|  |  |
| --- | --- |
| تقنيات عالية الأداء التعدين البيانات لكشف التسلل | Research Title (Arabic) |
| High Performance Data Mining Techniques for Intrusion Detection | Research Title (English) |
| الحوسبة عالية الأداء، والبيانات والتعدين ، وكشف التسلل | Research Topic (Arabic) |
| High performance computing, data mining, intrusion detection | Research Topic (English) |
| جامعة سنترال فلوريدا | Publisher (Arabic) |
| University of Central Florida | Publisher (English) |
| 2004 | Publishing Year (Arabic) |
| 2004 | Publishing Year (English) |
|  | ISBN |
| الكتلة الحوسبة، المعالجة المتوازية، واستخراج البيانات، كشف التسلل، وأمن الشبكات | Key Words (Arabic) |
| Cluster computing, parallel processing, data mining, intrusion detection, network security | Key Words (English ) |
|  | Journal Name, or (Conference + place and date being held) |
|  | Volume No. or Issue No. and the Number of Pages in case it has been published in a scientific journal |
| تحول النمو السريع لأجهزة الكمبيوتر في الطريقة التي تم تخزين المعلومات والبيانات. مع هذا النموذج الجديد من الوصول إلى البيانات ، ويأتي يتعرض للتهديد من هذه المعلومات للمستخدمين غير المصرح به وغير مقصودة. وقد نظم العديد المتقدمة التي التدقيق في البيانات عن الانحراف عن السلوك العادي لمستخدم أو نظام ، أو البحث عن توقيع المعروفة في البيانات. وتسمى هذه الأنظمة كما كشف التسلل أنظمة (الهوية). هذه الأنظمة تستخدم تقنيات مختلفة تتراوح بين الأساليب الإحصائية لتعلم الآلة الخوارزميات.   أنظمة كشف التسلل استخدام بيانات مراجعة الحسابات التي تم إنشاؤها بواسطة أنظمة التشغيل ، وبرامج التطبيق أو أجهزة الشبكة. هذه المصادر إنتاج كمية ضخمة من قواعد البيانات مع عشرات الملايين من السجلات في نفوسهم. لتحليل هذه البيانات ، يتم استخدام البيانات التعدين وهي عملية مفيدة لأنماط من حفر بكميات كبيرة من المعلومات. يشكل عقبة رئيسية في هذه العملية التي طغت على استخراج البيانات التقليدية والخوارزميات التعلم عن طريق حجم وتعقيد معظم البيانات المتاحة. وهذا يجعل من هذه الخوارزميات غير عملي للمهام الحرجة مثل وقت كشف التسلل بسبب تنفيذ كبيرة  الوقت.   نهجنا تجاه هذه المسألة يجعل من استخدام تقنيات عالية الأداء استخراج البيانات لتسريع هذه العملية عن طريق استغلال التوازي في خوارزميات استخراج البيانات الموجودة والأجهزة الأساسية. سنظهر كيف يمكن أن تستخدم عالية الأداء والحوسبة المتوازية لتوسيع نطاق خوارزميات التنقيب عن البيانات للتعامل مع قواعد البيانات الكبيرة ، والسماح للعنصر التنقيب عن البيانات للبحث مجموعة أكبر بكثير من أنماط ونماذج من منصات الحوسبة التقليدية والخوارزميات سيسمح.   نحن في تطوير خوارزميات التنقيب عن البيانات الموازية التي parallelizing آلة القائمة تقنيات التعلم باستخدام الحوسبة العنقودية. وتشمل هذه الخوارزميات backpropagation متوازية وموازية غامض ARTMAP شبكات العصبية. نحن تقييم الأداء من النماذج المتقدمة من حيث تسريع أكثر من الخوارزميات التقليدية ، ومعدل التنبؤ وكاذبة معدل الانذار. وأظهرت النتائج التي توصلنا إليها أن backpropagation التقليدية والخوارزميات ARTMAP غامض يمكن أن تستفيد من تقنيات الحوسبة عالية الأداء مما يجعلها مناسبة تماما لمثل المهام الحرجة الوقت لكشف التسلل. | Research Abstract (Arabic) |
| The rapid growth of computers transformed the way in which information and data was stored. With this new paradigm of data access, comes the threat of this information being exposed to unauthorized and unintended users. Many systems have been developed which scrutinize the data for a deviation from the normal behavior of a user or system, or search for a known signature within the data. These systems are termed as Intrusion Detection Systems (IDS). These systems employ different techniques varying from statistical methods to machine learning algorithms.  Intrusion detection systems use audit data generated by operating systems, application softwares or network devices. These sources produce huge amount of datasets with tens of millions of records in them. To analyze this data, data mining is used which is a process to dig useful patterns from a large bulk of information. A major obstacle in the process is that the traditional data mining and learning algorithms are overwhelmed by the bulk volume and complexity of available data. This makes these algorithms impractical for time critical tasks like intrusion detection because of the large execution  time.  Our approach towards this issue makes use of high performance data mining techniques to expedite the process by exploiting the parallelism in the existing data mining algorithms and the underlying hardware. We will show that how high performance and parallel computing can be used to scale the data mining algorithms to handle large datasets, allowing the data mining component to search a much larger set of patterns and models than traditional computational platforms and algorithms would allow.  We develop parallel data mining algorithms by parallelizing existing machine learning techniques using cluster computing. These algorithms include parallel backpropagation and parallel fuzzy ARTMAP neural networks. We evaluate the performances of the developed models in terms of speedup over traditional algorithms, prediction rate and false alarm rate. Our results showed that the traditional backpropagation and fuzzy ARTMAP algorithms can benefit from high performance computing techniques which make them well suited for time critical tasks like intrusion detection. | Research Abstract (English) |