1. 摘要 Abstract
2. 介绍 Introduction

新型冠状病毒COVID-19引起的肺炎2020年在全世界流行，造成了数以十万计的无辜生命死亡，致至少数千万人感染。世界卫生组织和各国政府采取了各种措施控制疫情的蔓延，但半年时间过去，仅有少数国家的疫情真正得到了控制，每天不断攀升的确诊和死亡人数触目惊心。人类在面临疫情时，显得格外渺小和束手无策，由于COVID-19远高于SARS和MERS的传染能力和高达14天的潜伏期，病毒在缺少防护的人群之间肆意传播，例如一个确诊患者在几小时的飞行途中就可以感染整个机舱的乘客，而这些被称为密切接触者的乘客又将病毒带下飞机，在工作单位和家庭中继续传播。更让人担忧的是无症状感染者的存在，这类人群感染了病毒，但却不表现丝毫的症状，仍以健康人的身份活动，直到一些偶然的机会才被发现，而在此期间已经接触了大量的人群。

与疫情初期各国防疫和医疗部门疲于救治重症确诊患者，而忽略了轻症患者以及大量的密切、间接接触者不同，随着医疗资源逐渐恢复正常，防疫经验逐步累积。目前世界上多国政府已经有能力在做好救治的同时，一旦发生新的局部疫情，就立即采取行动大规模对可能的接触者进行核酸检测，甚至对整个地区的人做检测。

传染病防治的一项重要任务就是流行病学调查，即追溯确诊患者在过去一段时间里（对于COVID-19一般是14天，一些国家甚至更久）的行踪，从而排查出他的多级密切接触者，是确定核酸检测目标对象的重要手段。流行病学调查主要依靠患者的自述，虽染目前很多地方也借助智能手机的传感器以及APP里诸如交易记录之类的信息进行调查，但是效果仍不理想，难免有漏网之鱼。6月，中国北京以农贸市场为中心的疫情之所以能被很快扑灭，得益于首位确诊患者的优秀的记忆力，他准确回想了自己过去14天的行踪，并提供了农贸市场这一关键信息，但专家似乎也很难考证患者的叙述是否完整和真实。另外，由于此次的疫情爆发出现在北京一个非常重要的市场，指向型极强，给后续排查接触者带来了一些便利。但事实上，这样的幸运事件并不总是发生，任何政府都不允许防疫工作建立在个人的运气之上，假设此次患者不能配合完成优秀的流行病学调查，或者患者停留的地方小而杂，恐怕就会给防疫工作带来很多不确定的挑战。因此，我们必须要拥有一套更好的追踪方法，不依靠个人的能力，与地点无关，准确安全，适合各类群体使用。

目前，各国政府和科技公司都在尝试提供科学的流行病学调查辅助方法。比较著名的是中国的健康码，各个省份甚至各个城市都推出了自己的健康码，这在中国被证明是一种有效的方式，因为在中国几乎所有的公共场所、小区都需要凭借健康码出入，因此一个参与社会生活的人必须申请，并自动纳入“联防联控”体系中。健康码大致分为三个登记，红色表示确诊患者或确诊患者的密切接触者和来自疫情高风险地区的人群，这类人在中国一般都住院治疗和由政府统一组织隔离，在社会生活中受到严格监管，不能出入任何公共场所，需要完成至少14天的隔离或治疗。黄色表示从疾病中康复的人员、确诊的病患的一般接触者或来自中风险地区的人员，这类人群根据属地的政策一般需进行7天的隔离观察，在社会生活中也受到比较严格的限制。绿色表示完全健康的人群，也是绝大多数人群的健康码颜色，指的是完全没有疫情接触史的人群或者得到了医疗机构认可从红黄码中隔离结束的人群，绿码持有者可以无障碍的在属地活动，不应受到歧视或阻拦。健康码的使用流程，以作者所在的辽宁省为例，一般是，首先申请者要拥有一台智能手机，下载所在地政府所规定的APP或者小程序，按照里面的要求填写个人信息以及疫情有关的信息，包括是否去过疫情高风险区，是否接触过确诊患者，以及自身是否有疫情相关症状等，之后系统根据这些信息生成对应颜色的码。使用的办法一般包括两种，一是向工作人员出示健康码以进入一些公共场所，更为有效的办法是用手机扫描贴在场所处的二维码，从而不仅可以验证现有健康码的颜色也可以将行踪进行记录。这种方法在中国大面积推广且取得了一定的效果，在防疫过程中起到了重要作用。然而，它存在几个非常重要的缺陷。第一、许多老年人、儿童没有智能手机或者即使有也在操作上有很大的障碍，据报道，截至2020年3月，中国网民规模为9.04亿，互联网普及率达64.5%。也就是说，中国大约有5亿人是不上网的，除此之外，在2019年中国60周岁及以上人口有25388万人，这其中，接触过网络的只有6056万人，而能熟练使用智能手机的占比更是少之又少，这还没有考虑到尚不能独立使用电子设备的儿童，即使填报流程可以由他人代劳，但是在使用时动作缓慢，堵塞在公共场合门口，极大影响了后面的人流，这时往往会采用手动登记或者录音的办法，从而增大了社交风险，为病毒传播提供了额外的机会。第二、由于工作人员不足，或是忙于测温工作，导致对健康码的检查存在漏洞，主要包括，（1）是否有码：部分人没有健康码却跟随人流混入场所，也存在因为手机没有电量导致确实不能使用的情况（2）码的时效性：一些人为了图方便出示截图（3）出示码而不扫码：这是在实际操作中最常出现的漏洞，这只能证实当前的健康状态，而无法在疫情发生后进行追溯。第三、由于网络环境导致APP无法调用，从而滞留大量人群引发聚集风险。第四、由于各地政府的政策不一致，导致每到一个地方就要申请新的健康码，甚至同一人到不同的省市申请到的健康码结果不同的情况，而在健康码互认的问题上更是矛盾多多，据报道多次出现了武汉市民持有绿码却在外地寸步难行的尴尬情况。另外，在用户使用的友好程度上，现行的健康码也并不令人满意，首先就是一次出行需要反复的打开手机，而由于佩戴口罩又不能快速解锁手机，再调用软件扫码和展示码，这一过程每天可能要占据10余分钟的时间。其次，APP的安全性令人担忧，虽是由政府掌控个人关键信息，但是难以保证其他恶意软件通过手机上的漏洞窃取到健康码APP的内容。如果说以上提到的还都是一些小的缺点和不足，那么由于智能手机性能、APP权限、网络故障和人为因素导致的关键信息遗漏问题就可能大大影响流行病学调查的准确性和完整程度，进而导致十分依赖于此的防疫工作功败垂成。

而从三月起就成为疫情新的震中的美国，截至7月中旬确诊病例已经接近400万，并且丝毫没有放缓的趋势，COVID-19已经在美国肆虐开来，而由于文化习惯，美国民众似乎并不喜欢佩戴口罩，这给疫情传播提供了极佳的条件，也为确诊患者的密切基础者追踪带来了极大的困难。因此，五月份两大科技巨头苹果和谷歌联合退出了一种基于蓝牙的密切接触者追踪系统，工作原理仍基于智能手机和APP，且要求设备的蓝牙保持开启，系统运行时，手机就会周期性地弹出一段小的、唯一的、匿名的代码，这段代码是从手机的唯一ID中提取出来的。其他手机会接收并记住这段代码，建立一个记录他们接收到的代码和接收到它们的时间的日志。相反，设备也会接收其他设备发来的代码。为了保护隐私，这段代码还会每隔一段时间更新一次。当使用该系统的人确诊患病后，他可以选择将自己的确诊信息上传到服务器上，其他用户定期从服务器上下载确诊信息，并与本地所存储的他人代码比对，如果有匹配项，则意味着自己与确诊患者曾密切接触过，从而决定自己的下一步医治计划。此类系统在隐私保护的原理上似乎没有漏洞，可能引起的隐私泄露问题尚在讨论，但单就其防疫效果来讲是很差的。首先，要求这些甚至反对口罩的美国民众下载此APP并需要打开蓝牙功能就是很困难的，出于文化和历史的原因，愿意花时间和精力去做防疫工作的民众并不占绝对多数，如果不是参与社交的每一个人都开启了这项服务，那整个过程的效果就大打折扣，极端来讲，在一个社交场合恰好有一个人没有开启服务，而恰恰这个人是确诊的患者，则其他所有人的参与都是徒劳。第二、依靠患者自主上传的方式并不可靠，确诊者极有可能因为主观或是客观原因没有上传确诊信息，那整个环节就失效了。第三、即使前序流程顺利，当一名用户获知自己成为了密切接触者时，他极有可能因为是一名无症状感染者或仍存在潜伏期没有发病而拒绝前往医疗点检测和治疗，并继续参与日常社交活动传播病毒。可怕的是，由于此系统的保密机制，只要用户不前往医院，外人甚至是政府医疗机构无法获知用户的身份和接触情况，也不能及时为其提供最佳救治方案。如果面对的是比COVID-19更致命的病毒，其中一环的缺失就将引起巨大的后果。如果再综合考虑其本身的工作机制和无线通信技术，该方法更是存在耗电较大、不同机型间的融合以及蓝牙传输距离限制导致无法获取稍远一点的接触者信息的问题，据报道，病毒不仅可以通过已知的飞沫方式传播，也可能通过气溶胶传播到稍远一些的地方，而这个距离很有可能已经超过了蓝牙的传输范围。此外，该方法仅限第一级密切接触者的追踪，而无法掌握次一级密切基础者的密切接触者，需知在各国报道的病例中，存在不少跨层的传染途径。

其他的方法还包括借助运营商网络的，比如中国国务院、信通院与三大运营商合作的防疫大数据行程码，可以根据手机使用连接基站的情况判断持有者到访过哪座城市，甚至哪个区。此码的作用更多体现在远程出行中，并不针对特定的确诊患者的密切接触，而是从地区的风险评级上判断一个人可能的健康程度。与健康码相比，此系统似乎不需要用户的主动参与，但是在实际使用中其灵敏度不高，例如短暂停留一个城市时，并没有该城市的记录，而需要至少拨通一次电话才可以。相反，也出现了仅仅在火车上路过却被记录在案的情况。另外一个风险是，一些投机者可以利用双卡的优势，完美的避开轨迹被记录，从而给追溯带来困难。还有一些地方的政府设计的被动的基于位置的追踪系统，这种方法也作为一部分包含在前文所述的中国健康码中，APP利用手机的GPS、Wi-Fi等功能，对手机进行定位，从而获取持有者的轨迹，但是无线定位技术本身存在的误差和场景局限性是制约此方法的障碍，而且只要其依靠智能手机工作，就仍然逃不掉被其他恶意软件窃取信息的威胁。一旦手机电量耗尽、用户私自关闭了某些功能，都将导致数据收集和疫情追踪的失败。

我们无法仅以此判断哪种追踪措施更好，因为在不同文化背景下生活的人群会对同样的事情有截然不同的看法，因此也会产生差异化的应对方式，就像对待是否应该佩戴口罩这件事，似乎在一些国家已经争论了半年多的时间仍然没有结论。

* 1. 密切接触者追踪 (Contact tracing)

因为与病毒感染者密切接触的人自身感染和进一步感染他人的风险更高，所以在接触感染者后密切观察这些接触者将有助于接触者获得护理和治疗，并将防止病毒的进一步传播。这种监控过程称为接触者追踪。**[https://www.who.int/news-room/q-a-detail/contact-tracing]**。或者解释为，接触者追踪是指确定可能与感染者有过接触的人，并随后收集有关这些接触者的进一步信息的过程**[Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing]** 。

目前广泛使用的contact tracing方法需要确诊患者清晰地、准确地回忆起确诊之前一段时间内所去过的场所以及所接触的人。但是这种情况常常是可遇不可求的，因为对于普通人来说准确地回忆起自己许多天内的活动轨迹以及时间是很困难的事情。而且在许多情况下回忆起所接触过的人是不可能完成的事情，比如在超市排队结账、搭乘地铁等，在这种情况下，周围的人几乎都是陌生人。所以，靠传统的contact tracing是很难做到高效率地排查密切接触者。

2020年以来，世界各国、各个公司相继推出了各自的密切接触者追踪方法和应用程序contact tracing apps，总体上分为三类：中心化的、去中心化的、混合式的 **[A Survey of COVID-19 Contact Tracing Apps]**。在中心化架构中，中央服务器在执行核心功能方面起到关键作用，如存储加密的PII信息、生成匿名TempIDs、风险分析和密切接触者通知等。同时又是服务器的过高权限会导致隐私问题。中心化架构的密切接触者追踪app里比较知名的有新加坡的TraceTogether、澳大利亚的CovidSafe、中国的各种健康码app等。与集中式架构相反，在去中心化的架构中，跟踪过程是由应用程序用户在他们的设备上本地执行的。这样可以缓解一些隐私问题，但是却给智能手机的配置提出了较高的要求，用户的手机需要完成较多的计算。去中心化架构的密切接触者追踪app的例子有Apple/Google Exposure Notification APIs、PACT等。混合式架构则是寻求在中心化架构和去中心化架构中取得平衡。但是，目前仅有少量的混合式架构app被提出，比如DESIRE和EpiOne，而且，已有的混合式架构的app需要手机和服务器进行较多的数据通信，也会导致手机耗电量增大，并且需要消耗较多的运营商流量。

而且，目前已有的三种架构的contact tracing app基本都是安装在智能手机上、基于蓝牙接口进行设备间感知和通信的。这些app之间使用蓝牙信号进行信息交互，比如广播和接收蓝牙信息，将信息保存在手机本地，并通过运行商网络将部分数据上传到服务器。使用蓝牙就不可避免的需要面对蓝牙的缺点，蓝牙有传输距离有限、不同设备间协议不兼容的缺点。而且需要使用者始终保持智能手机的蓝牙功能处于活跃状态，这样会很大程度地增大手机的耗电率、减少手机的使用寿命。并且，如之前所述，有很大一部分人不具有智能手机、不会操作智能手机，或是他们的智能手机的性能不足以支持长时间的蓝牙活动。

新加坡政府推出的TraceTogether Token物理设备将密切接触者追踪的功能从智能手机分离出来，作为一个独立的物理设备使用，作为手机app的替代品、减少了手机的负担。但是，该设备仍然使用蓝牙信号进行通信，仍然摆脱不了蓝牙信号所带来的局限性。

* 1. LoRa:

物联网通信技术繁多，从传输距离上可划分成两类：第一类是短距离通信技术，例如ZigBee、Wi-Fi、Bluetooth等，典型的应用场合如智能家居；第二类是低功耗广域网(Low Power Wide Area Network, LPWAN)，典型的应用为智能抄表系统。LPWAN技术又可根据工作频段分为两类：一类工作在非授权频段，如LoRa、SigFox等，此类技术无统一标准，也没有大型运营商提供支持，功能、规模完全自定义实现；第二类工作于授权频段下，3GPP支持的2/3/4G蜂窝通信技术，如全球移动通信系统(Global System for Mobile Communication , GSM)、长期演进(Long Term Evolution, LTE)和基于蜂窝的窄带物联网(Narrow Band Internet of Things, NB-IoT)等。

其中的LoRa 是一种低功耗广域网技术。它由法国格勒诺布尔的 Cycleo 公司开发，并被 LoRa 联盟的创始成员 Semtech 公司收购，并且在构建可扩展物联网的过程中使用了该 技术。LoRa 技术是基于扩频调制技术派生的啁啾扩频技术。

近年来，低功率广域网 (Low Power Wide Area, Networks LPWAN) 逐渐引起科学界 越来越多的关注，有望成为未来可扩展物联网(IoT)中重要的一部分。其新的发展旨在解决主要构成物联网设备的电池寿命、共存和通信范围的限制。

LoRa 因其开放的规范和网关基础设施而备受关注，这与其他 LPWAN 技术的封闭 设计和管理网关基础设施完全不同，所以发布后不久就吸引了科学界和工业界的大量关注，并汇集了具有强大潜力的许多研究人员的多边努力，他们将共同开发物联网设备的 全球覆盖。

LoRa 的流行源于一系列的功能和性能承诺。首先，LoRa 使用了未经授权的工业、 科学和医疗无线电(ISM)波段，保证了数公里的通信距离和长达数年的电池寿命。另外， LoRa 添加了一种可以变化的啁啾扩频(CSS)技术。Semtech 声称，除了多路径和多普勒 效应以外，该技术还具有很强的抗干扰能力。不仅如此，LoRa 还声称支持高信道容量， 这意味着单个信道支持来自多个设备的同步传输，这与传统的 LPWAN 技术形成了鲜明对比。尽管 LoRa 的物理层是 Semtech 收购的一项专有技术，但其中只有部分内容对公众开放。

LoRa 的优势在于改变了传输功耗和传输距离之间的平衡，彻底改变了嵌入式无线 通信领域的局面。它给人们呈现了一个能实现远距离、长电池寿命、大系统容量、低硬 件成本的全新通信技术，而这正是物联网所需要的。

* 1. 区块链:

区块链技术起源于2008年中本聪的《比特币：一种点对点电子现金系统》**[Bitcoin: A Peer-to-Peer Electronic Cash System]**，区块链诞生自中本聪的比特币。区块链可以在没有中央权威的情况下，对彼此的协作创造信任。区块链技术适用于一切缺乏信任的领域，也许在未来会成为全球人类文明信任的基石。目前，随着区块链技术的成熟和演进，区块链的应用场景不再局限于比特币，比如，以“以太坊”为代表的新一代区块链技术正在开始构建一个全新的去中心化互联网架构。

2020年以来，有许多学者提出将区块链技术应用在密切接触者追踪系统中。将区块链应用到密切接触者追踪系统中可以解决现有的许多隐私问题和安全问题，比如：区块链可以从技术设计上提供一种保护隐私的解决方案，而不是依赖于在一个集中的系统中遵守法规或法律；区块链技术结合加密和匿名技术的使用，可以进一步保护用户的身份；区块链在本质上是非区域性的，因此为COVID- 19大流行跟踪和控制提供了一个合适的全球获取平台；区块链透明的特点可以防止公众被当局或其他第三方故意误导**[BeepTrace: Blockchain-enabled Privacy-preserving Contact Tracing for COVID-19 Pandemic and Beyond]**。

但是，将区块链大规模应用到边缘计算上仍然面临着许多问题，比如性能受限、同步问题、高并发问题等。为了保证区块链的安全性和可靠性，区块链不可避免的需要应用PoW，这样会大幅提高移动终端设备的消耗、降低移动设备的工作效率。因此，PoW的计算不宜直接让移动终端设备完成，应该由专门的边缘计算服务器代替完成**[When Mobile Blockchain Meets Edge Computing]**。这样一来，我们便可以把消耗比较大的PoW的计算过程从性能较差的移动终端设备分离出来，同时在一定程度上保证了区块链的安全性和可靠性。

* 1. Contribution

本文抛开文化观念问题，完全从应对COVID-19甚至将来有可能发生的更为严重的传染病疫情角度出发，试图提出一种新的密切接触者追踪方法，我们认为一个好的解决方案应该包含以下特点：

1. 简单易用：我们希望新的办法可以降低对使用者的能力要求，无论是老年人还是儿童，要求尽可能少的操作；
2. 低耗耐用：如果必须依靠智能手机，应当尽可能降低因电量耗尽带来的尴尬，或许应该有更高效且低成本的替代品；
3. 不得不用：新的方法要所有人的参与，无论是出于对自身的防护还是对家庭社会的责任，你应该像佩戴口罩一样使用它，政府和医疗机构也需要始终监督，从而第一时间对有风险得人群进行检测和隔离；
4. 可靠好用：确诊患者接触信息是流行病学调查的重要依据，必须要保证数据的可靠性，任何人不能借助漏洞规避追踪，不能擅自脱离系统或是对已生成数据进行篡改；
5. 通用：一套方法适用整个地区，不需要换一个地方换一个方案。

因此，本文提出了一种LoRa-based Blockchain-Enabled privacy-preserving contact Tracing (LoBeTrace)，可以满足以上的所有要求。**首先**，仅需要在领取时进行一次登记即可，将个人的ID、联系方式信息录入相关的政府机构，今后的使用中再也不需要任何其他操作，避免了PII信息的泄露。**其次**，对于终端用户，由于LPWAN的超低功耗，一个设备可以工作数年以上，不仅可以供应对COVID-19使用，即使不幸遇到新的疫情，也可以拿出来继续使用。而且单个设备的成本非常低廉，体积足够小且可以根据需求制作成钥匙、卡甚至挂饰，并不给用户带来忘记携带难以携带的困扰，方便大规模推广。我们吸纳中国健康码的设计，为设备设置了同样三种颜色的指示灯，同时省去了进行拿出手机-解锁-扫码-展示的复杂环节，只需要拿出设备或者将设备挂在衣服上，工作人员就可以看见你的颜色并采取相应措施。也不需要担心老人或者儿童不会使用，并确保了每个人的参与。我们的设计是密闭的，仅记录了必要的联络信息，且由于他的工作不需要依靠任何其他智能设备，避免了被窃取的可能性。**另外**，由于没有任何可修改操作的可能性，只要携带设备参与了室内环境的社交活动，你的室内访问记录就会被保存。我们还采用了区块链技术防止数据被篡改，向任何想徇私枉法影响防疫工作的人说不。而且，用户的信息以匿名形式呈现，属于混合式的方法，可以在很大程度上保证用户的信息(PII)不被泄露。尊重了用户的隐私权。**最后**，这套方案有着很好的通用性，不需要申请多个健康码，用户就可以畅通的出行到部署此系统的地区。

1. 基于区块链的踪迹记录存储结构

区块链技术涉及数学、密码学、互联网和计算机编程等很多科学技术问题。从应用视角来看，简单来说，区块链是一个分布式的共享账本和数据库，具有去中心化、不可篡改、全程留痕、可以追溯、集体维护、公开透明等特点。这些特点保证了区块链的诚实与透明，为区块链创造信任奠定了基础。而区块链丰富的应用场景，基本上都基于区块链能够解决信息不对称问题，实现多个主体之间的协作信任与一致行动。区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。随着物联网技术的发展，已经有许多区块链与物联网结合的应用实例出现，但是这仍然面临着许多的、涉及各方面的问题**[On blockchain and its integration with IoT. Challenges and opportunities]**。

区块链与密切接触者追踪结合在一起，可以提供很高的安全性、可靠性、数据透明度，并加强隐私保护，这些是目前现有的许多方案无法做到的。区块链的提高安全性与可靠性的方法有分布式账本(**Distributed ledger**)、非对称加密(**asymmetric encryption**)、共识机制(**Consensus mechanism (CM)**)。另外，智能合约区块链(**smart contract blockchain**)还有智能合约(**smart contract**)技术来增强区块链的可靠性。其中共识机制是通过节点的投票，在很短的时间内完成对交易的验证和确认。对一笔交易，如果利益不相干的若干个节点能够达成共识，我们就可以认为全网对此也能够达成共识。区块链作为一种按时间顺序存储数据的数据结构，可支持不同的共识机制。共识机制是区块链技术的重要组件，它的目标是使所有的诚实节点保存一致的区块链。现有的区块链共识机制包括：工作量证明机制(**Proof of Work (PoW)**)、权益证明机制(**Proof of Stake (PoS)**)、股份授权证明机制(**Delegated Proof Of Stake(DPoS)**)和Pool验证池(**Verify the Pooling**)、基于有向无环图的工作量证明机制(**Direct Acyclic Graph (DAG) based CM (DAG-based CM)**)。区块链上采用不同的共识机制，在满足一致性和有效性的同时会对系统整体性能产生不同影响，一般从安全性、可扩展性、性能效率、资源消耗四个维度评价各共识机制的技术水平。

工作量证明(**Proof-of-Work (PoW)**)是一种对应服务与资源滥用、或是阻断服务攻击的对策。一般是要求用户进行一些耗时适当的复杂运算，并且答案能被服务方快速验算，以此耗用的时间、设备与能源做为担保成本，以确保服务与资源是被真正的需求所使用。与**PoW**不同，权益证明机制(**PoS**)要求证明人提供一定数量加密货币的所有权即可。权益证明机制的运作方式是，当创造一个新区块时，矿工需要创建一个“币权”交易，交易会按照预先设定的比例把一些币发送给矿工本身。权益证明机制根据每个节点拥有代币的比例和时间，依据算法等比例地降低节点的挖矿难度，从而加快了寻找随机数的速度。股份授权证明机制(**DPoS**)与董事会投票类似，该机制拥有一个内置的实时股权人投票系统，就像系统随时都在召开一个永不散场的股东大会，所有股东都在这里投票决定公司决策。基于DPoS机制建立的区块链的去中心化依赖于一定数量的代表，而非全体用户。在这样的区块链中，全体节点投票选举出一定数量的节点代表，由他们来代理全体节点确认区块、维持系统有序运行。Pool验证池(**verify the pooling**)不需要依赖代币就可以工作，在成熟的分布式一致性算法(Pasox、Raft)基础之上，可以实现秒级共识验证，更适合有多方参与的多中心商业模式。

共识机制的使用会消耗很多资源，这对于密切接触者追踪这种资源受限的系统而言代价很高。而且，PoW和PoS的使用会使得延迟边长，严重影响区块链的并发性，不利于系统的大规模应用。例如，比特币的吞吐量通常限制为每秒7笔交易(TPS)，以太坊的吞吐量通常为每秒20到30笔交易**[The bitcoin mining network-trends, marginal creation cost, electricity consumption & sources]**。比特币的确认延迟一般为60分钟，以太坊的确认延迟一般为3分钟，这对于LoBeTrace系统来说过于长。况且，LoBeTrace大量使用了物联网技术，会产生内部的传输延迟、处理延迟等其他延迟。因此，我们需要降低PoW或PoS的苛刻的计算复杂度。同时，为了保证区块链的安全性和可靠性，又因为“只有特定的网关才能参与维护区块链”这一特点，我们在降低LoBeTrace的工作量证明的计算复杂度的同时，人为限制网关的计算能力，使得它不可能具有篡改区块链的超过50%的算力。这样，我们既可以保证低延迟的要求，又可以避免区块链的安全性受到威胁。

2015年9月，Sergio Demian Lerner等人首次提出了DAG-Chain的概念**[DagCoin: a cryptocurrency without blocks]** ，之后，DAG彻底改变了区块链技术。由于其优化的验证机制、高可扩展性、对物联网和多方参与的支持，DAG正迅速overshadowing传统区块链架构。当前应用了DAG的代表项目，最知名的无疑是DagCoin、ByteBall和Nano。传统的区块链架构对于物联网场景来说非常昂贵，因为它们的工作量证明算法非常昂贵。A DAG (Directed Acyclic Graph) based blockchain architecture having properties such as tieing knots of provenance validation on various progressive levels of the trees can be address this problem **[A comparative analysis of DAG-based blockchain architectures]**。因此，DAG based-CM 将会非常适用于LoBeTrace。

1. LoRa-based blockchain-enabled privacy-preserving contact tracing

在本部分中，我们将详细描述工作流程并解释关键概念。在接下来的部分中，我们将首先介绍系统中涉及的实体、它们的角色以及它们如何相互配合工作。接下来我们会描述contact tracing framework的工作流程，**……**

* 1. Entities, functions, and interfaces

下面，我们定义了LoBeTrace的各个组成部分，并逐一解释它们的角色和接口(roles and interfaces):

* 踪迹区块链(**trace blockchain**)，是由许多网关设备共同维护的、用于保存用户在各个公共场所的活动轨迹信息的区块链。它接受新的网关设备的注册，便于LoRaTrace扩大使用范围。该区块链是公开的区块链，用户可以通过专门的网站下载该区块链来解决一些可能发生的纠纷问题。踪迹区块链是用于记录用户踪迹数据(**trace data**)的区块链，只保存近21天(504小时)内产生的区块，一条踪迹记录(**trace record**)存储在一个区块中，包括：产生踪迹的时间(**trace generation time**)、用户匿名(**user pseudonym**)、产生踪迹的地理位置(**trace generation location**)、上传该踪迹的LoRa网关的签名(**LoRa gateway signature**)、上一个块的哈希值(**previous hash**)、本块的哈希值(**hash**)。产生踪迹的时间是以时间戳形式记录的，精确到分钟； 用户匿名信息是在字符串“ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789”内随机挑选16个字符组成的新的16位字符串，字符可重复、有顺序差异，因此匿名信息重复的概率极低，可以忽略不计；产生踪迹的地理位置信息是字符串类型，属于官方指定的识别代码；签名信息以字符串形式保存，可用于验证该记录的合法性；
* 风险名单服务器 (**risky names server**)，是用于维护风险名单(**risky name list**)的服务器。该服务器提供一个网站，可供用户和网关查询和读取。风险名单是用于记录和通知用户的名单，其中记录了具有与确诊患者接触风险的用户所使用过的匿名。该名单以集合类型保存，可以保证每个匿名信息不重复，减少了数据的冗余，有利于在计算能力较弱的移动设备和固定设备使用。
* 固定设备(**fixed device**)，是由政府分配给各个室内公共场所（一般为建筑）的设备，每个建筑物在入口处的楼层的几何中心位置安装一个固定设备，这样可以大幅减少用户进入过建筑物但是没有被检测到的概率。固定设备由电源供电，可以收发SF=7和SF=9的LoRa信号。固定设备会周期性地通过SF=7的LoRa信号广播唤醒信息(**wakeup message**)，并接收各个移动设备发来的匿名信息(**pseudonym message**)，并加上时间戳、地理位置信息作为一条踪迹记录(**trace record**)存在内存中。之后，固定设备会将固定一段时间内收到的所有匿名信息(16字符字符串)处理后通过SF=9的LoRa信号发送给LoRa网关。固定设备同时也将从网关下载的风险名单伴随着唤醒信息发送给各个用户。由于LoRa包的大小(size)限制，固定设备每次发送一条踪迹信息，并等待来自LoRa网关的ACK信息以及伴随ACK信息发来的风险名单。每次发送一条踪迹信息后该设备会等待ACK信息5秒钟时间，如果5秒时间后设备未收到来自LoRa网关的ACK信息，则再次发送并继续等待ACK信息5秒钟，如此反复直到接收到ACK信息。当固定设备接收到来LoRa网关的ACK信息和风险名单后，设备会将踪迹记录列表(**trace record list**)中的开头一条记录(最近发出并接收到ACK的一条踪迹信息)删除，并把接收到的新的风险名单保存到内存中，更新本地的风险名单。
* 移动设备(**mobile device**)，是相关医疗机构售卖给用户的设备，由用户随身携带。每个用户持有一个移动设备，移动设备有一个指示灯(**indicator light**)来表示用户健康状态。移动设备轻便小巧、便于携带，尽量减少携带该设备给用户带来的不便影响。移动设备会周期性地更新现用匿名(**present pseudonym**)、保存21天内使用过的匿名。移动设备会在收到来自固定设备的唤醒信息后，将现用匿名信息通过LoRa信号发送给固定设备，之后进入固定时长的休眠模式(**sleeping mode**)。移动设备只能在休眠模式结束后才能接收唤醒信息并发送现用匿名信息。移动设备在接收到来自固定设备的风险名单后，根据名单内容改编自身的指示灯颜色。移动设备接收并识别固定设备发送的唤醒信息，并保存伴随唤醒信息发来的风险名单(16字符的字符串组成的列表)。之后该移动设备向固定设备发送现用匿名(16字符字符串)作为回复，并同时将最近21天使用过的匿名信息从存储中调入内存，与风险名单进行匹配，如果匹配成功且本来用户的感染风险级别为无风险(绿色)，则更改感染风险级别为有风险(黄色)。密切接触者用户(指示灯黄色)需要前往特定的医疗机构检查，并且诊断专家可以获取存储在移动设备中的所有近21天使用过的匿名。若某个用户被确诊为患者，则诊断专家可以将设备的用户的感染风险级别更改为确诊(红色)。若该用户经检查后确定没有检测出病毒，则诊断专家可以将设备的用户的感染风险级别更改为无风险(绿色)
* LoRa网关(**LoRa gateway**)，是接收LoRa信号、读取匿名信息、上传到因特网的设备，是连接LoRa物联网部分与因特网部分的桥梁。它是固定地理位置上的计算性能强、信号收发性能强的设备，接收范围较大，可以接收区级行政单位范围内的来自固定设备的基于SF=9、频段为CN470的LoRa的踪迹信息。LoRa网关接收固定设备发来的匿名信息列表并缓存在本地。之后，LoRa网关将每条匿名信息签名、处理成区块并上传到踪迹区块链上，同时具有验证踪迹信息、维护区块链的功能。此时，移动设备通过LoRa发出的匿名信息被成功上传到了踪迹区块链中。一个LoRa网关一般可以覆盖周围7~9所大型公共建筑。
* 诊断专家(**Diagnosticians**)，是对密切接触者(**close contacts**)进行诊断来确定是否是感染者的机构。诊断专家有权根据密切接触者的感染情况更改他们的移动设备的指示灯颜色。确诊患者(**diagnosed patients**)需要将自己的移动设备存储的21天内使用过的匿名信息提供给诊断专家，之后，诊断专家将患者的所有匿名信息上传。应用程序通过患者的匿名信息去查询密切接触者的匿名信息，并将这些查询到的匿名信息发送给固定设备。诊断专家可以操作风险名单，比如，他们可以上传患者匿名名单、删除风险名单中的匿名信息。

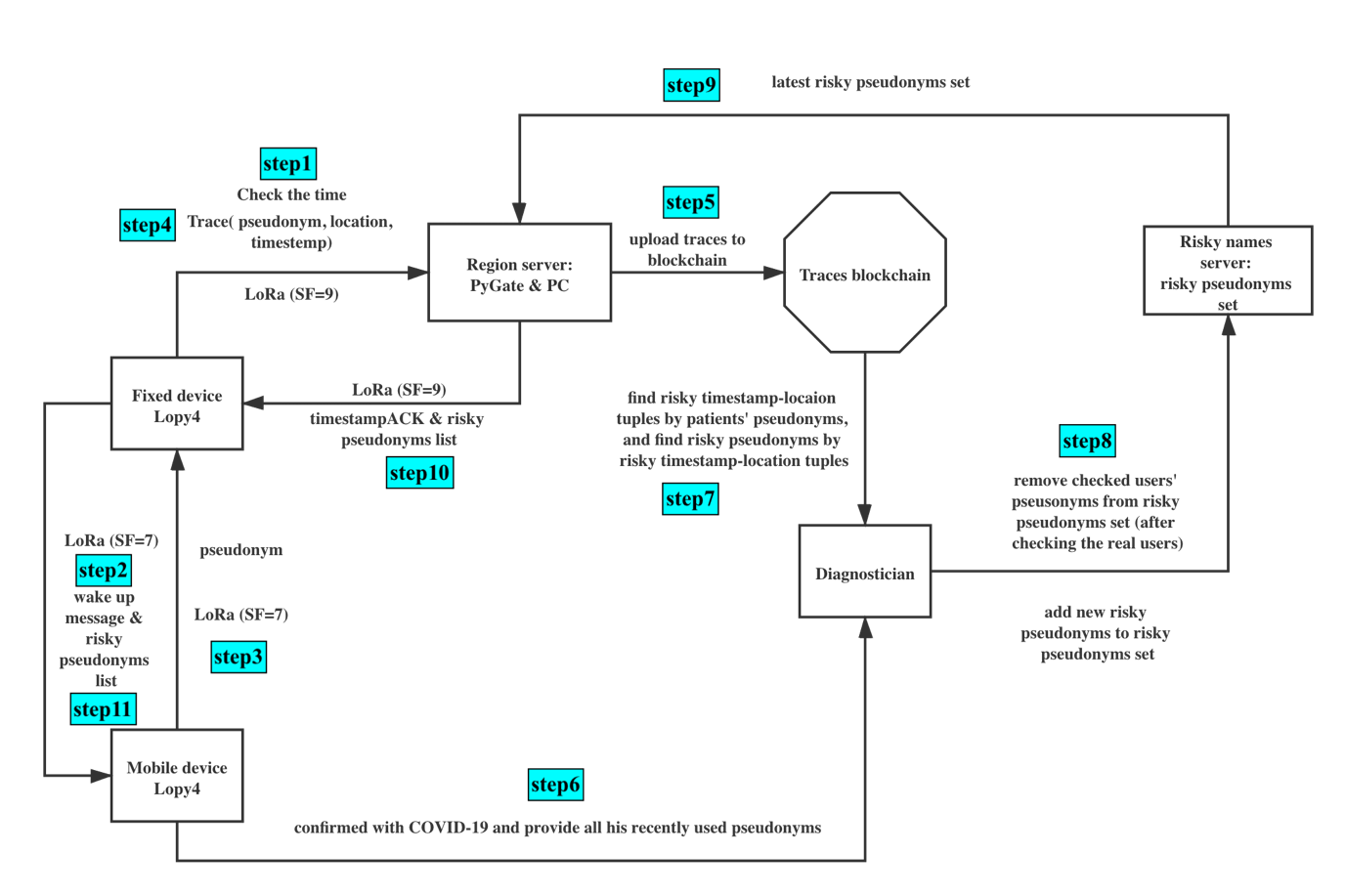


Fig 1. LoBeTrace工作流程图

* 1. Workflow of LoBeTrace

**Step1:**

固定设备总共运行两个进程：接收兼处理进程和发送进程；

首先会通过SF=9的、频段为CN470的LoRa信号向LoRa网关发送时间戳请求(**timestamp request**)，以从LoRa网关获取当前时间戳。只有获取了当前的时间之后，固定设备才能开始正常工作，否则不能进行正常工作，即在脱离了LoRa网关的情况下固定设备就无法工作。

公共场所的工作人员需要通过使用移动设备向固定设备发送若干数据包以供固定设备确定接受范围，以免接收到场所外的无关信息。在此期间，工作人员需要将移动设备和固定设备调节到范围检测模式(**Scope detection mode**)，然后移动设备会不断地每隔0.5s发出一个数据包。第一次采集数据需要工作人员手持移动设备在公共场所内和场所边缘移动，第二次采集数据需要工作人员手持移动设备在公共场所外移动。之后，固定设备会使用KNN算法计算出一个RSSI中间值。之后，固定设备会根据收到的来自移动设备的信号的RSSI值来判断是否是自己管辖的范围内的设备发出的信号。

**Step2:**

固定设备会周期性地每5s向周围环境发射基于SF=7、频段为CN470的LoRa信号的唤醒信息，起到唤醒移动设备的作用、并尽量减少疏漏。

**Step3:**

移动设备总共运行四个进程：维护匿名的进程、发送进程、接收兼处理进程、指示灯控制进程。移动设备会周期性地自动更换或者手动更换匿名信息。自动更换周期为随机2-4小时，使用随机周期的原因是为了在一定程度上克服**Wireless device tracking**[**A suvey …… Apps**]。移动设备会将最近21天使用过的所有匿名存储在设备中。 当某个匿名信息的存在时长超过21天后，会被检测出来并删除。当最新的匿名信息存在时长满2-4h后，移动设备会产生一个新的匿名。手动更换匿名则只需要用户按动移动设备上的按钮即可更换现用匿名。手动更换匿名的目的是为了避免**Enumeration Attack** [**A suvey …… Apps**]。

当移动设备进入固定设备的广播范围后，便可以接收到固定设备发送的唤醒信息和风险名单。移动设备接收到唤醒信息之后便回复匿名信息，它回复给固定设备的信息包括该用户的16字符匿名、标识发送设备种类的代码。为降低移动设备的消耗，移动设备不接收ACK信息。之后，移动设备进入10分钟的休眠期，在此期间，移动设备不接收和发送任何信息。

**Step4:**

之后固定设备就可以接收到移动设备发来的含有匿名信息，并将接收到该匿名信息的时间、信息包含的匿名、自身的地理位置信息结合在一起，构成一条踪迹记录保存，之后再周期性地向对应的LoRa网关发送本地的未上报的踪迹信息。我们将固定设备接收匿名信息和上传踪迹记录以分时的形式进行，即每接收3分钟的匿名信息，就用1分钟来上传本地暂存的踪迹记录到LoRa网关。上传踪迹记录期间不发送唤醒信息，不接收新的匿名信息。分时工作可以提高固定设备的工作效率，使得设备只需要一根LoRa天线，减少成本。尽量减小周期时长可以减少疏漏。

固定设备发送给LoRa网关的信息包括一条踪迹记录、标识发送设备的种类代码、对应LoRa网关的编号、10位的信息序列号。使用序列号的作用是供固定设备识别是否收到对应的ACK信息。每次发送一条踪迹记录后，固定设备会等待接收来自对应的LoRa网关的ACK信息。如果固定设备在发送后5秒内未收到ACK信息，则它会重复发送并继续等待ACK信息，直到接收到ACK信息才会发送下一条踪迹记录，其中重复发送的信息的序列号相同以供LoRa网关查重。

**Step5:**

LoRa网关可以接收更大范围内的来自固定设备的基于SF=9、频段为CN470的LoRa的踪迹信息，接收成功后立刻返回ACK、风险名单、对应的序列号，并把该序列号保存在临时黑名单(**temporary blacklist**)中。如果接收到具有与临时黑名单中的踪迹相同序列号的踪迹信息，则每次都返回ACK信息且不再重复上传该踪迹记录，而是只在第一次时上传该踪迹记录到区块链。

LoRa网关总共运行三个进程：区块链维护进程、接收兼处理进程、发送进程。LoRa网关是连接LoRa无线网通信和因特网的网关类设备，也具有验证签名、维护区块链的功能。LoRa网关接收来自固定设备的踪迹记录并将该记录签名、上传到踪迹区块链。最终在区块链中一个区块所包含的内容有：踪迹记录、网关代码(**gateway code**)、签名、本区块的哈希值、上一区块的哈希值；

每次LoRa网关接收到新的踪迹记录后，该LoRa网关会使用自己的私钥对将该踪迹信息进行签名、上传到区块链并通知其他联网的LoRa网关设备。每个LoRa网关都存有、更新所有其他LoRa网关的公钥。其他LoRa网关设备会验证该区块的合法性，如果验证成功则加入自己维护的区块链中，否则丢弃该信息。通过区块链技术可以在极大程度上提密切接触者追踪的可靠性，避免被不法分子篡改或私自利用。

**Step6:**

诊断专家具体是指每个可以对流行病病毒进行检测的医疗机构。当风险名单上的风险用户来到医疗机构检查后，无论是否患病，医疗机构都需要将其所使用过的所有匿名信息从风险名单中删除。

当有用户经诊断专家检查并确诊后，诊断专家有权获取该用户最近21天内使用过的所有匿名，并将该患者的移动设备指示灯改为红色。

**Step7:**

之后，诊断专家将这些匿名信息在踪迹区块链中进行比对，找到患者所有的踪迹记录，并将与该患者在同一时间范围内、同一地点出现的其他踪迹记录筛选出来作为风险踪迹记录(**risky trace record**)。我们认为在患者的踪迹记录的时间戳的前后1小时内出现的踪迹记录均算作风险踪迹记录，因为每个人进入某个建筑通常情况下都会逗留数个小时。

**Step8:**

诊断专家将提取出这些风险踪迹记录对应的匿名信息作为新增的风险名单上传到风险名单服务器。

**Step9:**

LoRa网关会周期性地通过因特网访问风险名单服务器并获取最新的风险名单，来更新本地的风险名单。之后，LoRa网关会将风险名单与ACK信息一同发送给固定设备。

**Step10:**

固定设备会接收到来自LoRa网关的风险名单后会将该风险名单发送给移动设备，自己则不对风险名单做任何处理。

**Step11:**

固定设备会将集合类型的风险名单信息与唤醒信息一起发送给移动设备，之后移动设备会将收到的风险名单和自己近21天内使用过的匿名名单进行匹配，如果匹配成功并且指示灯颜色本来为绿色，则指示灯颜色变为黄色。指示灯颜色为黄色的用户可以自愿地但是有义务地去医疗机构进行诊断，减少传播风险。

当患者痊愈后，诊断专家可以将该用户的设备指示灯改为绿色。



1. (b)

Fig 2. a）移动设备和固定设备 b）LoRa网关

1. Prototype of LoBeTrace for Contact Tracing

To demonstrate the feasibility and practicability of LoBeTrace, we implement a prototype system. 如图2(a)所示，原型机(**prototype**)使用Pycom LoPy4和Pycom ExpansionBoard3.1作为移动设备和固定设备，使用2000mAh的LiPo电池供电。实际应用情况下，固定设备使用电源供电，因为实验场所原因，实验中我们同样使用电池供电。**天线信息？？？**。如图2(b)所示，原型机使用Pycom LoPy4和Pycom Pygate与MacBook 2018 A1989相连接作为LoRa网关设备，Pycom设备和MacBook之间通过有线连接。其中，移动设备和固定设备之间、固定设备和LoRa网关之间通过频段为CN470的Raw LoRa信号交流，LoRa网关之间通过Wi-Fi信号接入Internet交流。LoRa网关的PC部分安装了和操作区块链有关的应用程序，起到维护区块链的作用。网关的PC部分显示器上可以显示区块链的区块添加信息。我们将风险名单服务器运行在腾讯云服务器上，设有专门的外网IP地址和端口号供各个网关访问、获取信息，设有特殊的端口号供诊断专家修改风险名单。原型机使用普通笔记本电脑作为诊断专家，运行有一个可以实现为风险名单添加、删除匿名列表等操作的程序。



Fig 3. 实验场所：大连理工大学开发区校区教学区，分为A、B、C三个建筑



Fig 4. 实验人员手持移动设备在建筑A内行走

如图3所示，我们的实验过程在大连理工大学开发区校区教学区内完成。教学区由三栋建筑物A、B和C组成。建筑结构为混凝土建筑，墙平均厚度约为0.5米。在实验中，移动设备均由实验人员手持，并且每个实验人员在建筑物内指定的范围内随机走动，如图4所示。

因为很难做到真正测试21天的数据，为了节省时间，我们按比例加快了实验速率。在实验中，我们将每个实验的时间压缩到大约30分钟。我们增大了固定设备发送唤醒信息的频率、减少了移动设备的休眠时间、增大了固定设备上传的频率、增大了LoRa网关上传踪迹记录到区块链中的速率。

并且，在实验中我们通过在随机生成的匿名中增加可以用于识别移动设备的前缀，以区分各条踪迹记录对应的移动设备。在实际应用中则不增加前缀。

我们进行了以下性能评估实验来测试丢包率、系统上传速率。

试验一，是使用单个移动设备和单个场所设备测试丢包率和上传速率。首先，我们先启动并配置好LoRa网关和固定设备，网关和固定设备均放置在建筑A内2层几何中心处。然后启动一个移动设备。其中，固定设备每6s发送一次唤醒信息，总共广播300次唤醒信息。固定设备的上传数据频率为每3s上传一次。移动设备的休眠时间为1s，每个移动设备的唤醒次数不限。实验人员在一层、二层、三层之间随机走动。理想情况下，区块链最终保存有300个区块。实验结果为区块链中共有283个区块，丢包率约为5.7%。

**Fig 5.**

试验二，是使用多个移动设备和单个场所设备测试丢包率和上传速率。我们一共做了四次，分别使用了5、6、7、8个移动设备。首先，我们先启动并配置好LoRa网关和固定设备，网关和固定设备均放置在建筑A内2层几何中心处。然后以10s为间隔，陆续启动所有移动设备，各个移动设备的活动范围分布如图5(a)**(平面图)**所示。其中，固定设备每3s发送一次唤醒信息，广播唤醒信息的次数不限。固定设备的上传数据频率为每3s上传一次。移动设备的休眠时间为20s，每个移动设备可以被唤醒25次。8名实验人员分别在一层、二层、三层的指定范围内随机走动。理想情况下，区块链最终应分别保存有125、150、175、200个区块。实验结果如图5(b) )**(启动移动设备数量-丢包率折线图)**和图5(c)**(地点-丢包率分布图)**所示，通过结果图可以看出，**……，**丢包率随着并发工作的移动设备的数量的增多而增大。

**Fig 6.**

试验三，是使用多个移动设备和多个场所设备测试丢包率和上传速率。首先，我们先启动并配置好LoRa网关和三个固定设备。网关放置在三栋建筑的几何中心处，即教学区中心花坛处。三个固定设备分别放置在A、B、C建筑的2层几何中心处，如图6所示。然后，我们在A区以10s为间隔陆续启动2个移动设备，2名实验人员分别在一层、二层随机走动。我们在B区以10s为间隔陆续启动2个移动设备，2名实验人员分别在一层、二层随机走动。我们在C区以10s为间隔陆续启动2个移动设备，2名实验人员分别在一层、二层随机走动。其中，固定设备每3s发送一次唤醒信息，广播唤醒信息的次数不限。固定设备的上传数据频率为每3s上传一次。移动设备的休眠时间为20s，每个移动设备可以被唤醒60次。理想情况下，区块链最终保存有360个区块。实验结果为区块链中共有314个区块，丢包率约为12.8%。

**Fig 7.**

试验四，是测量诊断专家在输入一个匿名到风险匿名名单后，到对应移动设备指示灯变色所需要的平均时间间隔。在实验4的基础上，我们随机选取区块链中的某个踪迹记录内的匿名信息，通过诊断专家对应的应用程序将该匿名信息添加到风险名单服务器的风险名单中。之后，我们等待对应的移动设备的指示灯颜色变化并记录这之间的时间间隔。该实验重复40次，并计算平均时间间隔。实验结果如图7**(每一次的时间间隔柱状图)**所示，平均响应时间间隔为12.6分钟。

通过试验一、二、三可以看出，我们设计的LoBeTrace原型机在每个固定设备对应少量并发工作的移动设备时，丢包率比较低；随着每个固定设备对应的并发工作的移动设备的数量的增多，丢包率开始增大，但是仍然在可接受范围内。并且，随着移动设备的位置越靠近固定设备管辖的范围的边缘，丢包率越高，原因为移动设备在边缘时发送的LoRa信号的RSSI值接近固定设备检测的临界值，有时会被算作范围外。

但是，因为我们的实验是按比例加快了频率的，所以在真正的实践中，频率会慢得多、并发性会减弱很多，相应地丢包率也会减小许多。

以COVID-19传染模型为例，在疫情防范较为严格的国家比如中国大陆，the average Infected-Suspected ratio is 2.399**[CoronaTracker: World-wide COVID-19 Outbreak Data Analysis and Prediction]**，namely one person will most likely infect 2.399 other people, and those 2.399 people will each go on to infect 2.399 more people and so forth. 参照实验三的结果，按丢包率为12.8%计算，假设每人每日在室内公共场所内的时间为6小时，移动设备每次的休眠时长为10分钟。最初有一名患者在室内公共场所，则一天结束后，**……怎么算？**

Pycom运行microPython。按每移动设备每两小时更换一次匿名信息计算，每个匿名为16个字符的字符串，大小约为65 bytes。每21天最多产生252个匿名信息，共计需要16380 bytes的存储空间来保存，约16KB。可见，该方案对移动设备所需要的存储空间和内存的要求很小，数据传输量也很小。因此，该方案的成本和消耗极低。

**电池寿命……，一次充电\*\*\*mAh可正常工作……**

1. CONCLUSIONS

本文提出了一种基于LPWAN和区块链的密切接触者追踪的解决方案，具有低功耗、低成本、便携、易于普及、可靠性高、安全性好、可追溯性强等特点。我们将密切接触者追踪程序从智能手机中分离出来，参考新加坡政府推行的TraceTogether使用单独的物理设备。我们用当下比较流行的LoRa信号代替蓝牙信号，降低了成本、减少了消耗，不再需要用户时刻保证手机蓝牙处于活跃状态导致频繁充电，为终端用户在最大程度上提供了方便。而且，我们使用了去中心化的区块链方法，很大程度上降低了单点故障和不法分子篡改信息的风险。用户的公共场所活动轨迹最终会被上传至区块链中，由于用户的踪迹记录是匿名的，匿名与**Personally Identifiable Information (PII)** 的对应只能在用户的移动设备中查到，无法被其他人得知，因此隐私保护性很强。而且，区块链可以是公开的，可以为解决纠纷或在某些法律问题上提供证据。但是，目前的方案仍有不足，比如，传统的区块链在大规模应用时会面临并发性差的问题，可以用最新流行的DAG-Based Blockchain替代传统的区块链来解决该问题。另外，由于物联网终端设备计算能力较差，我们需要在边缘云服务器(本方案中用LoRa网关来实现)的权限和终端设备的权限中取得一个平衡，既能保证安全性、可靠性，又能最大程度上提高计算速度。在后续的工作中，我们将改进方案、弥补这些不足。