تایید هویت دستگاه های بیسیم بااستفاده از اثر انگشت فرکانس رادیویی با مدل بندی الگوریتم های یادگیری ماشین

حسین درفکی [1] *، سعید سرآبادان [2]، میثم رئیس دانایی [3]

1-*کارشناسی ارشد برق-مخابرات سیستم ، دانشکده رادار، سونار، دانشگاه جامع امام حسین(ع)، hossein_dorfaki@yahoo.com
2- دکتری ریاضی، دانشکده علوم، دانشگاه جامع امام حسین(ع)، s.sarabadan@yahoo.com
3- دکتری برق ، دانشکده رادار، سونار و لیدار، دانشگاه جامع امام حسین(ع)، mraeesdanaee@ihu.ac.ir

چکیده

در این مقاله سعی گردیده تا تغییر هویت دستگاه‌های مجاز از دستگاه‌های غیرمجاز که تلاش در تغییر هویت خود جهت ورود به شبکه ی محلی را دارند شناسایی کرده و از ورود دستگاه‌های غیر مجاز به این شبکه امن جلوگیری شود.امروزه روش‌های احراز هویت [3] متعددی مانند احراز هویت مبتنی بر رمز عبور، احراز هویت مبتنی بر گواهی و .. وجود دارد[1]. روش ارائه شده ، احراز هویت بر اساس اثرانگشت فرکانس رادیویی دستگاه های بیسیم و مدل بندی آنها با الگوریتم های یادگیری ماشین [4] می باشد . در این پژوهش از 8000 داده [5] که دارای 204 ویژگی [6] می باشند، توسط دستگاه گیرنده رادیویی Hack RF One در مرکز شهید باقری سازمان جهاد خودکفایی سپاه انجام پذیرفته، استفاده گردیده و سپس از سه الگوریتم یادگیری ماشین ، Support vector Machin و K-Nearest Neighbors و Logistic Regression برای مدل بندی داده ها و تایید احراز هویت استفاده شده است.در آخر با 1000 داده تست مدل ها مورد ارزیابی قرار گرفته و با یکدیگر مقایسه شده اند و بهترین مدل انتخاب شده است.نتایج بدست آمده در بهترین حالت در الگوریتم SVM حدود 87% و احراز هویت در KNN حدود64% و احرار هویت در Logistic حدود65% بدست آمده است. زبان برنامه نویسی استفاده شده پایتون [7] و کد ها در گیت هاب . . . موجود می باشد. با توجه به تفاوت ها و خصوصیات فیزیکی یک دستگاه در زمان ساخت ،که باعث تغییر در فرکانس، فاز و دامنه سیگنال های ارسالی می شود این روش موثر بوده و می تواند دستگاه های غیر مجاز را با احتمال حدود87% از دستگاه های غیر مجاز متمایز نماید.

3/5 cm

3/5 cm

[3] Authentication
[4] Machine Learning
[5] Data
[6] Features
[7] Python

واژگان کلیدی: *احراز هویت، اثر انگشت فرکانس رادیویی، هوش مصنوعی ، یادگیری ماشین،*

Authentication of wireless devices using radio frequency fingerprints by modeling machine learning algorithms

Abstract:

In this article, an attempt has been made to identify the identity change of authorized devices from unauthorized devices that are trying to change their identity to enter the local network and to prevent unauthorized devices from entering this secure network. Today, there are many authentication methods such as authentication based on There are passwords, certificate-based authentication, etc.[1,2]. The presented method is authentication based on radio frequency fingerprints of wireless devices and their modeling with machine learning algorithms. In this research, 8000 data which have 204 features were collected and used by the Hack RF One radio receiver in the Shahid Bagheri center of the IRGC self-sufficiency jihad organization, and then three machine learning algorithms, Support vector Machin, K-Nearest Neighbors, and Logistic Regression has been used for data modeling and authentication verification. Finally, with 1000 test data, the models have been evaluated and compared with each other, and the best model has been selected. The results obtained in the best case in the SVM algorithm are about 87 % and authentication in KNN is about 64% and authentication in Logistic is about 65%.

The programming language used is Python and the codes in GitHub ؟؟؟ is available

According to the differences and physical characteristics of a device at the time of construction, which causes changes in the frequency, phase and amplitude of the transmitted signals, this method is effective and can detect unauthorized devices with a probability of about 87% of unauthorized devices Distinguish.

Keywords: authentication, radio frequency fingerprint, artificial intelligence, machine ,learning

1- مقدمه

ازآنجایی‌که استفاده از دستگاه‌های بیسیم در حال گسترش و ادغام شدن در جنبه‌های مختلف زندگی ما هستند، ایمن‌سازی آن‌ها در برابر حملات سایبری بسیار مهم تلقی می شود. احراز هویت هر دستگاه بیسیمی را می‌توان با استفاده از یک یا چند روش زیر اجرا کرد: (الف) رمزهای عبور، (ب) کلیدهای از قبل به اشتراک گذاشته‌شده (ج) دستگاه‌های رمزنگاری کلید عمومی.

یکی از اولین رویکردها برای احراز هویت دستگاه‌های بیسیم استفاده از یک شناسه منحصربه‌فرد، مانند آدرس MAC[1] یا آدرس IP، برای شناسایی و احراز هویت دستگاه بود. بااین‌حال، این روش در برابر جعل و سایر حملات آسیب‌پذیر است.

با افزایش تعداد دستگاه‌ها، نگرانی‌های امنیتی جدیدی ظاهرشده است. به‌عنوان‌مثال، کارخانه‌هایی که از دستگاه‌های بیسیم در یک شبکه محلی برای دستیابی به راندمان بالاتر در تولید استفاده می‌کنند و به آن‌ها وابسته هستند، اگر یک حمله مشخص، باعث اختلال در عملکرد کارخانه شود، با ضرر بزرگی مواجه خواهند شد. این مسائل زمانی که دستگاه‌های امنیتی سنتی (به‌عنوان‌مثال، فایروال) بحرانی‌تر و آسیب‌پذیرتر از قبل می‌شوند، بیشتر نمایان می‌گردد.

برای رفع این آسیب‌پذیری‌ها، پروتکل‌های احراز هویت جدیدی توسعه‌یافته‌اند. برای مثال، پروتکل (LwM2M)[2] برای مدیریت دستگاه و جمع‌آوری داده استفاده می‌شود و شامل ویژگی‌های امنیتی مانند احراز هویت متقابل و رمزگذاری است. پروتکل دیگری که معمولاً استفاده می‌شود MQTT[3] است[1] . این پروتکل دستگاه‌ها را قادر می‌سازد تا با استفاده از مدل انتشار-اشتراک با یکدیگر و با سرورهای مرکزی ارتباط برقرار کنند و از طریق استفاده از گواهی‌ها و کلیدها از ارتباط امن پشتیبانی می‌کند.

احراز هویت دستگاه‌های بیسیم در سال‌های اخیر به چالشی حیاتی تبدیل‌شده است و پروتکل‌ها و فناوری‌های مختلفی برای رفع این چالش توسعه‌یافته‌اند. هدف این راه‌حل‌ها ارائه روش‌های احراز هویت امن و قابل‌اعتماد برای دستگاه‌های بیسیم است که به آن‌ها امکان می‌دهد به‌طور مؤثر و کارآمد عمل کنند و درعین‌حال خطر حملات سایبری را به حداقل برسانند. به‌طورکلی می‌توان گفت احراز هویت، فرآیند تأیید هویت یک کاربر یا دستگاه است. در زمینه ارتباطات بیسیم احراز هویت تضمین می‌کند که فقط دستگاه‌های مجاز بتواند به داده‌های حساس دسترسی داشته باشد یا دستگاه‌های دیگر را در شبکه کنترل کند. روش‌های سنتی احراز هویت، مانند احراز هویت مبتنی بر رمز عبور، به دلیل قدرت پردازش و حافظه محدود برای دستگاه‌های بیسیم مناسب نیستند.

انگشت‌نگاری RF روشی است که از ویژگی‌ها منحصربه‌فرد سیگنال‌های RF برای شناسایی دستگاه‌ها استفاده می‌کند. هر دستگاه سیگنال‌های RF را منتشر می‌کند که منحصر به آن دستگاه است. این سیگنال‌ها می‌توانند به‌عنوان اثرانگشت برای احراز هویت دستگاه استفاده شوند. الگوریتم‌های هوش مصنوعی می‌توانند این اثرانگشت RF را تجزیه‌وتحلیل کنند تا تشخیص دهند که آیا دستگاه معتبر است یا خیر.

[1] Media Access Control
[2] Lightweight M2M
[3] Message Queuing Telemetry Transport

2- اثرانگشت فرکانس رادیویی

اثرانگشت فرکانس رادیویی به ویژگی‌های منحصربه‌فردی اشاره دارد که می‌توان از آن‌ها برای شناسایی و تمایز دستگاه‌های بی‌سیم مانند گوشی‌های هوشمند، دستگاه‌های اینترنت اشیا و روترهای Wi-Fi بر اساس سیگنال‌های فرکانس رادیویی ارسالی استفاده کرد. این اثرانگشت از ویژگی‌های فیزیکی، اجزای داخلی و پارامترهای عملیاتی دستگاه‌ها گرفته‌شده است.هنگامی‌که یک دستگاه بی‌سیم از طریق شبکه‌های بی‌سیم ارتباط برقرار می‌کند، سیگنال‌های فرکانس رادیویی را منتشر می‌کند که دارای ویژگی‌های متمایز بر اساس عواملی مانند طراحی آنتن دستگاه، ویژگی‌های تقویت‌کننده قدرت[1]، اجزای الکترونیکی، سیستم‌عامل[2] و تنظیمات نرم‌افزار است. این ویژگی‌ها باعث تغییر در قدرت سیگنال[3]، مدولاسیون[4]، فاز[5] و سایر پارامترهای فیزیکی می‌شود با تجزیه‌وتحلیل این اثرانگشت، می‌توان دستگاه‌های بی‌سیم را شناسایی و طبقه‌بندی کرد. این تجزیه‌وتحلیل را می‌توان برای اهداف مختلفی ازجمله تشخیص دستگاه بی‌سیم، ردیابی مکان، امنیت و احراز هویت و بهینه‌سازی شبکه استفاده کرد. یکی از روش های متداول در تحلیل داده های ذکر شده برای ابعاد بالا ، استفاده از الگوریتم های یادگیری ماشین جهت مدل بندی این اثر انگشت ها می باشد.

2-1- روش های رایج استفاده از اثر انگشت RF

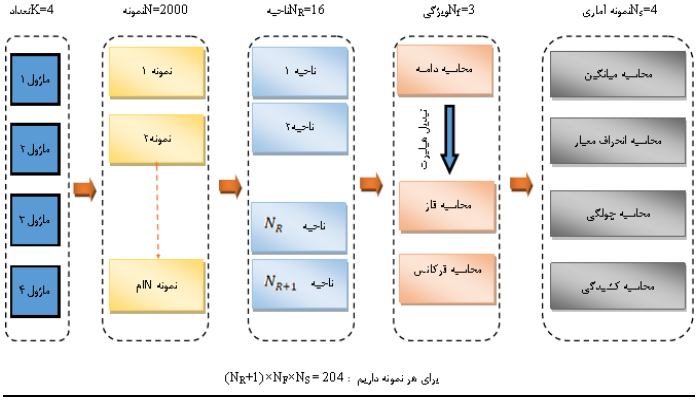
با توجه به داده های دریافتی در روش های اثر انگشت فرکانس رادیویی می توان این روش ها را به صورت زیر دسته بندی نمود. اثر انگشت RF بر اساس اطلاعات مکان[2] ، بر اساس سیگنال انتقال[3] ، بر اساس خطای مدولاسیون RF ، بر اساس مدل سازی فیزیکی[].

2-2- استفاده از هوش مصنوعی در اثر انگشت فرکانس رادیویی

در سال 2018، مرچند و همکاران[4] از شبکه عصبی کانولوشنال (CNN) برای طبقه‌بندی سیگنال‌های بی‌سیم برای شناسایی دستگاه‌های اینترنت اشیا با دقت تشخیص 92.29 درصد در هفت دستگاه ZigBee و استحکام کانال بالا استفاده کردند. در ادامه پژوهش[5] یک الگوریتم شناسایی منبع تابش مبتنی بر شبکه عصبی را پیشنهاد شد که می‌توانست بر روی دستگاه‌های بیسیم با محدودیت منابع اجرا شود و قادر به عملکرد تشخیص بهتری شود. برای حل مشکل شناسایی مربوطه در این مورد پیشنهاد شده است، ترکیب‌های مناسب از دست دادن آنتروپی متقاطع[6]، تلفات مرکزی[7] و تلفات بازسازی و همچنین فضای متریک فاصله مناسب برای یادگیری نمایش فضای ویژگی معنایی سیگنال، معرفی شده به طوری که ویژگی‌های معنایی حداقل فاصله بین کلاسی بزرگ‌تری نسبت به حداکثر فاصله درون کلاسی داشته باشند.[6] در 2021، [7]از سیگنال های جمع آوری شده واقعی ADS-B برای ساخت مجموعه داده، استفاده و مدل یادگیری عمیق برروی داده ها اعمال گردید و نتایج را با معیارهای ارزیابی مورد بررسی قرار دادند، که از همگرایی[8] بالایی برخوردار بود.

[1] Power Amplifier
[2] operating system
[3] Signal strength
[4] Modulation
[5] Phase
[6] Cross entropy
[7] Central losses
[8] Convergence

در این پژوهش سعی گردیده با جمع آوری یک مجموعه داده که از فرآیند تولید اثرانگشت فرکانس رادیویی و توسط دستگاه HACK RF ONE که در مرکز شهید باقری جمع‌آوری گردید و با استفاده از الگوریتم های یادگیری ماشین تجزیه و تحلیل سیگنالهای رادیویی ساطع‌شده انجام پذیرد و فرآیند احراز هویت انجام گردد. این مدل مورد ارزیابی با داده های تست قرار گرفت و نتایج مورد قبولی حاصل گشت که در فصل چهار به آن پرداخته می شود.



2-3- استخراج ویژگی و احراز هویت

در حال حاضر با استخراج ویژگی های مختلف فرکانس رادیویی و روش های متفاوت اقدام به احراز هویت دستگاه های بیسیم می شود[8] . که عبارت اند از: 1- قدرت سیگنال انتقال 2- خطای مدولاسیون RF 3- مدل سازی فیزیکی 4- سیگنال انتقال و...[9]

در این پژوهش با استفاده از ویژگی های آماری سیگنال انتقال که منجر به یک بانک اطلاعاتی گسترده می گردد و استفاده از روش های مختلف یادگیری ماشین، سعی در یافتن مدلی معتبر جهت احراز هویت دستگاه های بیسیم نموده.

3- الگوریتم یادگیری ماشین در اثر انگشت فرکانس رادیویی

الگوریتم های های یادگیری ماشین به دو خانواده بزرگ با ناظر [1] و بدون ناظر [2] تفکیک می شود ،که به واسطه داشتن برچسب در هر آزمایش این تفکیک انجام می پذیرد. همچنین الگوریتم های یادگیری ماشین با ناظر به دو دسته رگرسیون و طبقه بندی تقسیم می شوند. دراین پژوهش با توجه به نوع داده های ما که ویژگی های استخراج شده از فرکانس رادیویی با استفاده از تحلیل های آماری برای یک دستگاه خاص می باشد از روش طبقه بندی در الگوریتم های ماشین استفاده می شود. الگوریتم های استفاده شده در این پژوهش شامل: لجستیک رگرسیون [3]،k نزدیک‌ترین همسایه [4] و ماشین بردار پشتیبان [5] می باشد که با روش های ارزیابی در طبقه بندی مورد مقایسه قرار گرفته اند. در ذیل به شرح مختصری از این الگوریتم های می پردازیم.

3-1- لجستیک رگرسیون

این الگوریتم برای عملیات طبقه‌بندی دوکلاسه استفاده می‌شود،که در آن مدل یاد می‌گیرد احتمال یک ورودی متعلق به یک کلاس خاص را پیش‌بینی کند. رگرسیون لجستیک یک مدل ریاضی را آموزش می‌دهد که ویژگی‌ها ورودی را با احتمال کلاس مثبت ترسیم می‌کند. این کار، با برازش یک تابع سیگموئید به داده‌ها به دست می‌آید.

3-2- k نزدیک‌ترین همسایه

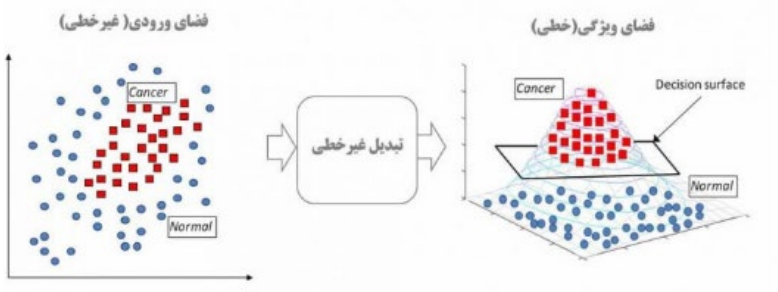
[1] Supervis
[2] Unsupervisor
[3] Logistic Regression
[4] K-Nearest Neighbors
[5] Support Vector Machine

این الگوریتم بر اساس این فرض طراحی شده که چیزهای مشابه در نزدیکی یکدیگر وجود دارند، این الگوریتم نمونه ها را در مرحله آموزش ذخیره می کند و تا زمانی که نمونه های آزمایشی دریافت نشود کاری انجام نمی دهد.
پارامتر کا نقش مهمی در این الگویتم ایفا می کند چرا که مقادیر کا مختلف به نتایج دسته بندی متفاوت منجر می شود[10] .

3-3- ماشین بردار پشتیبان

الگوریتم ماشین های بردار پشتیبان برای مسائل دسته بندی، رگرسیون و تشخیص نقاط دورافتاده استفاده می شود. در این الگوریتم مرز تصمیم با حاشیه ای حداکثری بین کلاس ها ایجاد می شود. در SVM بهترین خط برای دسته بندی نقاط با بیشترین حاشیه انتخاب می گردد، الگوریتم ماشین بردار پشتیبان بر اساس یک مسئله بهینه سازی است که به دو صورت دوگانه و اولیه قابل تعریف است، فرم اولیه زمانی ترجیح داده می شود که نیازی به اعمال ترفند هسته [1] برای داده ها نداشته باشیم و مجموعه داده بزرگ باشد، اما ابعاد هر نقطه داده کوچک باشد. در مقابل، هنگامی که داده ها ابعاد بزرگی دارند و ما نیاز به استفاده از ترفند هسته داریم ، فرم دوگانه ترجیح داده می شود. در این پژوهش با توجه به ماهیت داده از شکل دوم یعنی استفاده از هستها صورت می پذیرد. انواع توابع هسته در SVM :

هسته لاپلاسین : فرمول $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right)$	هسته چند جمله ای : فرمول $k(x_i, x_j) = (x_i . x_j)^d$
هسته گوسی: فرمول $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$	هسته سیگموئید: فرمول $(x_i, x_j) = \tan(ax^T y + c)$



3-4- معیار های ارزیابی مدل طبقه بندی

برای ارزیابی عملکرد یک مدل طبقه‌بندی،از مقایسه پیش‌بینی‌های آن در برابر برچسب‌های واقعی،در یک مجموعه داده استفاده می‌شود. معیارهای متفاوتی در این زمینه وجود دارد که از مهم ترین آنها می توان به دقت (Accuracy) و صحت(Precision) و امتیاز F1 (میانگین هارمونیک) اشاره کرد، که ازمولفه های ماتریس برهم ریختگی [2] استفاده می شود.

4- نتایج

نتایج بدست آمده از این پژوهش را در دو بخش ارائه نموده ایم: در بخش اول، تمرکز برروی مدل های بکار رفته روی دیتا ها و همچنین پارامترهای موثر در این مدل ها و نتایج حاصل از آنها می باشد.در بخش دوم با استفاده از معیارهای ارزیابی در مدل

[1] kernel
[2] Confusion matrix

های طبقه بندی و استفاده از داده های تست که قبلا مدل ها آنها را ندیده اند، توانسته ایم بهترین مدل با بهینه ترین پارامتر را انتخاب کنیم. این نتایج با موارد مشابه این موضوع [11] مقایسه گردیده و از صحت لازم برخوردار است.

4-1- بررسی مدل ها روی داده های اثر انگشت فرکانس رادیویی

در این پژوهش تعداد 10000 نمونه ی آزمایشی با 204 ویژگی در نظر گرفته شده است که 10 درصد از این داده ها به عنوان داده های تست انتخاب شده است. مدل های یادگیری ماشین بکار رفته عبارت اند از لجستیک رگرسون،کا نزدیک ترین همسایگی و ماشین بردار پشتیبان می باشد. برای آموزش این مدل ها از 9000 نمونه ی آزمایشی تعداد 7500 نمونه را جهت آموزش ماشین و 1500 نمونه را جهت ارزیابی[1] استفاده نموده ایم. در هرکدام از این مدل ها هایپر پارامترهایی وجود دارد که با توجه به داده های موجود می تواند مدل را بهینه تر و یا برعکس گمراه نماید. به همین دلیل جداولی تهیه گردیده و با پارامترهای مختلف و مقایسه ی آنها، بهترین مدل با پارامتر مربوطه انتخاب گردیده است. در مدل لجستیک رگرسیون از الگوریتم های بهینه سازب newton-cholesky ب newton-cg ب liblinear ب lbfgs استفاده شد و نتایج به صورت زیر می باشد:

LogisticRegression (*penalty='l2', max_iter=200*)	Algorithm to use in the optimization problem are ب newton-cholesky ب newton-cg ب liblinear ب lbfgs		
	precision	**recall**	**f1-score**
M1	0.50	0.72	0.59
M2	0.65	0.47	0.54
M3	0.77	0.67	0.72
M5	0.68	0.65	0.67
macro avg	0.65	0.63	0.63
weighted av	0.65	0.63	0.63
accuracy	0.65		

در مدل **k نزدیک‌ترین همسایه** از الگوریتم های بهینه ساز ب kd_tree ب ball_tree ب auto استفاده شد و متریک در نظر گرفته شده ، متر اقلیدسی می باشد و نتایج مشابه جدول قبل می باشد.

در آخر از مدل **ماشین بردار پشتیبان** استفاده شد و با در نظر گرفتن کرنل های ب sigmoid ب rbf ب poly ب linear و مفادیر C متفاوت توانستیم یک مدل مناسب با همگرایی مورد قبول را انتخاب کنیم.

Support Vector Classification (*C=1.0,2.0,3.0*)	kernel type to be used in the algorithm are ب sigmoid ب rbf ب poly(3,4) ب linear		
	precision	**recall**	**f1-score**
M1	0.80	0.83	0.81
M2	0.83	0.81	0.82
M3	0.91	0.88	0.89
M5	0.96	0.97	0.96
macro avg	0.87	0.87	0.87
weighted av	0.87	0.87	0.87
accuracy	0.87		

[1] validation

با توجه به اینکه تعداد نمونه ها در این دیتا زیاد می باشد ، براین عقیده هستیم که می توان با طراحی یک شبکه عصبی عمیق مناسب ، مدلی با هوشمندی و دقت بالاتر طراحی نمود.

[1] Neven Nikolov; Ognyan Nakov; Daniela Gotseva, "Research of MQTT versus LwM2M IoT communication protocols for IoT," *Communication and Energy Systems and Technologies (ICEST),* 28 July 2021.

[2] Jung Ho Lee; Taehun Kim; Beomju Shin; Changsoo Yu;, "RF Signal Strength Modeling in Indoor Environments for Cost-Effective Deployment of Fingerprinting Technology," *IEEE,* 11 April 2022.

[3] Qi Cheng; Jie Li; Xiaoli Gao; Huaqi Fan; Ting Jian, "Radio Frequency Transmitter Identification Based on Fingerprinting and Convolutional Neural Network," *IEEE,* 18 July 2022.

[4] K. Merchant, S. Revay, G. Stantchev and B. Nousain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE Journal of Selected Topics in Signal Processing,* p. 2018, 160ـ167.

[5] M. McGinthy, L. Wong and A. Michaels, "Groundwork for neural network-based specific emitter identification authentication for IoT," *IEEE Internet of Things Journal,* p. 6429ـ6440, 2019.

[6] Y. Tu, Y. Lin and H. Zha, "Large-scale real-world radio signal recognition with deep learning," *Chinese Journal of Aeronautics,* pp. 1-14, 2021.

[7] Jialan Shen; Jingchao Li; Haijun Wang; Cheng Cong;, "ADS-B Signal Recognition Method Based On Entropy Feature Fusion," *IEEE,* 24 November 2021.

[8] C. Chen, H. Wen, J. Wu, A. Xu, Y. Jiang, H. Song and C. Chen, "Radio Frequency Fingerprint-Based Intelligent Mobile Edge Computing for Internet of Things Authentication," *Sensors,* 2019.

[9] S. L. X. Z. a. L. Z. H. Wen, "A framework of the PHY-layer approach to defense against security threats in cognitive radio networks," *IEEE Netw, pp,* pp. 34-39, 2013..

[10] m. Vozan, Data structures (computer science) -- Mathematical models, tehran, 1400.

[11] ""https://onlinebme.com/svm," 30 02 1400. [Online].

[12] H. Chih-Wei, C. Chih-Chung and L. Chih-Jen, "A Practical Guide to Support Vector Classification," *Department of Computer Science,* May 19, 2016.

[13] j. hall;, M. Barbeau and E. Kranakis, "Detection of transient in radio frequency fingerprinting using signal phase. Wireless and Optical Communications," pp. 8-13, 2003.

[14] D. Reising, C. Joseph , k. Farah and L. Daniel , *Pre-print: Radio Identity Verification-based IoT,* 2020.