frac{P_{x_i \mid m_i}(x_i \mid m_i)}{P_{r_i \mid m_i}(r_i \mid m_i)}$

$n(B \cap C) = I(r_i; m_i \mid x_i) = E_x(I(r_i; m_i) \mid x_i) = \sum_{x_i \in A} P_{x_i}(x_i) \sum_{m_i \in B} \sum_{r_i \in C} P_{r_i, m_i \mid x_i}(r_i, m_i \mid x_i) \log \frac{P_{r_i, m_i \mid x_i}(r_i, m_i \mid x_i)}{P_{r_i \mid x_i}(r_i \mid x_i) P_{m_i \mid x_i}(m_i \mid x_i)}$. [1]

Achieving only $a$ information or only $b$ information or only $c$ is not the attacker's focus since the attacker seeks to establish a relationship between the sets and draw a conclusion. Applying mutual information is essential when trying to enhance the uncertainty or entropy of training datasets to make it very challenging for would-be membership inference attackers.

## VI. LATTICE-BASED ACCESS CONTROL

The lattice access control describes an upper and lower bound called layers or lattice, identifying the security clearance of designated subjects and their corresponding security classification objects. [31] [41] [42] Information flow is granted when the security designation of subjects is aligned with the designation of objects. [31] [41] [42]

The partial ordered set $(S, R)$ is called a lattice if and only if it is a meet semilattice and a join semilattice. [6] [18] [22] [38] In a Hasse diagram, for all $x, y$ members of $S$ $(\forall \, x, y \in S)$, the Greatest Lower Bound(GLB) of $x, y$ not equal null $(\emptyset)$ or phi $(\phi)$ $(GLB(x, y) \neq \emptyset \, or \, \phi)$ and for all $x, y$ members of $S$ $(\forall \, x, y \in S)$, phi $(\phi)$, the Least Upper Bound (LUP) of $x, y$ not equal to null $(\emptyset)$ $(\forall \, x, y \in S, (\phi) \neq \emptyset)$. [18] [38] In a complete lattice, the partial ordered set $(S, R)$ is considered a complete lattice provided for every subset $M$ of set $S$ has both a "meet" (greatest lower bound (GLB)) and a "join" (least upper bound (LUB)) in the partial ordered set $(S, R)$. [18] [38]

## VII. PROBLEM

Inference attacks may be reconnaissance for a main attack or as a primary attack. Inference attacks are done to exploit enterprises in the case of malicious actors or to have a competitive advantage in the case of competitors. [15]

### A. The BLP Model's Scenario (Military Environment) of Inference Attack

Consider a hypothetical three armies from three different countries, Yorogo, Bolga and Tamale. The Bolga army is attacking Yorogo. The president of Yorogo requested technical and logistic support from Tamale to defend itself from the invasion. Guided by Sun Tzu's war principle, discovering the vulnerabilities of the adversary, having the capabilities to exploit the adversary's vulnerabilities and yet not knowing whether the platform to use to exploit the

adversary is safe and usable, victory is not guaranteed. [19] [44]

The political elites and the army of Tamale then set up units to support Yorogo. These units include logistics supply (supply of weapons), geotechnical advisory (offer advisory on the geology and terrain of Yorogo lands), meteotechnical advisory (offer advisory on meteorology of Yorogo), and political and military advisory (for advising the political elites and directing the military units of Yorogo). Yorogo then planned a counterattack on the invading army of Bolga with the help of Tamale to retake the lost part of its lands.

Let $E$ represent the logistics supply unit, $F$ represents the geotechnical advisory unit, $G$ represent the meteotechnical advisory unit, and $R$ represents the political and military advisory unit. Let $R = (E, F, G)$, the security classifications of the information flow of the overall hierarchical structure of Tamale support to Yorogo be Top secret $(T_s)$, Secret$(S)$, Classified/Confidential $(C)$ and Unclassified $(U)$, $M = (T_s, S, C, U)$. In a typical Bell-LaPadula lattice representation, the given Hasse diagram of the partial ordered set (M,R) is given in $fig.1$ and $fig.2$.

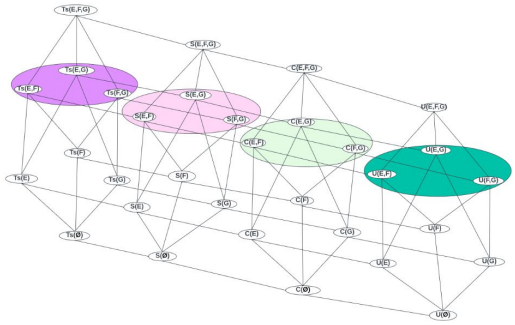

Fig. 1: A complete BLP lattice of bottom-up military information flow with highlighted sections where property inference can occur.

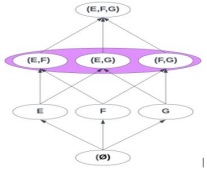

Fig. 2: A simplified BLP lattice that indicates the direction of information flow

Information flows upward in the above Hasse lattice diagram representation (from $U$ to $T_s$). Consider an officer working at the $T_s$ $(E, F)$ level and is a mole or insider attacker (consciously or unconsciously) leaking information to the invading army of Bolga. Assuming the mole obtained his information using the sensitive information he/she is previewed to by his/her position in the security clearance level $(T_s(E, F))$ and performing a property inference attack to obtain other information, he/she is not previewed to. The following steps will be undertaken.

Step1.

At $T_s(E, F)$, the mole has no possibility of getting information from $T_s(G), S(G), C(G)$ and $U(G)$ per the logic of the above Hasse diagram and based on the BLP model. Hence, the foe would execute the first inference attack on that segment.

Step2.

In the Hasse (or lattice) diagram above and a typical BLP model, information restriction decreases downward from $T_s$ to $U$. Meaning that information at the lower levels is loosely controlled. Hence, the priority target for the attacker will be $U(G)$ and $C(G)$. Note: In real life, information at the $U(G)$ level is open to the public (from publicly available records) or is made available to the public (through the declassification of $C(G)$). Therefore, many resources will not be expended to obtain information at level $U(G)$. However, the reliability of the G information increases upward for the mole. Hence, the focus of the property inference attack on G will be $C(G)$ because $C(G)$ relates more readily to the entire support of Tamale to Yorogo.

Step3.

Technically, from the lattice (Hasse) diagram above, phi $(\phi)$ constitutes a meet (Greatest Lower Bound (GLB)) to all lattice connections from $T_s$ to $U$. Phi$(\phi)$ could represent a meeting hall, a social media platform, or any gathering where ideas are shared. The mole, being on a mission, could befriend subjects at the $C(G)$ with the motive of obtaining information from an unsuspecting $C(G)$ subject (officer). Through casual interactions or chats, the mole will look out for comments relating to $C(G)$ from the $C(G)$ level subject (officer) and compare it with the publicly available $U(G)$ level information and then weeding out what he (mole) deems chuff (mole looks for $C(G) \cap U(G)$ information) and then aggregating that with information at his (mole) disposal at the Ts (E, F) clearance level$(T_s(E, F) \cup (C(G) \cap U(G))$.

Step 4

The aggregated information $(T_s(E, F) \cup (C(G) \cap U(G))$ is then used to perform a property inference attack on $T_s(E, F, G)$ by predicting the sort of directives that officers at the $T_s(E, F, G)$ clearance level will give to the political and army sectors of Yorogo regarding the counter-offensive attack on the army of Bolga. These directives could be when (time or month) to carry out the counter-offensive attacks, the weapons and machinery supply time, available manpower, mode of attack, directions to launch an attack and possible vulnerabilities of the ally (Yorogo) army and

the adversary (Bolga) army, based on topological, geological, and meteorological information.

Thus, defeating Sun Tzu's principle which states that we keep secret our attack direction and gives the enemy the daunting task of securing all directions with fewer opposing forces. [19] [44] In fig 1 and 2 of the Hasse diagram above of the Bell-LaPadula lattice, the $T_s(E,F), T_s(E,G), T_s(F,G), S(E,F), S(E,G), S(F,G), C(E,F), C(E,G), C(F,G), U(E,F), U(E,G)$,and $U(F,G)$ mostly play a redundancy role to $T_s(E,F,G)$, $S(E,F,G), C(E,F,G)$,and $U(E,F,G)$ respectively. For example, in the absence of $T_s(E,F), T_s(E,F,G)$ still gets $T_s(E,F)$ component from $T_s(E,G) \cup T_s(F,G)$. Similarly, in the absence of any of the combinations, that component could still be realised from the remaining combinations. The highlighted sections in the Hasse diagram of the BLP lattice mostly provide redundancy and do not necessarily ensure confidentiality or security.

## VIII. WAY FORWARD

The following are the proposed ways forward to help reduce the complexity of the security structure, aid rapid implementation of security, and to troubleshoot security issues.

### A. Breaking the lattice

Breaking the lattice refers to the artificial removal of the portion of the BLP lattice that plays redundant roles, as stated above, by replacing the humans in the redundant portion of the lattice with an automated system using an artificial intelligence system called an agent. The agent is expected to play a similar role as the humans by connecting and coordinating the activities of the upper and lower levels of the lattice, thereby hindering the knowledge the humans at the section replaced would have gotten and, thus, reducing property inference attacks possibilities.

### B. Administrative isolation and Need-to-know

This has to do with the design of office structure in a way that prevents people working on different independent portions of the security clearance levels of the enterprise from sharing the same office space, which could lead to the sharing of data, leading to property inference attacks through the aggregation of data. Allowing isolated users access to only data required for their specific task

## IX. THE SECURITY CONUNDRUM

Artificially breaking the lattice by introducing agents to replace humans at specific sections in the lattice's hierarchical structure, which facilitates property inference attacks, introduces new vulnerable layers. These agents would constitute an aggregated data point in a section in the lattice hierarchical structure. An adversary being able to determine the existence of an individual's record from an aggregated location time series risks the confidentiality of the entire records in the dataset of the agents. [39] The agents introduced will become the primary target for attacks, including MI attacks. Securing the agents against MI constitutes a new challenge. The agents are expected to use ML and deep learning (DL) algorithms. The selection of algorithms will be based on factors such as support for dynamism, defense against inference attack possibilities, overhead cost in processing time (latency) and speed. Researchers have developed numerous algorithms. [12] [5] [8] [9]These algorithms have their corresponding strengths and weaknesses.

### A. Countermeasures to MIA

Defensive measures against MIA attacks are broadly grouped into confidence score masking, regularisation, knowledge distillation, and differential privacy (DP). [26] An approach used to mask the actual confidence score obtained from the target classifier from the adversary's view, thereby effectively limiting the adversary's capabilities to carry out a black-box MI attack, is termed confidence score masking. [26] Regularisation refers to reducing the degree of overfitting in ML to mitigate MI attacks. [26] Knowledge distillation uses larger model results to train smaller models for knowledge to be passed to the smaller model. [26] [3] [24] The theoretical privacy information assurance obtained through a probabilistic approach is called DP. [26] [14] Over the years, researchers have explored several approaches to mitigate MIA attacks in ML and DL algorithms or models. [25] [7] [30]

### B. Dynamism, Overhead cost and Speed(latency)

The traditional Bell-LaPadula access control model is not dynamic. [45] [48] Policies and regulations often do not change once implemented and when the system is running or undergoing transition, which is a problem. [45] [48] The static nature of the traditional BLP model makes it susceptible to possible attacks [45] [48], including MI attacks. Employing ML and DL to implement the BLP confidentiality model seeks to introduce dynamism. [45] [48] The proposed agents are expected to use this dynamism and other features in the machine learning models or algorithms to increase performance and fend off inference attacks. In a production environment, speed matters, as it affects cost accruing from delays, affects the ease of use, which affects

the strict adherence to the implementation security policies and determines the efficacy of the supposed security policies and security architecture of the system.

In [45] [48] research, the MaxENT-BLP and CRFs-MaxENT-BLP have demonstrated some significant advantages, considering the results of other known DL and traditional ML models. However, latency comparison between the MaxENT-BLP, CRFs-MaxENTP-BLP models, and traditional ML was not done. Latency comparison is essential when adopting a model for the proposed agents. In this research, a simulation of three models (the SVM, Decision Tree, and MaxENT) was undertaken to evaluate the prediction accuracy of each, precision and the computational overhead cost, using a heart disease dataset of size 70,000 rows and performance against inference attacks using malware dataset.

## X. Simulation Results and Conclusion

Below are the results of three classifiers from the heart disease dataset, Dataset name: heart-data.csv Author: Kuzak Dempsy Source: https://data.world/kudem and Malware dataset [46]. The heart disease dataset is used to assess the performance of models on everyday working data, and the malware dataset is used to assess the performance of models against MI attacks.

| **Heart disease dataset** | | | | |
|---|---|---|---|---|
| | **Acc** | **Preci** | **F1 sco** | **T.in min.** |
| **D.Tree** | 73.414 | 76.378 | 71.807 | 0.011 |
| **SVM** | 69.838 | 74.317 | 72.522 | 4.858 |
| **MaxENT** | 70.316 | 74.371 | 71.282 | 0.969 |

TABLE I: Heart disease dataset simulation results

| **Malware dataset** | | | | |
|---|---|---|---|---|
| | **Acc** | **Preci** | **F1 sco** | **T. in min.** |
| **D.Tree** | 98.395 | 98.305 | 98.281 | 0.072 |
| **SVM** | 98.978 | 98.331 | 98.367 | 2.674 |
| **MaxENT** | 98.971 | 98.414 | 96.558 | 0.001 |

TABLE II: Malware dataset simulation results

| **Area under curve score** | | |
|---|---|---|
| **Model** | **H.disease-dataset** | **Mal.dataset** |
| **D.Tree** | 0.793 | 0.988 |
| **SVM** | 0.732 | 0.988 |
| **MaxENT** | 0.724 | 0.973 |

TABLE III: AUC Score

An auc value from an roc curve is a gauge metric of the achievement of a ML model. [28] An auc value is the most potent metric for assessing the prognostic achievements of a given ML mode. [28] Assessing the AUC values of the above models, the Decision Tree has the highest value of $0.793$ for the heart disease dataset and a marginally lower value of $0.996$ than the MaxENT for the malware datasets, and the MaxENT got the least $0.724$ for the heart disease dataset and slightly above the Decision Tree for malware datasets with $0.973$. Again, assessing the time cost, the average time cost for prediction in both heart disease and Malware datasets for the Decision Tree is $0.0415$ $((0.011{+}0.072)/2)$ minutes, and the MaxENT is $0.485$ $((0.969 + 0.001)/2)$ minutes.

Further, in terms of predictive accuracy, the Decision Tree has the highest accuracy of $73.414$ in the heart disease dataset, with the MaxENT being the least $70.316$, while with the malware dataset, the MaxENT slightly outperformed the Decision Tree with an accuracy of $98.971$ and $98.395$, respectively, a marginal difference of $0.576$. Given the main factors considered in adopting a desired machine learning model for the proposed agent, the Decision Tree fits the requirement much better than the other models. A secured system requires a thorough and detailed defense in all facets. Inference attacks, though an old form of security threat, have evolved. Security architecture models need to be constantly improved.

## References


[1] Ahmed Al-Ani, Mohamed Deriche, and Jalel Chebil. A new mutual information based measure for feature selection. *Intelligent Data Analysis*, 7(1):43–57, 2003.

[2] Dominic Ayamga. Telecommunication fraud prevention policies and implementation challenge, 2018.

[3] Jimmy Ba and Rich Caruana. Do deep nets really need to be deep? *Advances in neural information processing systems*, 27, 2014.

[4] D Elliot Bell, Leonard J LaPadula, et al. Secure computer systems: Mathematical foundations. Technical report, Citeseer, 1973.

[5] Adam Berger, Stephen A Della Pietra, and Vincent J Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

[6] G Birkhoff. Lattice theory. *American Mathematical Society*, 1940.

[7] Myria Bouhaddi and Kamel Adi. Mitigating membership inference attacks in machine learning as a service. In *2023 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 262–268. IEEE, 2023.

[8] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[9] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[10] Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramer. Membership inference attacks from first principles. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1897–1914. IEEE, 2022.

[11] Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song. The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX security symposium (USENIX security 19)*, pages 267–284, 2019.

[12] Gonzalo Martínez Ruiz de Arcaute, José Alberto Hernández, and Pedro Reviriego. Assessing the impact of membership inference attacks on classical machine learning algorithms. In *2022 18th International Conference on the Design of Reliable Communication Networks (DRCN)*, pages 1–4. IEEE, 2022.

[13] Mahesh R Dube and Shantanu K Dixit. Comprehensive measurement framework for enterprise architectures. *International Journal of Computer Science & Information Technology (IJCSIT)*, 3(4):71–92, 2011.

[14] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

[15] Csilla Farkas and Sushil Jajodia. The inference problem: a survey. *ACM SIGKDD Explorations Newsletter*, 4(2):6–11, 2002.

[16] Karan Ganju, Qi Wang, Wei Yang, Carl A Gunter, and Nikita Borisov. Property inference attacks on fully connected neural networks using permutation invariant representations. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security*, pages 619–633, 2018.

[17] Simson Garfinkel et al. *De-identification of Personal Information:*. US Department of Commerce, National Institute of Standards and Technology, 2015.

[18] Alexander Garver and Thomas McConville. Lattice properties of oriented exchange graphs and torsion classes. *Algebras and Representation Theory*, 22:43–78, 2019.

[19] James Gimian and Barry Boyce. *The Rules of Victory: How to Transform Chaos and Conflict (Strategies from the Art of War)*. Shambhala Publications, 2009.

[20] Neil Zhenqiang Gong and Bin Liu. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 979–995, 2016.

[21] Neil Zhenqiang Gong and Bin Liu. Attribute inference attacks in online social networks. *ACM Transactions on Privacy and Security (TOPS)*, 21(1):1–30, 2018.

[22] George Grätzer. *General lattice theory*. Springer Science & Business Media, 2002.

[23] Shon Harris. *CISSP all-in-one exam guide*. McGraw-Hill, Inc., 2010.

[24] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[25] Hongsheng Hu, Zoran Salcic, Gillian Dobbie, Yi Chen, and Xuyun Zhang. Ear: an enhanced adversarial regularization approach against membership inference attacks. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

[26] Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, Philip S Yu, and Xuyun Zhang. Membership inference attacks on machine learning: A survey. *ACM Computing Surveys (CSUR)*, 54(11s):1–37, 2022.

[27] Vincent C Hu, David Ferraiolo, D Richard Kuhn, et al. *Assessment of access control systems*. US Department of Commerce, National Institute of Standards and Technology . . . , 2006.

[28] Jin Huang and Charles X Ling. Using auc and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3):299–310, 2005.

[29] Jinyuan Jia and Neil Zhenqiang Gong. {AttriGuard}: A practical defense against attribute inference attacks via adversarial machine learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 513–529, 2018.

[30] Charles Kowalski, Azadeh Famili, and Yingjie Lao. Towards model quantization on the resilience against membership inference attacks. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3646–3650. IEEE, 2022.

[31] NV Narendra Kumar and RK Shyamasundar. A complete generative label model for lattice-based access control models. In *Software Engineering and Formal Methods: 15th International Conference, SEFM 2017, Trento, Italy, September 4–8, 2017, Proceedings 15*, pages 35–53. Springer, 2017.

[32] Saeed Mahloujifar, Esha Ghosh, and Melissa Chase. Property inference from poisoning. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1120–1137. IEEE, 2022.

[33] Bishop Matt. Computer security: art and science, 2002.

[34] Muhammad Naveed, Seny Kamara, and Charles V Wright. Inference attacks on property-preserving encrypted databases. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 644–655, 2015.

[35] Sergei Obiedkov, Derrick G Kourie, and Jan HP Eloff. On lattices in access control models. In *Conceptual Structures: Inspiration and Application: 14th International Conference on Conceptual Structures, ICCS 2006, Aalborg, Denmark, July 16-21, 2006. Proceedings 14*, pages 374–387. Springer, 2006.

[36] Junhyoung Oh and Kyungho Lee. Data de-identification framework. *Computers, Materials & Continua*, 74(2), 2023.

[37] Gunnar Peterson. Security architecture blueprint. *Arctec Group, LLC*, 2007.

[38] VBVN Prasad, T Ramarao, TS Rao, T Nageswara Rao, and K Prasad. Some basic principles on posets, hasse diagrams and lattices. *Test Engineering and Management*, 83(2):10771–10775, 2020.

[39] Apostolos Pyrgelis, Carmela Troncoso, and Emiliano De Cristofaro. Knock knock, who's there? membership inference on aggregate location data. *arXiv preprint arXiv:1708.06145*, 2017.

[40] Ahmed Salem, Yang Zhang, Mathias Humbert, Pascal Berrang, Mario Fritz, and Michael Backes. Ml-leaks: Model and data independent membership inference attacks and defenses on machine learning models. *arXiv preprint arXiv:1806.01246*, 2018.

[41] Ravi S. Sandhu. Lattice-based access control models. *Computer*, 26(11):9–19, 1993.

[42] Ravi S Sandhu. Role-based access control. In *Advances in computers*, volume 46, pages 237–286. Elsevier, 1998.

[43] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[44] Tzu Sun. *The art of war*. Hachette UK, 1994.

[45] Zhuo Tang, Xiaofei Ding, Ying Zhong, Li Yang, and Keqin Li. A self-adaptive bell–lapadula model based on model training with historical access logs. *IEEE Transactions on Information Forensics and Security*, 13(8):2047–2061, 2018.

[46] Christian Urcuqui-López and Andrés Navarro Cadavid. Framework for malware analysis in android. *Sistemas y Telemática*, 14(37):45–56, 2016.

[47] Michael E Whitman and Herbert J Mattord. *Management of information security*. Cengage Learning, 2019.

[48] Li Yang, Jin Wang, Zhuo Tang, and Neal N Xiong. Using conditional random fields to optimize a self-adaptive bell–lapadula model in control systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(7):4505–4519, 2019.

[49] Wei Zhang, Yaping Lin, Jie Wu, and Ting Zhou. Inference attack-resistant e-healthcare cloud system with fine-grained access control. *IEEE Transactions on Services Computing*, 14(1):167–178, 2018.

[50] Benjamin Zi Hao Zhao, Aviral Agrawal, Catisha Coburn, Hassan Jameel Asghar, Raghav Bhaskar, Mohamed Ali Kaafar, Darren Webb, and Peter Dickinson. On the (in) feasibility of attribute inference attacks on machine learning models. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 232–251. IEEE, 2021.