فصلنامه علمی-ترویجی پدافند غیرعامل
سال ششم، شماره ۲، تابستان ۱۳۹۴، (پیاپی ۲۲): صص ۲۶-۱۳

چارچوبی برای تجسم حملات سایبری مبتنی بر ادغام اطلاعات

علی‌جبار رشیدی[۱]، کوروش داداش‌تبار احمدی[۲]، منوچهر اکبری[۳]

تاریخ دریافت: ۹۳/۰۶/۲۶
تاریخ پذیرش: ۹۳/۱۰/۱۳

چکیده

هدف اصلی الگوریتم‌های تجسم، ارتقا عملکرد سامانه آگاهی وضعیتی سایبری است که این امر به‌واسطه پر کردن شکاف‌های موجود بین سیستم‌های تشخیص نفوذ و به‌کارگیری مکانیزم‌های دفاعی امنیتی صورت می‌گیرد. در این مقاله طرحی ارائه خواهد شد تا بتوان وضعیت آینده فضای سایبری را تجسم نمود و اقدام پیش‌دستانه‌ای را با بهبود آگاهی وضعیتی به اجرا گذاشت. طرح مذکور بر اساس ادغام اطلاعات سطح بالا است. این سیستم قادر به تجسم رد حملات بدون نیاز به دانش قبلی درباره معماری شبکه یا قالب‌های حملات ایستا است و از دو بخش معماری ادغام اطلاعات و معماری تجسم حملات سایبری تشکیل شده است. در بخش تجسم حملات، روشی برای تجسم حملات چندمرحله‌ای پیشنهاد شده‌است تا تحلیل‌گران فضای سایبر با اولویت‌بندی حملات به‌جای صرفاً واکنش‌های منفعلانه، اقدامات کنش‌گرایانه‌ای را در قبال اقدامات آتی یک حمله سایبری ارائه دهند. درنهایت، عملکرد سیستم توسط معیارهای مختلفی ازجمله میزان دقت تجسم، ارزیابی شده است.

کلیدواژه: حملات سایبری، ادغام اطلاعات، آگاهی وضعیتی، تجسم، مدل مارکوف با طول متغیر.

۱-دانشیار دانشگاه صنعتی مالک اشتر و مرکز پژوهشی علوم و فناوری پردازش و ادغام اطلاعات تهران، ایران
۲-استادیار دانشگاه صنعتی مالک اشتر و مرکز پژوهشی علوم و فناوری پردازش و ادغام اطلاعات تهران، ایران
۳-کارشناس ارشد دانشگاه صنعتی مالک اشتر و مرکز پژوهشی علوم و فناوری پردازش و ادغام اطلاعات تهران، ایران- akbarimanochehr@mut.ac.ir- نویسنده مسئول

۱- مقدمه

برای تجسم فعالیت‌های حاصله از نفوذ یک حمله سایبری ابتدا باید بتوان با استفاده از فناوری‌های مختلف، حجم بالایی از هشدارهای خام تولید شده توسط سیستم‌های تشخیص نفوذ را درک، اثرات آنها را ارزیابی و ردهای[1] نفوذ را تجسم نمود. ازجمله فناوری‌های لازم برای کاهش عدم قطعیت در این حجم انبوه داده‌ها، فناوری‌های مبتنی بر ادغام اطلاعات است [۳]. ادغام اطلاعات، همبستگی هشدارهای متعلق به یک حمله را حفظ کرده و آگاهی وضعیتی حاصل از فعالیت‌های نفوذ که توسط مهاجم عمداً مبهم شده یا به‌طور غیرمستقیم، توسط معماری شبکه پیچیده شده را بهبود می‌بخشد [۲]. هدف این پژوهش این نیست که تحلیل‌گر را به‌طور کلی از فرآیند تحلیل حذف کند بلکه این سیستم از میان مقدار انبوه اطلاعات در قالب هشدارهای نفوذ، اطلاعات را غربال می‌کند، رویدادهای مهم را تشخیص داده و به هم ربط می‌دهد، اقدامات محتمل آینده را تجسم می‌کند، اثر اقدامات فعلی و آینده مهاجم را ارزیابی و درنهایت این اطلاعات را به تحلیل‌گر ارائه می‌دهد تا تحلیل‌گر بتواند اقدامات لازم برای کاهش آسیب‌های وارده را انجام دهد.

اصولاً تجسم یک حمله، دنباله‌ای از هشدارهایی است که از همبستگی ردهای حمله ایجاد می‌شود. این همبستگی، با استفاده از موتورهای ادغام اطلاعات ایجاد می‌شوند [۳]. در این تحقیق از دنباله فوق به همراه مدل مارکوف با طول متغیر برای تجسم آینده محتمل استفاده شده است. رویکرد پیشنهادی این است که آینده محتمل که لزوماً آشکار نیست یا قبلاً توسط تحلیل‌گر در نظر گرفته نشده را فاش نماید.

در فضای سایبری می‌توان به چالش‌های زیادی ازجمله عدم قطعیت بالا و توانایی‌های متنوع مهاجم اشاره کرد. علاوه بر این با توجه به این‌که مبدأ و مقصد حملات سایبری مشخص نیست و فرصت‌های پیش‌روی مهاجم نامشخص است مسئله تجسم دشوارتر نیز می‌شود. برای این‌که بتوانیم مسئله تجسم اثرات حملات سایبری را پوشش دهیم این مقاله را به ۷ بخش تقسیم کرده‌ایم که در بخش اول مقدمه و ضرورت تحقیق شرح داده شده است. بخش ۲ کارهای مرتبط در این حوزه را تشریح می‌کند. در بخش ۳، موارد مربوط به با محیط ادغام اطلاعات، توضیح داده می‌شود. بخش ۴ به معماری کلی سیستم تجسم می‌پردازد. بخش ۵ تلاش صورت گرفته و ملاحظات مربوط به طراحی سیستم تجسم بلادرنگ را تعریف می‌کند، این بخش همچنین به مدل‌های ریاضی و الگوریتم‌هایی که اساس کار سیستم محسوب می‌شود، می‌پردازد. بخش ۶ با ارزیابی الگوریتم پیشنهادی از طریق شبیه‌سازی به تجزیه و تحلیل پرداخته و در انتها در بخش ۷ نتیجه‌گیری آورده شده است.

1– Track

۲-کارهای مرتبط

ادغام اطلاعات، نه‌تنها برای امنیت سایبری بلکه در بسیاری دیگر از حوزه‌های نرم‌افزاری هنوز در مراحل ابتدایی است و کار ترویج آن هنوز هم وجود دارد [۳]. باس [۴] از نیاز به ادغام اطلاعات برای پر کردن خلأ کشف نفوذ، دفاع می‌کند. از آن زمان به بعد، کارهای زیادی (به‌عنوان مثال [۱۵-۵]) در همبستگی هشدارهای سیستم تشخیص نفوذ انجام شده تا آگاهی وضعیتی بهتری از حملات سایبری فراهم شود [۱۶].

اغلب کارهای انجام‌شده در حوزه تجسم نیز مبتنی بر الگوهای حمله‌ای است که از پیش تعریف شده‌است. برای مثال گیون و لی [۱۷] یکی از اولین مدل‌های تجسم حمله را پیشنهاد کردند که در این طرح از همبستگی هشدارها برای پیش‌بینی حملات استفاده شده است. آنها در پژوهش خود دنباله‌های حمله را با استفاده از شبکه‌های بیزین[2] ایجاد کردند. هولسوپل و شانچی [۱۸]، در مقاله "ارزیابی تهدید علیه داده‌ها و اطلاعات شبکه"[3] اطلاعات قبلی را برای تجسم آینده محتمل، با تکیه بر دانش هم‌بندی شبکه و دنباله حمله مشاهده شده، اضافه کردند. نتایج نشان داد که ارزیابی تهدید علیه داده‌ها و اطلاعات اقدامات حمله آینده را تا زمانی که حمله بخشی از یک حمله هماهنگ نیست و شامل هیچ تهدید خودی نباشد با دقت پیش‌بینی می‌کند. نتیجه درجات تهدید از این کار و کار مشابه توسط آرنس و همکارانش [۱۹] نشان داد که موجودیت‌های شبکه به‌احتمال قوی در حال تبدیل‌شدن به اهداف بعدی در یک حمله بودند؛ اما فاوا و همکارانش [۲۰] با جای‌گذاری هشدارهای سیستم‌های تشخیص نفوذ در داخل مدل "صحنه نبرد سایبری"[4] به استنتاج منطقی و محتوایی برمبنای آسیب‌پذیری‌های آشکارشده پرداختند. آن‌ها از هشدارهای همبسته‌شده به‌عنوان ردهای منحصربه‌فرد برای پیش‌بینی حملات استفاده کردند در ادامه، بایرز و شانچی [۳] یک سیستم یادگیرنده دیگر ارائه کردند که قادر به تجسم حملات، بدون نیاز به دانش قبلی بود و

2– Bayesian Network

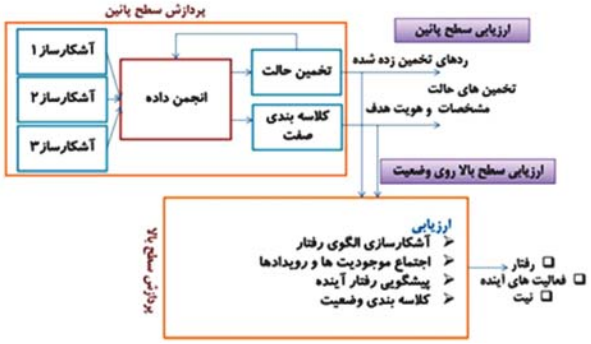
3– TANDI

۴-این مدل یک نوع نمایش امنیتی از شبکه را برای ارتباط دهی بین میزبان‌ها، سرویس‌ها و اولویت‌ها در یک شبکه کامپیوتری ارائه می کند

می‌توانست حملاتی که تاکنون رخ نداده بودند را نیز پیش‌بینی کند. در واقع ما میخواهیم کار بایرز و شانچی را ادامه دهیم. در این تحقیق از روشی به نام مدل مارکوف با طول متغیر [1] [۲۱] برای استخراج الگوهای رفتاری از دنباله‌های حمله، بدون نیاز به تعریف الگوهای حمله به‌صورت مجزا استفاده خواهیم‌کرد.

۳- ادغام اطلاعات

در دسته‌بندی که انجمن جهانی ادغام منابع [2]، از ادغام اطلاعات ارائه نموده‌است، ادغام اطلاعات به دو دسته اطلاعات سطح بالا [3] و سطح پایین [4] تقسیم شده است. در ادغام سطح پایین اطلاعات، موضوعاتی چون طبقه‌بندی، شناسایی و ردگیری هدف جزء بخش‌های اصلی محسوب می‌شوند، این درحالی است که در ادغام سطح بالا، اثر وضعیت و پالایش فرایند ادغام مهم‌ترین موضوعات هستند [۲۲]. ادغام سطح پایین با داده‌هایی چون مکان، جنبش‌شناسی، نوع و ویژگی هدف سروکار دارد. علاوه بر این پردازش سیگنال، هویت‌شناسی هدف و تخمین حالت شئ برای ردگیری، شناسایی و طبقه‌بندی در این بخش به اجرا درآورده می‌شود. ادغام سطح بالا با اطلاعات انتزاعی نظیر تهدید، نیت و اهداف روبرو است و با استفاده از کنترل، درک وضعیت و روابط حاکم بر محیط ادغام، رفتار، نیات و فعالیت‌های آتی یک رویداد استخراج می‌شود [۲۳]. تمایز بین دو دسته ادغام اطلاعات سطح بالا و پایین را می‌توان در شکل (۱) مشاهده نمود.



شکل ۱– مؤلفه‌های اصلی سیستم ادغام اطلاعات سطح پایین و بالا [۲]

برای معماری ادغام اطلاعات در این تحقیق از مدل آگاهی وضعیتی اندسلی استفاده شده است که این مدل را تشریح خواهیم کرد. اندسلی ادغام اطلاعات را در سه سطح درک [5]، فهم [6] و تجسم مطرح نموده است [۲۴] که به‌صورت زیر تشریح می‌گردند:

الف- سطح اول، درک: درک نشانه‌ها، یک موضوع حیاتی است و بدون درک اولیه از اطلاعات مهم، احتمال شکل‌گیری تصویری نادرست از یک موضوع به‌شدت افزایش می‌یابد. در این سطح به این سؤال پاسخ داده می‌شود که واقعیت‌های فعلی چیست؟

ب- سطح دوم، فهم: آگاهی وضعیتی مفهومی فراتر از درک یا توجه صرف به اطلاعات است، بلکه یکپارچه‌سازی تکه‌های مختلف اطلاعات، تعیین ارتباط میان آن‌ها و اهداف کاربر را دربر می‌گیرد. این موضوع درست شبیه تفاوت میان سطح بالایی از درک مطلب نسبت به خواندن لغات تنها است. درمجموع این سطح به این سؤال پاسخ می‌دهد که چه چیزی اتفاق می‌افتد؟

ج- سطح سوم، تجسم: در بالاترین سطح آگاهی وضعیتی که مبتنی بر بالاترین سطح درک از وضعیت است، توانایی آینده‌نگری رویدادهای مربوط به هر وضعیت مطرح خواهد بود. این توانایی با استفاده از تجسم رویدادهای پویای جاری و حرکت به سمت رویدادهای آتی مورد انتظار، امکان تصمیم‌گیری به‌موقع را فراهم می‌کند [۲].

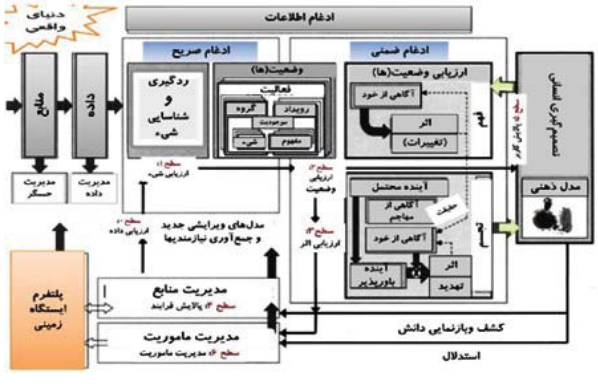
با خلاصه‌سازی رفتار مشاهده‌شده در معماری ادغام اطلاعات به دنبال تجسم برای تبدیل حالت یا وضعیت فعلی به یک مجموعه از وضعیت‌های آینده محتمل هستیم که اثر هر یک از وضعیت‌های آینده محتمل نیز ارزیابی‌شده تا به آگاهی وضعیتی کمک کند [۳].

با آگاهی وضعیتی می‌توان به درک و فهم وضعیت جاری فضای سایبری موردحمله پرداخت، وضعیت آینده را تجسم نمود و اقدام پیش‌دستانه‌ای صورت داد. علاوه با بهبود آگاهی وضعیتی سایبری از طریق ادغام اطلاعات، قادر به تفکیک و تلخیص اطلاعات مفید جهت ارائه به تحلیل‌گران خواهیم بود. این اطلاعات برای کاهش عدم قطعیت و افزایش اعتماد در تصمیم‌گیری کنار هم قرار می‌گیرند و موارد غیرمطمئن و درصد عدم اطمینان هر مورد نیز دقیقاً مشخص می‌شود تا تحلیل‌گر بتواند با بهره‌گیری از آگاهی ایجادشده تصمیم‌گیری بهتری را به اجرا بگذارد. همچنین با استفاده از آگاهی وضعیتی و ارزیابی وضعیت، روابط بین موجودیت‌ها کشف‌شده، رویدادها تعیین و فعالیت‌های مهم و بامعنا، مشخص خواهند شد. به‌طورکلی آگاهی وضعیتی از مشاهده‌ها و دانش موجود یا ذخیره‌شده شکل‌گرفته و با بهره‌برداری همزمان و اشتراک مساعی بین آن دانش‌ها، مدل آگاهی وضعیتی بهبود می‌یابد [۲].

به خاطر درک ناقص از معماری ادغام اطلاعات و عدم استفاده

1– Variable Length Markov Model
2– http://www.isif.org
3– High-level information fusion
4– low-level information fusion
5– Perception
6– Comprehension

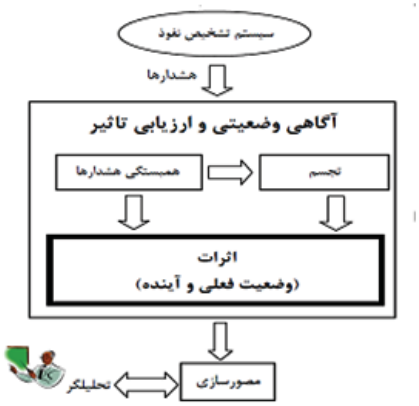
درست از آن در دفاع سایبری تاکنون موفقیت‌های چشم‌گیری در به‌کارگیری فرایند ادغام اطلاعات در فضای سایبری کسب نشده است. علاوه به‌دلیل پیچیدگی بالای این حوزه نیاز است تا ادغام اطلاعات از منابع مختلف در سطح وسیع صورت گیرد بنابراین باید الگوریتم‌های ترکیبی یا جدیدی از ادغام اطلاعات را به‌کار گرفت که آگاهی لازم را در سطوح مختلف ایجاد نماید و این آگاهی نیز محتوی محور باشد. با توجه به این‌که سطوح مختلفی برای ادغام اطلاعات وجود دارد از الگوریتم‌های مختلفی برای ادغام در سطوح مختلف استفاده می‌شود بنابراین با استفاده از استنتاج‌های حاصل از ترکیب اطلاعات می‌توان به سؤالات مختلف به‌صورت شفاف برای کسب آگاهی وضعیتی بهتر پاسخ داد. در این تحقیق با درک صحیح ارتباط بین مؤلفه‌های پایه و سطوح ادغام اطلاعات در دفاع سایبری، چگونگی ایجاد ابزارهای جدید دفاع سایبری مبتنی بر ادغام چند سطحی و پاسخ به سؤالات در سطوح مختلف برای ایجاد آگاهی وضعیتی بهتر معماری مناسب برای این حوزه به‌مانند آنچه در شکل (۲) نشان داده شده ارائه شده است [۲۵]. در شکل (۲) که نمایی از معماری ارزیابی وضعیتی ادغام اطلاعات ترکیبی مناسب حوزه سایبری نشان داده شده است، فعالیت‌های مربوط به یک وضعیت که کاربر با آن درگیر است، نظیر استدلال (درک)، ارزیابی (فهم) و پیش‌بینی وضعیت آینده (تجسم) مشخص شده است [۲]. از این معماری برای طراحی سیستم تجسم حملات سایبری استفاده شده است.



شکل۲- معماری ادغام اطلاعات در تجسم حملات سایبری [۲]

۴-معماری امنیت سایبری

این بخش معماری کلی امنیت سایبری (شکل ۳) را تشریح می‌کند و چشم‌اندازی از سیستم تجسم پیشنهادی ارائه می‌دهد.



شکل۳- نمایی از چارچوب کلی طرح در حوزه تحقیق

این سیستم به‌دنبال افزایش آگاهی وضعیتی از طریق ارزیابی و برآورد آینده محتمل با در نظر گرفتن مؤلفه‌هایی چون توانایی، فرصت، نیت و رفتار است. سیستم بالا از میان مقدار اطلاعات انبوه در قالب هشدارهای نفوذ، اطلاعات را غربال می‌کند، رویدادهای مهم را تشخیص و به هم ربط می‌دهد، اعمال آینده محتمل را بلادرنگ تجسم می‌کند، تأثیر اقدامات فعلی و رو به جلوی مهاجم را ارزیابی و این اطلاعات را به تحلیل‌گر ارائه می‌دهد.

با توجه به این‌که پیگیری و تجسم حملات سایبری متکی بر گزارش دقیق و به‌موقع از فعالیت‌های مشکوک است، میزبان‌ها و شبکه‌های مبتنی بر سیستم‌های تشخیص نفوذ زمانی که فعالیت‌های مشکوکی را مشاهده می‌کنند هشدارهایی را تولید می‌کنند و در نتیجه این کار حجم انبوهی از هشدارها به وجود خواهدآمد که برای حل این مشکل روش‌های متعددی برای تجمیع هشدار و همبسته‌سازی هشدار پیشنهاد شده است. همبسته‌سازی هشدار، فرایندی است که اساساً هشدارهای سیستم‌های تشخیص نفوذ را به‌عنوان ورودی گرفته و یک نمای سطح بالا از حملات در حال وقوع را تهیه می‌کند و آن‌ها را در مجموعه‌های سفارش داده‌شده سازماندهی می‌کند. این مجموعه‌ها می‌توانند برای نشان دادن خط سیرهای مجازی حملات سایبری در نظر گرفته شوند. بعد از همبسته‌سازی، این مجموعه‌ها برای ارزیابی اثرات فعلی و تجسم آینده مورد بررسی قرار می‌گیرند. ارزیابی تهدید و تأثیر، دلیلی برای تفاوت بین حملات خطرناک با نگرانی زیاد از تهدیدهای خوش‌خیم و کم‌خطرتر است.

این کار قصد دارد تا با اجرای آموزش الگوریتم و پیش‌بینی برای یک تعداد دلخواه از تعاریف فیلد هشدار، سیستم را گسترش دهد. این سیستم به‌جای از پیش آموزش دیدن، به‌طور مداوم یاد

می‌گیرد و پیش‌بینی‌های قبلی را که به‌عنوان گام‌های حمله جدید مشاهده شده‌اند تأیید یا رد می‌کند. درنهایت با تجسم این اقدامات آینده محتمل از الگوهای حمله، سیستم آسیب‌پذیری‌های بالقوه شبکه را که ممکن است سریعاً بهره‌برداری شود معلوم می‌کند.

۵- تجسم

گرچه دو اصطلاح تجسم[1] و پیش‌بینی[2] در بسیاری از منابع تحقیقی یکسان فرض شده‌اند اما باید توجه داشت که تعریف تجسم با پیش‌بینی متفاوت است. ازلحاظ نظری، پیش‌بینی حدس‌هایی است که مشخص می‌کند در آینده چه اتفاقی خواهد افتاد، درحالی‌که تجسم صرفاً چیزهای که در آینده اتفاق خواهد افتاد را مشخص می‌کند [۲۶]. به بیان ساده‌تر، پیش‌بینی به حدس‌هایی گفته می‌شود که با احتمالات متفاوتی در آینده رخ خواهند داد یعنی پیش‌بینی فرایند برآورد موقعیت‌های ناشناخته است. یک پیش‌بینی درواقع یک پیش‌گویی در مورد رویدادهای آینده در اختیار می‌گذارد ولی تجسم، زیرمجموعه‌ای از مجموعه پیش‌بینی‌ها است که در آینده به‌احتمال‌زیاد اتفاق خواهند افتاد. برای توضیح بیشتر می‌توان گفت پیش‌بینی و تجسم عبارتند از مجموعه حالاتی که بر اساس اطلاعات به‌دست‌آمده احتمال رخ دادن آن‌ها وجود دارد. از این مجموعه، حالاتی که با توجه به دانش قبلی احتمال رخداد آن‌ها بسیار بیشتر است را جدا کرده و اتفاق افتادن آن را مجسم می‌کنیم به این معنا که در مورد این حالات که احتمال رخداد بیشتری دارد تصمیم‌گیری‌هایی را اتخاذ می‌کنیم، به این مورد تجسم گفته می‌شود. این تمایز مهم است چون تعیین آینده‌ای دقیق، همیشه با توجه به رخداد تصادفی یا دانش ازدست‌رفته، مشخص نیست. تجسم به تحلیل‌گر اجازه می‌دهد تا به شرایط احتمالی آینده نگاه کند، پیش از آن به آینده محتمل مراجعه کند و برای نتایج مختلف آماده شود. پس لازم است که یک آینده محتمل را تجسم کنیم.

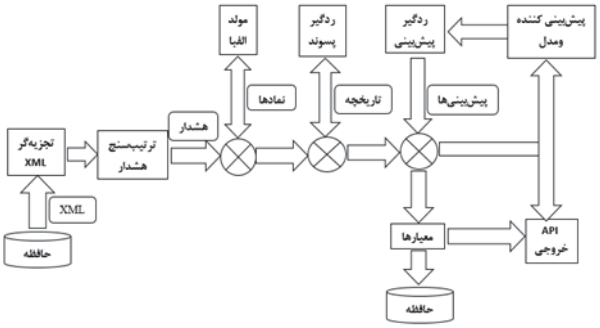
۵-۱- معماری تجسم

فرایند ادغام با استخراج الگوهای رفتاری همراه است که با تبدیل مجموعه‌ی منظم از هشدارهای سیستم تشخیص نفوذ به دنباله‌ی ساده‌ای از نمادها شروع می‌شود. هشدارهای ورودی با توجه به برچسب‌های XML مطابق رد خود تجزیه می‌شوند، برچسب‌های انتخاب‌شده به نمادهای الفبا نگاشت می‌شوند و ردهای حمله متناظر و درختان پسوندی[3] ایجاد می‌شوند [۳].

هدف مدل درخت پسوندی که درواقع نمایشی ادغام شده از ردهای مشاهده شده است، یکپارچه‌سازی رفتارهای مهاجم در یک مدل است. برای هر فیلد هشدار، هرکدام از نگاشت نمادها، درختان پسوندی و تاریخچه رد حملات به‌صورت جداگانه ایجاد و استفاده می‌شوند. در معماری مدنظر از مدل مارکوف با طول متغیر (VLMM) برای تجسم هر رد حمله مبتنی بر تاریخچه و درخت پسوندی منحصربه‌فرد استفاده شده است [۳].

شکل (۴) جریان منطقی داده از سیستم تجسم را نشان می‌دهد که با ورود وقایع نفوذ در سمت چپ شروع می‌شود و خروجی در سمت راست است. در این کار از داده‌های ورودی ذخیره‌شده در فایل‌های XML برای اهداف شبیه‌سازی استفاده می‌شود که می‌تواند یکپارچگی با سیستم‌های دیگر را تسهیل کند. ابتدا، کلاس تجزیه‌کننده XML فایل‌های XML را به هشدار ترجمه می‌کند، سپس به‌ترتیب زمان توسط ترتیب سنج هشدار[4]، سازماندهی می‌شوند و در بقیه سیستم که توسط رابط کاربر گرافیکی فرماندهی می‌شود، تزریق می‌شوند.

برای ترجمه از فیلد توصیفات هشدار به نمادها از کلاس مولد الفبا (که تعاریف نماد را ایجاد و ذخیره می‌کند) استفاده شده است. در مرحله بعد، ردگیر پسوند[5]، نمادهای جدید را به تاریخچه رد خود مرتبط می‌کند و این دنباله تکمیل‌شده (نماد جدید همراه با نمادهای پیشین مربوطه خود) به پیش‌بینی‌کننده و مدل عبور می‌کند. مدل درخت پسوندی با دنباله جدید آموزش دیده و پیش‌بینی‌کننده به‌احتمال زیاد نماد بعدی را پیش‌بینی می‌کند. این پیش‌بینی پس‌ازآن در ردگیر پیش‌بینی[6] ذخیره‌شده و سپس توسط کلاس‌های سنجه[7] برای تعیین درستی پیش‌بینی، (زمانی که نمادهای بعدی می‌رسند) استفاده می‌شود.



شکل۴- جریان داده منطقی از سیستم تجسم

1– Projection
2– Prediction
3– Suffix Tree

4– Alert Sequencer
5– Suffix Tracker
6– Prediction Tracker
7– metric

جدول ۱- یک مجموعه داده با ۱۲ هشدار به شکل دنباله هشدار [۳]

Time	Track	Destination IP	Group	Destinations
۱۲:۵۶:۰۳	۴۹۰۶۲	۱۰۰,۲۰,۲۰۰,۱۵	ExtIntrusionRoot	WEB-MISC Invalid HTTP Version String
۱۲:۵۷:۰۹	۴۹۳۳۲	۱۰۰,۵,۱۱۱,۱۶۶	ExtIntrusionRoot	SHELLCODE x86 NOOP
۱۲:۵۸:۱۱	۴۹۲۶۲	۱۰۰,۱۰,۲۰,۴	ExtIntrusionOther	(http_inspect) BARE BYTE UNICODE ENCODING
۱۲:۵۸:۴۵	۴۹۲۶۵	۱۰۰,۱۰,۲۰,۳	ExtIntrusionRoot	NETBIOS SMB IPCS Unicode share access
۱۲:۵۸:۵۹	۴۹۲۶۱	۱۰۰,۱۰,۲۰,۳	ExtScaning	ICMP L3retriever Ping
۱۲:۵۹:۳۷	۴۹۰۶۶	۱۰۰,۱۰,۲۰,۴	ExtIntrusionRoot	WEB-MISC Chunked-Encoding transfer attempt
۱۲:۵۹:۳۷	۴۹۰۶۶	۱۰۰,۱۰,۲۰,۴	ExtIntrusionOther	(http_inspect) OVERSIZE CHUNK ENCODING
۱۲:۵۹:۳۸	۴۹۰۶۶	۱۰۰,۱۰,۲۰,۴	ExtIntrusionOther	(http_inspect) BARE BYTE UNICODE ENCODING
۱۲:۵۹:۴۸	۴۹۲۹۴	۹۲,۶,۸۵,۱۰۳	Extfiltration	(portscan) TCP Portscan
۱۲:۵۹:۵۰	۴۹۰۶۷	۱۰۰,۵,۱۱۱,۲۰۸	IntScan	ICMP PING NMAP
۱۲:۵۹:۵۷	۴۹۰۷۰	۱۰۰,۱۰,۲۰,۴	ExtIntrusionRoot	WEB-MISC cross site scripting attempt
۱۳:۰۰:۲۱	۴۹۰۷۰	۱۰۰,۱۰,۲۰,۴	ExtIntrusionOther	(http_inspect) OVERSIZE CHUNK ENCODING

۵-۲- مولد الفبا[1]

در سیستم تجسم شکل (۴)، به فهرستی از هشدارهای از قبل تعریف‌شده و معنای نمادهای الفبا نیازی نیست. چراکه فیلدهای هشدار تعریف‌شده توسط کاربر[2] که مستقلاً به یک نماد در الفبای مربوطه خود نگاشت می‌شوند یک فضای نماد (Ω) را تشکیل می‌دهند. همچنین در این سیستم با مشاهده یک مقدار فیلد هشدار جدید به‌طور خودکار یک نماد جدید بامعنی متناظر خود در یک نقشه هش[3] ایجاد می‌شود.

جدول (۲) نگاشت توصیفات را برای دوازده هشدار تعریف‌شده در جدول بالا به نماد را نشان می‌دهند.

جدول ۲- الفبای توصیفات برای دنباله هشدار مثال [۳]

نماد	مقدار
1	WEB-MISC Invalid HTTP Version String
2	SHELLCODE x86 NOOP
3	(http_inspect) BARE BYTE UNICODE ENCODING
4	NETBIOS SMB IPCS Unicode share access
5	ICMP L3retriever Ping
6	WEB-MISC Chunked-Encoding transfer attempt
7	(http_inspect) OVERSIZE CHUNK ENCODING
8	(portscan) TCP Portscan
9	ICMP PING NMAP
10	WEB-MISC cross site scripting attempt

پس‌ازاین‌که فیلدهای مربوط به هشدارهای ورودی به نمادها تبدیل می‌شوند، به ردهای مربوطه خود اضافه می‌شوند.

ردهای حمله به‌موازات تعریف الفبا در فضای نماد ساخته می‌شوند،

این ردها نشان‌دهنده همان حمله چندمرحله‌ای از منظر توصیف حمله، زیرشبکه مقصد و غیره هستند. زمانی که یک هشدار می‌رسد، سیستم به نماد مربوط به هر فیلد هشدار نگاه می‌کند و نماد را به دنباله‌ی هشدارهای n اضافه می‌کند، $s= \{X_1,X_2,\ldots,X_n\}$ که در این مجموعه، X_i متعلق به الفبای فیلد هشدار مربوطه است و برابر با X_{i-1} نمی‌باشد. برای مثال، جدول ۱ رشته‌ای انتخاب شده از ۱۲ هشدار اول از یک مجموعه داده نمونه به‌منظور بررسی فضای الفبا و دنباله رد حمله را نشان می‌دهد [۳].

۵-۳- ایجاد درخت پسوندی با استفاده از رفتار گذشته یک حمله

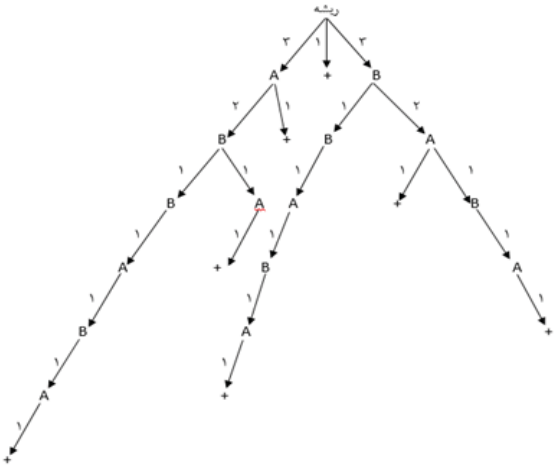
برای پیش‌بینی حرکت بعدی مهاجم، آخرین رفتار از حمله را در درخت پسوندی برای استفاده توسط الگوریتم‌های پیش‌بینی اضافه می‌کنند. برای هر الفبا، درختان جداگانه ایجاد می‌شوند و هر درخت شامل تمام ردها برای الفبای مربوطه خود است.

رسیدن بلادرنگ هشدارها، نیازمند یک الگوریتم جدید است و درخت پسوندی با دنباله نسبی متناهی به‌جای یک دنباله کامل آموزش داده می‌شود. الگوریتم آموزش درخت پسوندی به‌جای استفاده از یک دنباله طولانی از مشاهدات، یک مجموعه از دنباله‌ها

1– Alphabet
2– User-Defined
3– Hash map

با طول متناهی را لحاظ می‌نماید. دنباله‌های طول متناهی، آغاز و پایان نمادهای دنباله را تعریف می‌کنند.

برای مثال یک درخت ساخته شده روی دنباله‌ی *,A,B,B,A,B,A,+ که در آن نماد + پایان نمادهای دنباله است، در شکل (۵) نشان داده شده‌است. نمادهای A و B را به‌ترتیب به‌عنوان نماینده توصیف هشدارهای WEB–IIS nsiislog.dll access و WEB-MISC Invalid HTTP Version String در نظر بگیرید. یال‌ها براساس تعداد دفعاتی که درخت پسوندی از طریق آن شاخه پیموده شده‌اند وزن‌دار هستند. برای مثال A,B,A,+ تنها یک بار و A,B دو بار در دنباله اتفاق افتاده است.



شکل ۵- درخت پسوندی برای یک دنباله متناهی *,A,B,B,A,B,A,+

۵-۴- ایجاد درخت پسوندی توسعه یافته برای کمک به پیش‌بینی رفتارهای نوظهور

بدون پیش آموزش، این سیستم خود-یادگیر[1] به‌ناچار نمادهای جدیدی را مشاهده خواهد کرد. با این فرض برای پیش‌بینی وقوع نمادهای جدید یک تعریف نماد به‌عنوان نماد جدید به‌فهرست نمادها اضافه می‌شود و آموزش مدل درخت پسوندی با یک دنباله اضافی انجام خواهد شد. هنگامی که نماد جدیدی رخ می‌دهد، ابتدا درخت پسوندی به‌صورت عادی آموزش می‌بیند و پس از آن دوباره با الحاق تعریف "نماد جدید" خاص با تاریخچه درخت پسوندی (به‌جای نماد واقعی که رخ داده‌است) آموزش می‌بیند. توجه داشته باشید که هردوی دنباله‌ها با استفاده از الگوریتم ایجاد درخت پسوندی با در نظر گرفتن تنها نماد آخر آموزش می‌بینند. شکل(۶) شبه‌کد برای رسم درخت پسوندی توسعه یافته را نشان می‌دهد.

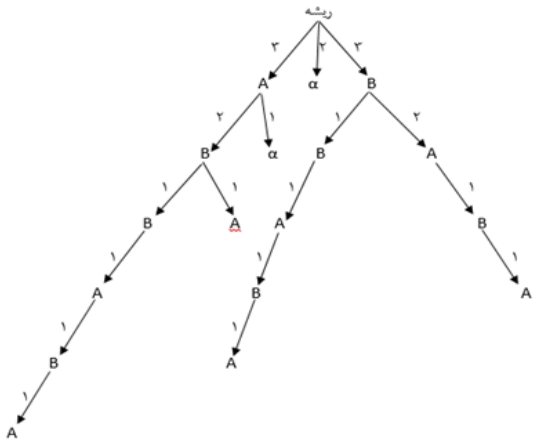
1– self –learning

```
۱   Learn new.alert (string sequence)
۲   Child [All node[ ] = ]
۳   For offset from 1 to sequence.Length do
۴   Current node = root
۵   For i from offset to sequence.Length do
۶   Check = sequence[i]
۷   If check Exist in child [current node]
۸   current node = check
۹   current node.incident edge frequency +=1
۱۰  else
۱۱  check = creat branch (current node, sequence[i ]to
    sequence [sequence.Length -1 ,]int i, sequence.Length - i)
۱۲  break
۱۳  end if
۱۴  end for
۱۵  end for
۱۶
۱۷  creat branch (parent, string newsequence, int i, int Length)
۱۸  descendent  =[۰]parent
۱۹  For j from 1 to Length  do
۲۰  New node = newsequence [j]
۲۱  New node. incident edge frequency = 1
۲۲  New node parent = descendent[j-1]
۲۳  descendent[j =] new node
۲۴  Child  [newsequence (j-1), i = ]Child  [newsequence (j-1), i
    . ]Concat new node
۲۵  If newnode not exist in newsequence ]۰[to newsequence[j-1
    ]
۲۶  If node(α) exist in child  [newsequence (j-1), i  ]
۲۷  Increase edge frequency
۲۸  Else
۲۹  New node = node(α)
۳۰  New node. incident edge frequency = 1
۳۱  New node parent = descendent [j-1]
۳۲  Child  [newsequence (j-1), i = ] Child  [newsequence (j-1), i
    . ]Concat node(α)
۳۳  End if
۳۴  i += 1
۳۵  End if
۳۶  End for
۳۷  Return descendent [length]
```

شکل ۶- شبه‌کد درخت پسوندی توسعه یافته مختلف

آموزش دیدن دنباله A,B,B,A,B,A,* که قبلا برای آموزش درخت پسوندی بلادرنگ استفاده می‌شد را در نظر بگیرید. دنباله باید به عنوان A و A,B و A,B,B و A,B,B,A و غیره آموزش داده شود. در نظر بگیرید α به‌عنوان یک نمادجدید خاص تعریف شده است. هنگامی که A می‌رسد (نماد A قبلا مشاهده نشده است)، درخت با دنباله A و سپس با دنباله α آموزش داده خواهد شد. هنگامی که هشدار بعدی می‌رسد چون B یک نماد جدید است پس درخت با دنباله A,B و سپس با دنباله A,α آموزش داده می‌شود. هنگامی که B دوباره به‌عنوان هشدار بعدی می‌رسد، آن یک نماد جدید نیست و بنابراین درخت به‌صورت عادی با دنباله A,B,B آموزش می‌بیند.

نتیجه آموزش درخت پسوندی برای دنباله A,B,B,A,B,A,* با استفاده از الگوریتم نماد جدید در شکل (۷) نشان داده شده است.



شکل ۷ - درخت پسوندی توسعه یافته برای دنباله A,B,B,A,B,A,*

همانگونه که در شکل (۴) نشان داده شده است هنگامی که رفتار نماد جدید در درخت پسوندی ادغام شد، مدل مارکوف با طول متغیر قادر به پیش‌بینی رخداد یک نماد از جمله رخداد یک نماد جدید (مانند هر نماد دیگری) خواهد بود.

۵-۵- پیش‌بینی گام بعدی حمله با استفاده از مدل مارکوف با طول متغیر

از وضعیت فعلی درخت پسوندی و الگوریتم مدل مارکوف با طول متغیر برای پیش‌بینی اقدامات آینده استفاده می‌شود. مرتبه nام مدل مارکوف محتوا محدود بیانگر یک گام حمله جدید با n مشاهده قبلی در رد حمله است:

$$P^n\left\{x_{(t+1)}\middle|x_{(t-n+1)},\ldots\ldots,x_t\right\} \quad (۱)$$

احتمال P^n با استفاده از وزن- یال‌های فرزند درخت پسوندی با توجه به در نظر گرفتن مشاهدات قبلی محاسبه شده‌است.

احتمالات هر پیش‌بینی تا مرتبه nام بالا رفته سپس با وزن هر سطح که با استفاده از روش احتمال گریز به‌دست‌آمده ترکیب می‌شود. احتمال گریز در واقع احتمال برخورد با یک کاراکتر قبلا مشاهده نشده در یک سطح از درخت پسوندی است که برای هر سطح با e_j نشان داده می‌شود. به عبارت دیگر احتمال گریز را می‌توان احتمال نادیده گرفتن یک نماد در یک سطح بیان کرد که همان احتمال فرار کردن یا گریز یک نماد در آن سطح می‌باشد (و به همین دلیل به این نام مشهور گردیده است) چنانچه احتمال گریز در سطح j را با e_j نمایش دهیم سه روش زیر برای محاسبه آن وجود دارد که به‌صورت روابط (۲) و (۳) و (۴) می‌باشند:

$$e_j = \frac{1}{c_j+1} \quad (۲) \qquad e_j = \frac{q_j}{c_j+q_j} \quad (۳) \qquad e_j = \frac{q_j}{c_j} \quad (۴)$$

که در آن c_j برابر با تعداد کل دنباله‌هایی به‌طول j در jامین سطح درخت است یا به‌عبارت دیگر می‌توان گفت c_j برابر با مجموع وزن (تعداد تکرار) یال‌های فرزندان در سطح j است. همچنین q_j برابر با تعداد کاراکترهایی است که در سطح j تنها یک‌بار دیده شده‌اند. قابل ذکر است که هیچ اثباتی برای برتری هریک از این روش‌ها بر دیگری وجود ندارند و ما در سیستم خود از همان رابطه اول استفاده خواهیم کرد که در آن $e_{-1}=0$ می‌باشد. دلیل استفاده از احتمال گریز در واقع حل مشکل فرکانس صفر است که این مشکل زمانی که مرتبه‌ها افزایش می‌یابد و رشته‌هایی که تا پیش از این دیده نشده‌اند، مشاهده می‌شوند پیش می‌آید. به‌طور کلی، از احتمال گریز برای دادن وزن به هر مرتبه از درخت استفاده می‌شود که به‌طور منطقی هرچه به سمت مرتبه‌های پایین‌تر حرکت می‌کنیم وزن‌های اختصاص یافته به هر سطح افزایش خواهد یافت (برای دادن تاکید بیشتر به مرتبه‌های پایین‌تر). اوزان معادل برای هرسطح را می‌توان با رابطه (۵) و (۶) و استفاده از احتمال گریز محاسبه نمود:

$$W_j = (1 - e_j) \times \prod_{k=j+1}^{L} e_k \,,\ -1 \le j \le L \quad (۵)$$

$$W_L = (1 - e_L) \quad (۶)$$

احتمال ترکیب شده برابر با مجموع وزن‌ها در $P^o(x)$ است:

$$P(x) = \sum_{(o=-1)}^{L} W_o \times P^o(x) \quad (۷)$$

که در آن L طول بلندترین شاخه برای دنباله مشاهده‌شده s در درخت پسوندی و o بیانگر مرتبه درخت پسوندی است. به‌عبارت دیگر، الگوریتم با محاسبه و ترکیب کردن پیش‌بینی‌های مدل مارکوف تا زمانی که طولش به‌طوری که دنباله مورد نظر در درخت پیدا نشود، ادامه دارد.

احتمالات برای یک مدل مرتبه 0 (P^0) بیانگر احتمال نسبی یک نماد بر اساس وزن یال‌های فرزندش از گره ریشه است که به‌صورت تعداد نرمال‌شده‌ای از تمام نمادهای ممکن در الفبا نشان داده می‌شود. یک مدل مرتبه منهای یک به‌صورت $P^{-1}(X)=1/|\Omega|$ که برای همه استفاده می‌شود، وزن مساوی را به هر نماد در الفبا می‌دهد. برای مثال، در شکل (۷) که از سه نماد A و B و α تشکیل شده‌است مدل مرتبه منهای یک به هر کدام از این نمادها وزن یکسان ۱/۳ اختصاص می‌دهد.

در نهایت، احتمالات به‌دست‌آمده برای هر یک از نمادها، به شکل نزولی طبقه‌بندی می‌شوند و اگر نماد رویداد بعدی مطابق با یکی از نمادهای پیش‌بینی باشد، تجسم درست در نظر گرفته می‌شود.

۶- پیاده‌سازی و نتایج بحث

۶-۱- طراحی آزمایش

در این کار، پردازش تمام هشدارها به‌صورت بلادرنگ به ترتیب زمان (هر هشدار در یک زمان) پیاده‌سازی می‌شود. سپس سیستم نمادهایی ایجاد می‌کند و مجموعه‌های پیش‌بینی گام بعدی برای هر رد حمله را تولید می‌کند. سادگی محاسباتی نسبی در مدل درخت پسوندی و پیش‌بینی‌کننده مدل مارکوف با طول متغیر برای دست‌یابی به عملکرد بلادرنگ در حجم بالای هشدار مهم است.

این کار برای انتخاب داده‌ها برای آزمایش، از چندین مجموعه‌داده آزمایشی جدید در محیط مجازی یا شبیه‌سازی با در نظر گرفتن نیاز برای حقیقت زمینه استفاده می‌کند که در آن هیچ پیش فیلتری از هشدارها انجام نشده است. هشدارها به‌طور طبیعی در هر تعریف فیلد هشدار فیلتر شده‌اند سپس آن‌ها به نمادها تبدیل می‌شوند و به ردها (اگر نماد تکراری نباشد) اضافه می‌شوند.

۶-۱-۱- محیط مجازی SKAION

آزمایش‌های تجسم حمله با استفاده از یک مجموعه داده تولید شده توسط حمله‌های چندمرحله‌ای نوشته شده در یک شبکه مجازی انجام می‌شود. شبکه مجازی شامل هفت زیرشبکه داخلی (هر کدام یک شماره از فضای آدرس کاربر دارند)، ۲۲ سرور خارجی و ۲۴ سرور داخلی است. برای مثال سرورها عبارتند از سرور Web IIS، سرور تبادل MS، سرورهای FTP وVPN که در حال اجرا روی سیستم عامل‌های مختلف لینوکس و ویندوز هستند.

مجموعه‌داده SKAION2 شامل پنج مجموعه از سناریوهای حمله است که یک مجموعه‌ای از ۱۹۹۰۸ هشدار و ۲۵۵۹ رد حمله ایجاد می‌کند. محدوده سناریوهای حمله از سرریز CGI، خروج داده، سایت‌های تقلبی تا انکار سرویس متفاوت است و در اهداف حمله هم متفاوت می‌باشد. پیام‌های هشدار توسط snort، Dragon، Apache و IIS تولید شده است. یک سیستم عامل واقعی باید یک مولفه‌ی پیش‌پردازشی هم ترازی داده داشته باشد که پیام‌های هشدار تولید شده توسط انواع مختلف سیستم‌های تشخیص نفوذ را یکسان‌سازی کند. گام یکسان‌سازی روی مجموعه‌داده مورد استفاده برای این پژوهش انجام نشده است، در عوض، فقط هشدارهای snort مورد استفاده قرار گرفته است.

مجموعه‌داده SKAION3 هشت سناریو بیشتر از SKAION2 روی یک شبکه مشابه فراهم می‌کند.

۶-۱-۲- شبکه شبیه‌سازی شده

ایجاد مجموعه‌های هشدار توسط شبیه‌ساز یک روش جای‌گزین برای سناریوی SKAION مبتنی بر VMWARE است که به نوبه خود اجازه می‌دهد این روش سریع‌تر پیاده‌سازی شود. مدل شبیه‌سازی رویداد گسسته برای تولید داده هشدار سیستم تشخیص نفوذ برای آزمایش ابزارهای آگاهی وضعیتی سایبری توسعه داده شد.

این سیستم تجسم با ۱۰ مجموعه از داده‌ها در این محیط شبیه‌سازی شده مورد آزمایش قرار گرفت و پارامترهای نهانی و کارایی متنوعی برای دقت تجسم آزمایش شدند. هرچه مقدار نهانی بیشتر باشد، تعداد مراحل هدف میانی تولید شده توسط شبیه‌ساز کمتر است. به‌طور مشابه، مقادیر کارایی بالاتر با تعداد مراحل کلی کمتری مرتبط است. یک مهاجم ماهر به احتمال زیاد یک مسیر مستقیم‌تر به هدف (سواستفاده صحیح را انتخاب می‌کند) را پیش‌رو می‌گیرد و رد کمتری که توسط سیستم‌های تشخیص نفوذ شناسایی شده‌اند را از خود برجای می‌گذارد.

۶-۲- آزمایش- دقت پیش‌بینی SKAION2

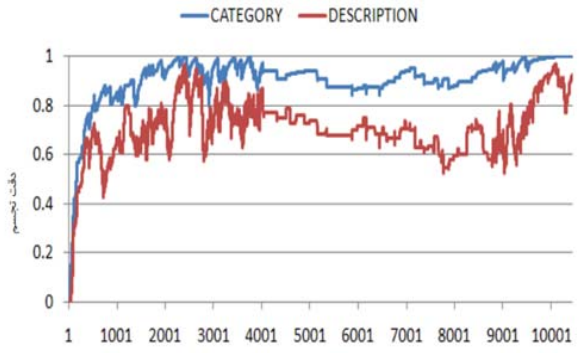
جدول ۳ به‌طور متوسط دقت تجسم در تمام مجموعه‌داده برای توصیف حمله، دسته حمله، پروتکل شبکه و زیرشبکه مقصد که برای تعریف فضای نماد استفاده می‌شود را نشان می‌دهد. دقت تجسم درصدی از نمادهای اتفاق افتاده بعدی است که در یک مجموعه تجسم متشکل از نمادها با بالاترین احتمال با توجه به مدل مارکوف با طول متغیر قرار می‌گیرند؛ بنابراین، در ردیف TOP-3 یک تجسم درست در نظر گرفته می‌شود. اگر رویداد مشاهده شده یکی از سه پیش‌بینی با بالاترین احتمال باشد. انتخاب سه اختیاری است و می‌تواند به هر تعداد معقول تغییر کند تا منعکس کننده تعداد آینده محتمل یک تحلیلگر برای ارزیابی باشد.

اعداد نشان‌داده‌شده در ردیف ایده‌آل بیانگر بهترین دقت پیش‌بینی نظری ممکن برای هر الفبا است؛ به‌عبارت دیگر، پیش‌بینی‌های نادرست برای این مورد ایده‌آل زمانی رخ می‌دهد که یک نماد جدید به‌عنوان رویداد بعدی ظاهر شود. اعداد نشان‌داده‌شده در ردیف ایده‌آل برای مقایسه عملکرد سیستم پیشنهادی زمانی که از تعاریف فضای نماد متفاوت استفاده می‌کند قابل استفاده است. ردیف پیش‌بینی‌ها در جدول ۳ تعداد کل پیش‌بینی‌هایی که سیستم برای هر تعریف الفبا دارد را نشان می‌دهد. از آنجا که فقط انتقال‌ها در ردهای حمله منعکس شده است (و هیچ تکراری منعکس نمی‌شود) این تعداد کمتر از توانایی بالقوه خود که بیشتر از ده هزار هشدار اصلی است، می‌باشد.

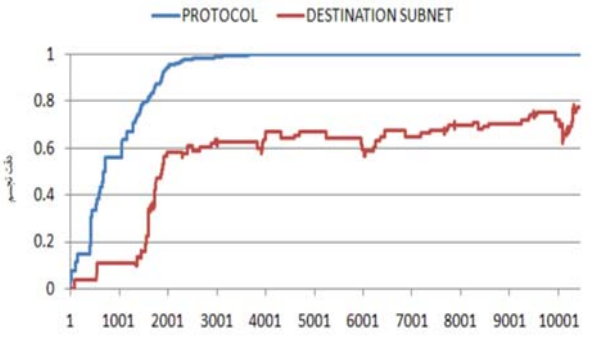
شکل ۸ و ۹ میانگین دقت پیش‌بینی به‌دست‌آمده سیستم را با توجه به تعداد کل هشدارهای تزریق شده نشان می‌دهد. نخست، دوره‌های گذار اولیه است که در آن سیستم به اندازه کافی هشدار برای ساخت یک مدل دقیق را دریافت نکرده است، پس از حدود ۲۰۰۰ هشدار، بیشتر نمادهای الفبا در مجموعه‌داده رخ داده است و نرخ دقت پیش‌بینی نسبت به مقادیر نشان‌داده‌شده در جدول ۴ افزایش می‌یابد. این مجموعه از نتایج نشان می‌دهد که این سیستم هرچه با هشدارهای بیشتری آموزش می‌بیند بهتر اجرا می‌شود. شکل ۱۰، درصد تعداد کل نمادهای شناخته شده که به سیستم تجسم تزریق شده‌اند را نشان می‌دهد. تعداد کلی نمادها در طول میانه‌ی چندین سناریو ثابت است اما سناریوی آخر منجر به نمادهای جدید برای الفباهای زیرشبکه مقصد و توصیفات می‌شود.

جدول ۳- SKAION2 آمار دقت تجسم کلی

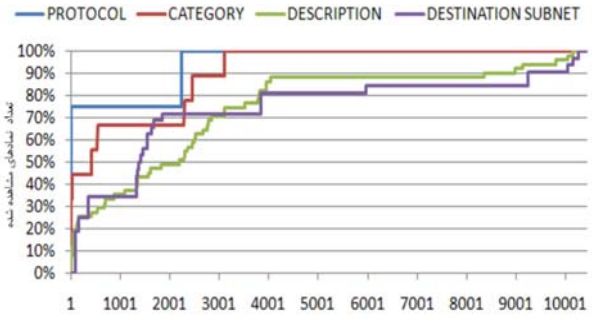
	توصیفات	پروتکل	زیرشبکه مقصد	دسته
TOP-3	٪۷۵,۴	٪۱۰۰	٪۷۱,۱	٪۹۴,۶
ایده آل	٪۹۲,۶	٪۱۰۰	٪۹۲,۶	٪۹۹,۲
تعداد پیش‌بینی‌ها	۱۵۹۶	۳۲۱	۱۲۱	۱۳۱۹
تعداد نمادها	۵۱	۴	۳۳	۹



شکل ۸- SKAION2 دقت دسته و توصیفات



شکل ۹- SKAION۲ دقت زیرشبکه مقصد و پروتکل



شکل ۱۰- SKAION2 درصد نمادهای مشاهده شده

۶-۳- آزمایش- دقت پیش‌بینی SKAION3

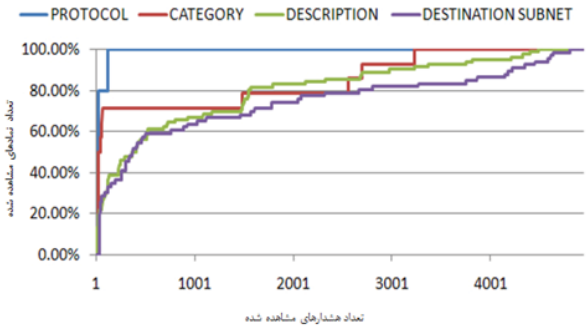
جدول ۴ میانگین دقت تجسم در کل مجموعه‌داده SKAION3 با فضای نماد مشابه با آزمایش اول را نشان می‌دهد (توصیف حمله، دسته حمله، پروتکل شبکه و زیرشبکه مقصد).

جدول ۴- SKAION3 آمار دقت تجسم کلی

	توصیفات	پروتکل	زیرشبکه مقصد	دسته
TOP-3	٪۷۲,۲	٪۹۸,۸	٪۵۵,۱	٪۸۷,۱
ایده آل	٪۹۰,۰	٪۹۸,۸	٪۸۸,۳	٪۹۹,۱
تعداد پیش‌بینی‌ها	۲۱۴۵	۱۷۱	۴۰۳	۱۳۲۵
تعداد نمادها	۸۱	۴	۶۵	۱۳

توجه داشته باشید در حالی که مجموعه داده SKAION3 حاوی هشدار کمتری نسبت به SKAION2 (۴۹۴۴ در مقابل ۱۰۴۲۵) است، پیش‌بینی‌های بیشتری هنگام پردازش SKAION3 ایجاد می‌شود (۲۱۴۵ در مقابل ۱۵۹۶) که نشان‌دهنده هشدارهای تکراری کمتری در مجموعه‌داده SKAION3 است.

شکل ۱۱، درصد تعداد کل نمادهای شناخته‌شده که به سیستم تجسم به‌عنوان هشدار تزریق می‌شوند را نشان می‌دهد. به غیر از پروتکل، هر سناریو همچنان نمادهای جدیدی برای توصیف حمله، دسته حمله و فضای زیرشبکه مقصد دارد.



شکل ۱۱- SKAION3 درصد نمادهای مشاهده شده

نتایج پیش‌بینی توصیفات برای مجموعه داده شبیه‌سازی‌شده در هر سناریو در جدول ۵ نشان داده شده است. نتایج نشان می‌دهد که حملات پنهان برای پیش‌بینی کمی سخت‌تر هستند، از سوی دیگر نشان داده می‌شود که حملات کارامد برای پیش‌بینی آسان‌تر هستند. یک حمله‌ی غیرکارآمد ممکن است رفتار واقعی مهاجمان را با درنظر گرفتن انحرافات غیرضروری، قبل از اقدام برای یک هدف درست منحرف کند.

به‌طور کلی، میزان دقت برای این مجموعه‌داده شبیه‌سازی شده کمتر از SKAION2 و SKAION3 به‌دست آمده‌است. ماهیت این مجموعه‌داده شامل محیط شبیه‌سازی، مدل شبکه امن و پایگاه‌داده بزرگی از حملات نظری در دسترس است که به‌طور مستمر به عنوان نماد توصیفات جدید معرفی شده است که توسط یک میزان دقت ایده‌آل کم منعکس شده است.

علاوه بر پردازش سناریوها به‌طور مستقل، آموزش درخت پسوندی و یادگیری سیستم در سراسر سناریوها انجام شد. این نتایج در جدول ۶ نشان داده شده است. دقت ۳ توصیف بالایی بهبود نیافته است، با این حال نرخ توصیف ایده‌آل بهبود یافته است. این بهبود به احتمال زیاد باتوجه به معرفی اولیه اغلب نمادها توسط اولین سناریوها در این گروه‌ها است.

در شکل ۱۲ مشاهده می‌شود که الفباهای پروتکل، دسته و زیرشبکه مقصد به‌طور کامل پس از دو سناریوی اول تعریف شده‌اند، درحالی‌که هر سناریوی بعدی نمادهای منحصر به‌فردی در الفبای توصیفات معرفی می‌کند.

جدول ۶- شبکه شبیه‌سازی شده: آمار دقت تجسم کلی

	توصیفات	پروتکل	زیرشبکه مقصد	دسته
TOP-3	٪۴۰,۴	٪۹۱,۹	٪۷۸,۵	٪۶۲,۴
ایده آل	٪۸۵,۱	٪۹۲,۱	٪۹۴,۴	٪۹۳,۱
تعداد پیش‌بینی‌ها	۵۲۸۰	۵۲۱	۲۱۱۵	۵۰۳۱
تعداد نمادها	۲۶۸	۴	۸	۱۱

جدول ۵- ماتریس سناریوهای شبیه‌سازی شده برای توصیفات

ویژگی / سناریو	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J
کارآمدی	۱,۰	۰,۸	۰,۶	۰,۴	۰,۲	۱,۰	۰,۸	۰,۶	۰,۴	۰,۲
پنهانی بودن	۰,۸	۰,۸	۰,۸	۰,۸	۰,۸	۰,۴	۰,۴	۰,۴	۰,۴	۰,۴
TOP 3	٪۴۲,۹	٪۴۳,۸	٪۳۸,۵	٪۳۷,۹	٪۳۷,۳	٪۴۶,۶	٪۴۳,۳	٪۴۲,۱	٪۴۱,۸	٪۳۷,۳
ایده‌آل	٪۶۵,۷	٪۶۳,۵	٪۶۴,۵	٪۶۷,۱	٪۷۸,۸	٪۶۸,۲	٪۷۲,۸	٪۶۵,۰	٪۷۴,۹	٪۸۰,۵
هشدارها	۵۰۱	۵۶۴	۶۷۴	۷۵۹	۱۴۳۶	۴۶۳	۵۲۰	۶۴۷	۸۶۴	۱۳۷۴
ردها	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
نمادها	۶۲	۷۲	۷۶	۸۵	۹۲	۵۵	۵۷	۷۷	۶۸	۸۷

۶-۵- آزمایش- پیش‌بینی نمادهای جدید

این پیاده‌سازی برای پیش‌بینی حمله‌ای که قبلا هرگز مشاهده نشده و در حال رخدادن است تلاش می‌کند. سیستم دقیقا پیش‌بینی نمی‌کند که یک حمله جدید اتفاق خواهد افتاد، بلکه پیش‌بینی می‌کند که نماد جدیدی رخ خواهد داد. هنگامی که این موارد رخ می‌دهند، در درخت پسوندی شاخه‌های منجر به این نماد به‌روزرسانی خواهند شد. اگرچه از چنین مواردی تعداد کمی وجود دارد، اما ترکیب این نماد در مدل مارکوف با طول متغیر برای هشدار دادن راجع به یک حمله جدید اجازه می‌دهد. در این آزمایش، سیستم قادر به پیش‌بینی ۸ رخداد از ۵۱ رخداد از روش‌های حمله جدید (توصیفات) برای مجموعه SKAION2 و ۱۶۴ از ۲۶۸ رخداد برای مجموعه شبیه‌سازی شده است.

درصد نتایج که در جدول ۷ نشان داده‌شده به‌صورت موثر دقت پیش‌بینی‌ای که قبلا ارائه شده است را ندارد، اما این نتایج نشان می‌دهد که سیستم نه تنها قادر به آموزش و پیش‌بینی حملاتی که مشاهده شده‌است می‌باشد بلکه هشدارهایی از حملات جدید ارائه می‌کند.

۶-۶- آزمایش- پیش‌بینی با مجموعه داده جدید

در جدول ۸ یک مثال از نتایج الگوریتم تجسم نشان داده شده‌است که درسمت چپ این شکل، هشدارهای واقعی رسیده از سیستم‌های تشخیص نفوذ و در سمت راست هشدارهای پیش‌بینی شده توسط الگوریتم نشان داده شده است. قابل ذکر است که این مثال از حملات اجراشده بر روی شبکه VMWARE که در کل از ۱۱۱۳ دنباله حمله از ۴۷۲۳ هشدار تشکیل شده، ایجاد شده است.

۷- نتیجه‌گیری و محدودیت‌ها

این تحقیق یک روش برای تجسم نفوذهای شبکه مبتنی بر ادغام اطلاعات و مدل‌سازی رفتاری نفوذهای مرتبط‌شده مشاهده شده، پیشنهاد می‌کند. تجسم صفات خاص حمله از جمله توصیف حمله، یک تحلیل از آینده محتمل فراهم می‌کند که قادر به دخالت دستی یا پردازش توسط سیستم‌های دفاعی پیگیری خودکار می‌باشد.

سیستم تجسم حمله سایبری که در اینجا ارائه شده است چند بخش مهم دارد. اولین بخش، پیاده‌سازی مدل‌سازی حمله و الگوریتم‌های تجسم به‌صورت بلادرنگ است که با یک رابط کاربر

جدول ۷- درصد پیش‌بینی نمادهای جدید

	SKAION2	SKAION3	شبیه سازی
توصیفات	٪۱۵,۷	٪۳۳,۳	٪۶۱,۲
پروتکل	٪۰,۰	٪۰,۰	٪۷۵,۰
زیرشبکه مقصد	٪۰,۰	٪۱۰,۸	٪۵۰,۰
دسته	٪۰,۰	٪۰,۰	٪۳۶,۴

جدول ۸- مثال پیش‌بینی هشدارها

هشدارهای پیش‌بینی شده	هشدارهای تولید شده توسط اقدامات حمله
1.(no prediction)	1. K (http_inspect) Oversize Request-URI Directory
2. J WEB-MISC Invalid HTTP Version String	2. F (http_inspect) Bare Byte Unicode Encoding
3. J WEB-MISC Invalid HTTP Version String	3. A ICMP PING NMAP
4. H ICMP L3retriever Ping	4. H ICMP L3retriever Ping
5. A ICMP PING NMAP	5. J WEB-MISC Invalid HTTP Version String
6. J WEB-MISC Invalid HTTP Version String	6. J WEB-MISC Invalid HTTP Version String
7. J WEB-MISC Invalid HTTP Version String	7. A ICMP PING NMAP
8. H ICMP L3retriever Ping	8. H ICMP L3retriever Ping
9. F (http_inspect) Bare Byte Unicode Encoding	9. H ICMP L3retriever Ping
10. I NETBIOS SMB-DS ILC$ unicode share access	10. I NETBIOS SMB-DS ILC$ unicode share access
11. A ICMP PING NMAP	11. A ICMP PING NMAP
12. H ICMP L3retriever Ping	12. H ICMP L3retriever Ping

گرافیکی و موتور شبیه‌سازی بلادرنگ تکمیل می‌شود، همچنین سیستم می‌تواند نتایج شبیه‌سازی رویداد محور برای مجموعه داده‌های ضبط شده را فراهم کند. یادگیری به‌طور مستمر، سیستم را قادر می‌سازد تا با تکنیک‌های حمله جدید منطبق شود و الگوها به خوبی با معماری شبکه تغییر کنند. توسعه این سیستم تجسم بلادرنگ در یک قسمت کلیدی از معماری امنیتی سایبر کلی قرار خواهد گرفت.

برای کارهای آینده، روش ارزیابی تجسم فعلی که برای تعیین صحت، تنها به گام بعدی در یک رد حمله توجه می‌کند می‌تواند برای آینده‌های تجسم شده مورد تجدیدنظر قرار گیرد که در یک دلتای زمانی معقول رخ دهد. همچنین، نرخ دقت پیش‌بینی و خصوصیات داده‌ها نیز وابسته به واقع‌گرایی از آزمایش مجموعه‌داده‌ها می‌باشد. در نهایت اگرچه آزمایش داده‌ها به‌منظور مدل کردن شبکه‌های واقعی و حملات است، اما سناریوهای واقعی متفاوت هستند و نمی‌توانند یک نمایش دقیق را تضمین کنند.

۸- مراجع

۱. داداش تبار، کوروش؛ رشیدی، علی‌جبار؛ شیرازی، "ارائه الگوی برای بهبود آگاهی وضعیتی سایبری مبتنی بر ادغام اطلاعات در جنگ شبکه مدار، " ارائه شده در ششمین کنفرانس ملی جنگ الکترونیک، دانشگاه جامع امام حسین (ع)، (۱۳۹۲).

۲. داداش تبار، کوروش؛ رشیدی، علی‌جبار؛ شیرازی، "ارائه مدلی برای تجسم حملات سایبری در چارچوب آگاهی وضعیتی سایبری، " ارائه شده در دومین کنفرانس ملی دفاع سایبری، دانشگاه جامع امام حسین (ع)، اردیبهشت (۱۳۹۳).

3. S. R. Byers and S. J. Yang, "Real-time fusion and projection of network intrusion activity," presented at the Information Fusion, 2008 11th International Conference on, pp. 1–8, **(2008).**

4. T. Bass, "Intrusion detection systems and multi-sensor data fusion," *Commun. ACM*, vol. 43, no. 4, pp. 99–105, **(2000)**.

5. E. P. Blasch, D. Lambert, P. Valin, M. M. Kokar, J. Llinas, S. Das, C. Chong, and E. Shahbazian, "High level information fusion (hlif): Survey of models, issues, and grand challenges," *Aerosp. Electron. Syst. Mag. IEEE*, vol. 27, no. 9, pp. 4–20, **(2012)**.

6. F. Cuppens and A. Miege, "Alert correlation in a cooperative intrusion detection framework," presented at the Security and Privacy, 2002 IEEE Symposium on, pp. 202–215, **(2002)**.

7. P. Ning, Y. Cui, and D. S. Reeves, "Constructing attack scenarios through correlation of intrusion alerts," presented at the Proceedings of the 9th ACM conference on Computer and communications security, pp. 245–254, **(2002)**.

8. F. Valeur, G. Vigna, C. Kruegel, and R. A. Kemmerer, "Comprehensive approach to intrusion detection alert correlation," *Dependable Secure Comput. IEEE Trans. On*, vol. 1, no. 3, pp. 146–169, **(2004)**.

9. P. Ning, Y. Cui, D. S. Reeves, and D. Xu, "Techniques and tools for analyzing intrusion alerts," *ACM Trans. Inf. Syst. Secur. TISSEC*, vol. 7, no. 2, pp. 274–318, **(2004)**.

10. D. Xu and P. Ning, "Alert correlation through triggering events and common resources," presented at the Computer Security Applications Conference 2004, 20th Annual, pp. 360–369, **(2004)**.

11. M. Sudit, A. Stotz, and M. Holender, "Situational awareness of a coordinated cyber attack," presented at the Defense and Security, pp. 114–129, **(2005)**.

12. S. T. King, Z. M. Mao, D. G. Lucchetti, and P. M. Chen, "Enriching Intrusion Alerts Through Multi-Host Causality," presented at the NDSS, **(2005)**.

13. A. Valdes and K. Skinner, "Probabilistic alert correlation," presented at the Recent Advances in Intrusion Detection, pp. 54–68, **(2001)**.

14. S. Mathew, D. Britt, R. Giomundo, S. Upadhyaya, M. Sudit, and A. Stotz, "Real-time multistage attack awareness through enhanced intrusion alert clustering," presented at the Military Communications Conference, pp. 1801–1806, **(2005)**.

15. J. J. Salerno, M. Sudit, S. J. Yang, G. P. Tadda, I. Kadar, and J. Holsopple, "Issues and challenges in higher level fusion: Threat/impact assessment and intent modeling (a panel summary)," presented at the Information Fusion (FUSION), 2010 13th Conference on, pp. 1–17, **(2010)**.

16. S. J. Yang, A. Stotz, J. Holsopple, M. Sudit, and M. Kuhl, "High level information fusion for tracking and projection of multistage cyber attacks," *Inf. Fusion*, vol. 10, no. 1, pp. 107–121, **(2009)**.

17. X. Qin and W. Lee, "Attack plan recognition and prediction using causal networks," presented at the Computer Security Applications Conference, pp. 370–379, **(2004)**.

18. J. Holsopple, S. J. Yang, and M. Sudit, "TANDI: Threat assessment of network data and information," presented at the Defense and Security Symposium, pp. 62420–62429, **(2006)**.

19. A. Arnes, F. Valeur, G. Vigna, and R. A. Kemmerer, "Using hidden Markov models to evaluate the risks of intrusions: system architecture and model validation," *Lect. Notes Comput. Sci.*, pp. 145–164, **(2006)**.

20. D. Fava, J. Holsopple, S. J. Yang, and B. Argauer, "Terrain and behavior modeling for projecting multistage cyber attacks," presented at the Information Fusion, 2007 10th International Conference on, pp. 1–7, **(2007)**.

21. D. Fava, "Characterization of cyber attacks through variable length markov models," **(2007)**.

22. G. Toth, M. M. Kokar, K. Wallenius, K. B. Laskey, M. Sudit, M. Hultner, and O. Kessler, "Higher -level information fusion: Challenges to the academic community," presented at the Proceedings of the 11th International Conference on Information Fusion, Cologne, Germany, **(2008)**.

23. E. Blasch, E. Bosse, and D. A. Lambert, " High-Level Information Fusion Management and Systems Design," Artech House, **(2012)**.

24. M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” Hum. Factors J. Hum. Factors Ergon. Soc., vol. 37, no. 1, pp. 32–64, **(1995)**.

25. L. E. Chase, “Integration of Cyber Situational Awareness into System Design and Development,” DTIC Document, **(2009)**.

26. J. Holsopple, S. J. Yang, M. Kuhl, D. Hall, R. Nagi, S. Shapiro, M. Sudit, B. Panulla, M. Kandefer, and P. Seyed, “National Center for Multisource Information Fusion,” DTIC Document, **(2009)**.

A New Framework for Projection of Cyber-Attacks Based On Information Fusion

A. J. Rashidi [1]
K. Dadashtabar Ahmadi[2]
M. Akbari [3]

Abstract

The main objective of projection algorithms is to improve the performance of cyber situational awareness. This improvement could be performed by removing the existing gap between intrusion-detection systems and cyber defense mechanisms. In this paper, we propose a system that could be able to project the cyberspace future state, improve the cyber situational awareness and finally execute the prevention action against the potential threats. The proposed system is based on high level information fusion. Projecting attack tracks without any requirement to prior knowledge about network architecture or static attack guidance templates is another considerable capability of this system. The system consists of two separate parts: information fusion and cyber-attack projection. In second part of the system, a new method for projecting multi-stage attacks have been proposed that enable the cyberspace analysts to prioritize attacks and perform appropriate actions against future attack steps. Finally, the system performance is evaluated by different criteria such as accuracy.

Key Words: *Cyber-attacks, information fusion, situational awareness, projection, variable length markov model*

1-Assistant Professor of Malek Ashtar University
2-Instructor of Malek Ashtar University
3- MS Candidate of Malek Ashtar University (akbarimanochehr@yahoo.com) - Writer-in-Charge