مقایسه روش های ماشین بردار و شبکه های عصبی در سیستم های تشخیص نفوذ

نویسنده: امین داستانپور – amindastanpour@gmail.com

چکیده: استفاده از سیستم‌های شبکه‌ای یکی از روند تکنولوژی در سال‌های اخیر می‌باشد، و انتظار استفاده‌کنندگان از شبکه دارا بودن شبکه امن برای پردازش داده‌ها و ارتباطات می‌باشد.در این خصوص تشخیص نفوذ به‌عنوان یکی از مسائل مهم در امنیت شبکه برای پژوهشگران بوده است. یکی از روش‌های تشخیص نفوذ استفاده از الگوی رفتاری شبکه می‌باشد که از یادگیری ماشین برای تشخیص نفوذ در رفتار شبکه استفاده می‌شود.در این مقاله روش‌های یادگیری ماشین زیر برای تشخیص حملات شبکه استفاده می شود: ماشین بردار (***SVM***) و شبکه‌های عصبی مصنوعی (***ANN***). هدف از این مطالعه مقایسه نتایج ماشین بردار و شبکه‌های عصبی مصنوعی و سپس نتایج ماشین بردار و شبکه‌های عصبی مصنوعی با سایر الگوریتم‌های استفاده‌شده برای تشخیص نفوذ که اخیراً پژوهشگران مورداستفاده قرار داده‌اند را شامل می‌شود. برای دستیابی به نتایج، در این مقاله، (***KDD CPU99***) برای ارزشیابی روش‌ها مورداستفاده قرارگرفته است.

کلیدواژه: امنیت شبکه , تشخیص نفوذ,عصبی مصنوعی شبکه , ماشین بردار پشتیبانی , یادگیری ماشین

۱ – مقدمه

امروزه اینترنت رایج‌ترین وسیله ارتباطی برای مردم است. درنتیجه، هر کس یک کانال یا شبکه امن را برای اهداف ارتباطی خود انتظار دارد. در سال‌های اخیر مطالعات متعددی در ارتباطات امن برای ایمن‌سازی داده‌های و ارتباطات انجام‌شده است[۱]. سیستم تشخیص نفوذ (IDS) یکی از ابزارهایی است که توسط مدیران شبکه برای محافظت از شبکه‌ها علیه فعالیت‌های ناشناخته اعمال می‌شود. محدودیت سیستم‌های امنیتی شبکه این است که تنها قادر به شناسایی حملات از قبلاً شناخته‌شده است و لازم است که اغلب این پایگاه داده‌ها به‌روزرسانی شود [۲]. علاوه بر این، ویژگی‌های زیادی را باید موردتوجه قرار داشته باشد تا بتواند امنیت شبکه را ایجاد کند پس این امر به ترافیک زیادی در شبکه منجر می‌شود و توزیع داده‌ها به‌شدت ناسازگار خواهد بود. بنابراین،در شناخت رفتار طبیعی شبکه امری است چالش پذیر. روش‌های مختلف هوش مصنوعی برای غلبه بر این مشکلات خاص استفاده‌شده است [۳]. هدف از یادگیری ماشین، بهبود عملکرد دستگاه ، یادگیری و کشف شرایطی است که احتمالاً در طول زمان تغییر می‌کند.

در زمینه تشخیص نفوذ، الگوریتم‌های یادگیری ماشین، ورودی مرجع را برای یادگیری الگوهای حملات را اعمال می‌کند و سپس، الگوریتم‌ها بر اساس حملات ناشناخته برای انجام تشخیص واقعی در سیستم‌های شبکه اعمال می‌شوند [۴]. علاوه بر توانایی چنین الگوریتمی‌هایی, برای شناخت الگوهای جدید حمله در شبکه بسیار مناسب عمل می‌کنند به این صورت که توانایی شناخت حملات جدید را بر اساس داشتن دانش قبل به انجام می‌رسانند. همچنین آن‌ها قادر به پاک‌سازی مجموعه داده‌ها با ویژگی‌های نامناسب و غیرفعال را دارا می‌باشند. توانایی ذکرشده منجر به دستیابی به فرآیند تشخیص بهینه‌شده و تنها شامل چندین ویژگی اصلی در مجموعه داده است [۵]. هدف از یادگیری ماشین کشف، یادگیری، و سپس سازگاری با شرایطی است که ممکن است تغییر کند، بنابراین بهبود عملکرد دستگاه یکی از مهم‌ترین چالشی پژوهشی هست. در زمینه تشخیص نفوذ، ورودی ارجاعات در الگوریتم‌های یادگیری ماشین به کار می‌رود تا بتوانند الگوهای حمله را بیاموزند. پس‌ازآن، الگوریتم‌ها در حملات ناشناخته برای انجام فرآیند تشخیص واقعی موردبررسی قرار می‌گیرند. علاوه براین, این الگوریتم‌ها توانایی تشخیص الگوی جدید حمله را نیز دارا می‌باشند. آن‌ها همچنین می‌توانند

www.kahrobaonline.ir

داده‌های مجموعه را از ویژگی‌های نامناسب و غیرفعال پاک کنند و درنتیجه تنها ویژگی‌های نامناسب و غیرفعال پاک کنند و درنتیجه تنها ویژگی‌های کلیدی در مجموعه داده‌ها وجود دارد را در فرآیند تشخیص بهینه‌سازی کنند [۶] برخی از روش‌های مهم یادگیری ماشین عبارتند از شبکه عصبی مصنوعی (ANN)، ماشین بردار (SVM).

شبکه عصبی مصنوعی (ANN) یکی از محبوب‌ترین تکنیک‌های یادگیری ماشین است و برای حل مسائل طبقه‌بندی و رگرسیون مورداستفاده قرار می‌گیرد [۷]. شبکه عصبی مصنوعی توانایی یادگیری ماشین و شناخت الگوهای را دارا هست. به‌عنوان‌مثال، در تشخیص سیستم عصبی، داده‌های ورودی ممکن است به فعال کردن مجموعه‌ای از نورون‌های شبکه عصبی مصنوعی در قسمت ورودی برای نشان دادن حمله و یا یک فعالیت نرمال منجر شود. شبکه عصبی مصنوعی دارای تعدادی مزیت است, یکی از این مزایا شناخته‌شده به‌عنوان‌یکی از توانایی آن‌ها در یادگیری از مشاهده مجموعه داده‌ها هست که در این روش شبکه عصبی مصنوعی به‌عنوان‌یک ابزار تقریبی برای توابع تصادفی استفاده می‌شود [۸]. شبکه‌های عصبی دارای سه سطح ارتباطات هستند, لایه اول نورون‌های ورودی است. این نورون‌ها داده‌ها را به لایه بعدی ارسال می‌کنند. دومین لایه نتیجه نورون‌ها را در لایه سوم می‌فرستد و لایه سوم خروجی است که کاربرد شبکه عصبی برای طبقه‌بندی و سازمان‌دهی مجدد داده‌ها استفاده می‌شود [۹].

یکی از محبوب‌ترین روش‌های یادگیری ماشین, ماشین بردار است و برای حل مشکلات رگرسیون و طبقه‌بندی مورداستفاده گرفته است [۱۰]. برای هر یک از داده‌های ورودی داده‌شده در یک ماشین بردار فقط یک سری داده‌های ورودی خاص را می‌پذیرد و سپس پیش‌بینی می‌کند که خروجی توسط کدام‌یک از دو کلاس احتمال دارد که ایجاد بشود (همچنین به‌عنوان طبقه‌بندی خطی باینری شناخته می‌شود). با توجه به مجموعه‌ای از نمونه‌های آموزشی، هرکدام به‌عنوان‌یکی از دودسته (Attack یا Normal) می‌باشند که در تشخیص حمله، بردار ماشین مسئول پیش‌بینی اینکه آیا داده‌های جدید در دسته داده‌های عادی یا گروه حمله قرارگرفته‌اند [۱۱]. پس درنتیجه بردار ماشین در سازمان‌دهی مجدد و طبقه‌بندی داده‌ها بسیار مفید است.

۲- پیشینه تحقیق

در مطالعات قبلی محققان تلاش کرده‌اند با استفاده از روش‌های مختلف مانند FASVFG, LCFS ، FFSA , MMIFS و الگوریتم کلونی مورچه این مشـــکل را حل کنند. جدول ۱ ا این روش ها را به‌طور خلاصـــه بیان می‌کند.

جدول ۱: پیشینه پژوهشی

نویسنده	روش کار	اهداف
Bin Luo et al. [12]	رویکرد تولید ویژگی تجسم مبتنی بر چهار زاویه‌ای (FASVFG)	ارزشیابی فاصله بین نمونه‌ها در ۵ کلاس طبقه‌بندی‌شده
Abraham et al.[13]	طبقه‌بندی‌های مبتنی بر قانون فازی	چارچوب سیستم‌های تشخیص نفوذ توزیع‌شده (DIDS)
Amiri et al.[14]	الگوریتم انتخاب پیشرو(FFSA) انتخاب ویژگی همبستگی خطی (LCFS) تغییر ویژگی اطلاعات متقابل (MMIFS)	پیشنهاد یک مرحله انتخاب ویژگی، که عموماً در هر تشخیص نفوذ قابل‌اجرا است
Li et al.[15]	الگوریتم کلونی مورچه	یک مدل مطلوب IDS با راندمان بالا و دقت بالا ارائه می‌دهد
Dastanpour et al.[16]	پیشنهاد یک انتخاب ویژگی بر اساس الگوریتم ژنتیک	بهبود میزان تشخیص با تعداد کمتر از ویژگی‌های

۳- تحلیل داده‌ها

در این مطالعه، داده‌کاوی (KDD CUP 1999) برای ارزشیابی روش‌های ذکرشده در قسمت قبل مورداستفاده قرارگرفته است [۱۷]. این مجموعه داده به دلیل جامع بودن آن استفاده‌شده است و همچنین بهترین مجموعه داده برای بررسی عملکرد IDS به شمار می‌رود [۱۸]. ۲۲ نوع حمله در ای داده‌کاوی وجود دارد که می‌توان آن‌ها را به ۴ گروه تقسیم کرد: پروب کردن، U2R، R2L، و DOS با جزئیات زیر: [۱۹]

www.kahrobaonline.ir

پروب کردن : شبکه توسط این نوع حملات جمع‌آوری داده‌ها در مورد میزبان موردنظر هست.

U2R: دسترسی غیرمجاز به ریشه (کاربر فوق‌العاده محلی). این حملات به‌عنوان حملاتی شناخته می‌شود که مهاجم می‌تواند به سیستم دسترسی پیدا کند و از آسیب‌پذیری‌ها برای به دست آوردن مجوزهای کلیدی بهره‌برداری کند.

R2L: دسترسی غیرمجاز از یک دستگاه از راه دور. در این نوع حمله، برخی از بسته‌ها در شبکه برای دسترسی به شبکه به‌عنوان‌یک کاربر شناخته‌شده و محلی ارسال می‌شوند.

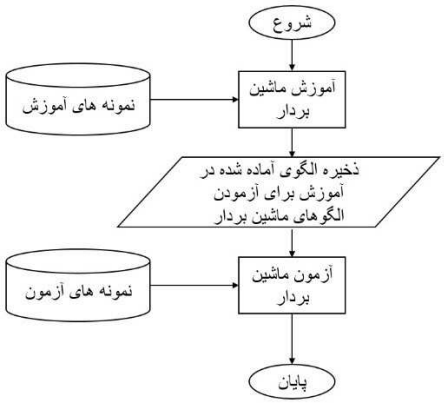
DOS: انکار سرویس این نوع حمله برای درک رفتار کاربر استفاده می‌شود. این نوع حمله نیاز به‌صرف برخی از منابع محاسباتی و حافظه دارد.

۴- روش کار

در این قسمت شرح مفصلی از طراحی و پیاده‌سازی سیستم تشخیص نفوذ توسط سیستم‌های ماشین بردار و شبکه عصبی ارائه می‌گردد. طراحی این روش شامل الگوریتم یادگیری ماشین برای تشخیص است و به دنبال آن‌یک توضیح مختصر از اجرای آن. بعضی از فرایندهای اولیه موردنیاز برای انجام وظایف خود که در این قسمت بحث می‌شود را دارا هست.

۴-۱- ماشین بردار

SVM یک روش یادگیری ماشین است که بر اساس نظریه‌ی آماری VC و اصل کمینه‌سازی ریسک ساخته‌شده است و می‌تواند مشکلات عملی غیرخطی و حداقل‌های محلی را حل کند. SVM می‌تواند مشکل فضای چندبعدی را به مسئله فضای کم بعدی تبدیل کند. پس بنابراین، هنگام برخورد با مسئله فضایی چندبعدی خطی طبقه‌بندی‌شده می‌تواند برای طبقه‌بندی بهینه استفاده شود. سیستم SVM دو جنبه دارد. یک جنبه آموزش است که در این مرحله مدل SVM با آموزش داده‌های نمونه به دست می‌آید. جنبه دیگر آزمون است که در این مرحله، اطلاعات دیگر شناسایی و طبقه‌بندی‌شده توسط مدل SVM است. نمودار جریان کلی سیستم SVM به شکل ۱ نشان داده‌شده است.

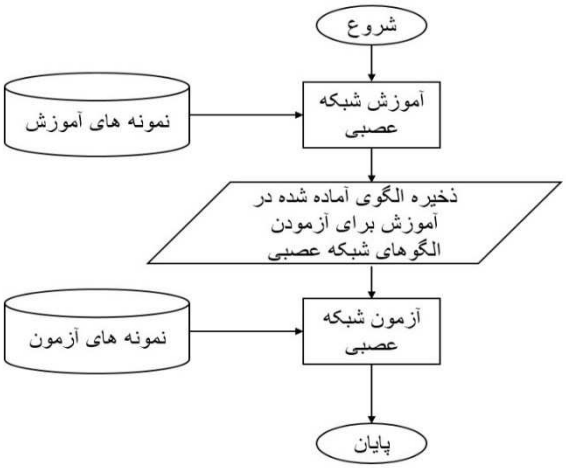


شکل ۱ : نمودار جریان کلی سیستم SVM

۴-۲- شبکه عصبی

در یادگیری ماشین و علم شناختی، ANN ها خانواده‌ای از الگوریتم‌های یادگیری آماری هستند که الهام گرفته از بیولوژیکی (سیستم‌های عصبی مرکزی حیوانات، به‌ویژه مغز) هستند و برای تخمین یا تقریبی توابع استفاده می‌شود که می‌تواند بر تعداد زیادی از ورودی‌ها بستگی داشته باشد. ANN به‌طورکلی به‌عنوان سیستم‌های "نورون" به همدیگر متصل می‌شود که می‌توانند ارزش‌ها را از ورودی‌ها محاسبه کنند و توانایی یادگیری ماشین و همچنین تشخیص الگو را دارند. مانند سایر روش‌های یادگیری ماشین سیستم‌هایی که از داده‌ها یاد گرفته‌اند را برای حل طیف وسیعی از وظایف مورداستفاده قرار می‌دهند که برای حل آن‌ها با استفاده از برنامه‌ریزی عادی مبتنی بر قانون سخت است. شکل ۲ الگوریتم ANN را نشان می‌دهد.

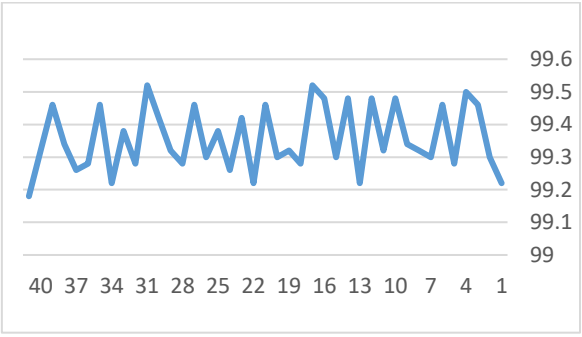
www.kahrobaonline.ir



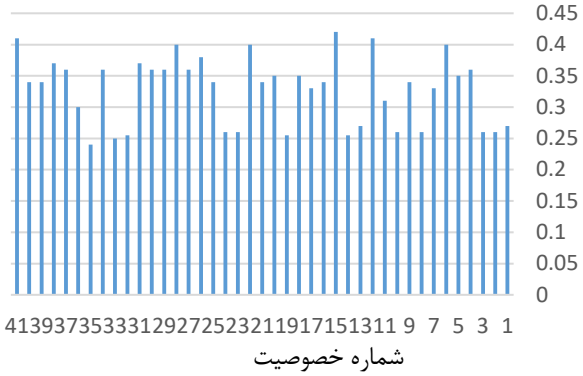
شکل ۲ : نمودار جریان کلی سیستم ANN

۵- بحث و نتیجه‌گیری

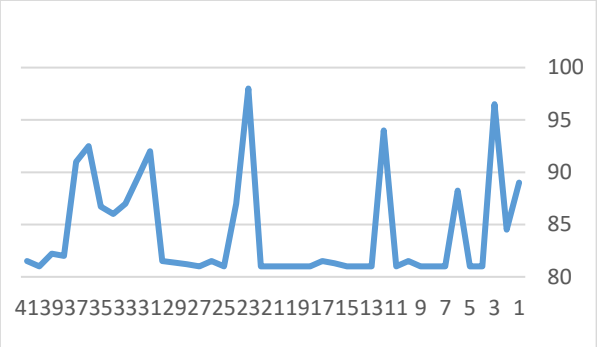
در این مقاله ابتدا SVM و ANN برای طبقه‌بندی و به رسمیت شناختن داده‌ها به دو گروه عادی و حمله مورداستفاده قرار می‌گیرند و سپس مقایسه با یکدیگر در تشخیص نفوذ و تشخیص اشتباه مورد مقایسه قرار می‌گیرند. و در آخر با الگوریتم‌های مورد ذکر در قسمت قبل که در بیشینه بیان‌شده است را مورد مقایسه قرار می‌گیرند. در شکل ۳ ماشین بردار در تشخیص نفوذ و در شکل ۴ تشخیص اشتباه ماشین بردار نمایش داده می‌شود. . در شکل ۵ شبکه عصبی در تشخیص نفوذ و در شکل ۶ تشخیص اشتباه شبکه عصبی نمایش داده می‌شود. . در شکل ۷ ماشین بردار و شبکه عصبی در تشخیص نفوذ و در شکل ۸ تشخیص اشتباه ماشین بردار و شبکه عصبی مورد مقایسه قرار می‌گیرند.



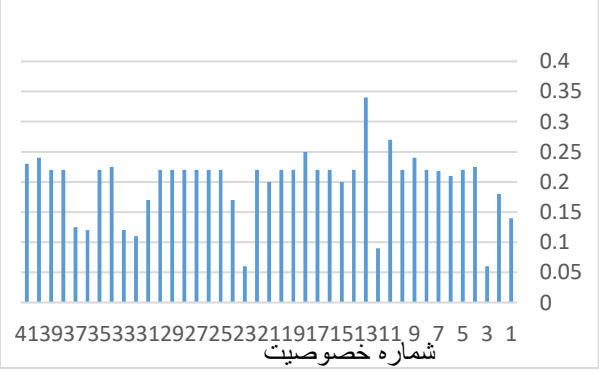
شکل ۳ : روند تشخیص نفوذ ماشین بردار



شکل ۴ : روند تشخیص نفوذ اشتباه ماشین بردار(محور عمودی درصد تشخیص اشتباه است)

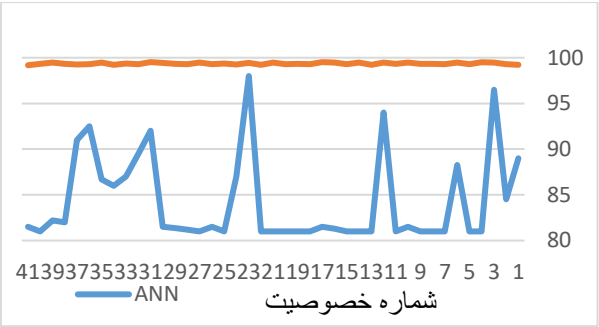


شکل ۵ : روند تشخیص نفوذ شبکه عصبی(محور عمودی درصد تشخیص اشتباه است)

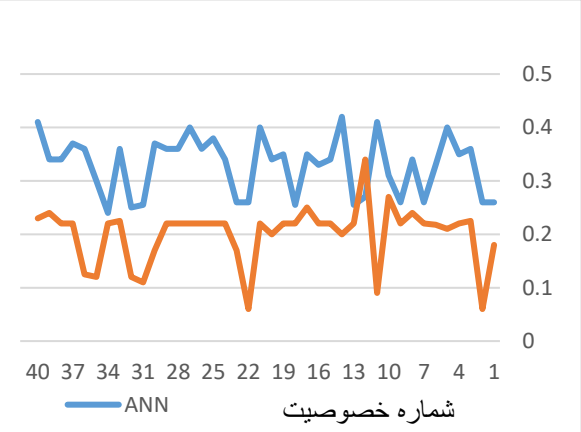


شکل ۶ : روند تشخیص نفوذ اشتباه شبکه عصبی(محور عمودی درصد تشخیص اشتباه است)

www.kahrobaonline.ir



شکل ۷ : مقایسه ماشین بردار و شبکه عصبی در پارامتر تشخیص نفوذ(محور عمودی درصد تشخیص اشتباه است)



شکل ۸ : مقایسه ماشین بردار و شبکه عصبی در پارامتر تشخیص اشتباه

در جدول ۲ با مقایسه الگوریتم‌ها نشان می‌دهد که ماشین بردار باعث عملکرد بهتر با تعداد کمتر از تشخیص اشتباه هست. هنگامی‌که SVM با ANN مقایسه می‌شوند، می‌توان تشخیص داد که ماشین بردار عملکرد نسبتاً بهتری در مقایسه با شبکه عصبی در داده‌های KDD CUP 99 دارا هست. اگرچه با استفاده از الگوریتم‌های دیگر می‌توان میزان تشخیص بالا را به دست آورد،ولی در ماشین بردار می‌توانند به میزان پایین تشخیص با تعداد کمتر از تشخیص اشتباه روبه‌رو شد.

جدول ۲ : مقایسه الگوریتم‌های ماشین بردار و شبکه عصبی با دیگر الگوریتم‌های استفاده‌شده برای سیستم‌های تشخیص نفوذ

اسم روش استفاده‌شده	درصد تشخیص	درصد تشخیص اشتباه
LCFS	٪۹۷,۹۹	٪۰,۶۴
FFSA	٪۹۷,۹۲	٪۰,۴۱
MMIFS	٪۹۷,۸۹	٪۰,۵۲
fuzzy rule based	٪۹۷,۹۸	٪۰,۸۱
FASVFG	٪۹۴	٪۱,۳
ANN	٪۹۸	٪۰,۲۵
SVM	٪۹۹,۴۶	٪۰,۰۶

در این مطالعه SVM و ANN برای سیستم تشخیص نفوذ استفاده می‌شود و در مقایسه با یکدیگر برای نشان دادن کارایی بالاتر با درصد تشخیص اشتباه پایین‌تر موردمطالعه قرارگرفته‌اند. نتایج نشان می‌دهد که در مقایسه با روش‌های دیگر، بیشترین میزان تشخیص توسط ماشین بردار به‌دست‌آمده. در این مطالعه، مجموعه‌ای از آزمایش‌ها با استفاده از مجموعه داده KDD cup 99 برای تشخیص چهار نوع حملات شبکه انجام شد. انتخاب ماشین بردار برای میزان تشخیص مناسب در سیستم تشخیص نفوذ پیشنهاد می‌شود. برای شناسایی حملات مؤثر با اشتباه کمتر، SVM با ۹۹,۴۶ درصد تشخیص و با ۰,۰۶ درصد تشخیص اشتباه در بهترین روش با مقایسه دیگر روش‌ها در دیتابیس KDD CUP 99 و ANN با ۹۸درصد تشخیص و با ۰,۲۵ درصد تشخیص اشتباه در رده دوم برترین روش با مقایسه دیگر روش‌ها در دیتابیس KDD CUP 99 قرار دارد.

مراجع

[1] A. Ayyagari, T. M. Aldrich, D. E. Corman, G. M. Gutt, and D. A. Whelan, "Context aware network security monitoring for threat detection," ed: Google Patents, 2015.

[2] S. Scott-Hayward, G. O'Callaghan, and S. Sezer, "SDN security: A survey," in Future Networks and Services (SDN4FNS), 2013 IEEE SDN For, 2013, pp. 1-7.

[3] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE

www.kahrobaonline.ir

visualization strategy," Expert Systems with Applications, vol. 41, pp. 4139-4147.۲۰۱۴ ,

[13] A. Abraham, R. Jain, J. Thomas, and S. Y. Han, "D-SCIDS: Distributed soft computing intrusion detection system," Journal of Network and Computer Applications, vol. 30, pp. 81-98, 2007.

[14] F. Amiri, M. R. Yousefi, C. Lucas, A. Shakery, and N .Yazdani, "Mutual information-based feature selection for intrusion detection systems," Journal of Network and Computer Applications, vol. 34, pp. 1184-1199, 2011.

[15] W. Li, "Using genetic algorithm for network intrusion detection," Proceedings of the United States Department of Energy Cyber Security Group, vol. 1, pp. 1-8, 2004.

[16] A. Dastanpour and R. A. R. Mahmood, "Feature selection based on genetic algorithm and SupportVector machine for intrusion detection system," in The Second International Conference on Informatics Engineering & Information Science (ICIEIS2013), 2013, pp. 169-181.

[17] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89 ,pp. 117-123, 2016.

[18] I. Levin, "KDD-99 classifier learning contest: LLSoft's results overview," SIGKDD explorations, vol. 1, pp. 67-75, 2000.

[19] M. K. Siddiqui and S. Naahid, "Analysis of KDD CUP 99 dataset using clustering based data mining," International Journal of Database Theory and Application, vol. 6, pp. 23-34, 2013.

Communications Surveys & Tutorials, vol. 18, pp. 1153-11.۲۰۱۶ ,۷۶

[4] J. Singh and M. J. Nene, "A survey on machine learning techniques for intrusion detection systems," International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, pp. 4349-4355, 2013.

[5] C. Modi, D. Patel, B .Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," Journal of network and computer applications, vol. 36, pp. 42-57, 2013.

[6] N. A. Alrajeh and J. Lloret, "Intrusion detection systems based on artificial intelligence techniques in wireless sensor networks," International Journal of Distributed Sensor Networks, vol. 9, p. 351047, 2013.

[7] A. Dastanpour, S. Ibrahim, and R. Mashinchi, "Using genetic algorithm to supporting artificial neural network for intrusion detection system," in The international conference on computer security and digital investigation (ComSec2014), 2014, pp. 1-13.

[8] A. Dastanpour, S. Ibrahim, R. Mashinchi, and A. Selamat, "Comparison of genetic algorithm optimization on artificial neural network and support vector machine in intrusion detection system," in Open Systems (ICOS), 2014 IEEE Conference on, 2014, pp. 72-77.

[9] E. Hodo, X. Bellekens, A. Hamilton, P.-L. Dubouilh, E. Iorkyase, C. Tachtatzis, et al., "Threat analysis of IoT networks using artificial neural network intrusion detection system," in Networks, Computers and Communications (ISNCC), 2016 International Symposium on, 2016, pp. 1-6.

[10] R. A. Amin Dastanpour, "Feature Selection Based on Genetic Algorithm and Support Vector Machine for Intrusion Detection System," The Society of Digital Information and Wireless Communications (SDIWC), 2013.

[11] A. Dastanpour, S. Ibrahim, R. Mashinchi, and A. Selamat, "Using gravitational search algorithm to support artificial neural network in intrusion detection system," SmartCR, vol. 4, pp. 426-434, 2014.

[12] B. Luo and J. Xia, "A novel intrusion detection system based on feature generation with

www.kahrobaonline.ir