(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104216983 B
(45)授权公告日 2019.03.01

(21)申请号 201410441434.5

(22)申请日 2014.09.01

(65)同一申请的已公布的文献号
申请公布号 CN 104216983 A

(43)申请公布日 2014.12.17

(73)专利权人 北京邮电大学
地址 100876 北京市海淀区西土城路10号
专利权人 无锡北邮感知技术产业研究院有限公司

(72)发明人 芦效峰 鲁鹏 李睿凡 李蕾 袁彩霞 刘咏彬 曲昭伟 李晖

(51)Int.Cl.
G06F 16/95(2019.01)
G06F 21/60(2013.01)

(56)对比文件
CN 102163230 A,2011.08.24,
CN 102508918 A,2012.06.20,
CN 101917513 A,2010.12.15,
CN 102163230 A,2011.08.24,
CN 101841529 A,2010.09.22,
Ukil A,Bandyopadhyay S, Pal A.IoT-Privacy: To be private or not to be private.《IEEE INFOCOM 2014 - IEEE Conference on Computer Communications Workshops》.2014,123-124.

审查员 乔帅

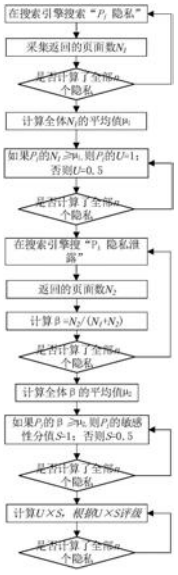
权利要求书1页 说明书3页 附图1页

(54)发明名称

基于采集搜索引擎数据的隐私信息评级方法

(57)摘要

本发明公开了一种基于采集搜索引擎数据的隐私信息评级方法,包括以下步骤:第一步从搜索引擎采集数据确定每个隐私信息的普遍性分值U,第二步从搜索引擎采集数据确定每个隐私信息的敏感性分值S,第三步根据U×S计算结果确定隐私信息的安全等级。本发明提供的评级方法使用的数据从搜索引擎采集,数据来源于使用搜索引擎的庞大的用户群,因此评级结果不依赖于个人的经验或意见,评级结果具有公正性;同时本发明提供的隐私信息评级方法不针对特定的隐私内容,既可以评定全体隐私信息,也可用于评定应用系统中有限数量的隐私信息。



CN 104216983 B

1.基于采集搜索引擎数据的隐私信息评级方法,其特征在于,评级包括以下步骤:

第一步:从搜索引擎采集数据确定每个隐私信息的普遍性分值U;

$$U=\begin{cases}1, N_{1}(x) \geq \frac{\sum_{i=1}^{T} N_1(i)}{T} \\ 0.5, N_{1}(x) < \frac{\sum_{i=1}^{T} N_1(i)}{T}\end{cases}$$

$N_1(i)$是查询第i个隐私信息的页面数量,$N_1(x)$是查询当前隐私信息的页面数量,T是查询隐私信息的总数;

第二步:从搜索引擎采集数据确定每个隐私信息的敏感性分值S;

$$S=\begin{cases}1, \frac{N_2(x)}{N_1(x)+N_2(x)} \geq \frac{\sum_{i=1}^{T} \frac{N_2(i)}{N_1(i)+N_2(i)}}{T} \\ 0.5, \frac{N_2(x)}{N_1(x)+N_2(x)} < \frac{\sum_{i=1}^{T} \frac{N_2(i)}{N_1(i)+N_2(i)}}{T}\end{cases}$$

$N_1(i)$是查询第i个隐私信息的页面数量,$N_2(i)$是查询第i个隐私信息泄漏的页面数量,$N_1(x)$是当前隐私信息的页面数量,$N_2(x)$是当前隐私信息泄漏的页面数量,T是查询隐私信息的总数;

第三步:根据U×S计算结果确定隐私信息的安全等级,如果U×S=0.25,则等级为1级;如果U×S=0.5,则等级为2级;如果U×S=1,则等级为3级。

基于采集搜索引擎数据的隐私信息评级方法

技术领域

[0001] 本发明涉及一种信息评级方法，更特别地说，是通过数据统计的方式确定隐私信息安全等级的方法。

背景技术

[0002] 随着网络技术、物联网技术的发展，个人隐私信息安全受到越来越严重的威胁。

[0003] 随着网络技术和互联网的发展，人们的工作和日常生活越来越离不开互联网，个人的资料和隐私信息都直接或间接地存储在网络上。随着隐私在个人电脑和互联网的网站内长时间大量地积累，隐私的经济效益越来越大，互联网上大量的病毒、木马和黑客对个人隐私信息的窃取和破坏行为越来越多。一旦个人隐私被窃取或泄露，其传播范围和造成的影响是无法预计的，对个人及社会造成的损失是巨大的。

[0004] 物联网应用在使用过程中需要收集各种信息，物联网可以通过摄像头、GPS、各类传感器(声音、光、温度、烟等)、RFID等设备全面感知环境信息和个人的行为状态，而这些信息中包括个人的隐私信息，因此物联网对个人隐私具有主动获取的能力。物联网往往更深入地应用于各个行业，以及国防、军事等国家核心利益领域，具有巨大的商业和政治利益。在商业和政治利益的驱动下，物联网病毒制造、黑客入侵等行为更具有目标性和危害性，因此物联网中隐私信息安全也面临巨大威胁。

[0005] 目前很多技术文献提出了各种隐私保护技术，例如基于混淆原理的对位置隐私的保护，基于编码的对身份隐私的保护技术等，但是目前各种隐私保护技术存在的问题是对各种隐私保护技术的研究是相互独立的，都是针对单独某个隐私内容的保护，在提出保护该隐私的技术时，不考虑其他的隐私如何保护。一些技术人员提出了隐私信息分类方法，如根据隐私信息产生的不同方式而分成静态、动态和派生三类，目前关于隐私信息分类方法的缺陷是仅提出了分类方法，却不管不同隐私之间在敏感性和安全性方面的区别。

[0006] 实际上，一个IT应用中往往涉及大量的不同的隐私，但是针对不同隐私的保护技术实现的难易程度、成本和占用系统资源是不同的，很多时候由于建设系统成本有限和运行效率的考虑，我们不能保护全部的隐私，此时就需要将隐私区分为不同的安全等级，这样可以在成本和技术受限条件下优先保护重要的隐私。

发明内容

[0007] 本发明目的在于至少解决现有隐私保护的不足之一，提供一种进行系统性地隐私信息间比较的方法。实际上，一个IT应用中往往涉及大量的不同的隐私信息，但是针对不同隐私的保护技术实现的难易程度、经济成本和占用系统资源多少是不同的，很多时候由于建设系统成本有限和运行时效率的考虑，我们不能保护全部的隐私内容。此时就需要将隐私区分为不同的安全等级，这样可以在成本和技术受限条件下首先确保重要隐私的安全。

[0008] 为实现上述目的，本发明采取如下技术方案：

[0009] 第一步：确定每个隐私信息的普遍性分值。隐私信息的普遍性反映了全部人中多

少人认为某信息是隐私信息。在搜索引擎搜索“隐私项隐私”，可以得到用户搜索返回的总页面数N_1。取全体隐私信息项的N_1的平均值为阈值μ_1，如果某隐私的N_1大于等于阈值μ_1，则该隐私的普遍性较高，分值为1，否则说明该隐私的普遍性一般，分值为0.5。

[0010] 第二步：确定每个隐私信息的敏感性分值。隐私信息的敏感性是指隐私信息对个人而言的私密程度，隐私信息的敏感性反映该隐私信息泄露后对个人及社会造成后果的严重程度。在搜索引擎搜索“隐私项隐私泄露”的内容，可以得到用户搜索返回的总页面数N_2，然后计算该隐私相关的所有查询结果中N_2所占的比例β。取全体隐私信息项的β值的平均值为隐私信息的敏感性的阈值μ_2，如果某隐私的β值大于等于该阈值μ_2，则该隐私的敏感性较高，分值为1，否则说明该隐私的敏感性一般，分值为0.5。

[0011] 第三步：确定隐私信息的安全等级。

[0012] 隐私信息的安全等级依据威胁对隐私信息侵害后所造成的后果决定，后果则由隐私信息的普遍性和隐私信息敏感性共同作用。

[0013] 隐私信息的安全等级分值＝隐私信息的普遍性分值×隐私信息的普遍性分值

[0014] 将隐私普遍性是一般但是敏感性是较高的隐私和普遍性是较高但是敏感性是一般的隐私视为同一安全等级，则隐私信息的安全等级分值分别是：1，0.5，0.25。将安全等级分值分0.25的隐私评定为1级隐私，将安全等级分值分0.5的隐私评定为2级隐私，将安全等级分值分1的隐私评定为3级隐私。隐私信息的安全等级越高，则该隐私越重要，需要首先被保护。

[0015] 与现有技术相比，本发明具有以下优势：

[0016] 1、本发明提供一种全面地进行隐私信息评级方法，该方法不针对特定的隐私内容，适用于全部隐私信息；

[0017] 2、本发明提供的评级方法使用的数据是从搜索引擎或搜索网站收集，由于使用搜索引擎的用户数量巨大，因此评级不依赖于少数人的意见或经验，评定结果具有真正的公正性；

[0018] 3、本发明提供的是一种评级方法，而不是评级结果，因此在实际应用中，即使某应用只涉及用户的部分隐私信息，即隐私信息的数量有限，仍然可以在有限数量的隐私信息中确定出1、2、3级隐私，因此本发明具有灵活而广阔的应用领域。

附图说明

[0019] 图1是方法流程图。

具体实施方式

[0020] 下面将结合附图对本发明做进一步的详细说明。

[0021] 第一步：确定每个隐私信息的普遍性分值U，具体处理为：

[0022] 步骤101：在搜索引擎输入：隐私项P隐私；

[0023] 步骤102：统计返回该查询的查询结果数量，记作N_1；

[0024] 步骤103：重复步骤101、102直到统计了全部的隐私信息；

[0025] 步骤104：计算全体隐私信息项的N_1的平均值，该平均值就是阈值μ_1；

[0026] 步骤105：如果隐私P的$N_1 \geq \mu_1$，则该隐私的普遍性较高，隐私普遍性分值U＝1；否则

说明该隐私的普遍性一般，U＝0.5；

[0027] 步骤106：重复步骤105，直到计算了全部隐私的隐私普遍性分值U。

[0028] 第二步：确定每个隐私信息的敏感性分值S，具体处理为：

[0029] 步骤201：在搜索引擎输入：隐私项P隐私泄露；

[0030] 步骤202：统计返回该查询的查询结果数量，记作N_2；

[0031] 步骤203：计算N_2占(N_1+N_2)的比例β，$\beta = \frac{N_2}{N_1+N_2}$；

[0032] 步骤204：计算全体隐私信息的β值的平均值，该平均值就是隐私信息敏感性阈值μ_2；

[0033] 步骤205：如果隐私P的$\beta \geq \mu_2$，则该隐私的敏感性较高，隐私的敏感性分值S＝1，否则说明该隐私的敏感性一般，隐私的敏感性分值S＝0.5；

[0034] 步骤206：重复步骤205，直到计算了全部隐私的隐私敏感性分值S。

[0035] 第三步：确定隐私信息的安全等级，具体处理为：

[0036] 步骤301：计算隐私信息的U×S；

[0037] 步骤302：确定隐私信息的安全等级，如果隐私的U×S＝0.25，该隐私评定为1级隐私，如果隐私的U×S＝0.5，该隐私评定为2级隐私，如果隐私的U×S＝1，该隐私评定为3级隐私；

[0038] 步骤303：重复步骤301、302，直到全部隐私评级完成。

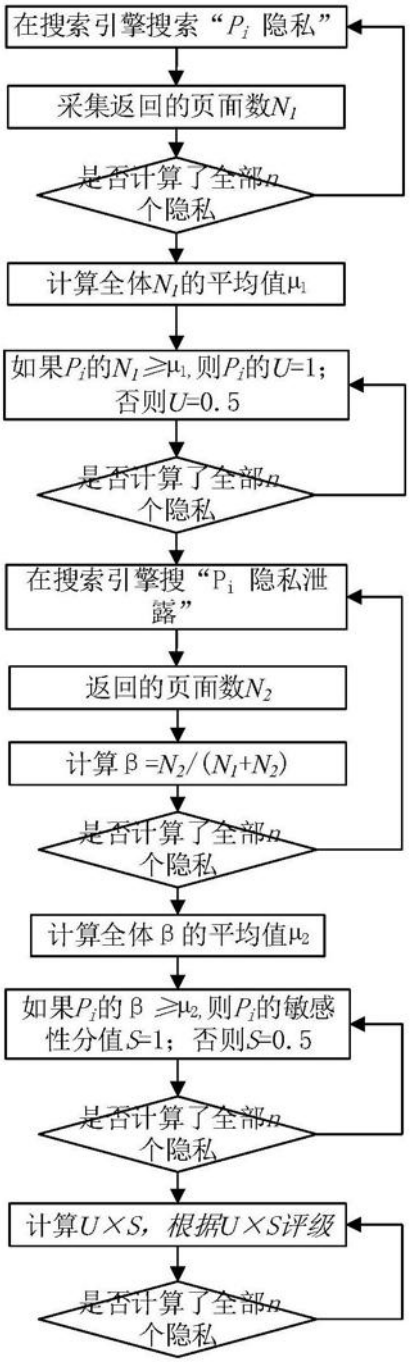

在搜索引擎搜索“P_i 隐私”
采集返回的页面数N_1
是否计算了全部n个隐私
计算全体N_1的平均值μ_1
如果P_i的$N_1 \geq \mu_1$，则P_i的U=1；否则U=0.5
是否计算了全部n个隐私
在搜索引擎搜“P_i 隐私泄露”
返回的页面数N_2
计算$\beta = N_2/(N_1+N_2)$
是否计算了全部n个隐私
计算全体β的平均值μ_2
如果P_i的$\beta \geq \mu_2$，则P_i的敏感性分值S=1；否则S=0.5
是否计算了全部n个隐私
计算$U \times S$，根据$U \times S$评级
是否计算了全部n个隐私

图1