Controller of Certifying Authorities
Ministry of Electronics & Information Technology
Government of India

G20 भारत 2023 INDIA

IEEE COMPUTER SOCIETY Bangalore Chapter

IAS IEEE INDUSTRY APPLICATIONS SOCIETY Linking Research to Practice Bangalore Chapter

4TH INTERNATIONAL CONFERENCE ON

PUBLIC KEY INFRASTRUCTURE AND ITS APPLICATIONS (PKIA 2023)

SEPTEMBER 8-9TH, 2023

An Architecture for Risk-Based Authentication System in a Multi-Server Environment

PRAMILA R M, CHRIST, PUNE
SAMIKSHA SHUKLA, CHRIST, BANGALORE

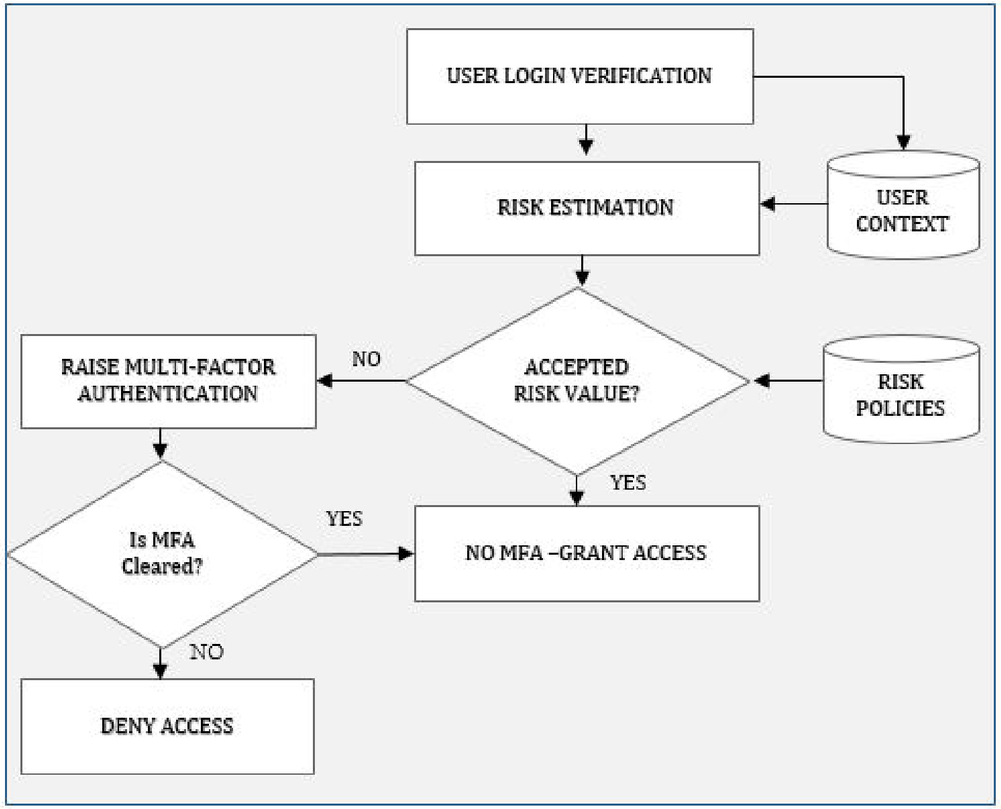
Abstract

- *Machine Learning and Adoptive Authentication is one of the advanced security solution to detect user behavior anomalies but very less work is proposed in this area. The existing works does not focus on extracting new information from AA features neither recommends efficient technique involved in developing models for user anomaly detection during the first few login attempts.*
- *The work proposes a design of Risk–based architecture to estimate risk for the user during the initial login process and also when the user's data is extracted enough for prediction in a multi-server environment.*

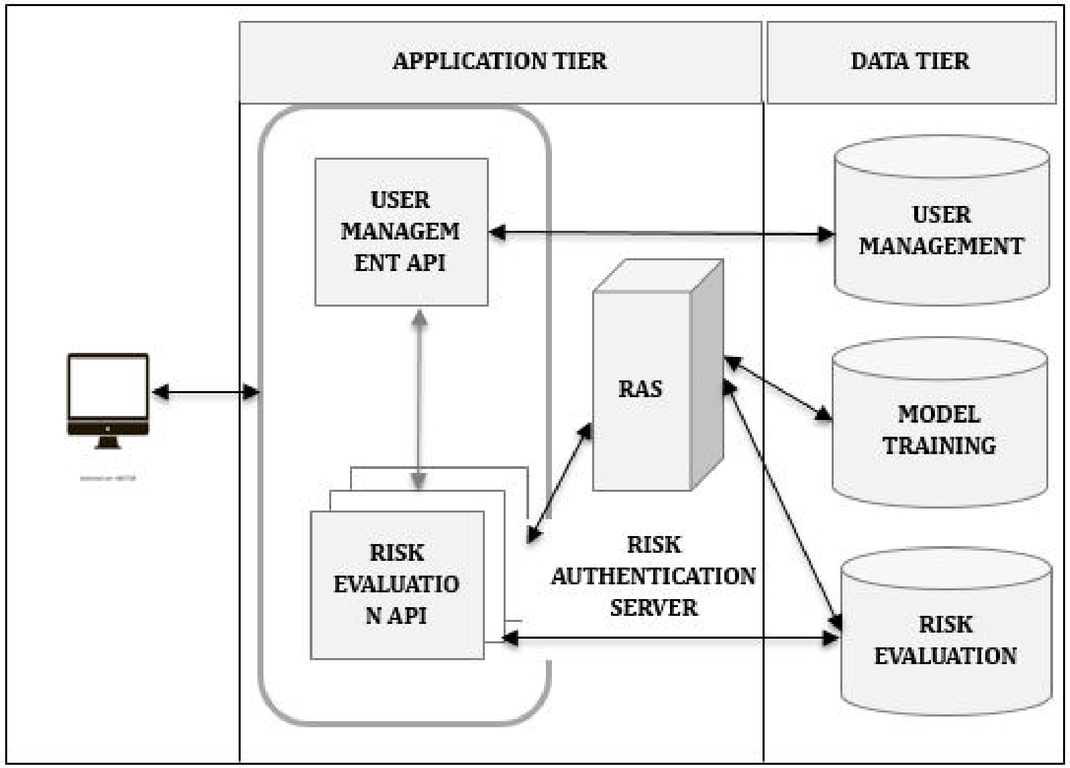
Related Work

- *Misbahuddin et al. - Design of Risk based authentication system using machine learning environment.*
- *Solano et al. - Risk-Based Static Authentication in Web Applications with Behavioral Biometrics and Session Context Analytics.*
- *Ding et al. - User identity authentication and identification based on multi-factor behavior features*
- *Martin et al. - An approach to detect user behaviour anomalies within identity federations*
- *Wiefling et al. - Pump Up Password Security! Evaluating and Enhancing Risk-Based Authentication on a Real-World Large-Scale Online Service*

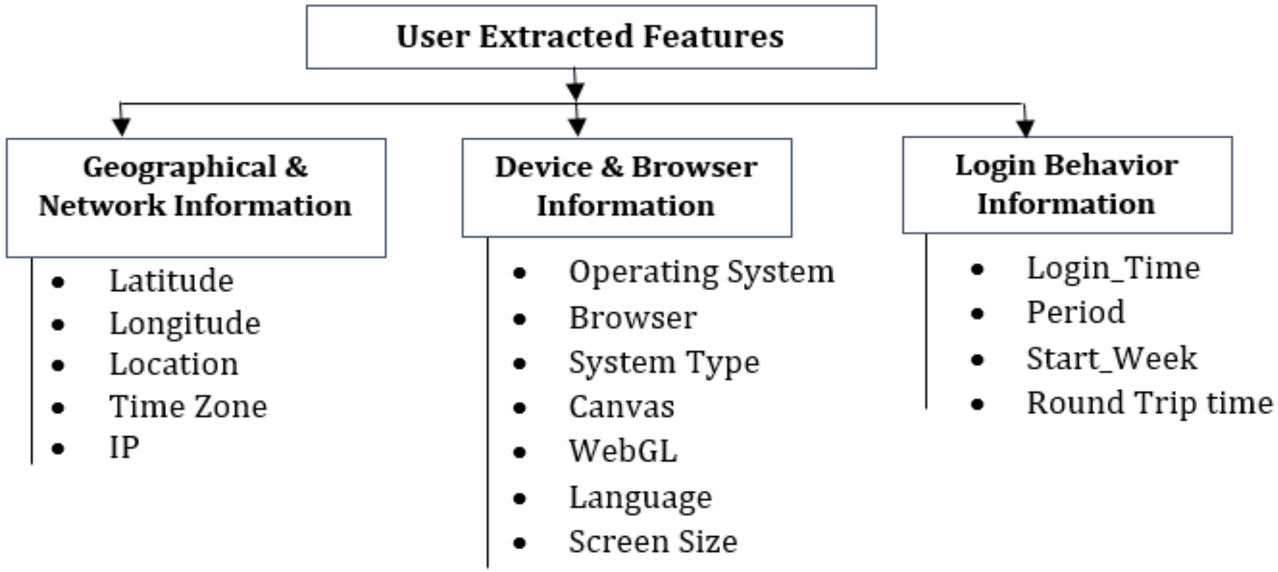
Proposed Work – Control Flow Architecture



Proposed Work – Control Flow Architecture



BROWSER AND DEVICES FINGERPRINTING ATTRIBUTES



Category 1: Geographical Information & Network Information

Category 2: Devise and Browser Information

Category 3: Login Behavior Information

Risk Estimation Methodology

$$\mathrm{LoR_u} = \sum_{i=1}^{n} UPS_i * UPW_i$$

-UPW_i is a user parameter weight of the contextual feature
-UPS_i is a probability of occurrence of a value in a feature
The value of UPS_i is the following:

$$UPS_i = \begin{cases} 1 & \text{if } p(x_i) \in [0 ; e_1] \\ 0 & \text{if } p(x_i) \in [e_2 ; 1] \end{cases}$$

Where:

- e_1 and e_2 are thresholds

Risk Estimation Methodology

Conditional Evaluation at Level 2

CG4: Category 1 and Group 1& 2 Features: IP, Time Zone, Location, Latitude& Longitude.

CG5: Category 2 and Group 1& 2 Features: Canvas, Browser, Operating System, Language, Screen Size, System Type, WebGL.

CG6: Category 3 and Group 1& 2 Features: Period, RTT, Login Time, Start Week

Risk Estimation Condition at Level 2:

If $LoR_u(CG4) > t_4$ or $LoR_u(CG5) > t_5$ or $LoR_u(CG6) > t_6$ Then raise MFA

Feature Weights and Threshold

Table 1: RBA Configurations for Level 1 Authentication

		Risk Level Behavior	
		LLS	**HLS**
Feature Categories	Geographical & Network	>0.5<1	1
	Device & Browser	>0.4<0.7	>=0.7
	Login Behavior	>0.5<1	1

Table 2: RBA Configurations for Level 2 Authentication

		Risk Level Behavior	
		LLS	**HLS**
Feature Categories	Geographical & Network	> 0.2 < 0.4	>= 0.4
	Device & Browser	> 0.15 < 0.25	>= 0.25
	Login Behavior	> 0.15 < 0.3	>= 0.3

Use Cases for Threshold Deciding Factor

Geographical & Network Information

1. IP
2. Latitude-Longitude
3. Location & Latitude-Longitude
4. Location & Latitude-Longitude & IP
5. Time Zone & Location & Latitude-Longitude
6. Time Zone & Location & Latitude-Longitude & IP

Login Behavior Information:

1. Period
2. Login Time
3. Start Week

Device & Browser Information

1. Screen-Size
2. Language
3. Operating System & Browser
4. Operating System & Canvas & WebGL
5. Browser & Canvas & WebGL
6. Browser & Screen-Size
7. System Type & Canvas & WebGL
8. System Type & Browser & Operating system & Canvas & WebGL & Screen Size

Results of Level 1 Authentication

User ID	Time Zone	IP Address	Canvas	Browser	Operating System	Period	RTT	Total Risk Score
rbadmin	IST	123.252.201.190	4eef-------	Chrome 113.00.00	Windows 10	0	0.4012673	NIL
rbadmin	AST	123.252.201.191	4eef-------	Chrome 113.00.00	Windows 10	1	0.4352673	(1.0, 0.0, 1.0)
rbadmin	IST	123.252.201.190	4eef-------	Chrome 113.00.00	Windows 10	0	0.4252673	(0.0, 0.0, 0.0)
rbadmin	IST	132.252.201.191	4eef-------	Chrome 113.00.00	Windows 10	0	0.4152673	(0.5, 0.0, 0.0)
rbadmin	IST	132.252.201.191	4eef-------	Chrome 113.00.00	Ubuntu 7.10	0	0.5352673	(0.5, 0.3, 1.0)
rbadmin	IST	132.252.201.191	4eef-------	Chrome 113.00.00	Windows 10	1	0.5152673	(0.0, 0.0, 0.5)
rbadmin	IST	123.252.201.190	5def----	Chrome 112.00.00	Windows 10	0	0.4152673	(0.5, 0.7, 0.0)
rbadmin	IST	123.252.201.190	4eef-------	Chrome 113.00.00	macOS Ventura 13.4	0	0.4452673	(0.0, 0.3, 0.0)

Results of Level 2 Authentication

UserID	Time Zone	IP Address	Lat& Long	Location	Canvas	Browser	Operating System	Language	Screen Size	WebGL	System Type	Period	RTT	Login Time	Start Week	Total Risk Score
rbadmin	IST	123.252.20 1.190	20.0760:72.8	IndiaMahara	5eefcac4260	Chrome 112.00.0 0	Windows 10	en-US,en	824:1537	6ebabd8644	Deskto	1	0.45267	17.30.18	2	NIL
rbadmin	IST	123.252.20 1.191	20.1160:72.8	IndiaMahara	5eefcac4260	Chrome 112.00.0 0	Windows 10	en-US,en	824:1538	6ebabd8644	Deskto	0	0.41267	17.39.04	2	(0.02, 0.0, 0.7)
rbadmin	IST	123.252.20 1.190	21.1160:73.9	IndiaMahara	5eefcac4260	Chrome 112.00.0 0	Windows 10	en-US,en	824:1540	6ebabd8644	Deskto	1	0.42287	17.38.04	4	(0.2, 0.0, 0.15)
rbadmin	IST	123.252.20 1.190	21.1260:74.9	IndiaMahara	5cefcac4260	Chrome 113.00.0 0	Windows 10	en-US,en	824:1539	6bbabd8644	Laptop	0	0.40397	17.31.10	2	(0.18, 0.55, 0.4
rbadmin	IST	123.252.20 1.191	20.1260:72.8	IndiaMahara	4cefcac4260	Chrome 112.00.0 0	Ubuntu 7.10	en-US,en	800:1440	6cbabd8644	Deskto	0	0.44287	17.32.07	3	(0.4, 0.65, 0.15
rbadmin	IST	123.252.20 1.190	20.1160:72.9	IndiaMahara	5eefcac4260	Chrome 112.00.0 0	Windows 10	en-US,en	824:1540	6ebabd8644	Deskto	0	0.84235	17.31.05	2	(0.2,0.0,0.4)

RBA Updated Estimation Methodology

$LoR_u = \sum_{i=1}^{n} UPS_i * UPW_i$

where:

-UPW_i is a user parameter weight of the contextual feature

-UPS_i is a user parameter score that can be calculated based :

Type 1: $p(x_i)$ probability of occurrence of a value in a feature

Type 2: $fv(x_i)$ feature value computed using a different methodology (refer to Table 3)

Finally, the value of UPS_i is the following:

$$UPS_i = \begin{cases} 1 \text{ if } p(x_i) \mid fv(x_i) \in [0 ; e_1] \\ 0 \text{ if } p(x_i) \mid fv(x_i) \in [e_2 ; 1] \end{cases}$$

Where:

- e_1 and e_2 are thresholds

$LoR_u = \sum_{i=1}^{n} UPS_i * UPW_i$ (2)

where:

-UPW_i is a user parameter weight of the contextual feature

-UPS_i is a user parameter score that can be calculated based :

Type 1: $o(x_i)$ occurrence of feature in the database

Type 2: $fv(x_i)$ feature value computed using a different methodology (refer to Table 3)

Finally, the value of UPS_i is the following:

$$UPS_i = \begin{cases} 1 \text{ if } o(x_i) \mid fv(x_i) \in [0 ; e_1] \\ 0 \text{ if } 0(x_i) \mid fv(x_i) \in [e_2 ; 1] \end{cases} \quad (2)$$

Where:

- e_1 and e_2 are thresholds

Models Results Estimation Methodology

	MODEL -1		MODEL -2	
	FRR	FAR	FRR	FAR
USER 1	0.28	0.14	0.31	0.43
USER 2	0.24	0.00	0.25	1.00
USER 3	0.52	0.00	0.57	0.75
USER 4	0.50	0.50	0.70	0.00
AVERAGE	0.39	0.16	0.46	0.54

Weightage for the Features (Public Dataset)

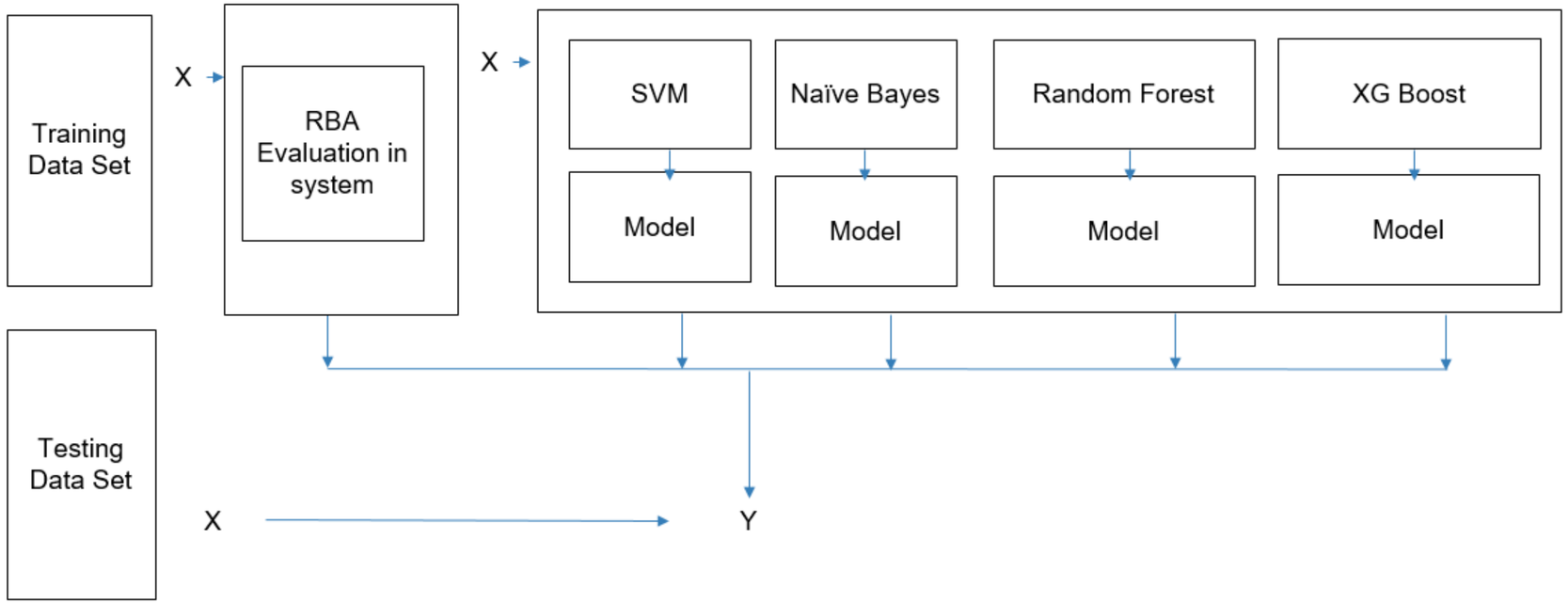
Level 1

FEATURES	WEIGHTAGE
IP	2
ASN	3
COUNTRY	2
REGION	3
DEVICE TYPE	2
OPERATING SYSTEM	5
BROWSER	3
RTT	8
LOGIN TIME	2

Level 2

FEATURES	WEIGHTAGE
IP	2
ASN	3
COUNTRY	2
REGION	3
DEVICE TYPE	1
OPERATING SYSTEM	4
BROWSER	3
USER AGENT	2
RTT	6
LOGIN TIME	2
PERIOD	2

ML Models


Training Data Set
X
RBA Evaluation in system
X
SVM
Naïve Bayes
Random Forest
XG Boost
Model
Model
Model
Model
Testing Data Set
X
Y

ML Model Results – F1 Score

	NB	RF	XGB	SVM
Imbalanced Data	85	81	81	84
SMOTE(OHE)	87	94	96	93
SMOTE(M-Estimate)	89	98	99	98
SMOTE(James-Stein)	87	99	98	99
SMOTE(Target)	92	98	99	97

Conclusion

- *In the proposed architecture the main idea is to provide an effective authentication system in a multi server during initial login process and also when sufficient data is collected from user. The first phase uses a NON ML model for authentication and the second phase uses ML model for prediction. The proposed architecture identifies risk at two levels based on the data. The Control flow has proven to be effective with the sample dataset.*

Controller of Certifying Authorities
Ministry of Electronics & Information Technology
Government of India

G20 भारत 2023 INDIA

THANK YOU

भगत संगणन विकास केंद्र
सीडैक CDAC
॥ज्ञानादेव तु कैवल्यम्॥
CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING

IEEE COMPUTER SOCIETY
Bangalore Chapter

IAS IEEE INDUSTRY APPLICATIONS SOCIETY
Linking Research to Practice
Bangalore Chapter