



دورة إخفاء هوية البيانات والتقنيات المعززة للخصوصية PETs لمشاركة البيانات



دورة إخفاء هوية البيانات والتقنيات المعززة للخصوصية PETs لمشاركة البيانات

مقدمة:

دورة إخفاء هوية البيانات والتقنيات المعززة للخصوصية PETs لمشاركة البيانات دورة تمتد أسبوع، موجهة لفرق علم البيانات ومشاركة البيانات والنشر الإحصائي وهندسة الخصوصية، وتنتهي بإعداد خطة نشر البيانات واختيار التقنيات المعززة للخصوصية لجهة افتراضية. فكل من الوحدات التي تملك بيانات جزئية ثمينة ترفض طلبات الشركاء أو تسلم مستخلصات حذفت منها الأسماء فقط، لأن أحدا لا يقيس مدى سهولة إعادة التعرف على الأفراد. ويعد المشاركون مجموعات البيانات أو يعتمدونها في عملهم، ويتدربون في مختبرات تطبيقية على معالجة المعارف شبه المباشرة واختبار الهجمات ومعايرة الضجيج. وتقدم كور كونسبت هذه الدورة.

أهداف الدورة:

- تصنيف سمات مجموعة البيانات إلى معارف مباشرة ومعارف شبه مباشرة وقيم حساسة وفق تصنيف التقنيات في ISO/IEC 20889، وهو المعيار الدولي لمصطلحات إزالة الهوية
- تطبيق هرميات التعميم والحجب وتسقيف القيم المتطرفة وتبديل السجلات للوصول إلى مستويات مستهدفة من إخفاء الهوية بأسلوب k وتنوع l وتقارب t
- قياس مخاطر إعادة التعرف بنماذج هجوم المدعي والصحفي والمسوق واختبارات الربط مع مصادر مساعدة
- معايرة آليات الخصوصية التفاضلية والحساسية وتوزيع قيمة إيسيلون على الأعداد والمجاميع وواجهات الاستعلام المنشورة
- المقارنة بين بيئات التنفيذ الموثوقة والحوسبة الآمنة متعددة الأطراف والتشفير المتماثل الشكل والتعلم الاتحادي للتحليل المشترك دون تبادل البيانات الخام
- إعداد خطة نشر البيانات واختيار التقنيات المعززة للخصوصية التي تربط كل سيناريو مشاركة بمسار نشر ودرجة مخاطر متبقية

الفئة المستهدفة:

- فرق علم البيانات المسؤولة عن إعداد المستخلصات والملفات التحليلية للشركاء والباحثين
- فرق مشاركة البيانات والوصول إليها المسؤولة عن تقييم الطلبات الواردة واعتماد شروط النشر
- فرق الإنتاج الإحصائي المسؤولة عن نشر جداول التكرار وجداول المقادير والبيانات الجزئية للاستخدام العام
- فرق هندسة الخصوصية المسؤولة عن اختيار تقنيات الحماية من الإفصاح وضبط إعداداتها
- فرق خدمات بيانات البحث المسؤولة عن تشغيل بيئات الوصول المحكوم وفحص المخرجات

محاوّر الدورة:

اليوم 1: أسس مخاطر الإفصاح وتوصيف مجموعات البيانات

- مقارنة أنواع الإفصاح عن الهوية والسمات والعضوية
- نظرة على تصنيف تقنيات إزالة الهوية في ISO/IEC 20889
- وسم المعارف المباشرة وشبه المباشرة والسمات الحساسة
- الأسماء المستعارة مقابل إخفاء الهوية عبر التجزئة بالمفتاح وخزائن الرموز
- توصيف قائمة الإصدارات بقياس التفرد وتكرارات العينة

اليوم 2: نماذج الخصوصية وبنية مسارات النشر

- سياق إطار ISO/IEC 27559 لإزالة الهوية وعتبات المخاطر
- فئات التكافؤ في إخفاء الهوية بأسلوب k واختيار الحد الأدنى للمجموعة
- تنوع t وتقارب t في مواجهة هجمات التجانس
- طيف مسارات النشر من الملفات المفتوحة إلى الوصول عبر البيئات المعزولة
- ربط مبادئ الخصوصية في ISO/IEC 29100 بقرارات النشر

اليوم 3: مختبر معالجة البيانات الجزئية وحماية الجداول

- هرميات التعميم للعمر والرمز البريدي ورموز المهن
- الحجب الموضوعي وتسقيف القيم وتبديل السجلات في البيانات الجزئية
- الحجب الأولي والثانوي للخلايا في جداول المقادير بأداة tau-Argus
- التقريب المضبوط ومعالجة الأعداد الصغيرة في أداة sdcTable
- مقاييس فقدان المعلومات وفحوص الصلاحية التحليلية

اليوم 4: اختبار الهجمات والخصوصية التفاضلية والتقنيات التشفيرية المعززة للخصوصية

- محاكاة هجمات الربط باستخدام السجلات العامة المساعدة
- اختبار هجمات استنتاج العضوية وإعادة البناء على البيانات المجمعة
- آليتا لابللاس وغاوس مع الحساسية وتوزيع قيمة إيسيلون
- المفاضلة بين بيانات التنفيذ الموثوقة والحوسبة الآمنة متعددة الأطراف
- التشفير المتمثل الشكل والتعلم الاتحادي للتحليل بين الجهات

اليوم 5: تطبيقات مختبرية وإعداد خطة نشر البيانات واختيار التقنيات المعززة للخصوصية

- حالة سجلات دخول المستشفى ومعالجة البيانات الجزئية ودرجة المخاطر
- حالة معاملات التجزئة وتصميم نشر الاستعلامات بالخصوصية التفاضلية
- فحص نتائج الباحثين الخارجة من البيئة المعزولة بأداة SACRO
- مصفوفة قرار مسارات النشر والتقنيات المعززة للخصوصية لخمسة طلبات مشاركة
- استكمال خطة نشر البيانات واختيار التقنيات المعززة للخصوصية ومناقشتها مع الأقران

المهارات التي ستكتسبها:

- تحليل المعارف شبه المباشرة
- تصميم هرميات التعميم
- حجب خلايا الجداول
- اختبار هجمات الربط
- معايرة ميزانية الخصوصية
- اختيار تقنيات الحوسبة الآمنة
- قياس فقدان المعلومات
- تصميم مسارات النشر

لماذا تحضر هذه الدورة:

- تسليم خطة نشر البيانات واختيار التقنيات المعززة للخصوصية إلى مالك البيانات ولجنة اعتماد مشاركة البيانات، متضمنة المعالجة ودرجات المخاطر ومسارات النشر
- البت في إمكانية إتاحة مجموعة البيانات المطلوبة بيانات جزئية للاستخدام العام أو عبر بيئة معزولة محكومة أو مجاميع مضاف إليها ضجيج فقط
- تفادي حوادث إعادة التعرف ورفض الشراكات والملفات المحجوبة بإفراط التي تعقب الإصدارات غير المختبرة
- توجيه المحللين والإحصائيين في وسم المعارف شبه المباشرة وإعدادات الحجب وتوزيع قيمة إبسيلون في الإصدارات الدورية

الخاتمة:

بعد العودة إلى العمل يسلم المشاركون خطة نشر البيانات واختيار التقنيات المعززة للخصوصية إلى مالك البيانات ولجنة اعتماد مشاركة البيانات، فتستند الجهات إلى درجات المخاطر وإعدادات المعالجة ومسارات النشر الواردة فيها لقبول طلبات الشركاء أو ردها، وللاختيار بين الملفات المفتوحة والوصول عبر البيئات المعزولة والحوسبة على البيانات المشفرة، ولتحديد ميزانيات الخصوصية لخدمات الاستعلام المنشورة. وبعد أول إصدار وفق الخطة، تقارن الوحدة مخاطر إعادة التعرف المقيسة وفقدان المعلومات بالمستهدفات، ثم تعدل أحجام المجموعات وقواعد الحجب وقيم إسيلون.

الأسئلة الشائعة:

ما الذي يحتاجه المشاركون قبل دورة إخفاء هوية البيانات والتقنيات المعززة للخصوصية؟

يحتاج المشاركون إلى خبرة عملية في إعداد مجموعات البيانات أو الاستعلام عنها أو اعتمادها، وقدرة على قراءة الإحصاءات الوصفية الأساسية. وتستخدم المختبرات جداول إلكترونية وأدوات إحصائية مفتوحة، فتفيد خبرة بسيطة في R أو Python، مع توفير ملفات إرشادية لكل تمرين.

بم تختلف دورة إخفاء هوية البيانات والتقنيات المعززة للخصوصية عن دورات أنظمة الخصوصية أو أمن قواعد البيانات؟

تعلم هذه الدورة المعالجة التقنية لمجموعات البيانات واختبارها قبل النشر. أما دورات أنظمة الخصوصية فتتناول الواجبات النظامية والحوكمة وحقوق الأفراد، ودورات أمن قواعد البيانات تقوي المحركات والصلاحيات والتشفير، ولا تقيس أي منهما مخاطر إعادة التعرف أو تضبط الضجيج في الإحصاءات المنشورة.

هل يكفي حذف الأسماء وأرقام الهوية لإخفاء هوية البيانات قبل مشاركتها؟

لا يكفي. فتركيبات المعارف شبه المباشرة مثل تاريخ الميلاد والمنطقة والمهنة كثيرا ما تميز أفرادا بعينهم، وقد يعيد الربط مع المصادر العامة كشف الهويات. ويتطلب إخفاء الهوية معالجة هذه التركيبات وقياس المخاطر، واستخدام آليات مثل الخصوصية التفاضلية عند تكرار الاستعلامات.

ماذا يعود به المشاركون من دورة إخفاء هوية البيانات والتقنيات المعززة للخصوصية؟

يعود المشاركون بخطة نشر البيانات واختيار التقنيات المعززة للخصوصية، وفيها وسم السمات وإعدادات المعالجة ودرجات المخاطر ومسار نشر لكل طلب واختيار تقنية التحليل المشترك، إضافة إلى ملفات المختبر التي يمكن إعادة استخدامها على مجموعات بيانات وحدته.