التقنيات و الأساليب الفنية للإدارة والتحكم بقواعد البيانات عن بعد ومنع الاختراق

جدول المحتويات

[1. قواعد البيانات 2](#_Toc32069274)

[2. حماية قواعد البيانات 3](#_Toc32069275)

[3. أنواع الاختراقات في قواعد البيانات وطرق مواجهتها 4](#_Toc32069276)

[1. امتيازات وصول مفرطة Excessive Privileges 4](#_Toc32069277)

[2. حقن عبارات الاستعلام SQL Injection 5](#_Toc32069278)

[3. البرمجيات الخبيثة Malware 5](#_Toc32069279)

[4. سجل تدقيق ضعيف Weak Audit Trail 6](#_Toc32069280)

[5. عدم حماية النسخ الاحتياطية 6](#_Toc32069281)

[6. المصادقة الضعيفة Weak Authentication 6](#_Toc32069282)

[7. احتواء قواعد البيانات على نقاط ضعف وتهيئتها بشكل غير مناسب 6](#_Toc32069283)

[8. البيانات الحساسة غير المدارة بشكل جيد 7](#_Toc32069284)

[9. هجوم منع الخدمة Denial of Service 7](#_Toc32069285)

[10. الخبرة الأمنية المحدودة 7](#_Toc32069286)

[4. تقنيات وأساليب التحكم بقواعد البيانات لمواجهة الاختراقات 7](#_Toc32069287)

[5. قصور تقنيات وأساليب التحكم عن مواجهة الاختراقات في قواعد البيانات 9](#_Toc32069288)

[6. أنظمة كشف التسلل intrusion detection systems IDS 9](#_Toc32069289)

[1. تصنيف أنظمة كشف التسلل بناء على طريقة الكشف 10](#_Toc32069290)

[a. نظام كشف التسلل المستند إلى التوقيع ISDS 10](#_Toc32069291)

[b. نظام كشف التسلل المستند إلى الشذوذ AIDS 10](#_Toc32069292)

[7. تنفيذ طريقة حديثة لكشف التسلل إلى قواعد البيانات العلائقية 11](#_Toc32069293)

# قواعد البيانات

قاعدة البيانات او قاعدة المعطيات Database، هي مجموعة من عناصرِ البيانات المنطقية المرتبطة مع بعضها البعض بعلاقة رياضية، وتتكون قاعدة البيانات من جدول واحد أو أكثر. يتكون الجدول من سجل (صف) أو أكثر. يتكون السجل من حقل أو أكثر. نورد كمثال السجل الخاص بموظف معين يتكون من عدة حقول مثل رقم الموظف - اسم الموظف - درجة الموظف - تاريخ التعيين - الراتب - والقسم التابع له، وغير ذلك من بيانات الموظفين التي تخزن في جهاز الحاسوب عَلى نحو منظّم، حيث يقوم برنامج يسمى محرك قاعدة البيانات (database engine) بتسهيل التعامل معها والبحث ضمن هذه البيانات، وتمكين المستخدم من الإضافة والتعديل عليها. يتم استرجاع البيانات باستخدام أوامر من لغة الاستعلام حيث تعتبر معلومات تساعد في عملية اتخاذ القرار.

نظام إدارة قواعد البيانات Database Management System DBMS هو البرنامج الذي يتم من خلاله استرجاع البيانات، أو الإضافة أو التعديل عليها، أو حذفها، حيث يقوم البرنامج بالربط بين المستخدم وبين محرك قاعدة البيانات، لأداء تلك المهمة. فى حال وجود علاقة بين جداول قاعدة البيانات يسمى هذا بنظام قواعد البيانات العلائقية (نظام إدارة قواعد البيانات العلائقية).

الهدف الأساسي لقواعد البيانات هو التركيز على طريقة تنظيم البيانات وليس على التطبيقات الخاصة. أي أن الهدف الرئيسي لمصمم قاعدة البيانات هو تصميم البيانات بحيث تكون خالية من التكرار ويمكن استرجاعها وتعديلها والإضافة عليها دون المشاكل التي يمكن أن تحدث مع وجود التكرار فيها. يتم ذلك عن طريق ايجاد ثلاثة مستويات من التجريد أو النماذج لقواعد البيانات تسمى نماذج التطبيع (تسوية قاعدة البيانات)، ويقصد بها جعل تركيبة البيانات أقرب للطبيعة التصنيفية.

وهناك تركيبات لقواعد البيانات حسب نوع العلاقة الرياضية بين البيانات، ومنها:

* التركيب العلائقي : وهو اعتماد علاقة محددة بين عناصر البيانات، مثل أن تكون قيمة عنصر معتمدة على حاصل جمع عنصرين. وهذا التركيب هو أنجح التراكيب المطبقة في عالم قواعد البيانات المعلوماتية، وذلك بسبب إعطائه تنوع في نوع العلاقة بين البيانات، لأن احتمالية تنفيذ العلاقات فيه أكبر من اي تركيب اخر.
* التركيب الهيكلي : وهو اعتماد علاقة الهيكل التنظيمي بين عناصر البيانات، مثل أن يكون عنصرين مصنفين تحت عنصر واحد أو تابعين له.
* التركيب الهرمي : وهو اعتماد علاقة الهرم بين عناصر البيانات، مثل أن يكون كل عنصر مسؤول عن عنصر واحد فقط وليس أكثر.

# حماية قواعد البيانات

في عالمنا المعاصر، يتم إنشاء البيانات بسرعة كبيرة وبكميات كبيرة جدا حيث يتم تخزين هذه البيانات في قواعد البيانات. توفر قواعد البيانات سهولة وفعالية عالية لإدارة المعلومات المخزنة فيها حيث تتم كافة العمليات المتعلقة بمعالجة هذه البيانات عن طريق برنامج يطلق عليه اسم نظام إدارة قواعد المعطيات Database Management System. تعتبر بعض البيانات التي تخزنها الشركات والمنشأت في قواعد البيانات ذات سرية وأهمية عالية، كالبايانات التي تخص الحسابات المصرفية لزبائن البنوك، لذلك تأتي حماية البيانات المخزنة في قواعد المعطيات في أولويات الشركات عند تصميم قواعد بياناتها.

تعرف قاعدة البيانات الآمنة secured database بأنها قاعدة البيانات التي توفر السرية، السلامة، التوفر Confidentiality, Integrity, Availability. تتخذ المؤسسات تدابير مختلفة لضمان حماية قواعد البيانات الخاصة بها من التهديدات الداخلية والخارجية. يشمل أمان قاعدة البيانات حماية قاعدة البيانات نفسها، والبيانات التي تحتوي عليها، ونظام إدارة قاعدة البيانات الخاص بها، والتطبيقات المختلفة التي تصل إليها. يجب على المؤسسات تأمين قواعد البيانات من الهجمات المتعمدة مثل تهديدات الأمان الإلكتروني، وكذلك إساءة استخدام البيانات وقواعد البيانات من أولئك الذين يمكنهم الوصول إليها. في السنوات القليلة الماضية، ارتفع عدد انتهاكات البيانات بشكل كبير. بالإضافة إلى الضرر الكبير الذي تسببه هذه التهديدات لسمعة الشركة وقاعدة العملاء، هناك عدد متزايد من لوائح العقوبات على انتهاكات البيانات التي يجب على المؤسسات التعامل معها، مثل تلك الموجودة في اللائحة العامة لحماية البيانات GDPR وهي مكلفة للغاية. يعد الأمان الفعال لقاعدة البيانات أساسيا لحماية سمعة المنظمات والحفاظ على عملائها.

هناك ثلاث طبقات لأمان قاعدة البيانات: على مستوى قاعدة البيانات، على مستوى طبقة الوصول access level، وعلى مستوى المحيط perimeter level. يطبق الأمن على مستوى قاعدة البيانات داخل قاعدة البيانات نفسها، حيث تخزن البيانات. يركز أمان طبقة الوصول على التحكم في من يُسمح له بالوصول إلى البيانات. يحدد أمان قاعدة البيانات على مستوى المحيط من يمكنه الوصول إلى قواعد البيانات ولا يمكنه ذلك. كل مستوى من المستويات المذكورة يتطلب حلول أمنية مختلفة عن الحلول التي تتطلبها المستويات الأخرى.

# أنواع الاختراقات في قواعد البيانات وطرق مواجهتها

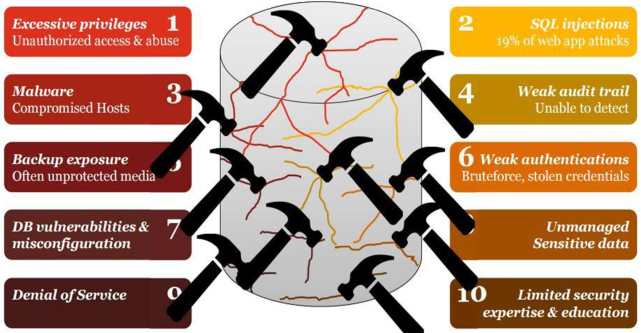
يمكن تصنيف الهجمات الرئيسية على قواعد البيانات كما هو موضح في الشكل 1. نوضيح هذه الهجمات فيما يلي:

## امتيازات وصول مفرطة Excessive Privileges

يتم إساءة استخدام امتيازات الوصول إلى قاعدة البيانات بعدة طرق، حيث يمكن للمستخدم إساءة استخدام الامتياز لغرض غير مصرح به. يوجد أشكال مختلفة لإساءة الاستخدام: إساءة استخدام الامتيازات المفرطة، إساءة استخدام الامتيازات المشروعة وإساءة استخدام الامتيازات غير المستخدمة. يعتبر هذا النوع من التهديد الأكثر خطورة لأن إساءة الاستخدام تتم من قبل المستخدمين المصرح لهم. إن منح أذونات مفرطة يمثل مشكلة لأن حوالي 80 ٪ من الهجمات على بيانات الشركة تنفذ في الواقع من قبل الموظفين الحاليين أو الموظفين السابقين. إن منح الكثير من الامتيازات أو عدم إبطال تلك الامتيازات في الوقت المناسب يجعل من السهل عليهم إساءة الاستخدام بشكل مقصود أو غير مقصود.

**تشمل الإجراءات المضادة لإساءة الامتياز**

1. سياسة التحكم في الوصول Access Control policy: عدم منح المستخدم امتيازات وصول غير ضرورية.
2. يمكن إيقاف إساءة استخدام الامتيازات المشروعة من خلال توفير سجل تدقيق Audit Trail.



الشكل 1 تهديدات قاعدة البيانات

## حقن عبارات الاستعلام SQL Injection

هو نوع من هجمات الحقن التي تجعل من الممكن تنفيذ عبارات SQL الضارة. تتحكم هذه العبارات في خادم قاعدة البيانات خلف تطبيق الويب. يمكن للمهاجمين استخدام ثغرات SQL Injection لتجاوز تدابير أمان التطبيق. حيث يمكنهم الالتفاف حول المصادقة والترخيص authentication and authorization لصفحة ويب أو تطبيق ويب واسترداد محتوى قاعدة بيانات SQL بأكملها. كما يمكنهم أيضًا استخدام SQL Injection لإضافة السجلات وتعديلها وحذفها في قاعدة البيانات.

**تشمل الإجراءات المضادة لحقن عبارات الاستعلام**

* 1. استخدام الإجرائيات Stored Procedures عوضا عن عبارات الاستعلام المباشرة.
  2. استخدام بنية MVC (Model, View, Control).

## البرمجيات الخبيثة Malware

يستخدم مجرمو الإنترنت، المتسللون والجواسيس هجمات متطورة تمزج بين تكتيكات متعددة - مثل رسائل البريد الإلكتروني للتصيد العشوائي والبرامج الضارة - لاختراق المنظمات وسرقة البيانات الحساسة. فيصبح المستخدمون الشرعيون الغير مدركين أن البرامج الضارة قد أصابت أجهزتهم؛ قناة لهذه المجموعات للوصول إلى الشبكات والبيانات الحساسة للمؤسسات.

**تشمل الإجراءات المضادة للبرمجيات الخبيثة**

1. تمكين جدار حماية Firewall واستخدام Antivirus

## سجل تدقيق ضعيف Weak Audit Trail

يجب أن يكون التسجيل الآلي للإجراءات والعمليات على قاعدة البيانات التي تتضمن بيانات حساسة جزءًا من أي نظام يستخدم قاعدة البيانات. يمثل الفشل في جمع سجلات التدقيق التفصيلية للأنشطة على قاعدة البيانات مخاطرة تنظيمية على العديد من المستويات. ستجد المنظمات ذات آليات تدقيق قاعدة البيانات الضعيفة أنها تتعارض بشكل متزايد مع المتطلبات الصناعية والحكومية التنظيمية.

لا تتمكن معظم آليات التدقيق من تحديد المستخدم المسؤول عن نشاط ما، بل تربط النشاط بحساب قد يحوي مجموعة من المستخدمين. كما يمكن للمستخدمين الذين لديهم حق الوصول الإداري إلى قاعدة البيانات، سواء تم الحصول عليها بطريقة مشروعة أو غير مشروعة، إيقاف تشغيل تدقيق قاعدة البيانات الأصلية لإخفاء نشاط احتيالي. لذلك، يجب أن يكون سجل التدقيق التدقيق ومسؤولياته منفصلة بشكل تام عن مسؤولي قواعد البيانات ومنصة خادم قاعدة البيانات.

**تشمل الإجراءات المضادة لسجلات التدقيق الضعيفة**

1. تنفيذ سجلات تدقيق مستندة إلى الشبكة. بحيث لا ينبغي أن يكون لهذه السجلات أي تأثير على أداء قاعدة البيانات، وأن تعمل بشكل مستقل عن جميع المستخدمين وتقدم بيانات متعلقة بأنشطة المستخدمين وليس أنشطة الحسابات.

## عدم حماية النسخ الاحتياطية

غالبًا ما تكون وسائط تخزين النسخ الاحتياطية غير محمية بشكل جيد من الهجوم. نتيجة لذلك، تضمنت العديد من انتهاكات الأمان سرقة أقراص النسخ الاحتياطية لقاعدة البيانات. علاوة على ذلك، يمكن أن يؤدي الفشل في تدقيق ومراقبة أنشطة المسؤولين الذين لديهم حقوق وصول منخفضة المستوى إلى المعلومات الحساسة إلى تعريض البيانات للخطر. إن اتخاذ التدابير المناسبة لحماية النسخ الاحتياطية للبيانات الحساسة ومراقبة المستخدمين الأكثر امتيازًا يعتبر من أفضل الممارسات الأمنية التي يجب اتباعها.

**تشمل الإجراءات المضادة للهجوم على النسخ الاحتياطية**

1. تشفير قواعد البيانات: يجب تخزين البيانات في صيغة مشفرة بما يتيح تأمين النسخ الحالية والنسخ الاحتياطية لقواعد البيانات، كما يجب تدقيق الأنشطة على البيانات الحساسة والتحكم في وصول المستخدمين إليها.

## المصادقة الضعيفة Weak Authentication

تسمح خطط المصادقة الضعيفة للمهاجمين بافتراض هوية مستخدمي قاعدة البيانات الشرعيين. لذلك، يعتبر تنفيذ المصادقة باستخدام كلمات المرور أوالمصادقة ثنائية العامل أمر لا بد منه.

## احتواء قواعد البيانات على نقاط ضعف وتهيئتها بشكل غير مناسب

من الشائع العثور على قواعد بيانات تحتوي على العديد من نقاط الضعف وغير محدثة بآخر التصحيحات Unpatched، أو اكتشاف قواعد بيانات لا تزال بتهيئتها الافتراضية. يعرف المهاجمون كيفية استغلال هذه الثغرات الأمنية لشن هجمات ضد المؤسسة. لسوء الحظ، تتطلب بعض التصحيحات جهدا كبيرا من المسؤولين عن قواعد البيانات وقد يتطلب تنفيذها وقتا كبيرا أيضا يصل إلى عدة أشهر. تبقى قاعدة البيانات خلال هذا الوقت عرضة للخطر.

**تشمل الإجراءات المضادة للهجوم نقاط الضعف**

1. لا ينبغي وجود حسابات افتراضية. يجب إنشاء حسابات باستخدام اسم مستخدم وكلمة مرور جديدين. كما يجب ملاحقة التصحيحات Patches التي تصدرها الشركات البرمجية باستمرار.

## البيانات الحساسة غير المدارة بشكل جيد

تكافح العديد من الشركات للحفاظ على جرد دقيق لقواعد بياناتها وكائنات البيانات الهامة الموجودة فيها. قد تحتوي قواعد البيانات المنسية على معلومات حساسة، ويمكن أن تظهر قواعد بيانات جديدة - على سبيل المثال ، في بيئات اختبار التطبيقات - دون اطلاع فريق الأمان. ستتعرض البيانات الحساسة في قواعد البيانات هذه للتهديدات إذا لم يتم تطبيق الضوابط والأذونات المطلوبة.

**تشمل الإجراءات المضادة للهجوم على البيانات الحساسة غير المدارة**

1. تشفير البيانات الحساسة في قاعدة البيانات.
2. تطبيق الضوابط والأذونات المطلوبة في قاعدة البيانات أو أي جزء منبثق عنها.

## هجوم منع **الخدمة** Denial of Service

منع الخدمة عبارة عن فئة هجوم عامة يتم فيها منع المستخدمين من الوصول إلى تطبيقات الشبكة أو بياناتها.

**تشمل الإجراءات المضادة لهجوم منع الخدمة**

1. تطبيق الإعدادات المناسبة لزيادة حجم قائمة انتظار اتصال TCP، تقليل فترة إنشاء الاتصال، واستخدام آليات ديناميكية لضمان عدم استنفاذ قائمة انتظار الاتصال.
2. استخدم نظام كشف التسلل عبر الشبكة IDS لأن هذا النظام يمكن أن يكتشف هجمات منع الخدمة ويستجيب لها تلقائيًا.

## الخبرة الأمنية المحدودة

تلعب الإجراءات الامنية غير التقنية دورا هاما. وفي بعض الحالات, لا تواكب ضوابط الأمن الداخلي نمو البيانات كما أن العديد من المؤسسات غير مجهزة للتعامل مع أي خرق أمني. غالبًا ما يكون هذا بسبب نقص الخبرة اللازمة لتطبيق الضوابط الأمنية أو فرض السياسات أو إجراء عمليات الاستجابة للحوادث.

**تشمل الإجراءات المضادة للخبرة الامنية المحدودة**

1. تثقيف المستخدمين أمنيا

# تقنيات وأساليب التحكم بقواعد البيانات لمواجهة الاختراقات

التحكم في الوصول هو مصطلح أمان يستخدم للإشارة إلى مجموعة من السياسات المتبعة لتقييد الوصول إلى المعلومات والأدوات. يحد التحكم في الوصول إلى المعلومات الولوج إلى البيانات والبرامج المستخدمة لمعالجة هذه البيانات. يُستخدم التحكم في الوصول إلى المعلومات بشكل شائع في أمان الكمبيوتر والشبكات. بعض الأمثلة تشمل:

• مستخدم يسجّل الدخول إلى حاسوبه المحمول باستخدام كلمة مرور

• مستخدم يفتح هاتفه الذكي عن طريق فحص بصمة الإبهام

• مستخدم Gmail يقوم بتسجيل الدخول إلى حساب بريده الإلكتروني

• موظف عن بعد يصل إلى شبكة صاحب العمل الداخلية باستخدام VPN

في جميع هذه الحالات، يتم استخدام البرنامج للمصادقة ومنح الترخيصAuthentication, Authorization للمستخدمين الذين يرغبون في الوصول إلى المعلومات الرقمية. يعد كل من المصادقة ومنح الترخيص عنصرين أساسيين في التحكم في الوصول إلى المعلومات.

باختصار، المصادقة Authentication هي ممارسة أمنية يتم من خلالها التأكد من هوية شخص ما، في حين أن منح الترخيص Authorization يتعلق بمستوى الوصول الممنوح لكل مستخدم. على سبيل المثال، عندما يسجل شخص في الفندق كزبون يطلب منه موظف الاستقبال تقديم جواز سفر لإثبات أنه بالفعل الشخص الموجود اسمه في الحجز. هذا مثال على المصادقة. بمجرد مصادقة موظف الاستقبال على الضيف، يحصل الضيف على بطاقة مفاتيح بامتيازات محدودة. هذا مثال على الترخيص. تمكن بطاقة المفاتيح الخاصة بالضيف من الوصول إلى غرفته ومصعد الضيوف والمسبح. لن تفتح بطاقة المفاتيح هذه غرف الضيوف الآخرين، ولن تمكنه من طلب مصعد الخدمة، لأن الضيف غير مصرح له بالوصول إلى تلك المواقع.

يحصل موظف التدبير المنزلي في الفندق على بطاقة مفاتيح بمستوى ترخيص أعلى؛ حيث يمكنهم الوصول إلى جميع غرف الضيوف، مصعد الخدمة، غرفة غسيل الملابس وصالة الموظفين. ولكنهم لا يستطيعون الوصول إلى بعض المناطق الحساسة، مثل مركز الأمان أو الخزانة النقدية. في الوقت نفسه، فإن رئيس الأمن في الفندق لديه بطاقة مفاتيح تمكنه من الوصول إلى أي جزء من الفندق.

تحتوي أنظمة الكمبيوتر والشبكات على عناصر تحكم مصادقة ومنح ترخيص مشابهة جدًا. فعندما يسجل المستخدم الدخول إلى حسابه الإلكتروني أو حسابه البنكي عبر الإنترنت، فإنه يستخدم اسم مستخدم وكلمة مرور يفترض معرفتهما من قبله فقط. يستخدم البرنامج هذه المعلومات لمصادقة المستخدم. تفرض بعض التطبيقات متطلبات مصادقة أكثر صرامة من غيرها؛ على الرغم من أن كلمة المرور كافية بالنسبة للبعض، فقد يتطلب البعض الآخر مصادقة ثنائية العامل أو حتى تأكيدًا حيويًا، مثل فحص بصمة الإبهام أو المسح الضوئي للوجه.

بمجرد المصادقة ، يمكن للمستخدم رؤية المعلومات المصرح لهم بالوصول إليها فقط. في حالة وجود حساب مصرفي عبر الإنترنت، يمكن للمستخدم فقط رؤية المعلومات المتعلقة بحسابه المصرفي الشخصي. يمكن لمدير الصندوق في البنك تسجيل الدخول إلى التطبيق نفسه ورؤية البيانات المتعلقة بالممتلكات المالية للبنك، وقد يكون لديهم أيضًا إمكانية الوصول إلى أداة لبيع وشراء الأوراق المالية نيابة عن البنك. نظرًا لأن البنك يتعامل مع معلومات شخصية حساسة للغاية، فمن الممكن ألا يتمكن أي شخص من الوصول غير المقيد إلى البيانات. حتى رئيس البنك أو رئيس الأمن قد يحتاج إلى مراجعة بروتوكول الأمان للوصول إلى البيانات الكاملة للعملاء.

# قصور تقنيات وأساليب التحكم عن مواجهة الاختراقات في قواعد البيانات

اختراق قاعدة البيانات هي عملية الوصول إلى المعلومات المخزنة على أجهزة الكمبيوتر أو الخوادم أو غيرها من الأجهزة بقصد المساس بالخصوصية أو الحصول على معلومات سرية. يعد اختراق قاعدة البيانات مشكلة متنامية لمستخدمي أجهزة الكمبيوتر الفردية وكذلك للشركات والمؤسسات الكبيرة. يمكن للاختراق أن يحدث من داخل أو من خارج الشركات، ويعتبر الحد من خطر الاختراق من الداخل على مستوى الشركات أمرا صعبا للغاية، لأن مسؤولي النظام والعاملين يملكون إمكانية الوصول إلى خوادم قواعد البيانات وأجهزة الكمبيوتر المكتبية والأجهزة الخارجية بما في ذلك أجهزة USB والهواتف الذكية وغيرها من الأجهزة المحمولة والقابلة للنقل.

يكمن السبب الجوهري لنسبة عالية من الخروقات في إهمال الموظفين والعاملين داخل الشركات. يمكن أن يحدث الاختراق بسبب شخص ما منح كلمة المرور الخاصة به لشخص آخر غير مصرح له، أو شخص تم تصيده من خلال الإيميل بواسطة برمجيات خبيثة Malware، أو قد يكون جهاز USB مفقود، أو شخص ما يسيء التعامل مع الملفات. لا تنطوي التهديدات دائمًا على خبث متعمد؛ في كثير من الأحيان، يمكن أن تعزى الاختراقات التي تحدث من داخل المنظمة إلى نقص معرفة الموظفين بأسس حماية البيانات، أو في بعض الأحيان، ببساطة إلى سلوك مهمل.

بغض النظر إذا كان الاختراق من الداخل متعمدا أو غير متعمد، فإنه سيؤدي حتما إلى الوصول إلى المعلومات السرية للشركة والمساس بخصوصية عملائها. في هذا النوع من الاختراقات التي تتم من الداخل، تعجز أساليب الحماية التقليدية كالتحكم بالوصول access control والمصادقة ومنح الصلاحيات وحتى تشفير قاعدة البيانات عن منع الاختراق في الوقت المناسب. لذلك لابد من استخدام آليات مساعِدة تحسن من أمان قواعد البيانات كأنظمة كشف التسلل Intrusion Detection Systems IDS.

# أنظمة كشف التسلل intrusion detection systems IDS

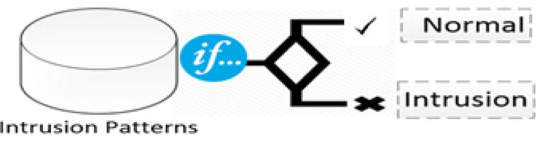
لا يمكن ضمان الأمان في قاعدة بيانات إلا إذا كان يفي بنموذج أمان ثلاثي: السرية، النزاهة والتوافر. السرية تمنع الوصول إلى المعلومات الحساسة من قبل أشخاص غير مصرح لهم. يضمن ذلك أن الأشخاص المعنيين فقط يمكنهم الوصول إلى البيانات بما يتوافق مع مستوى الأمان والامتيازات الخاصة بالوصول. تضمن النزاهة دقة البيانات والالتزام بها. يضمن ذلك عدم تغيير البيانات من قبل أي شخص غير مصرح له. يضمن التوافر أن المعلومات المطلوبة في متناول المستخدمين المعتمدين في جميع الأوقات. إن انتهاك أي من هذه المبادئ يؤدي إلى فقدان الأمن. لقد ثبت أن أنظمة قواعد البيانات تتعرض لهجمات أمنية من قبل مهاجمين داخليين أو خارجيين. المهاجمون من الداخل هم أشخاص داخل المنظمة مرخص لهم من قِبل المنظمة، لكن يتصرفون بنوايا خبيثة. المهاجمون الخارجيون هم أشخاص خارج المنظمة ممن حصلوا على وصول غير مصرح به إلى قاعدة البيانات من خلال استغلال الثغرات الأمنية. تركز تقنيات أمان قاعدة البيانات التقليدية –كالتحكم في الوصول والتشفير- بشكل أساسي على منع الهجمات من الخارج و بالكاد تساعد في تحديد التهديدات الداخلية. لذلك تم تصميم أنظمة كشف التسلل (IDS) لتحديد الخروقات الأمنية الناشئة عن كل من التهديدات الداخلية والخارجية.

## تصنيف أنظمة كشف التسلل بناء على طريقة الكشف

نظام كشف التسلل هو برنامج أو نظام يحدد النشاطات الضارة بأنظمة الكمبيوتر بما يسمح بالحفاظ على أمن النظام. يهدف IDS إلى تحديد الأنواع المختلفة للنشاطات الضارة على الشبكة أو الاستخدام المشبوه لأجهزة الكومبيوتر، والتي لا يمكن التعرف عليها بواسطة جدار الحماية التقليدي. يعد هذا أمرًا ضروريًا لتحقيق حماية عالية ضد النشاطات التي تعرض توفر البيانات أو سلامتها أو سريتها للخطر. يمكن تصنيف أنظمة IDS إلى مجموعتين أساسيتين بناء على طريقة الكشف: نظام كشف التسلل المستند إلى التوقيع Signature based Intrusion Detection System (SIDS) ونظام كشف التسلل المستند إلى الشذوذ Anomaly based Intrusion Detection System (AIDS).

### نظام كشف التسلل المستند إلى التوقيع ISDS

تعتمد أنظمة الكشف عن الاختراق بالتوقيع على تقنيات مطابقة الأنماط pattern matching techniques لإيجاد هجوم معروف؛ تُعرف هذه الأنظمة أيضًا باسم الكشف المعتمد على المعرفة أو الكشف عن إساءة الاستخدام. بمعنى آخر، عندما يتطابق توقيع الاختراق مع توقيع اقتحام سابق موجود بالفعل في قاعدة بيانات التوقيع، يتم تشغيل إشارة إنذار. بالنسبة إلى SIDS، يتم فحص سجلات المضيف للعثور على تسلسل من الأوامر أو الإجراءات التي تم تحديدها مسبقًا على أنها برامج ضارة. تم تصنيف SIDS في الأدبيات على أنها "اكتشاف قائم على المعرفة أو اكتشاف إساءة الاستخدام". يوضح الشكل 2 مفهوم عمل أنظمة SIDS. والفكرة الرئيسية هي بناء قاعدة بيانات للتوقيعات الاقتحامية ومقارنة توقيعات الأنشطة الحالية بالتوقيعات المخزنة ثم إثارة إنذار في حالة العثور على تطابق.



الشكل 2 مفهوم عمل أنظمة SIDS

توفر SIDS عادةً دقة كشف ممتازة لعمليات الاختراق المعروفة سابقًا. ومع ذلك، تواجه SIDS صعوبة في اكتشاف هجمات zero-day بسبب عدم وجود توقيع مطابق في قاعدة البيانات إلى أن يتم استخراج توقيع الهجوم الجديد وتخزينه.

### نظام كشف التسلل المستند إلى الشذوذ AIDS

أثار AIDS اهتمام الكثير من العلماء بسبب قدرته على التغلب على محدودية SIDS. في AIDS، يتم إنشاء نموذج عادي لسلوك النظام باستخدام التعلم الآلي machine learning أو الأساليب الإحصائية statistical-based أو الأساليب القائمة على المعرفة knowledge-based. يعتبر أي انحراف كبير بين السلوك المرصود والنموذج شذوذًا، يمكن تفسيره على أنه اقتحام. تقوم هذه التقنية على افتراض أن السلوك الضار يختلف عن سلوك المستخدم العادي. تصنف سلوكيات المستخدمين غير الطبيعيين التي تختلف عن السلوكيات القياسية على أنها تدخلات intrusions. يشتمل نظام كشف التسلل المستند إلى الشذوذ على المراحل التالية:

1. تعريف وبناء الملفات الشخصية الخاصة بكل مستخدم profiles والتي تمثل بدقة سلوك المستخدم "العادي" normal user، ويجب أن تحتوي هذه الملفات على معلومات عن الهجمات المتوقعة من قبل هذا المستخدم وهو ما يسمى بتحديد تواقيع الهجوم attack signature.
2. يتم بالاعتماد على هذه الملفات الشخصية استخلاص السمات السلوكية behavioral features التي سيتم الاعتماد عليها لبناء عناقيد المجموعات Clusters ولتصنيف classification سلوكيات المستخدمين في مرحلة الاختبار. يجب اختيار هذه السمات السلوكية بدقة بما يحقق أداء أعظميا للتقنيات المستخدمة في التجميع Clustering والتصنيف Classification.
3. بناء عناقيد المجموعات Clustering والتي تضم الملفات الشخصية للمستخدمين.
4. تدريب المصنف Classifier باستخدام الملفات الشخصية التي تم بناءها في المرحلة الأولى.
5. استخدام المصنف المدرَّب لتصنيف سلوكيات المستخدمين عن طريق حساب انحراف السلوك عن السلوك العادي المعرف في الملف الشخصي للمستخدم.
6. تبليغ المسؤولين الأمنيين عن حالة النظام وإطلاعهم على الإنذارات المولدة.
7. تطوير طريقة لمنع وإيقاف الاختراق عند الكشف عنه. هذه الميزة قد لا تكون متوفرة في نظام IDS، ولكن في حال توفرها يطلق على النظام Intrusion Detection and Response System (IDRS).

الميزة الرئيسية لأنظمة AIDS هي القدرة على تحديد هجمات zero-day بسبب كون التعرف على نشاط المستخدم غير الطبيعي لا يعتمد على قاعدة بيانات التوقيع. يطلق AIDS إشارة خطر عندما يختلف السلوك المدروس عن السلوك المعتاد. علاوة على ذلك، هناك فوائد مختلفة لأنظمة AIDS. أولاً، لديها القدرة على اكتشاف الأنشطة الضارة الداخلية. فإذا بدأ المتسلل في إجراء نشاط في حساب مسروق لم يتم تحديد هويته، سيؤدي ذلك إلى إطلاق إنذار. ثانياً، من الصعب للغاية بالنسبة لمجرمي الإنترنت معرفة سلوك المستخدم العادي دون إصدار تنبيه.

# مقاربات مختلفة لتنفيذ Intrusion Detection System

لتوضيح حدود وسعة مجال حماية قواعد البيانات نورد فيما يلي الأصناف المختلفة لأنظمة كشف التسلل التي تم العمل عليها حتى الآن، مع أمثلة على كل صنف:

## أنظمة كشف التسلل القائمة على التحليل الزمني temporal analysis

تركز هذه التقنية على السمات الزمنية temporal features مثل الفترة الزمنية بين عمليات المستخدم والمدة الزمنية لتلك العمليات. تستخدم الدراسة الواردة في [1] الانحراف المعياري عن المتوسط الزمني المبني من تواقيع زمنية لعمليات طبيعية normal للتحقق من القيم المتطرفة ضمن مجال زمني محدد مسبقًا في أنظمة قواعد البيانات التي تعمل في الزمن الحقيقي real time database systems.

عند تحديث كائن بيانات مؤقت في قاعدة بيانات تعمل في الزمن الحقيقي، يتم مقارنة وقت التحديث مع الوقت المتوقع للتحديث. يقوم النظام بالإبلاغ عن التحديث على أنه تسلل إذا لم تحقق المقارنة شروطا محددة.

## 2.7. أنظمة كشف التسلل القائمة على تحليل العلاقات والتبعيات بين أوامر المستخدم Dependency and relation analysis

تحسب هذه التقنية التبعيات Dependencies والعلاقات Relations بين المجموعات المختلفة من إجراءات المستخدم والبيانات التي يتم الوصول إليها لمعرفة الأعمدة والصفوف والجداول المستهدفة من قبل المستخدم بالإضافة إلى الأوامر التي يتم إصدارها من قبل المستخدم ومعالجتها معًا.

فعلى سبيل المثال، يقوم نظام DEMIDS [2] ببناء ملفات شخصية للمستخدمين profiles بناءً على نشاطهم عن طريق تحديد مجموعات العناصر المتكررة frequent itemsets المكونة من أزواج سمة/قيمة. في مرحلة الكشف، يحسب النظام المسافة بين نشاط المستخدم ومجموعات العناصر المتكررة المبنية في مرحلة التدريب، فإذا كانت هذه المسافة أكبر من عتبة محددة مسبقا يُعتبر النشاط اختراقا. يعتمد استخلاص السمات عادةً على التحليل النحوي syntactical analysis لأوامر المستخدم.

يعمل نظام كشف التسلل المطروح في [3] على قواعد بيانات يتم التحكم بالوصول إليها عبر Role-based access control. ينمذج هذا النظام استعلامات المستخدم ببنى معطيات يطلق عليها اسم الخماسيات quiplets. تتألف كل خماسية من خمس حقول:

(SQL Command, Projection Relation Information, Projection Attribute Information, Selection Relation Information and Selection Attribute Information)

ويمكن التعبير عن الخماسية بشكل مختصر كما يلي Q(c, PR, PA, SR, SA) حيث:

* c: هو الحقل الذي يحمل معلومات حول نوع الأمر أو الاستعلام الصادر عن المستخدم (Select, Insert, Update, etc.)
* PR: هو الحقل الذي يحمل معلومات حول العلاقات relations (جداول قاعدة المعطيات) المستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام.
* PA: هو الحقل الذي يحمل معلومات حول المعاملات attributes المستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام.
* SR: هو الحقل الذي يحمل معلومات حول العلاقات relations (جداول قاعدة المعطيات) المستخدمة في جزء االاختيار Selection من الأمر أو الاستعلام.
* SA: هو الحقل الذي يحمل معلومات حول المعاملات attributes المستخدمة في جزء الاختيار Selection من الأمر أو الاستعلام.

في هذه الدراسة يتم تعريف ثلاثة أنواع من الخماسيات كل نوع يحمل كمية مختلفة من المعلومات عن الأمر أو الاستعلام الصادر عن المستخدم، وفق ما يلي:

* **coarse quiplet أو c-quiplet** تمثل كل خماسية من هذا النوع بند log من بنود سجل تدقيق قاعدة المعطيات database audit log. تتكون كل خماسية من هذا النوع من الحقول التالية:

(SQL-CMD, PROJ-REL-COUNTER, PROJ-ATTR-COUNTER, SEL-REL-COUNTER, SEL-ATTR-COUNTER)

حيث:

* + SQL-CMD: الحقل الأول ويحتوي أمر SQL الصادر من المستخدم (Select, Insert, Update, etc.)
  + PROJ-REL-COUNTER: الحقل الثاني ويحتوي عدد العلاقات relations (جداول قاعدة المعطيات) المستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام.
  + PROJ-ATTR-COUNTER: الحقل الثالث ويحتوي عدد المعاملات attributes المستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام.
  + SEL-REL-COUNTER: الحقل الرابع ويحتوي عدد العلاقات relations (جداول قاعدة المعطيات) المستخدمة في جزء الاختيار Selection من الأمر أو الاستعلام.
  + SEL-ATTR-COUNTER: الحقل الخامس ويحتوي عدد المعاملات attributes المستخدمة في جزء الاختيار Selection من الأمر أو الاستعلام.

من الواضح أن كمية كبيرة من المعلومات القيمة في سجل قاعدة البيانات يتم تجاهلها من قبل c-quiplets. ومع ذلك، من المفيد النظر في مثل هذا التمثيل البدائي للبيانات لأنه مناسب في حالة وجود عدد قليل من الأدوار Roles المنفصلة. علاوة على ذلك، فإن تمثيلات أكثر تعقيدًا لبنود سجل قاعدة البيانات مبنية على تعريف الخماسيات من النوع c-quiplets.

* **Medium quiplet أو m-quiplet** تمثل كل خماسية من هذا النوع بند log من بنود سجل تدقيق قاعدة المعطيات database audit log. تتكون كل خماسية من هذا النوع من الحقول التالية:

(SQL-CMD, PROJ-REL-BIN [], PROJ-ATTR-COUNTER [], SEL-REL-BIN [],

SEL-ATTR-COUNTER [])

حيث:

* + SQL-CMD: الحقل الأول ويحتوي أمر SQL الصادر من المستخدم (Select, Insert, Update, etc.)
  + PROJ-REL-BIN []: الحقل الثاني وهو عبارة عن شعاع من القيم الثنائية. يساوي طول الشعاع إلى عدد العلاقات relations (جداول قاعدة المعطيات). يأخذ العنصر ith في الشعاع PROJ-REL-BIN [] القيمة 1 إذا كانت العلاقة ith في قاعدة البيانات مستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام، ويأخذ القيمة 0 فيما عدا ذلك.
  + PROJ-ATTR-COUNTER []: الحقل الثالث هو عبارة عن شعاع من القيم العددية. يساوي طول الشعاع إلى عدد العلاقات في قاعدة المعطيات. قيمة العنصر ith في الشعاع PROJ-ATTR-COUNTER [] يساوي إلى عدد المعاملات attributes للعلاقة ith المستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام.
  + SEL-REL-BIN []: الحقل الرابع هو عبارة عن شعاع من القيم الثنائية. يساوي طول الشعاع إلى عدد العلاقات relations (جداول قاعدة المعطيات). يأخذ العنصر ith في الشعاع SEL-REL-BIN [] القيمة 1 إذا كانت العلاقة ith في قاعدة البيانات مستخدمة في جزء الاختيار Selection من الأمر أو الاستعلام، ويأخذ القيمة 0 فيما عدا ذلك.
  + SEL-ATTR-COUNTER []: الحقل الخامس هو عبارة عن شعاع من القيم العددية. يساوي طول الشعاع إلى عدد العلاقات في قاعدة المعطيات. قيمة العنصر ith في الشعاع ٍSEL-ATTR-COUNTER [] يساوي إلى عدد المعاملات attributes للعلاقة ith المستخدمة في جزء الاختيار Selection من الأمر أو الاستعلام.

نلاحظ أن m-quiplets هي عبارة عن توسيع لـ c-quiplets، فهي تحتوي على كمية أكبر من المعلومات عن بنود سجل قاعدة المعطيات.

* fine quiplet أو f-quiplet تمثل كل خماسية من هذا النوع بند log من بنود سجل تدقيق قاعدة المعطيات database audit log. تتكون كل خماسية من هذا النوع من الحقول التالية:

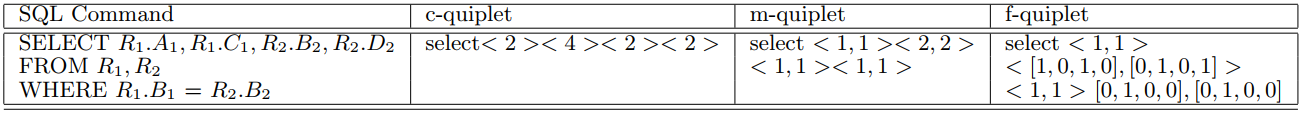
(SQL-CMD, PROJ-REL-BIN [], PROJ-ATTR-BIN [] [], SEL-REL-BIN [], SEL-ATTR-BIN [] [])

حيث:

* + SQL-CMD: الحقل الأول ويحتوي أمر SQL الصادر من المستخدم (Select, Insert, Update, etc.)
  + PROJ-REL-BIN []: الحقل الثاني وهو عبارة عن شعاع من القيم الثنائية. يساوي طول الشعاع إلى عدد العلاقات relations (جداول قاعدة المعطيات). يأخذ العنصر ith في الشعاع PROJ-REL-BIN [] القيمة 1 إذا كانت العلاقة ith في قاعدة البيانات مستخدمة في جزء الاسقاط Projection من الأمر أو الاستعلام، ويأخذ القيمة 0 فيما عدا ذلك.
  + PROJ-ATTR-BIN [] []: الحقل الثاني هو عبارة عن شعاع من n شعاع، حيث n هي عدد العلاقات في قاعدة البيانات. يأخذ العنصر PROJ-ATTR-BIN [i] [j] القيمة 1 إذا كان المعامل jth في العلاقة ith مستخدم في جزء الاسقاط Projection من الأمر أو الاستعلام، ويأخذ القيمة 0 فيما عدا ذلك.
  + SEL-REL-BIN []: الحقل الرابع هو عبارة عن شعاع من القيم الثنائية. يساوي طول الشعاع إلى عدد العلاقات relations (جداول قاعدة المعطيات). يأخذ العنصر ith في الشعاع SEL-REL-BIN [] القيمة 1 إذا كانت العلاقة ith في قاعدة البيانات مستخدمة في جزء الاختيار Selection من الأمر أو الاستعلام، ويأخذ القيمة 0 فيما عدا ذلك.
  + SEL-ATTR-BIN [] []: الحقل الخامس هو عبارة عن شعاع من n شعاع، حيث n هي عدد العلاقات في قاعدة البيانات. يأخذ العنصر PROJ-ATTR-BIN [i] [j] القيمة 1 إذا كان المعامل jth في العلاقة ith مستخدم في جزء الاختيار Selection من الأمر أو الاستعلام، ويأخذ القيمة 0 فيما عدا ذلك.

إن الخماسيات من نوع f-quiplet تحتوي الكمية الأكبر من المعلومات عن نشاطات المستخدم.

نورد في الشكل 3 مثالا عن أحد أنشطة المستخدم المسجلة في سجل تدقيق قاعدة البيانات والتمثيلات الثلاثة المذكورة آنفا الممثلة لهذا النشاط.



الشكل 3 مثال عن الأنواع الثلاثة من الخماسيات

لكشف الشذوذ في نشاطات المستخدمين الموجهة لقاعدة بيانات تستخدم التحكم في الوصول عبر توزيع المستخدمين على أدوار محددة role-based database access control بما يمكن من ربط نشاط كل مستخدم مع دور محدد يتم استخدام Naïve Bayes Classifier (NBC) وفق ما يلي:

في مرحلة التدريب يتم بناء المصنف لكل نوع من الخماسيات التي تمت دراستها بحيث يعتبر كل دور Role عبارة عن مجموعة أو cluster. في مرحلة الاختبار، يعتبر الاستعلام الصادر عن دور Role أنه اختراق إذا صنفه أي مصنف من المصنفات الثلاثة على أنه اختراق.

## 3.7. أنظمة كشف التسلل القائمة على تحليل محاذاة السلاسل Sequence Alignment

تقوم هذه التقنية بشكل أساسي بتحديد تسلسل عمليات المستخدم الشائعة (مثل الأوامر، معاملات جداول قاعدة البيانات، القيم التي تم الوصول إليها، إلخ). تتعرف أنظمة كشف التسلل من هذا النوع على سلسلة عمليات المستخدم المتكررة ذات الطول الكبير وتحددها ثم تقوم بتقسيمها إلى مجموعات فرعية أصغر لتصنيف تلك التسلسلات ومجموعاتها الفرعية كسلوك عادي للمستخدم. في مرحلة الكشف، تتم مطابقة كل تسلسل جديد للعمليات الصادرة عن المستخدم مع تسلسلات المستخدم المخزنة في فترة التدريب ومجموعاتها الفرعية. من خلال هذه المطابقة، يتم قياس كيفية اختلاف التسلسل الجديد عن التسلسلات المخزنة من أجل تقييم احتمالية كونها تدخلًا.

في الحل المقدم في [4]، يتم تحديد تسلسلات المعاملات والأوامر والجداول التي تم الوصول إليها لبناء ملفات تعريف المستخدمين profiles، حيث أن السمات المقترحة proposed features هي أنواع الأوامر (Select, Insert, Update, etc.) ، المعاملات الحساسة، جميع المعاملات، العمليات على المعاملات، ومزيج من جميع السمات. في مرحلة التعلم، يقوم النظام ببناء نماذج تسلسل مع إعطاء عتبة لتحديد الحد الأقصى لعدد الاختلافات. في مرحلة الكشف، يستخدم النظام أيضًا عتبة لحساب أكبر عدد من الاختلافات المسموح بها بين التسلسلات المختبرة وتلك التي تم حفظها في مرحلة التعلم، للنظر في التسلسل على أنه طبيعي أو غير طبيعي.

## 4.7. أنظمة كشف التسلل القائمة على دمج تحليل التبعية وتحليل محاذاة السلاسل Integrating Dependency with Sequence Alignment

في الدراسة الواردة في [5]، يتم إيجاد علاقات التبعية بين معاملات البيانات attributes على مستوى الإجراءات transactions عن طريق تحليل التسلسلات الأكثر احتمالية في عمليات المستخدم. حيث يفترض القائمون على الدراسة أن أي تعديل لقيمة معامل مرتبط بسلسة من العمليات الأخرى كقراءة أو كتابة معاملات أخرى. وبالتالي فإن كل عملية تعديل تعرف بثلاث مجموعات:

1. مجموعة القراءة The read set: وهي مجموعة المعاملات التي يتم قراءتها بسبب عملية التعديل.
2. مجموعة ما قبل الكتابة The pre-write set: وهي مجموعة المعاملات التي يتم كتابتها قبل أن تتم عملية التعديل وبسببها.
3. مجموعة ما بعد الكتابة The post-write set: وهي مجموعة المعاملات التي يتم كتابتها بعد أن تتم عملية التعديل وبنتيجتها.

يتم تصنيف كافة الإجراءات transactions التي لا تتبع قواعد التسلسل المحددة في مرحلة التدريب، على أنها اختراقات.

يحسّن العمل في [6] [7] الدراسة الواردة في [5] من خلال مراعاة حساسية المعاملات، أي إعطاء مقدار من الأهمية لكل معامل. يقترح العمل ثلاثة مستويات من حساسية المعاملات، وذلك بالنظر إلى تكرارها في الإجرارءات التي تم تحليلها: عالية ومتوسطة ومنخفضة. يتم استخدام خوارزمية تنقيب في المعطيات لاستخراج التبعيات بين معاملات قاعدة البيانات وإنشاء قواعد تعكس هذه التبعيات، وذلك بالنظر إلى تسلسل العمليات (القراءة والكتابة) وحساسية كل معامل. تصنف أي إجرائية لا تتبع هذه القواعد على أنها اختراقات.

تم اقتراح خوارزمية تعلم لتمثيل الإجراءات من خلال الرسوم البيانية الموجهة directed graphs التي تصف مسارات التنفيذ في [8]. يُنظر إلى الإجراءات الجديدة التي تنحرف عن مسارات التنفيذ المبنية في مرحلة التدريب على أنها تسلسلات غير مصرح بها. السمات features المستخدمة لبناء مسارات التنفيذ هي نوع الأمر (Select, Insert, Update, etc.) ، الكائنات المستهدفة (الجداول والأعمدة)، معاملات البيانات، والتي يتم الحصول عليها جميعًا من سجل تدقيق قاعدة المعطيات database audit log.

## 5.7. أنظمة كشف التسلل القائمة على التحليل الإحصائي Statistical analysis

يستخدم التحليل الإحصائي في العديد من DIDS لحساب إحصائيات نشاط المستخدم. يستخدم البحث في [9] الدالَّات الإحصائية على قيم مرجعية يتم الحصول عليها من البيانات في العلاقات (جداول قاعدة البيانات) ومن التغيرات في العلاقات (التغيرات في قيم الكائنات/المعاملات المعتبرة كقيم المرجعية) للكشف عن الشذوذ.

يعتمد العمل في [10] على حساب الدالات الإحصاءات مثل العدد، الحد الأقصى، الحد الأدنى، المتوسط والانحراف المعياري لكل معامل من مجموعة البيانات الناتجة أو المتأثرة بكل أمر صادر عن المستخدم. يتم تخزين هذه الإحصائيات في شعاع ذي بعد ثابت يسمى S-Vector. يتم بعد ذلك استخدام مجموعة S-Vectors لكل مستخدم لتطبيق تقنيات مثل التجميع clustering، naïve Bayes، SVM أو أشجار القرار decision trees من أجل الحصول على نماذج تمثل السلوك الطبيعي للمستخدم. في مرحلة الكشف عن التسلل، يتم تطبيق الانحراف الإحصائي لفحص كل أمر صادر عن المستخدم وتصنيفه على أنه طبيعي أو غير طبيعي.

## 6.7. أنظمة كشف التسلسل القائمة على نظرية المعلومات Information-Theoretic Analysis

تحسب تقنية التحليل النظري للمعلومات Information-theoretic analysis مقاييس مثل الإنتروبية entropy وكمية المعلومات المعلومات لتوصيف ملفات تعريف المستخدمين profiles ومقارنتها مع الإجراءات اللاحقة لمعرفة كيفية اختلافها واتخاذ قرار بشأنها كونها طبيعية أو غير طبيعية.

يصف العمل في [11] حلا قائما على هذه التقنية. تتكوّن السمات من مجموعة من بنود سجل التدقيق بحيث يكون لكل كائن بيانات n متغير (مثل عنوان IP، حجم الرسالة، إلخ). يُستخدم الإنتروبي كمقياس لاختلاف بيانات سجل التدقيق بعضها عن بعض؛ كلما كان الإنتروبي أصغر، قل عدد السجلات المختلفة عن بعضها البعض (أي، هناك كمية أكبر من التكرار)، مما يشير إلى مجموعات بيانات أكثرطبيعية في سجل التدقيق. تشير حقيقة تكرار العديد من الأحداث في مجموعة بيانات إلى احتمال ظهورها في المستقبل مما يشير إلى أنها ناتجة عن تصرف طبيعي من المستخدم.

## 7.7. أنظمة كشف التسلسل القائمة على تحليل قوالب الأوامر Command Template Analysis

تقوم أنظمة كشف التسلل المبنية وفق هذه التقنية باستخدام سجل تدقيق لأوامر المستخدم لتحليل كافة الأوامر الطبيعية للمستخدم وبناء قوالب تمثل النشاط النموذجي للمستخدم بشكل عام.

في [12] يقترح المؤلفون تطبيق تحليل القواعد grammar-based analysis باستخدام تقنيات التعلم الآلي المبنية على tree-kernel بدلاً من vector-based data الشائعة الاستخدام. يستخدم هذا الأسلوب التحليل الشجري لبنية SQL لربط الأوامر مع التطبيقات وللتمييز بين الأوامر الحميدة والخبيثة من خلال فحص التغييرات في بنية أشجار عبارات الاستعلام والأوامر. تم تطوير مقياس مسافة لقياس تشابه أوامر SQL باستخدام أشجار التحليل الخاصة بتلك الأوامر. يتم استخدام Support Vector Machine SVM في مرحلة التعليم وتطبيق التجميع Clustering للتمييز بين الأوامر الخبيثة والحميدة عن طريق كشف النقاط الشاذة. تعمل هذه الطريقة على تعزيز التشابه لسياق الأوامر الذي يتيح تحديد أقرب أمر طبيعي من عبارة ضارة، مما يساعد في تحليل السبب الجذري للاختراق.

# الاستجابة للاختراق ومنعه

يمكن تعريف الاستجابة للاختراق ومنعه بأنها القدرة على وقف عمل الاختراق عند حدوثه أو حتى قبل حدوثه. يمكن أن نلاحظ أن بعض أنظمة كشف التسلل المطروحة في الفقرة 7 تتيح منع الاختراق بشكل كامل، بينما يمكن للبعض الآخر تحقيق ذلك جزئيًا فقط.

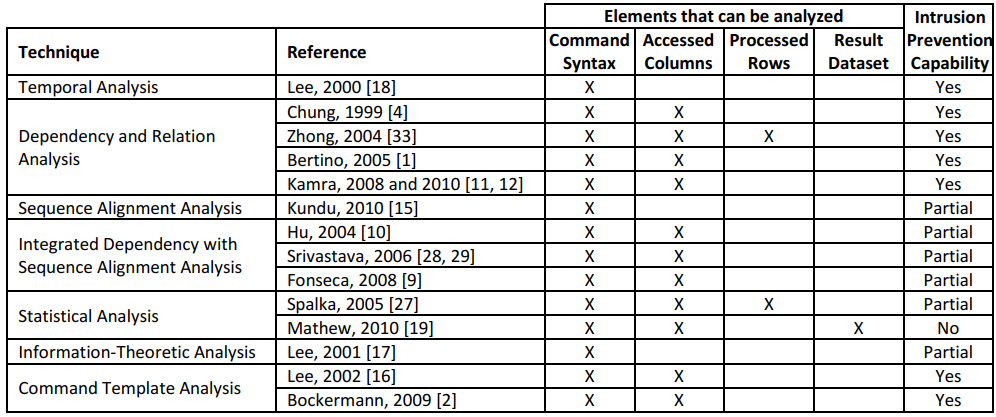
في [1]، تكتشف تقنية التحليل الزمني أي أوامر صادرة عن المستخدم تتطلب التنفيذ خارج مجال زمني محدد مسبقًا وبالتالي ترفض تنفيذها وتمنع تحقيق الاختراق. تتيح تقنية تحليل محاذاة التسلسل المستخدمة في [4] منع الاختراق من خلال تجنب إجراءات المستخدم عندما تكتشف تسلسلًا مريبًا من الإجراءات. ومع ذلك، فإن هذه التقنية تحتاج إلى انتظار كمية كبيرة من الإجراءات التي تشكل هذا التسلسل، مما يعني أنه من المحتمل أن يُكتشف التدخل بعد انتهاء تنفيذ بعض هذه الإجراءات، مما يجعل تقنية التحليل الزمني قادرة على منع الاختراق بشكل جزئي فقط.

الحلول المستندة إلى تحليل العلاقات والتبعيات كالواردة في [3] قادرة تمامًا على منع الاختراق، نظرًا لأنها تتحقق من صيغة كل أمر صادر عن المستخدم فإذا وجدت هذه الأوامر مشبوهة تقوم بإيقاف تنفيذها. الحلول التي تجمع بين مزيج من تحليل التبعية ومحاذاة التسلسل مثل [8] [5] [6] [7] قادرة على أداء منع الاختراق جزئياً فقط، لنفس الأسباب المشار إليها بشأن الحل المقترح في [4].

الحلول المقدمة في [10] [9] المستندة إلى التحليل الإحصائي، غير قادرة في الغالب على منع الاختراق، لأنها تعتمد في الغالب على تحليل التغييرات في البيانات أو نتائج الأوامر بعد تنفيذها. هذا يعني أنه يمكن لهذه الحلول اكتشاف التسلل بعد حدوثه. ومع ذلك، يمكن تكييف الحل في [9] للتحقق من البيانات الإحصائية المسبقة المتعلقة بصفوف جداول قاعدة البيانات المطلوب معالجتها، مما يتيح قدرات منع اختراق بشكل جزئي. لهذا السبب نفسه، فإن منهج تحليل نظرية المعلومات المقدم في [11] قد يحقق أيضًا منع اختراق بشكل جزئي.

يمكن للحلول المستندة إلى تحليل الأوامر والقوالب في [12] أن تُمكِّن تمامًا منع الاختراق بسبب نفس الأسباب المذكورة سابقًا لتحليل العلاقات والتبعيات في [11].

نلخص في الجدول 1 تقنيات أنظمة كشف التسلل في قواعد البيانات مع عناصر البيانات التي تستند إليها كل طريقة. في العمود الأخير نجد معلومات عن إمكانية كل طريقة لمنع الاختراق.



الجدول 1 تقنيات كشف التسلل

# تطبيق أنظمة كشف التسلل في قواعد البيانات

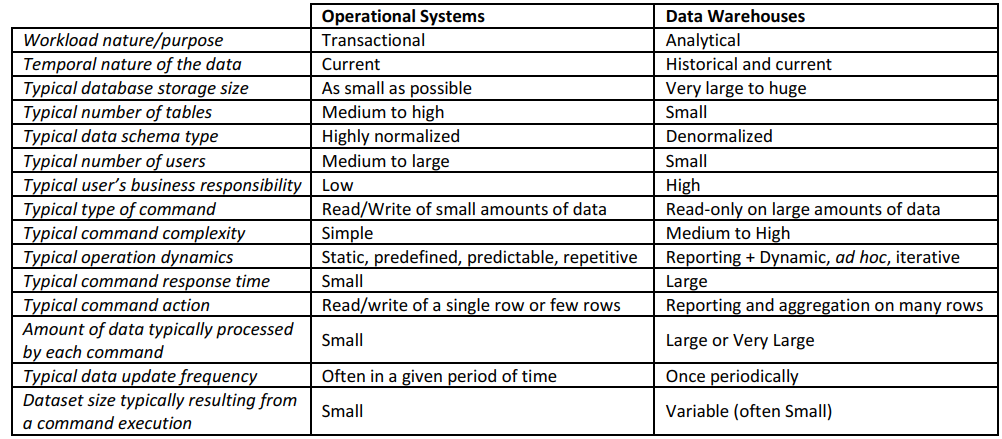
تعتمد قابلية تطبيق DIDS في أنظمة قواعد البيانات على البيئة التي يفترض أن تعمل ضمنها. يُعتبر فهم خصائص العمليات النموذجية التي تتم ضمن كل بيئة من بيئات العمل أمرًا حاسمًا لتحديد نوع أو فئة تقنيات كشف الاختراق ID التي يمكن أن تكون أكثر كفاءة نظرًا لطبيعة العمليات التي تتم ضمن هذه البيئة، وبالتالي تعتبر أكثر ملاءمة لقاعدة البيانات المحددة. يوجد نوعين أساسيين لبيئة العمل، نتكلم عنهما فيما يلي.

## 1.9. أنظمة قواعد البيانات الإجرائية transactional وأنظمة قواعد البيانات التحليلية analytical

تتكون أنظمة قواعد البيانات الإجرائية Transactional database systems عادةً من مجموعة من التطبيقات ومصادر البيانات التي تمكن من إنجاز المعاملات وتخزينها [13]. تم تصميم قواعد البيانات الإجرائية لإدارة البيانات المطلوبة في دعم المعاملات والإجراءات الصادرة عن المستخدم بدلاً من تحليل البيانات للمساعدة في اتخاذ القرارات. تحوي قواعد البيانات الإجرائية عادة العديد من المستخدمين الذين يقومون بقراءة وكتابة كميات صغيرة من البيانات. على سبيل المثال، يمكن أن يكون في النظام المصرفي ATM مئات أو آلاف المستخدمين الذين يمكنهم الوصول إلى أرصدة حساباتهم في نفس الوقت كما يمكنهم سحب أو تحويل مبلغ معين من المال. ومن خصائص هذه الأنظمة أنها لا تتطلب الاحتفاظ بالبيانات لفترات طويلة؛ تحتاج فقط إلى الرصيد الحالي وأحدث سجلات الحركة لتكون قادرة على تلبية طلبات المستخدمين والمعاملات بشكل كافٍ.

في المقابل، يتم استخدام الأنظمة التحليلية عادة من قبل عدد أقل من المستخدمين الذين يستفسرون عن كميات كبيرة من البيانات للحصول على معلومات تفيد في تحليل الأعمال للمساعدة في اتخاذ القرار. باستخدام نفس نظام الصراف الآلي المصرفي كمثال، فإن الفرق هو أن الأشخاص -من البنك- الذين يحتاجون إلى اتخاذ قرارات بشأن الأعمال (أي المديرين والإداريين، وما إلى ذلك) يريدون معرفة متوسط الرصيد للحسابات خلال الأشهر الستة الماضية داخل منطقة جغرافية معينة، على سبيل المثال، من أجل المساعدة في اتخاذ القرارات الاستراتيجية مثل فتح مكتب فرعي جديد أو تشجيع الناس على زيادة استثماراتهم من خلال تقديم أسعار فائدة أفضل. لتنفيذ هذا النوع من الاستعلام، يحتاج النظام إلى الاحتفاظ ببيانات الأرصدة لفترات طويلة، بالإضافة إلى أنه سيقرأ ملايين السجلات لجميع العملاء داخل منطقة معينة لحساب هذا المتوسط.

يؤدي هذا النوع من الإجراءات التحليلية إلى أنماط وصول إلى البيانات تستهلك الكثير من المتطلبات، بحيث إذا تم تشغيلها فوق قاعدة بيانات إجرائية يمكن أن تؤدي إلى حجب كميات كبيرة من البيانات واستهلاك الموارد الحسابية بطريقة يمكن أن تعرض توفر نظام الإجراءات للخطر، بحيث يجعله غير قادر على دعم الإجراءات الصادرة عن المستخدمين الآخرين. للتخفيف من استهلاك الموارد، والحد من المخاطر التشغيلية في التطبيقات الإجرائية التي تدعم الأعمال التجارية ولتوفير بنية بيانات محسنة لأغراض دعم القرار التحليلي، يتم استخدام مستودعات البيانات Data Warehouses (DWs). تفصل DWs بوضوح عمليات الأعمال التحليلية عن العمليات الإجرائية. وفقًا لـ [13]، يمكننا أن نلخص الاختلافات بين الأنظمة الإجرائية والأنظمة التحليلة في الجدول 2.



الجدول 2 الاختلافات بين الأنظمة الإجرائية والأنظمة التحليلية

## 2.9. تطبيق كشف التسلل في أنظمة قواعد البيانات

كما هو موضح في الجدول 1، تركز معظم DIDS على التحليل النصي Syntax Analysis لأوامر المستخدم (أي تحليل نص استعلام SQL لإنشاء ملفات تعريف المستخدمين profiles). كما هو موضح في [10]، فإن أبرز المشاكل التي تعترض هذا النوع من الأنظمة هي:

* قد تختلف صياغة استعلامات المستخدم الطبيعي اختلافًا كبيرًا، ولكن هذه الاستعلامات المتباينة كثيرا من ناحية الصياغة (النص) تنتج مخرجات "طبيعية" (أي غير مؤذية)، مما يؤدي إلى توليد إنذارات موجبة كاذبة false positives.
* قد يصيغ المهاجم استعلاماته الخبيثة لتختلف قليلاً عن صيغة الاستعلامات الواردة في الملفات الشخصية لسلوك المستخدم "الطبيعي"، ولكن ينتج عنها إخراج "غير طبيعي" (أي ضار وتدخلي)، مما يولد إنذارات سالبة كاذبة false negatives (أي تمر الهجمات دون أن يتم اكتشافها).

في قواعد البيانات حيث تحتوي عمليات المستخدم الطبيعي على عدد محدد من الأوامر المصاغة بشكل متمايز والتي يتم إصدارها بشكل متكرر، يكون من الممكن الاعتماد على تحليل صيغة أوامر SQL لتحقيق أنظمة كشف تسلل ذات كفاءة عالية، وهذا ما يحدث عادة في أنظمة قواعد البيانات الإجرائية transactional databases. ولكن، في أنظمة قواعد البيانات التحليلية مثل DW تكون الأوامر الصادرة عن المستخدم مخصصة للعمل المراد إنجازه، ولها أوقات تنفيذ متغيرة مع أنماط متغيرة للوصول إلى البيانات. وبالتالي، لا يمكن التنبؤ بها ولا يمكن حصرها في مجموعات منفصلة. وهذا ما يجعل التمييز بين الأوامر الطبيعية وغير الطبيعية في DWs مهمة صعبة للغاية. في قواعد البيانات التحليلية هذه، لا يمكن الاقتصار على منهجية تحليل صيغة الأوامر كما هو الحال في منهجية تحليل قوالب أوامر SQL أو منهجية التحليل الإحصائي للبيانات التي تم الوصول إليها (على سبيل المثال، الجداول أو الأعمدة التي يتم الوصول إليها). إن الاقتصار على هذا النهج في قواعد البيانات التحليلية سيحقق أنظمة كشف تسلل غير موثوقه.

بالنظر إلى خصائص الإجراءات التي يمكن أن تتم على قواعد البيانات التحليلية، فإن الاعتماد على التحليل الزمني Temporal Analysis الوارد في [1] غير كاف لتحقيق أنظمة كشف تسلل فعالة. يعود السبب في ذلك إلى أن الأوامر الصادرة عن المستخدم تحدث بتكرارات عشوائية وأزمنة تنفيذها غير قابلة للتوقع. في أنظمة قواعد البيانات الإجرائية Transactional databases، يحقق الاعتماد على التحليل الزمني أنظمة كشف تسلل فعالة للغاية خاصة إذا كانت أوامر المستخدم تحدث بتكرار محدد ويتطلب تنفيذها وقتا قابلا للتوقع.

يمكن الاعتماد على منهجية تحليل محاذاة التسلسل و منهجية الدمج بين تحليل التبعية وتحليل محاذاة التسلسل لتحقيق أنظمة كشف تسلل فعالة في قواعد البيانات التحليلية. وذلك لأنها لا تعتمد بشكل أساسي على تحليل صيغة أوامر المستخدم وإنما على تسلسل عملياته. ولكن الاعتماد على هذا النهج ينطوي على المخاطرة بفقدان كمية من المعلومات قبل اكتشاف الاختراق. وذلك لأن النظام في هذه الحالة يحتاج إلى تحليل عدة عمليات قبل أن يقرر كون الإجراء الذي يتم تنفيذه اختراقا أو تصرف طبيعي من مستخدم طبيعي.

# أنظمة كشف التسلل القائمة على تحليل نتائج الاستعلامات data-centric approach

في الدراسة الواردة في [10] يبني الباحثون منهجا في كشف التسلل قائم بشكل أساسي على ملاحظة مفادها أن جميع أنظمة كشف التسلل القائمة على تحليل صيغة أوامر SQL تقدم أداء متواضعا من ناحية الإنذارات الموجبة والسالبة الكاذبة. الفكرة الرئيسية في هذا المنهج هي أن أفضل طريقة للتمييز بين السلوكيات الطبيعية وغير الطبيعية هي النظر مباشرة إلى ما يحاول المستخدم الوصول إليه - نتيجة الاستعلام نفسه - بدلاً من كيفية التعبير عن الاستعلام. وبعبارة أخرى، فإن هذا المنهج الذي يركز على البيانات الناتجة يقدر دلالات الاستعلامات أكثر من طريقة تركيبها. فعندما يحاول شخص ضار من الداخل معرفة معلومات حول البيانات وعلاقاتها، فإن البيانات التي يتم الوصول إليها تختلف بالضرورة عن البيانات التي اعتاد المستخدم الطبيعي الوصول إليها. يحدث هذا الانحراف في هجمات جمع البيانات data harvesting attacks وكذلك في الهجمات المتخفية masquerading attacks. لذلك، قد يتجاوز التكامل بين منهج تحليل صيغة أوامر SQL ومنهج تحليل البيانات الناتجة عن هذه الأوامر معظم العقبات التي تعاني منها أنظمة كشف التسلل.

# References

|  |  |
| --- | --- |
| [1] | Lee, "Intrusion Detection in Real-time Database Systems via Time Signatures," *Real-time Technology and App,* 2000. |
| [2] | Chung, "DEMIDS: A Misuse Detection System for Database Systems," in *IFIP TC11 WG11.5 Conf. on Integrity and Internal Control in Information Systems*, 1999. |
| [3] | Kamra, "Detecting Anomalous Access Patterns in Relational Databases," *Springer VLDB Journal,* 2008. |
| [4] | Kundu, "Database Intrusion Detection Using Sequence Alignment," *International Journal of Information Security (9),* 2010. |
| [5] | Hu, "A Data Mining Approach for Database Intrusion Detection," *ACM Intern. Symposium on Applied Computing (SAC),* 2004. |
| [6] | Srivastava, "Database Intrusion Detection using Weighted Sequence Mining," *Journal of Computers,* 2006. |
| [7] | Srivastava, "Weighted Intra-Transactional Rule Mining for Database Intrusion Detection," in *Int. Pacific-Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 2006. |
| [8] | Fonseca, "Online Detection of Malicious Data Access using DBMS Auditing," *ACM Int. Symposium on Applied Computing (SAC),* 2008. |
| [9] | Spalka, "A Comprehensive Approach to Anomaly Detection in Relational Databases," in *IFIP Int. Conf. Data and Applications Security and Privacy (DBSec)*, 2005. |
| [10] | Mathew, "A Data-Centric Approach to Insider Attack Detection in Database Systems," in *International Conference on Recent Advances in Intrusion Detection (RAID)*, 2010. |
| [11] | Lee, "Information-Theoretic Measures for Anomaly Detection," *IEEE Symposium on Security and Privacy,* 2001. |
| [12] | Bockermann, "Learning SQL for Database Intrusion Detection using Context-Sensitive Modeling," in *International Conference on Knowledge Discovery and Machine Learning (KDML)*, 2009. |
| [13] | Kimball, The Data Warehouse Toolkit, Wiley & Sons, 2013. |