# بررسی چند روش ترکیبی در سیستم‌های تشخیص نفوذ و ارائه ترکیب‌های جدید

مهناز زارع[۱]، نفیسه موسی الرضایی گلیان[۲]، مرجان ناخدا[۳]

[۱] دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه بین المللی امام رضا، مشهد، ایران
Mahnaz.zare86@gmail.com

[۲] دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه بین المللی امام رضا، مشهد، ایران
n.moosarrezayi@gmail.com

[۳] دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه بین المللی امام رضا، مشهد، ایران
marjannakhoda@gmail.com

## چکیده

**با توجه به گسترش روز افزون شبکه‌های کامپیوتری و تسهیل حمله و نفوذ به آن‌ها به دلیل وجود ابزارهایی که دانش مورد نیاز برای حمله‌ها را کاهش می‌دهد، روزانه شاهد حملات فراوانی در سطح شبکه‌ها هستیم و سیستم‌های امنیتی نظیر فایروال‌ها و آنتی‌ویروس‌ها و غیره برای جلوگیری از این خطرات کافی نیست. یکی از روش‌های مقابله با چنین حملاتی استفاده از سیستم‌های تشخیص نفوذ می‌باشد و در میان روش‌های موجود تشخیص نفوذ، استفاده از روش‌های یادگیری ماشین، آن هم به صورت ترکیبی، به دلیل مزایای آنان بیشتر مورد استقبال قرار گرفته است. هدف از این پژوهش، بررسی و مقایسه چند روش ترکیبی موجود و ارائه دو روش ترکیبی جدید می‌باشد که در انتها نیز به مقایسه آن‌ها می‌پردازیم.**

## کلمات کلیدی

**تشخیص نفوذ، سوءاستفاده، ناهنجاری، مدل‌های ترکیبی، درخت تصمیم، ماشین بردار پشتیبانی یک کلاسه (SVM)، خوشه‌بندی، یادگیری ماشین.**

## ۱- مقدمه

برای ایجاد امنیت کامل در یک سیستم کامپیوتری، علاوه بر دیواره‌های آتش و دیگر تجهیزات جلوگیری از نفوذ، سیستم‌های دیگری به نام سیستم‌های تشخیص نفوذ[۱] مورد نیاز می‌باشند تا بتوانند در صورتی که نفوذگر از دیواره‌ی آتش، آنتی‌ویروس و دیگر تجهیزات امنیتی عبور کرد و وارد سیستم شد، آن را تشخیص داده و چاره‌ای برای مقابله با آن بیاندیشند. سیستم‌های تشخیص نفوذ را می‌توان از سه جنبه‌ی معماری، نحوه‌ی پاسخ به نفوذ و روش تشخیص طبقه‌بندی کرد. انواع مختلفی از معماری سیستم‌های تشخیص نفوذ وجود دارد که به طور کلی می‌توان آن‌ها را در سه دسته‌ی مبتنی بر میزبان(HIDS)، مبتنی بر شبکه (NIDS) و توزیع‌شده (DIDS) تقسیم‌بندی نمود [1].

الگوریتم‌های تشخیص نفوذ[۲] از جنبه روش تشخیص، در دو دسته کلی طبقه‌بندی می‌شوند: تشخیص سوءاستفاده[۳] و تشخیص ناهنجاری[۴]. الگوریتم‌های تشخیص سوءاستفاده حمله‌ها را برمبنا امضای حملات شناخته شده شناسایی می‌کنند. با وجود اینکه این نوع از سیستم‌های تشخیص نفوذ در تشخیص حمله‌های شناخته شده با درصد خطای پایین کارآمد هستند، اما نمی‌توانند حمله‌های جدیدی که ویژگی و خصوصیت مشابه‌ای با حمله‌های شناخته شده ندارند را شناسایی کنند. درمقابل، الگوریتم‌های تشخیص ناهنجاری برمبنا این فرضیه هستند که رفتار حمله کننده با رفتار یک کاربر نرمال متفاوت است؛ از این رو ترافیک‌های نرمال را آنالیز کرده و الگوهای

ترافیک نرمال را ایجاد می‌کنند. حال ترافیک ورودی را به عنوان حمله در نظر می‌گیرند اگر خصوصیات آن با الگوهای ترافیک نرمال متفاوت باشد. با وجود اینکه که الگوریتم‌های تشخیص ناهنجاری برای شناسایی حمله‌های جدید مناسب هستند اما در تشخیص حمله‌های شناخته شده به اندازه مدل‌های تشخیص سوء استفاده کارآمد نیستند[2]. به منظور حل معایب این دو روش تشخیص نفوذ معمولی، روش‌های تشخیص نفوذ ترکیبی از دو روش فوق ارائه شده است.

در اغلب سیستم‌های تشخیص نفوذ ترکیبی یک مدل تشخیص سوءاستفاده و یک مدل تشخیص ناهنجاری به طور مستقل آموزش دیده و سپس نتایج آن‌ها با هم جمع می‌شوند. به عنوان مثال، بعضی از سیستم‌های تشخیص نفوذ ترکیبی ترافیک ورودی را بعنوان حمله درنظر می‌گیرند اگر حداقل یکی از دو مدل تشخیص دهد که این ترافیک ارتباطی حمله است و یا سیستم‌های تشخیص ترکیبی دیگری هستند که ترافیک ورودی را به عنوان حمله در نظر می‌گیرند اگر هر دو روش تشخیص بدهند که ترافیک ورودی یک حمله است. در حالت اول سرعت کشف و شناسایی بهبود می‌یابد اما سیستم تشخیص نفوذ هنوز نرخ مثبت کاذب[5] بالایی خواهد داشت در حالی‌که در حالت دوم پیغام کاذب کاهش خواهد یافت اما ممکن است بسیاری از ترافیک‌های حمله را نادیده بگیرد [3].

ما در این تحقیق به جای بررسی روش‌هایی که فقط نتایج دو مدل تشخیص نفوذ را ترکیب می‌کنند، روش‌های ترکیب سلسله مراتبی را مورد مطالعه قرار داده‌ایم. در مقالات [3] و [4] دو نمونه جدید از مدل‌های ترکیبی بیان شده است که در ادامه توضیح داده می شوند. در [3] از درخت تصمیم و 1-class SVM[6] و در [4] از فیلترینگ و خوشه بندی به همراه 1-class SVM استفاده شده است.

در این مقاله در بخش دوم مروری بر کارهای گذشته و مقالات اشاره شده در بالا خواهیم داشت. در بخش سوم، حالات ممکن برای ترکیب روش‌های تشخیص نفوذ به کار برده شده در این دو مقاله را مورد بررسی قرار می‌دهیم. بعد از آن به ارزیابی ترکیب‌های پیشنهادی پرداخته و در انتهای مقاله نتیجه‌گیری را خواهیم داشت.

## ۲- کارهای مرتبط

تحقیقات گسترده‌ای در زمینه متدهای تشخیص نفوذ ترکیبی در جهت غلبه بر محدودیت‌های روش‌های تشخیص سوءاستفاده و ناهنجاری صورت گرفته است. در کل، در این تحقیقات از سه روش متفاوت برای ترکیب مدل تشخیص نفوذ سوءاستفاده و مدل تشخیص نفوذ ناهنجاری استفاده کرده اند: ۱- تشخیص ناهنجاری و سپس تشخیص سوءاستفاده، ۲- اعمال تشخیص ناهنجاری و تشخیص سوءاستفاده بصورت موازی و ۳- تشخیص سوءاستفاده و سپس تشخیص ناهنجاری.

در سال ۲۰۰۱، Barbara, Couto, Jajodia, Popyack, and Wu روش آنالیز داده و داده‌کاوی‌ای پیشنهاد دادند که در آن تشخیص سوءاستفاده به دنبال تشخیص ناهنجاری اعمال می‌شود. در این روش ابتدا مدل تشخیص ناهنجاری با استفاده از قوانین داده‌کاوی اتصالات مشکوک را تعیین می‌کند و سپس به یک مدل تشخیص سوءاستفاده جهت دسته بندی اتصالات به عنوان اتصال نرمال، اتصال حمله شناخته شده و یا اتصال حمله ناشناخته ارسال می‌شود.

روش‌های ترکیبی موازی از یک مدل تشخیص ناهنجاری و یک مدل تشخیص سوءاستفاده بصورت موازی استفاده می‌کنند. در سال ۲۰۰۵ Depren et al. مدل ترکیبی موازی را ارائه کرد که در آن هر مدل تشخیص نفوذ بصورت مستقل آموزش دیده و سپس یک سیستم پشتیبان تصمیم[7] نتایج دسته‌بندی هرکدام را با هم ترکیب می‌کند.

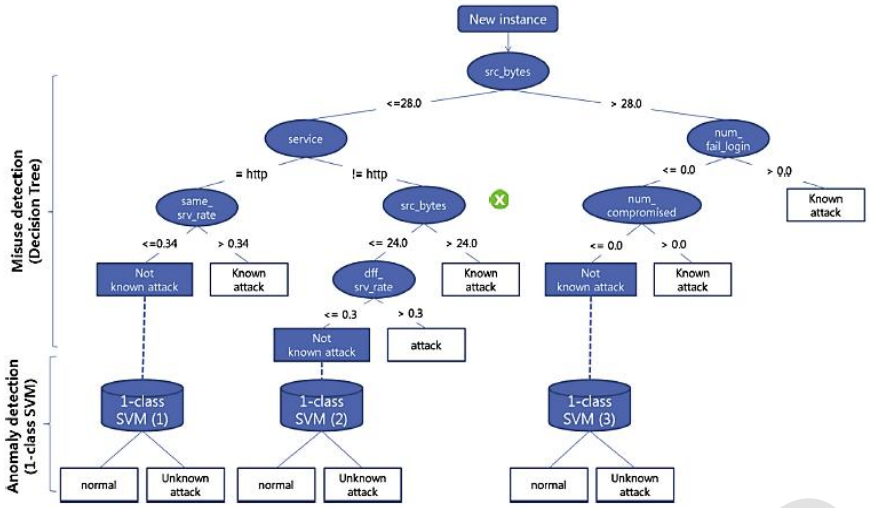
Hwang, Chen, and Qin در سال ۲۰۰۷، از روش تشخیص سوءاستفاده که توسط روش تشخیص ناهنجاری دنبال می‌شود برای طراحی یک سیستم تشخیص نفوذ ترکیبی استفاده کرد. به خاطر اینکه مدل تشخیص سوءاستفاده می‌تواند حمله‌های شناخته شده را با نرخ مثبت کاذب پایین تشخیص دهد و همچنین سریع‌تر از مدل تشخیص ناهنجاری عمل می‌کند؛ در ابتدا برای تشخیص حمله‌های شناخته شده استفاده شده و سپس مدل تشخیص ناهنجاری فقط بر روی اتصالات مشکوک باقی‌مانده اعمال می‌شود.

یک اصل مهم و چالشی در روش تشخیص ناهنجاری ایجاد پروفایل‌های نرمال است. اگر پروفایل‌ها خیلی بزرگ باشند؛ روش تشخیص ناهنجاری در تشخیص اکثر حملات شکست خورده و منجر به یک نرخ تشخیص پایین می‌شود. از سویی نیز اگر پروفایل‌ها بسیار محدود باشند؛ روش تشخیص ناهنجاری می‌تواند تقریبا همه حمله‌ها را شناسایی کند اما اکثر اتصالات نرمال را نیز به عنوان حمله دسته‌بندی می‌کند. که متاسفانه در تمامی روش‌های پیشین توجهی به این موضوع و برطرف کردن آن نکردند [3].

روش‌های ترکیبی مورد بررسی در ادامه (Kim et al. (2014) و Jungsuk Song et al.) و همچنین روش‌های پیشنهادی بر روی کاهش نرخ مثبت کاذب مدل تشخیص ناهنجاری با ترکیب مدل‌های تشخیص بصورت سلسله مراتبی، مجتمع شده در یک ساختار تجزیه شده تمرکز دارند.

در مدل ارائه شده توسط Kim et al. در سال ۲۰۱۴، سیستم تشخیص نفوذ ترکیبی به جای ترکیب نتایج دو روش تشخیص نفوذ، از درخت تصمیم[8] C4.5 برای ایجاد یک مدل تشخیص سوءاستفاده و از یک 1-Class SVM برای ایجاد یک مدل تشخیص ناهنجاری به صورت سلسله مراتبی استفاده می‌کند. در این روش، علاوه بر اصول کلی روش‌های تشخیص نفوذ، داده‌های آزمایشی نرمال با استفاده از روش تشخیص سوءاستفاده به زیر مجموعه‌های گسسته تجزیه شده و سپس بر روی هر قسمت یک روش تشخیص ناهنجاری اعمال می‌شود.

هنگامی که مجموعه داده آموزشی به زیر مجموعه‌های کوچکتر تجزیه می‌شود، زمان‌های آموزش و تست به میزان قابل توجه‌ای کاهش می‌یابند. به طور خاص، زمان تست باید به منظور کاهش سربار الگوریتم تشخیص جهت عمل کردن مدل در یک زمان واقعی مینیمم گردد. خلاصه‌ی مراحل فرایند تست و آموزش و همچنین دیاگرام روش Kim و همکارانش به ترتیب در جدول‌های شماره (۱)، (۲) و شکل شماره (۱) آمده است.




**شکل (۱): شمای کلی روش پیشنهادی Kim [3]**

**جدول (۱): فرایند آموزش روش تشخیص نفوذ ترکیبی Kim**

| | |
|---|---|
| ۱- | فراهم کردن یک مجموعه داده آموزشی شامل داده‌های نرمال و داده‌های حمله شناخته شده. |
| ۲- | ساخت یک مدل تشخیص سوء استفاده با استفاده از الگوریتم درخت تصمیم بر اساس مجموعه داده آموزشی. |
| ۳- | تجزیه کردن داده‌های آموزشی نرمال به زیر مجموعه‌های وابسته به ساختار درخت تصمیم.<br>داده‌های برگ‌های یکسان، به یک زیرمجموعه مشابه تعلق دارند. |
| ۴- | برای هر برگ نرمال از درخت تصمیم، ایجاد یک مدل تشخیص ناهنجاری با استفاده از الگوریتم 1-class SVM بر اساس یک زیرمجموعه داده نرمال برای برگ. |

**جدول (۲): فرایند تست روش تشخیص نفوذ ترکیبی Kim**

| | |
|---|---|
| ۱- | دریافت اتصالات ورودی و استخراج ویژگی‌های ترافیک از اتصال. |
| ۲- | بررسی اتصال به وسیله یک درخت تصمیم آموزش دیده با استفاده از ویژگی‌های استخراج شده ترافیک و تعیین اینکه اتصال حمله شناخته شده است یا نه؟ |
| ۳- | اگر درخت تصمیم اتصال را به عنوان حمله شناخته شده دسته‌بندی کرد، اتصال را مسدود کرده و به گام ۶ می‌رود؛ در غیر اینصورت، درخت تصمیم برگی را که اتصال متعلق به آنست را بررسی کرده و به گام ۴ می‌رود. |
| ۴- | بررسی اتصال با یک 1-class SVM آموزش دیده برای برگ مربوطه، با استفاده از ویژگی‌های ترافیک استخراج شده به منظور تشخیص اینکه آیا اتصال حمله ناشناخته است یا نه؟ |
| ۵- | اگر 1-class SVM اتصال را به عنوان یک اتصال پرت تشخیص داد، اتصال را مسدود کرده و یک آلارم به مدیر امنیت می‌فرستد؛ در غیر این صورت اجازه اتصال به شبکه حفاظت شده را می‌دهد. |
| ۶- | منتظر ورود اتصال بعدی می‌ماند. |

در روش Jungsuk Song et al. در سال ۲۰۱۳، فاز آموزش از سه مرحله اصلی تشکیل شده است: فیلترینگ، کلاسترینگ و مدلینگ. شکل (۲) شمای کلی روش و مراحل آن را نشان می‌دهد.

فاز آموزش ابتدا بیشتر داده‌های حمله را از داده‌های نرمال فیلتر می‌کند تا تضمین کند که بخش عمده هر کلاستر بدست آمده در مرحله بعد (کلاسترینگ) شامل داده‌های نرمال است و داده‌های آموزشی فیلتر شده را به k کلاستر که نشان دهنده داده‌های نرمال است تقسیم می‌کند. این خیلی مهم است که داده‌های ترافیک در خوشه‌های منحصر بفرد خوشه‌بندی شود تا دقت شناسایی فعالیت‌هایشان بالاتر رود.

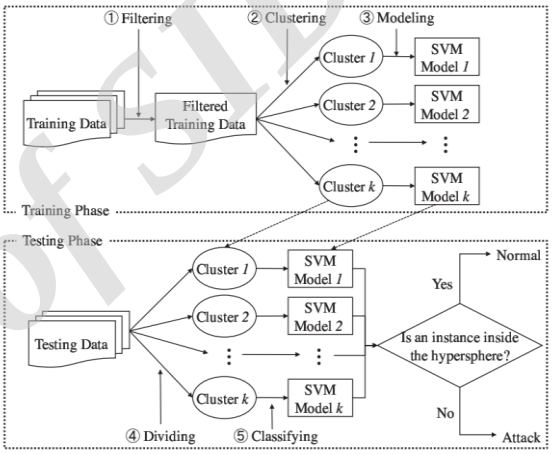
سپس SVM یک کلاسه به هر خوشه نرمال اعمال می‌شود و k مدل SVM در حالی‌که داخل آن‌ها فضاهای نرمال در نظر گرفته شده است، به صورت متوالی بدست می‌آید. SVM یک کلاسه یکی از معمول‌ترین متدهای هدایت نشده است و توانایی پیدا کردن ابر کره را دارد که در آن بیشتر داده‌ها قرار می‌گیرند. در این حالت همه داده‌هایی که در ابرکره قرار می‌گیرند می‌توانند به عنوان نرمال در نظر گرفته شوند زیرا بیشتر کره داده نرمال است.

در فاز تست ابتدا فاصله هر داده با مرکز k کلاستر تولید شده در مرحله آموزش محاسبه می‌شود و هر داده به نزدیک‌ترین خوشه نسبت داده می‌شود. داده نسبت داده شده به عنوان ورودی SVM متناظری که در فاز آموزش ساخته شده است در نظر گرفته می‌شود. اگر داده در محدوده نرمال SVM قرار گرفت، آن داده به عنوان نرمال در نظر گرفته می‌شود؛ در غیر این صورت حمله است.

بسیاری از داده‌های حمله در ابتدا فیلتر می‌شوند تا این نیاز را که هر کلاستری تعداد کمی حمله را در خود داشته باشد برآورده کند. در این فرآیند فیلترینگ، α که نرخ داده‌های حمله به کل داده‌هاست، نقش مهمی ایفا

می‌کند. در اینجا یک محدودیت مهلک وجود دارد و آن اینکه داده‌های اصلی ما هیچگونه برچسبی ندارند و ما نمی‌توانیم تعداد حملات را بدانیم ($\alpha$) و همچنین تعداد دسته‌های داده‌های نرمال($k$) را نمی‌دانیم. این تعیین نشدن $k$ و $\alpha$ باعث پایین آمدن کارایی می‌شود. این روش یک روش پیکربندی خودکار را برای $k$ و $\alpha$ می‌دهد. مراحل فیلترینگ و کلاسترینگ را در ادامه بررسی می‌کنیم.

اساس فیلتر بر این است که اگر داده‌ای نرمال باشد تعداد زیادی داده دیگر با خصوصیات مشابه آن وجود دارد. بنابراین داده‌ها به دو دسته تقسیم می‌شوند spars و dense (خلوت و متراکم). اگر داده‌ای حداقل یکی از ابعادش در قسمت خلوت باشد (یعنی صفت با تکرار کم) در گروه خلوت طبقه‌بندی می‌شود. در غیر این صورت عضو گروه متراکم می‌شود. گروه متراکم به عنوان داده‌های فیلتر در نظر گرفته می‌شوند و وارد فرآیند مدلینگ می‌شوند (گروه خلوت حذف می‌شود).




**شکل (۲): شمای کلی روش Song [4]**

در[8] در طی عمل خوشه‌بندی، داده‌های فیلتر شده (یعنی گروه متراکم) به $K$ خوشه تقسیم می‌شوند. در این روش $k$ از قبل تعیین شده در نظر گرفته شده است. اما مسئله اصلی این است که در واقع ما نمی‌توانیم تعداد دقیق خوشه‌های موجود در گروه متراکم را تعیین کنیم. برای رویارویی با این محدودیت یک روش جدید خوشه‌بندی توسط [9] ارائه شده که استخراج همه خوشه‌های نرمال را از گروه متراکم بدون از پیش تعیین کردن مقدار $k$ امکان‌پذیر می‌کند. که در آن دو مقدار زیر با مراجعه به کلاستر $C_j$ محاسبه می‌گردد:

$\Delta_j$: میانگین فاصله بین داده‌های کلاستر $C_j$ با مرکز آن $c_j$.

$\sigma_j$: انحراف استاندارد فاصله بین داده در کلاستر $C_j$ و مرکز آن $c_j$.

اما این روش جداسازی را نیز نمی‌توانیم مستقیما روی روش پیشنهادی اعمال کنیم زیرا نرخ داده‌های حمله به داده‌های نرمال در گروه متراکم بسیار پایین است. در بسیاری موارد گروه متراکم تنها شامل داده‌های نرمال است اما [9] فرض کرده هر خوشه شامل داده‌های پرت (یعنی حملات) نیز هست. از آنجایی‌که فاصله بین داده پرت و مرکز خوشه‌اش باعث افزایش $\Delta_j$ و $\sigma_j$ می‌شود، فرض آن‌ها باعث تخمین بیش از حد اندازه خوشه‌های نرمال می‌شود. بنابراین ما باید به دو شرط در طی فرآیند کلاسترینگ توجه کنیم:۱- کاهش معیار جداسازی ۲- جلوگیری از جداسازی بیش از حد.

برای برآورده شدن شرط اول، ما فقط $\Delta_j$ را به عنوان معیار جداسازی در روش پیشنهادی اتخاذ می‌کنیم. به عبارتی دیگر اگر فاصله بین داده و مرکز بزرگ‌تر از $\Delta_j$ بود؛ ما آن را به عنوان مرکز داده کلاستر جدید در نظر می‌گیریم. به عنوان اقدام متقابل برای جداسازی بیش از حد، ما بیشترین $\Delta$ رابه عنوان ملاک جداسازی اتخاذ کردیم که از تمام داده‌های درون گروه متراکم بدست می‌آید. فرآیند روش خوشه‌بندی پیشنهادی به این صورت است:

۱- آغاز: میانگین تمام داده‌های درون گروه متراکم مرکز خوشه اولیه می‌شود.
۲- محاسبه $\Delta$ : میانگین فاصله بین همه داده‌های درون گروه متراکم و مرکز خوشه اولیه.
۳- انتخاب: دورترین داده از بزرگ‌ترین کلاستر از نظر کمی (تعداد) در بین $k$ کلاستر میانی انتخاب می‌شود.
۴- جداسازی: اگر فاصله بین دورترین داده و مرکز کلاسترش بزرگتر از $\Delta$ بود، آن داده به عنوان مرکز خوشه جدید در نظر گرفته می‌شود.
۵- تخصیص: فاصله بین هر داده در گروه متراکم و $k+1$ کلاستر میانی محاسبه می‌شود و داده به نزدیک‌ترین خوشه تخصیص داده می‌شود.
۶- آپدیت کردن: مرکز هر خوشه با میانگین داده‌هایش جایگزین می‌شود.
۷- تکرار: مراحل انتخاب تا آپدیت تکرار می‌شوند تا زمانی‌که هیچ داده‌ای وجود نداشته باشد که ملاک جداسازی را در فاز جداسازی داشته باشد.

## ۳- ترکیب‌های پیشنهادی جدید

### ۳-۱- ابتدا روش Kim سپس Song هم در فاز آموزش و هم درتست

در این مدل ترکیبی، در فاز آموزش، ابتدا از درخت تصمیم بکار برده شده در مقاله kim برای جداسازی حملات شناخته شده استفاده می‌کنیم و سپس با کمک مدل مقاله Song از بین داده‌های باقیمانده، داده‌های نرمال را خوشه‌بندی کرده و بر روی آن‌ها 1-class SVM اعمال می‌شود. در فاز تست نیز، ابتدا با استفاده از درخت تصمیم حملات شناخته شده تشخیص داده می‌شود. اگر داده ورودی جزء حملات شناخته شده نباشد، برای آنالیز بیشتر ابتدا بررسی می‌شود که متعلق به کدام خوشه ایجاد شده در فاز آموزش می‌باشد و سپس بر روی آن یک 1-class SVM اعمال می‌شود. خلاصه مراحل فازهای آموزش و تست به ترتیب در جداول شماره (۳) و (۴) آمده است.

## ۳-۲- ابتدا روش kim سپس Song در آموزش، روش Song در تست

در این روش که فاز آموزش آن مانند روش قبلی است، در مرحله تست بجای روش ارائه شده در مقاله kim، از فاز تست مقاله Song استفاده شده است. جزئیات مراحل در جدول‌های (۳) و (۵) آمده است

در بخش بعدی روش‌های پیشنهادی خود را با روش‌های مقالات دیگر مقایسه می‌کنیم.

# ٤- ارزیابی

## ٤-۱- مقایسه روش‌های پیشنهادی با هم

همانطور که گفته شد روش‌های پیشنهادی دارای فاز آموزش یکسان هستند و تفاوت آنها در فاز تست می‌باشد، بنابراین برای مقایسه این دو روش کافیست فاز تست آنها را با هم مقایسه کنیم.

روش تست در مقاله Song بر این مبنا است که هر داده ورودی را با خوشه‌های ایجاد شده در مرحله آموزش مقایسه کرده و آن را در خوشه مربوطه قرار می‌دهد و سپس با کمک 1-class SVM همان خوشه به نتیجه نهایی در مورد حمله یا نرمال بودن داده ورودی می‌رسد. از آنجایی که در این روش تمامی داده‌ها وارد فرآیند خوشه‌بندی و سپس 1-class SVM می‌شوند؛ هم زمان بیشتری صرف می‌شود و هم دقت 1-class SVM نسبت به روش Song–Kim در فاز تست کمتر است. علاوه بر موارد ذکر شده، از آنجایی که درخت تصمیم با نرخ مثبت کاذب پایین‌تری قادر به شناسایی حملات شناخته شده است؛ روش Song–Kim دقت بالاتری دارد و همچنین نرخ خطا نسبت به فاز تست در مقاله Song کمتر است.

همچنین چون اساس قسمت فیلترینگ روش Song این بود که داده‌هایی را به عنوان حمله در نظر بگیرد که فاصله زیادی از سایر داده‌ها دارند، یعنی از بخش‌های متراکم در فاصله دورتری قرار گرفته‌اند، بنابراین ممکن است در این حالت نرخ منفی کاذب بیشتر از روشی باشد که ابتدا داده‌ها را با درخت تصمیم بررسی کرده و داده‌های حملات شناخته شده را از میان آن‌ها حذف می‌کند و سپس روی داده‌های باقی‌مانده عملیات فیلترینگ را اعمال می‌کند.

## ٤-۲- مقایسه روش‌های پیشنهادی با مقاله Kim در فاز آموزش

در مقاله Kim ابتدا داده‌های آموزشی توسط درخت تصمیم تجزیه شده و بر روی هر برگ نرمال از درخت تصمیم یک1-class SVM اعمال می‌شود. در حالی‌که در روش‌های پیشنهادی علاوه بر تفکیک داده‌های حمله شناخته شده توسط درخت تصمیم، سایر داده‌های حمله نیز با عملیات فیلترینگ جداسازی شده و داده‌های آموزشی به زیر مجموعه‌های کوچک‌تری تجزیه می‌شوند؛ که پیچدگی کمتری دارند و می‌توان کاهش تعداد داده‌های ورودی 1-class SVM که منجر به افزایش دقت 1-class SVM می‌شود را از مزایای روش‌های پیشنهادی بر شمرد.

از سویی در روش‌های پیشنهادی علاوه بر درخت تصمیم و 1-class SVM مراحلی دیگر برای تفکیک‌سازی بیشتر و افزایش دقت نیز اعمال می‌شود که منجر به افزایش زمان آموزش نسبت به روش Kim می‌شود و حال متناسب با شرایط و نیازمندی‌ها باید تعادلی بین سرعت و دقت برقرار شود.

با توجه به پردازش‌های پر هزینه و سربار حافظه، اغلب سیستم‌های تشخیص ناهنجاری برای آنالیزهای آفلاین طراحی می‌شوند. از این رو به منظور عمل کردن مدل تشخیص در زمان واقعی باید پیچیدگی زمانی و حافظه مدل تشخیص ناهنجاری در فاز تست به حداقل رسانده شود. حال با توجه به نتایج بدست آمده تست در روش Song برای کاربردهای بلادرنگ[۹] مناسب‌تر می‌باشد. اما در روش‌های آفلاین به دلیل اینکه پیچیدگی زمانی نسبت به کاربردهای بلادرنگ از اهمیت کمتری برخوردار است؛ پس تست با روش Song–kim می‌تواند نتایج بهتر و دقیق‌تری را برای ما به همراه داشته باشد.

## ٤-۳- ارزیابی نهایی

مقایسه‌ی روش های پیشنهادی با معیار دقت نشان می‌دهد که روش‌های پیشنهادی دقت بالاتری دارند و روش اول از همه دقیق‌تر است، چون در آموزش تعداد کمتری داده وارد SVM می‌شوند و در تست هم ابتدا داده‌های حمله شناخته شده جدا می‌شوند، پس تشخیص‌ها دقیق‌تر می‌شود.

همچنین با توجه به معیار زمان و هزینه، مشاهده می‌شود روش‌های پیشنهادی به دلیل بالا بردن تعداد مراحل مدلینگ، با وجود این‌که دقت را بالا می‌برند اما زمان و هزینه را نیز افزایش می‌دهند و به همین دلیل برای کاربردهای بلادرنگ مناسب نمی‌باشند.

و در نهایت بسته به نرخ داده‌های حمله (نرخ وجود داده‌های حمله نسبت به داده‌های نرمال در مجموعه داده‌ها) برای مجموعه داده‌هایی که داده‌های حملات آن کم است روش Song، برای مجموعه داده‌هایی که داده‌های حملات شناخته شده آن بیشتر است روش Kim و برای سایر مجموعه داده‌ها روش‌های پیشنهادی بهتر هستند.

# ٥- نتیجه گیری

در این مقاله ابتدا دو روش ترکیبی موجود را بررسی کرده سپس دو ترکیب جدید ممکن از آن‌ها را ارائه دادیم. پس از آن روش‌های موجود و ارائه شده را از جنبه‌های مختلف باهم مقایسه کردیم. لازم به ذکر است تصمیم‌گیری برای انتخاب بهترین روش به طور کلی ممکن نیست، چرا که بستگی مستقیم با کاربرد مدل (بلادرنگ یا آفلاین بودن)، ترکیب داده‌های موجود (نرخ وجود حملات در داده‌ها) و... دارد.

**جدول (۳): فرایند آموزش روش تشخیص نفوذ ترکیبی پیشنهادی بخش ۳-۱ و بخش ۳-۲**

| | |
|---|---|
| ۱- | فراهم کردن یک مجموعه داده آموزشی شامل داده‌های نرمال و داده‌های حمله شناخته شده. |
| ۲- | ساخت یک مدل تشخیص سوءاستفاده با استفاده از الگوریتم درخت تصمیم بر اساس مجموعه داده آموزشی. |
| ۳- | تجزیه کردن داده‌های اموزشی نرمال به زیر مجموعه‌های وابسته به ساختار درخت تصمیم. داده‌ها در برگ‌های مشابه به یک زیرمجموعه مشابه تعلق دارند. |
| ۴- | داده‌های هر برگ ابتدا فیلتر شده و داده‌های حمله آن کنار گذاشته می‌شوند و داده‌های نرمال باقی می‌ماند. |
| ۵- | داده‌های نرمال باقی‌مانده از مرحله ۴ خوشه‌بندی می‌شوند. |
| ۶- | روی هر خوشه 1-class SVM را اعمال می‌کنیم. |
| ۷- | اگر 1-class SVM اتصال را به عنوان یک حمله تشخیص داد، اتصال را مسدود کرده و یک آلارم به مدیر امنیت می‌فرستد؛ در غیر اینصورت اجازه اتصال به شبکه حفاظت شده را می‌دهد. |

**جدول (٤): فرایند تست روش تشخیص نفوذ ترکیبی پیشنهادی بخش ۳-۱**

| | |
|---|---|
| ۱- | دریافت اتصالات ورودی و استخراج ویژگی‌های ترافیک از اتصال. |
| ۲- | بررسی اتصال به وسیله یک درخت تصمیم آموزش دیده با استفاده از ویژگی‌های استخراج شده ترافیک و تعیین اینکه اتصال حمله شناخته شده است یا نه؟ |
| ۳- | اگر درخت تصمیم اتصال را به عنوان حمله شناخته شده دسته‌بندی کرد، اتصال را مسدود کرده و به گام ۶ می‌رود؛ در غیر اینصورت، درخت تصمیم برگی را که اتصال متعلق به آنست را بررسی کرده و به گام ۴ می‌رود. |
| ۴- | داده‌های باقی‌مانده در گام ۴ با خوشه‌ها مقایسه شده و هر داده وارد خوشه مربوط به خود می‌شود. |
| ۵- | سپس 1-class SVM اعمال شده بر روی خوشه مربوطه، نرمال یا حمله بودن اتصال را تعیین می‌کند و در صورت تشخیص حمله آن را مسدود می‌کند. |
| ۶- | منتظر ورود اتصال بعدی می‌ماند. |

**جدول (۵): فرایند تست روش تشخیص نفوذ ترکیبی پیشنهادی بخش ۳-۲**

| | |
|---|---|
| **۱-** | ابتدا با توجه به ویژگی‌های اتصال ورودی، خوشه مربوط به آن مشخص می‌شود. |
| **۲-** | سپس SVM اعمال شده برخوشه مربوطه، نرمال یا حمله بودن اتصال را تعیین می‌کند و در صورت تشخیص حمله آن را مسدود می‌کند. |
| **۳-** | منتظر ورود اتصال بعدی می‌ماند. |

## مراجع:


[1] Hung-Jen Liao, Chun-Hung Richard Lin, Ying-Chih Lin, Kuang-Yuan Tung., *Intrusion detection system: A comprehensive review*, Journal of Network and Computer Applications, Vol. 36, pp. 16-24, 2013.

[2] P. Garcia-Teodoro, J.Diaz-Verdejo, G. Macia-Fernandez, E.*Vazquez., Anomaly-based network intrusion detection: Techniques, systems and challenges*, Computers & Security, pp. 18-28, 2009.

[3] Gisung Kim, Seungmin Lee, Sehun Kim., *A novel hybrid intrusion detection method integrating anomaly detection with misuse detection*, Expert Systems with Applications, Vol. 41, PP. 1690–1700, 2014.

[4] Jungsuk Song, Hiroki Takakura, Yasuo Okabe, Koji Nakao., *Toward a more practical unsupervised anomaly detection system*, Information Sciences, Vol. 231, pp. 4-14, 2013.

[5] Barbara, D., Couto, J., Jajodia, S., Popyack, L., & Wu, N., *ADAM: Detecting intrusions by data mining*, In Proceedings of the IEEE workshop on information assurance and security, pp. 11–16, 2001.

[6] Depren, O., Topallar, M., Anarim, E., Ciliz, M. K., *An intelligent intrusion detection system for anomaly and misuse detection in computer networks*, Expert Systems with Applications, 29(4), PP. 713–722, 2005.

[7] Hwang, K., Chen, Y., Qin, M., *Hybrid intrusion detection with weighted signature generation over anomalous Internet episodes*. IEEE Transactions on Dependable and Secure Computing, 4(1), PP. 41–55, 2007.

[8] Jungsuk Song, Hiroki Takakura, Yasuo Okabe, Yongjin Kwon., *Unsupervised Anomaly Detection Based on Clustering and Multiple One-class SVM*, IEICE, Transactions on Communications E92-B (06), pp. 1981–1990, 2009.

[9] Jungsuk Song, Kenji Ohira, Hiroki Takakura, Yasuo Okabe, Yongjin Kwon., *A clustering method for improving performance of anomaly-based Intrusion Detection System*, IEICE Transactions on Information and Communication System Security E91-D (5), pp. 1282–1291, 2008.


## زیرنویس:

1 Intrusion Detection System

2 Intrusion Detection Algorithm

3 Misuse Detection

4 Anomaly Detection

5 False Positive

6 One-class Support Vector Machines

7 Decision Support System

8 Decision Tree

9 Real Time