Research Article, 2022, Volume 9, Number 1
http://www.jcomsec.org

Journal of Computing and Security

Persian Abstract

ارزیابی عملکرد الگوریتم‌های Apache Spark MLlib بر روی یک مجموعه داده‌های تشخیص نفوذ

رامین عاطفی نیا [آ]، محمود احمدی [آ,*]

[آ] گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه رازی، ایران.

اطلاعات مقاله

تاریخچه مقاله:
دریافت: 9 November 2021
اصلاح: 26 May 2022
پذیرش: 31 May 2022
انتشار آنلاین: 4 July 2022

کلمات کلیدی:
سیستم‌های تشخیص نفوذ، Apache Spark، MLlib، یادگیری ماشین

چکیده

افزایش استفاده از اینترنت، سرویس‌های وب و ظهور نسل پنجم فناوری شبکه سلولی (5G) همراه با ترافیک روزافزون اینترنت اشیا (IoT) باعث رشد استفاده جهانی از اینترنت خواهد شد. برای اطمینان از امنیت شبکه‌های آینده، سیستم‌های تشخیص نفوذ و پیشگیری مبتنی بر یادگیری ماشین (IDPS) باید برای شناسایی حملات جدید پیاده‌سازی شوند و ابزارهای پردازش موازی داده‌های بزرگ می‌توانند برای مدیریت مجموعه عظیمی از داده‌های آموزشی در این سیستم‌ها استفاده شوند. در این مقاله Apache Spark، یک پلت‌فرم محاسباتی خوشه‌ای همه‌منظوره و سریع برای پردازش و آموزش حجم زیادی از داده‌های ویژگی ترافیک شبکه استفاده شده است. در این کار، مهم‌ترین ویژگی‌های مجموعه داده CSE-CIC-IDS2018 برای ساخت مدل‌های یادگیری ماشین، و سپس از معروف‌ترین رویکردهای یادگیری ماشین، یعنی رگرسیون لجستیک، ماشین بردار پشتیبان (SVM)، سه دسته‌بندی مختلف درخت تصمیم و الگوریتم بیز ساده برای آموزش مدل با استفاده از حداکثر هشت عدد گره کارگر استفاده می‌شود. خوشه Spark ما شامل هفت ماشین است که به عنوان گره کارگر عمل می‌کنند، و یک ماشین به عنوان هر دو مدیر و کارگر پیکربندی شده است. از مجموعه داده CSE-CIC-IDS2018برای ارزیابی عملکرد کلی این الگوریتم‌ها در حملات بات نت و تنظیم فراپارامتر توزیع شده برای یافتن بهترین پارامترهای درخت تصمیم منفرد استفاده می‌شود. ما با استفاده از ویژگی‌های انتخابی با روش یادگیری در آزمایش‌های خود تا ۱۰۰٪ دقت دست یافته‌ایم.

* نویسنده مسئول.
آدرس‌های رایانامه: ramina@post.com (ر. عاطفی نیا)،
m.ahmadi@razi.ac.ir (م. احمدی)

ISSN: 2322-4460