2019.5.23 数据和试验驱动的网络演进

会议主题：数据和试验驱动的网络演进

会议时间：9：00--12：10

会议地点： U谷一期444

主持人：那我开始，第一位他是来自美国东北大学技术主管，Abhimanyu Gosain。其实是他第三次来跟我们分享，这个无线的一些测试平台。我们欢迎Abhimanyu Gosain博士。

Abhimanyu Gosain：大家.早上好，非常感谢大家今天来，也要感谢我们主办方再次邀请我来到这里。

我想你们也听说过，我们的这个项目，这是在美国的一个项目（英文），我给大家介绍一下这个项目的现状，我们在过去花了一些时间，在过去有哪些进展，以及我们所学到的一些东西，另外我们今后的一些方向。

我是Abhimanyu Gosain我是美国大学技术主管，我们主要是无线的研究的一个测试平台，这样一个项目。介绍一下我们这个项目以及它的历史，这个历史其实和我们昨天我想主会场当中讲的其他一些项目都是差不多，很多项目都是同期开展，主要是我们更关注于无线方面的测试。另外我们还有公司的合作，是得到国家科学基金会的支持，还有29个企业的联合体，每个都贡献了五千万美元，这是一个这样的资金池子来确保我们进行全面一个无线的测试平台。

为什么我们需要这样一个平台？关键就是对这个学术还有我们的企业之间缺乏这样的合作，因此我们需要面对未来一个平台，可以进行编程，能够进行我们无线的前沿的研究。比如说边缘的运算还有零运算，并不是基于大学的教学的，因为大学教学往往只是一到两个学生可以用，但是我们是使得行业还有我们学界可以联手起来，当然整个项目也是高度可控，高度可运作的，我们是建立一个城市的一个平台，所以它的规模非常大，能够给予到大家一定的保真度，才能够确保这个项目非常成功。另外也希望对整个行业有非常重要的影响，同时也对学术界有重要的影响。关键就是我们要加速，我们的基础性研究发现

确保我们有这样的投资，确保整个投资是合适的，我们也有这样的实力来有这笔资金，从29个企业提供这笔资金，这些公司其实都是涵盖了整个无线的系统IBGA还有设备制造商、芯片制造商、运营商、收集制造商等等，这就是我们行业因此我们也是非常兴奋，行业也投入了很多的资源，希望能够产生非常深远的影响。

我们看一下在过去一年，是2017年以来我们代表了我们项目的办事处，也是代表了我们资源科学基金会，还有行业联合体，这确实是非常漫长这样的流程，我们花了一年来招募团队，建立我们第一个平台，目前还在选择下一步该怎么走。我简单介绍一下在过去一年我们项目的一些进展，我们使用的一个缩写就是（英文），就是在盐湖城，这是在中西部有点朝西部，这是在美国的内陆，另外这个纽约城市叫COSMOS，这是在曼哈顿，所以有这两个。在COSMOS就是简单的介绍一下我们的一些技术，直接了当给大家介绍一下我们在过去一年技术的一些进步。

COSMOS我要讲到很多其他术语的缩写，还有哥伦比亚大学还有纽约市和我们一起合作这是一个非常重要，这是到带宽还有低延迟的项目，我们关键是在无线方面还有毫米波等等，还有28GHZ，这个28是在（英文），还有除了我们硬件之外我们还有很多的细节，除此之外，还有昨天在我们的其他的一些（英文），我当然可以看到有很多是可配置化的，还有兼容的，我们还可以和高带宽和我们毫米波融合在一起，有非常高容量的一个回传，确实可以驱动我们的无线创新，我们确实有一些无线的技术，这是我们的光鲜的技术。我们GBU都是可以获得光纤进行实验。

我们并不是仅仅研究而已，而是帮助研究者快速加速实验，这个是可编程的基础设施，我们可以看到这些都是开园，所有都是在线颁布的，能够让我们社区变得更强大。在底部右手，这是我们的部署这是大约在哥伦比亚校园区有九平方米这样一个区域，这是九平方英里一个空间，这是我们的部署，这是我们今后两年的部署在九平方英里展开。

我们的试点的项目也最近完成，这是黄颜色这是在哥伦比亚校区，确实是有不同的一个我们的图标，有这样三个这样的节点，有大型、中小、小型。大型他们是有大型边缘的预算，也有非常大的一个天线，安装在非常高的楼宇上面，我们叫做这样的一个区域。还有重点是街道路灯上面可以看到有些是阿莫斯特大街，我们来这个地区进行一些部署。还有其他一些，蓝色就是纽约大学，比如说哪里有个小岛，确实我们使得高带宽和低延时接着这样的走廊来进行测试。

我已经谈到了毫米波，我想关注于当然我希望更谈的是比较宽泛一点，并不是细节。对毫米波一个节点我们跟阿里信也有合作，有64个机缘，我们对无线电话除了RF之外，我们也需要有这样的阶段。还有是有这样非常大的一个带宽，然后进行这样的处理，不仅仅是机带，还有能让所有的信息传到你边缘的设备，这就是一些细节，我们想就是28集合资，像3.5这样一个中部的频率，送到两个阶段PGA再进行整合。我想这个也是一个缩写，这确实是一个RFO的一个非常大的处理能力。BSB能够快速进行这样一个处理，能够搜集大量一个数据流，这是一个象征控一个技传。每个有16个这样机缘，每个是16个机缘的波速，每个能分成16个有16个这样的机缘这个一个波素可以同时送到好几个用户里，这是我们的愿望。现在仅仅是这样一个流传到了3.5兆和次，再回到我们光纤网络，这个是在纤维上无线的频率，回到我们中心设备进行这样的处理，这是我们的硬件，这是软件的一个框架，我们能够给我们研究所所用。

这是28集合资当然还有另外一个我们在在测试床当中，有这个技术也是脸谱还有（英文）一个项目叫（英文），这是60集合资，美国我们也是允许这样的制造，最近我们也确实收到，这也是一个行业的个联合体像（英文）和（英文），他们业提供了。这是有36（英文）的馈线，这是是非常紧凑我们用了这样一个系统，主要是用了整个通道的声音，人们也知道第一件事要做我们如果有这样的频率我们就是要这个环境、通道进行建模，我们使用这个就在我们（英文）进行安装，然后通道一个建模，我们可以看到在60集合资在室外怎样运行，这当然是我们一个软件，我们有各种各样的团队。关键我们是可以看到每个一个变化都是在黑盒里面完成控制，可以进行电子话波速的导向，在整个系统当中可以完成。因此我们要为我们通道来进行表征，在COSMOS区域有这样一个。

接下去我们看一下我们光学医，这当然很多项目当中你们也看到比较熟悉，这是我们总体的部署，我们的一个光学的一个交换器，这样大型一个部署在COSMOS平台当中，这个是可再配置有好几个，所有这些都是，在一起有大型、中型和小型的节点，所有这些都是通过专门的光纤可以回收。这是有300多个光学交换器，这就是我们所说的我们启动的托博（音），在合适地方有大型和中型的节点，我们可以看到在旁边有不同的光纤分离器，从大转到小再转到更小，所以我们有足够灵活性，能够在不同的环境中进行实验。在这里我要忽视它是非常基本的一个子系统的集合资，也是60%这样的部署，我们早期可以看到我们的部署的一个计划，剩下40%是毫米波所赋能的是边缘，现场的无线电。

我们建立了大概有9个无线电的机站，在这一页上有基本的情况。在我们美国运营商也包含我们基础设施运营商，所有这些数据都是通过管钱系统传到我们一个叫做中央空间交换器的端口处。大家可以看到我们右侧显示的就是哥伦比亚大学部署情况，这里面的（英文）数量是有限的，这里都是和我们外部信道进行连接。这个3D的（英文）谁可以登陆这是非常重要的一点，也是我们关注的一点，这其中的成本代价是非常高昂的，所以我们也希望进行一个合理的配置，给我们用户进行一个比较好的体验。

下面关于这个平台的技能和相关的成本，我们每个平台构建成本都在两千万以上，其中大部分都是基于这个部署的基础设施，然后其中也涉及到我们软件配置情况。另外我也可以给大家提供一个网站地址，关于这个项目具体信息大家可以登陆这个网站获取详细信息，当然在中国也有相关研究工作。另外一个平台我们叫RENEW它是放置在盐湖城的平台，我们这里的平台它是一个可进行配置的生态系统，另外它可以提供一个不同的平铺的接入。在在（英文）大学，在这个大学我们运营商可以向我们机站发送大量的数据源。相关的数据会通过工作系统传输到我们的社区，同时为我们不同客户提供服务，这里的关键就是，就是我们大量的机站的建设，它这个容量达到50兆赫资到58集合资，实际上这是一个共享的流程，这也便于我们GDP的设置。

除了这个RAP之外，我们还有一个混合的计算，这个包含SCPU的处理情况。其中包含了180个元素都可以把我们这个服务提供给，已经购买这项服务的用户，最重要的就是我们为不同层级用户提供不同的软件服务，可以通过IPI、（英文）这个就是为用户开发的一种语言。最重要的一件事，我们这个边缘计算和云计算，这个希望我们构建相关软件的架构帮助我们。当然在美国这个也涉及到高昂的成本。

这是主要成本架构情况，我把所有组成要素都放在了这里，大家可以快速看到的一个就是所有的，就是这些元素都是移动式的也是构建在两个社区两个校园一个系统的情况，这为我们用户提供了专用的解决方案，这个其中也具有很大的可靠性，我们的城市和大学之间构建了这样一个联系。我们通过这样一个项目可以进行编程，如果我们需要对机站进行设施可以通过这样一个架构来实现。所以在最早期的时候我们可以通过部署来实现试验的目的，在最后端我们云端计算也是非常强大的，这是基于云端的软件配置。

如果没有去过盐湖城大家可以通过这个图片了解一下大概的情况，这个中间的红色是大概3.6个平方公里的一个位置，在右侧大家可以看到这是一个受热图，这个可以显示我们资源移动性的情况，这是在整个城市中的一个穿梭情况。在美国的话这是最大的一个部署规模，这个实现了全范围的一个城市范围的一个部署情况。给大家讲一下这个机站内部构成情况，这是针对大学构建的部署社区，当然我们出售的设备大家可以在这里看得到，如果大家有兴趣的话在最初进行连接的时候我们可以获得什么，SDR等等这些，大家有兴趣都可以向我了解。我们配置有64-128天线的配置，我们传输是通过光纤传输，我们还有一个非常酷的地方，就是我们有一个内置的时间同步。我们这个网络数量非常多的时候，非常重要一点就是要把时间实现同步，我们通过同样处理方式，确保我们时间在一个水平。具体的介绍大家可以看一下在这里，这两个是可以相互进行切换的一种模式，2.6、2.7等等一直到RSM（音）这些都是可以进行互换的。在小的时候我们也可以把这个玩具组装在一起，把小火车组装好了以后小火车在前面跑我们在后面跑，这就是我们信号采取同样类似的一个原则。当然在这个当中我们一态网是非常重要的。

这是另外一个不同的概念，给大家了解一下我们这个团队在过去几年完成的一些进展，我们建立了一个射频的机站，在（英文）设的机站，我们在U它（音）建立了一个机站。大家可以看一下下面这张图这是在我们U它大学设立的机站，这个机站是设立在屋顶的，我们机站可以实现不同的处理能力。大家可以通过这个机站各个功能，实现不同实验目的，也可以进行不同类型的组合，从左侧到右侧大家可以看到我们射频的频率可以根据实验需要进行调整，我们有带宽和宽带宽这样的选择响，我们有25个集合资的选择。我们通过左边连接方式传达到右边，当然这个过程中我们需要不断的学习。我们把光纤系统连接好了以后所有硬件也需要到位，当然我这一页上所显示其中就是一些代表性的硬件。

这个是用户所使用的一个配置情况，我们可以很快的把我们这个在硬件上布置我们的软件核心要素，这个是端对端的也是完全实现可以编程的一个过程，我们可以在这个云端实现各个技术相对应的解决方案。这是在COSMOS已经完成的一些工作，我们有（英文）一个架构，大家晚点有兴趣可以了解更多的详细情况。

我们这个（英文）系统是怎样进行运作的？我们如果登陆这个系统可以找到这样一个资源端，我们所有的事业以及移动性怎样进行部署，都可以进行调整，我们不需要针对每一个情况设立不同的语言，我们可以基于这个系统设置它的配置和进行相关的部署。除此以外，除了硬件以外，我们还有软件部署情况，我们有这个射频可以帮助我们实现射频层图片的处理，达到接口端。然后再基于我们基础设施，我们还需要制定一些相关政策支持，昨天也听到了中国移动关于这方面的介绍，我们还有这个软件（英文）一些情况。这些都是大家可以获得的一些服务。

（英文）在2019年1月份启动，我们有11个这样的实验已经完成，我们项目的数量也在逐渐增加，接下去我们看一下我们还有两个这样的一个平台，我现在关注于我们应用领域，这个运用领域我们也会有一些技术的要求。比如说像UIB（音）还有农业车、互联网等等，过不久我也可以跟他们谈一下，接下来两个新的平台应该是什么。花一些时间，给大家介绍一下我们确实有大规模城市之外的试验平台，我们是这样大型之外的测试台，当然现在我们还有一些差距，我们在整个实行过程当中我们要进行这种差距在什么地方。最近我们也有一个投资，这是来自国防局对我们的投资，这是关于Colosseum的项目，这是全球最大的测试平台。有256乘256通道的仿真，有128个无线的接点，有大量的FPGA、CPU我们国防部进行投资的。我们确实能够帮助解决各种各样新挑战。在我们全球移动通信大会，10月份也会介绍这样的项目。

好，这是我们一个基础架构也是非常类似，有很多互相连通的，这是完全可控制的环境，这样的环境也可以是移动可访问，遥控可接入，有非常好的一个无线的兼容性，也有很好一个调节能力。还有我们试验的控制力和框架，能够帮助我们进行控制，这确实是一个非常大规模仿真的环境，不久将会推出，20201月份我们社区将会用到。我已经说到关键并不是进行这样一个通道的一个模拟和通道一个实验，另外可以帮助你进行这样的探索，在无线网络当中进行协作，能够让我们进行协调。比如说谁将会发出哪个平铺（音），也可以应对协作当中的挑战。可以看到这是一个基础架构，我们可以通过基础架构满足我们原先设定的目标，这当然是一个非常好的。另外我们和国家科学基金会还有他的一些合作伙伴，来合作建成这样一个（英文），其实就是罗马圆形大剧场的项目。另外还有这个项目相关的一些链接，非常感谢大家仔细聆听，谢谢大家。

主持人：我们有时间给你们提两个问题。我有两个前端的问题，这个项目会持续多长时间？

Abhimanyu Gosain：这个项目将会一直持续到2022年，我要强调一点，这并不是说2022年之后它就是掉下悬崖，其实我们建立这样一个项目从第一天开始它其实就是可持续的，所有的这些基础设施，我们看到比如说来自于这个国家科学基金会，一直会支持到2022年，都是免费可以可以接入，2022年之后会有一个用户付费的模式，比如说我们还有政府还有行业，会提供资金。行业从第一天他们已经来支持我们的设施，比如说操作、维护、开放、实验，我想这个项目会持续。

主持人：谁会最后能获得这样的平台？

Abhimanyu Gosain：政府也会进行投资。比如说每隔五年他们都会好几百万每年的投资，最后还会持续运营。我们确实希望这样的平台能够给到社区，用户会付费然后给社区提供服务。

主持人：第二个问题关于它的开放度。每个人其实可以访问，我想这样的技术观5G，所以我前面说是不是中国也可以访问？

Abhimanyu Gosain：现在没有这样的限制访问这样一个政策，随着用户的增加我们这个平台更加有用，也更加有希望，会更多人访问，当然有的时候我们更倾向于美国的这些学术界或者是投资者，当然我们这个模型也可以复制的。我们也有欧洲这部分合作伙伴，他们也在研究这样的项目，他们也在做这样一些投资，在日本也对此感兴趣，我们有一天也会说服在车里有这样的无线平台。

主持人：谢谢您，第二位演讲者是Lourent Mathy来自我们比利时大学教授，他也是中国科学院讲座教授，他主要是关注网络化试验研究，他也参加了好几次大会，有请我们的教授非常感谢。

Lourent Mathy：感谢我们的主办方，感谢我们的朋友，感谢你再次邀请我来。今天我跟大家谈一谈，给大家介绍一下我们实验室跟德国做的一些实验合作。

要建立这样的数据面我们有两个数据，一个叫（英文），还有一个是容器，现在人们更关注容器。这个谷歌有各种各样容器的服务，比如说还有亚马逊还有谷哥还有（英文）他们都有，比如说你到互联网上去购买这样一个容器。这个优势在什么地方？关于容器的优势我会花一些时间，它是非常快速可以进行这样的时裂化，你在几百毫秒就可以启动，这就像电脑启动更快。它的内存足迹也是非常小，它站到100兆比特的空间，是非常高的密度，你可以在同样一个主机上面可以由100-1000个这样一个副本，这样的容器副本。如果你把虚拟机和容器比较一下，虚拟机是比较重的，你可以看这个照片他比较肥大比较笨重，这是虚拟机的情况。

对容器来说这个问题是，好，它的内核的API是非常大，所以很难对这个内核确保它的安全性。这张图这是有一些系统的调用，你可以看到系统调用的数量也是随着时间增加也有增加。（英文）这样一个机器，比如说有400这样一个系统性调用，后来也慢慢增加，后来越来越增加。所以你可能会有很多这样BUG，这样一个问题，这样一个单个一个不安全容器，也可能让在主机上所有其他的容器感染。其实像这样的容器，它其实就是像一个机上运行大的处理器，如果对（英文）攻击，或者（英文）这样一个攻击或者是（英文），这样的攻击。后来人们觉得这样一个容器对攻击来说很容易受到攻击，所以人们更愿意使用这样的虚拟机，因为它更安全。到底怎么做呢？我们是完全虚拟化还是容器呢？在虚拟化当中比如说你有像高级的、超级监督器还有我们硬件等等，全面的操作系统等等。对容器来说你有比如说单个核，然后容器会有切片，其实是一个单核的切片，其实容器就是一个切片。像一个出问题这样一个容器，它其实会让整个系统都出问题。

虚拟化给到你这样一个隔离，但是非常重，但是容器是分得比较清，但是隔离没有虚拟机那么强，有的时候你要喝下毒药，我在这里问的问题虚拟机是不是一定要成为重量的虚拟机是不是可以进行轻量化？其实这个是可以的，我给大家介绍一下UNikerenl，你们知道吗？其实我也不知道，我等下给大家解释一下。你可以看到这里有很多调性，你得到你的库和第三方的应用，问题就是不管什么样的运用，你需要整个核当中一小部分。

这个UNikerenl想法也非常简单，其实有很多地方都是没用的这些代码，这些代码也许有一些BUG可能会带来一些安全问题。我们只是把我们需要运行应用的部分保留下来，UNikerenl其实就是把我们需要东西留下来，其他的不要，我们其实可以看到它只有在我们运行运营当中东西，所以UNikerenl只是最最需要的东西在里面，我们是一种非常优先的一个模式来运行这个壳。

我想要么你是单独优先模式，要么是非优先模式，就这两种模式就行，我们也使用了，我们看一下这样一个UNikerenl是不是能够兑现诺言，能够实现非常好的性能呢？这个是我们进行虚拟化的时间，这是我们机线的情况，就是我们创造这样一个包括这样一个应用流程当中所需要的时间，这就是一个流程的创建。其实几个毫秒可以创造这样一个过程，Docker在这样的容器，其实几百个这样的毫秒完成了这个过程，它其实也不是太坏。还有虚拟机我们完全一个虚拟机它需要多长时间，这是一个（英文）分布器，当然花的时间很多。在这个盒子外面我们有虚拟机它实际上是非常重的一个东西，我们建立的有一些UNikerenl我们目标是对他们实现优化，在其中我们可以看到也实现了UNikerenl的优化，通过这样的优化可以实现这个速度。实际上稍微有点缓慢，创建这个容器的时间有些放缓。

下面一个问题就是，这个速度是否能够提高？如果我们在开发UNikerenl的时候，可不可以把这个性能优化的同时，来降低它的时间提高速度？我们所要构建的就是我们称之为轻制的虚拟系统。为了帮助大家来理解，我快速跟大家介绍一下，就是我的（英文），这是一个虚拟系统实际上。我们有什么呢？我们有（英文）还有这样一个域，还有在UNikerenl上面有虚拟机还有驱动器、虚拟化系统，在那里还有（英文）工具集合，可以帮助我们管理和配置，虚拟机的一个装置。除此以外安装了（英文）以后，最初的时候可以棒正所有的配置是到位，以及系统一个稳定性。另外还帮助我们保住系统中不同通信的情况，除此以外我们还有帮助我们支持这个虚拟机的一些元素，也就是我们（英文）这个平台上的这些设施。我们我们还有（英文）驱动器，在用户运用上可以帮助我们更好运用虚拟机。我们是在驱动器的前端，在前端它起到的作用就是，把这个流量带到传输到后端以及具有优先级的优先的域，然后把这个流量再传输到真实的驱动器。只有（英文）和我们这个设备进行对话。这一点并不是那么重要，但是我们需要提及的就是，在我们虚拟机（英文）的时候，我们首先要在设备上、系统上创建（英文）、安装（英文）。我们在虚拟机上发生了什么情况呢？它需要跟我们（英文）进行对话，我们后端（英文）跟我们对话。我们前端虚拟机也需要跟这个（英文）对话，在当天结束的时候发生的情况就是，我们大概有30（英文）的交易，这是在单个虚拟机上发生的一个情况。当然这个是需要花费大量时间，所以我们需要对这个虚拟机进行优化。当然如果我们想要提高这个速度，我们需要做的就是，把这个（英文）给去除掉，寻找新的替代物，我们取得了一些进展，具体内容我不过多详细接受，但是我们开发了一些系统，可以取代我们（英文），这个是可以帮助我们实现更好的通信对话。

我们做了什么工作呢？我们重新对系统进行了评价，然后进行了改进和改造。我们开发了不同的五种UNikerenl尽量都是最简版。实际上仅有UNikerenl解决不了问题，所以我们要有基本配制，我们有（英文）服务配置，它可以帮助我们和时间进行一个响应。我们UNikerenl还有一个TLSUNikerenl，它的工作就是可以帮助我们实现代理。具体样本尺寸是怎么样的？大家可以看到在白天的时间，在这个磁盘上它大概不到500个KB，在这个（英文）上面，这个磁盘上的容量是三点几个MB。TLS稍微大概大一点，在这个磁盘上不到4个MB。我们做了大量的试验工作，我们做增加了VM运行时间、运行次数，我们把这个实验的结果和在DOcker进行对比。我们创建了成千上百的数据，大家可以看到这个（英文）内容不足，这里面大概有3000个容器左右。大家可以看到在另外一方面我们这个（英文）可以创建大概8000个，我们这个其实是可以继续做下去，但是我们觉得无聊没有往下做。

我们虚拟机在这个过程当中实现了完全的一个隔离，这个是毫无疑问没有可以妥协的地方。我们这个LightVm可以在十秒钟进行完成，我们可以对我们规模进行快速的发展。UNikerenl如果使用一个改进的（英文）它的速率实际上是更快，以及重量也是更轻的。

我总结一下UNikerenl一个潜在优势，我们这个实力化的速度更快，然后这个环节的时间也更快，实际上这个时间可以缩短，为几个毫秒，或者更小。另外我们对于内存的要求比较低，只需要几个GB，这个是在单个服务器上的要求。除此以外也非常好的一个方面就是我们这个可以实现高密度的一个服务。我们测量过，我们每秒可以实现4个GB在Lightvm上实现的，是单个用户CPU上，是再一个CPU、Lightvm，一千到一百个lightvm。我们降低了受攻击的面，因为我们有内置了专用的内核。我们这个计算机的可信度更高，另外我们（英文）可以实现更好的隔离。我们可以通过内核进行重新配置，帮我们构建一个体积更小的规模。

UNikerenl它是最完美的方式，大家都想用UNikerenl，当然还是有一些小问题的，当下我们经过我们需要通过一种手段方式安装这种进行优化的UNikerenl。大家如果有兴趣可以找我的博士学生，还有我们其他的一些研究人员，他们正在致力于安装这样的UNikerenl系统。不同的实验室然后不同的系统都是需要进行专门开发来适应我们的需求。具体需要采用怎样的系统，什么样系统的内核？实际上这个是很难确定的。整个编程过程当中需要了解解码、编码，或者是使用这个（英文）来创建每种场景、情况下的需求，对每个应用程序我们需要做的就是，需要对我们这个目标应用程序进行不断的进行反复尝试。这对于研究人员来说这是一个好消息，因为我们只要有问题在我们才有发展的潜力。

我们有一个项目叫UNIKROFT，实际上是开园项目的名称，我们来自欧盟的授权在我们Unikraft项目内，我们要构建这样一个体系，可以帮助我们非专家人士可以在很快的时间内，快速的高效的构建他们所需要的类似于这种虚拟机等等这种系统。无论你的目标体系是什么，你可以通过这个Unikramft这个项目来构建项目体系，我们这个项目的目标就是可以帮我们降低时间，可以帮助我们构建不同目标平台，即使对于容器来说也是可以这样做容器的，隔离并不是一个很大的问题。或者（英文）我们也可以使用UNikerenl，也可以在我们CPU上进行实现。另外我们还可以提供简单的定制化的服务，同时不需要触发我们应用程序的代码。

具体是怎么样做的呢？我们有一个微实验室，我们把这个（英文）参与进来，这并不是一个非常大组成部分，大家可以看到我们有文件系统，所有元件都放在这个体系当中。你可以看到所有这些你需要都在这里，当然这并不是我的目标。过去我们是单一的操作系统，但现在我们要把它和比如说，编程还有其他一个操作系统组队，我想这样会更为可行，这也是我们一个梦想。我想关键就是魔鬼总是藏在细节之中，并不一定是很容易实现的，我们觉得这还是我们未来梦想，当然我们还有我们的微库（音）的一个平台，还有架构我们的平台。我们可以产生两元的代码，在这些架构当中产生，我们想法就是我们这样的应用，可以定一个API可以设置能够生效。我们还有像我们可以进行配置你所需要的库，然后这个系统也是你所需要的，还有最后就是你要做的就是部署，你可以根据不管什么样的目标或者什么样的应用，都可以建立这样的unikraft来运行它。

要做到这样我们要做很多代码的分析，我们现在花了很大的努力，来有这样的应用，然后建立这些工具，是通过这样的一个运算，如果你是有这样的（英文）和经验的话也都差不多，你是需要经历这些过程。比如说你要选出这样的平台，然后目标，然后进行编译，进行这样的一个编辑最后完成。

UNikerenl的目标有很多系统，当然还有五分钟就结束了。我们一个例子就是Python还有（英文）是建立在KVN上面，这个系统并不重要，我们看一下这个，这个KVN是不到1兆比特是非常小的。我们也建立了（英文）Unikraft用在容器上，这个是（英文）的创意，是可兼容的目标，这当然是非常详细的内容到底怎么做。

这里我们所有的就是，这个容器其实要比虚拟机更小，仅仅只有500K这样的大小。我们Unikraft这是我们的现状，这是我们需要的，还有红颜色这是以后的，我们现在有两个团队还有三个团队现在全职在完成这个项目。如果你想加入开园也欢迎你们加入。我们还有很多工作要做，还有一个是非常关键的事情，就是C库（音）很多都以来这样的C库（音）。如果要集成的话还要有这样的语言支持的话，我们其实都需要大家支持。

其中有一个我们想要做的就是，我们希望有更现代的一个版本，这个（英文）这个也是现在还没有完全完成，不管怎么说我们现在也看到了可喜的进步，我们有信心在接下来两年我们会建成这样的系统，比如说像（英文）也会有，我们会建成这样一个非常好的一个可金融一个项目，供我们使用者使用，大家有问题可以提问。

主持人：我们的听众有什么问题吗？我有一个问题，对我来说如果我的理解是正确的话，（英文）定制化的，我们听说过定制化的芯片、定制化的内核，当然是数据在我们一个影象片当中运行，会达到我们芯片当中，像这样的Unikraft你会有这样的（英文），然后运行对吗？这个也取决于这样一个系统调用，你在整个进展中你会调用。如果你漏用，漏了一些调用可能（英文）不会运转下去，这个其实也取决于，这个API的覆盖范围。

Lourent Mathy：我们有一些博士生他们也在研究这个话题，你可以通过数据分析来研究这样一个覆盖情况，当然是针对这样系统调用，我们当然有的时候不一定完全百分之一百保证这样的涵盖，我需要我们的信心，我们覆盖情况还是很好。有的时候有些地方我们还不太确信，我们是不是有一些技术，我们可以有一些代码，这个其实这个代码是要调用系统的调用，但是没有在我们里面，当然是不是我们代码是不是可以自动更新一下，可以有一个请求让它做一下，我们目前不知道如何去做，我们实验室当中也在讨论这个话题，有些人在全职负责这一点。我们（英文）需要更进一步提升和优化，我们这个（英文）也有更进一步优化，比如说（英文）方面的工作。像（英文）也是有非常大的一个代码，比如说我们有的时候，在（英文）超级监督器上做一些工作也是可以，也不一定像我们担忧一样非常严重。我们有的时候不应该去追两只兔子而关注一个问题来解决，所以我觉得这个（英文），我们还不能自动完整一个建造，现在我们还仅仅是集中精力，关注到这一点上。

主持人：谢谢你。我们第三位演讲者是DaliKaafar，是澳大利亚麦考瑞大学教授，让我们欢迎DaliKaafar教授。

DaliKaafar：非常感谢你们邀请我，我是跟大家介绍一下应用层。有一些数据所驱动的一些工作，有些工作是关于互联网的安全，还有网络的安全，我如果有时间我还要跟大家，有的时候你的浏览器每一次你访问你的喜欢的网页的时候，可能会有一些演换，有的时候200K就是20万个网站在全球，可能其中有些会有安全隐患。我并不知道中国网站哪些是合适。

我们先花一些时间来跟大家介绍一下，我们最近研究一些结果。我们在收集一些数据，这些数据我们就是要建立一个数据集，数据集是对于我们互联网当中安全有一些这样的一个纵向的一些洞见。我们在2016-2018年之间，我们都收集了一些数据，是互联网攻击的数据，我们希望能找到危险演进的逻辑，找到一些属性。我想有一些博士、研究生、研究者他们也会对这些数据感兴趣，我们有一个链接等会儿大家都会知道这个链接，我希望把网络安全性数据收集对大家有帮助。

这是看一下我们一些网络一个安全，还有这些攻击一个数据确实是很有冲动，这么多数据这当然也是过去数据储存的情况，现在当然不是这样，现在是像这样张照片一样。不管是在卧室或者是发电厂你的数据都会受到威胁，像这样网络安全真正存在一些事件，像2007-2017年，我们听到了很多网络攻击。比如说对发电机的一些攻击，电网的攻击还有我们像各种地方的攻击，像乌克兰电站攻击，所有这些数据造成损失是非常大的。你们可以看一下这些数据是非常大，像澳大利亚是一年整个网络攻击代价是45亿到140亿美元，在香港整个一年一个网络攻击损失是320亿美元，在全球来说2015年网络攻击是4000亿美元，这个就是给我们星球带来很大的问题。

我要跳过一些内容，我们是建立一个非常大规模一个数据集，我跟大家介绍一下我们如何实现目标。简而言之也是非常简单，你所做的就是先到那些黑名单，尽可能搜集这些黑名单，要搜集的越多越好，然后有一些比如说反病毒或者反这样一个恶意软件这些信息也增加进去，然后到这些平台当中收集，这些数据要反反复复增加，比如说我们会有一个荟萃数据BGP这样的路由器，这样一个数据搜集起来，关键就是我们要更进一步延伸。好，这就是我们所做的工作，我们一开始是从公共黑名单这里搜集起来，这当然是有14个或者15个这样一个公共的一个黑名单员，他们都会提供你在互联网当中比较有威胁的网站。另外我们还有互联网的报告，也看一下像这些公开黑名单资源一些历史，所以我们可能收集很多历史数据，后来我们产生这样一个报告单，比如说2007-2017这样一个报告清单，后来我们建立一个黑名单有很多也参加了这样一个恶性的攻击。如果我们想做的就是我们要拓展它，并不仅仅是在公共里，我们也看一下服务对这种在线的服务我们也进行这样的搜索。有这样64个在线网站病毒分析工具，这些工具都会有这样一个报告，所有这些都会告诉你对恶意供给IP在什么地方，另外还有一个时间，最后你会收集这些额外的IP这些IP也会给到这样的恶意的攻击，所以这些文件都要收集起来，像这样恶意软件或者恶意的攻击，或者还有一些参考还有推荐，所有这些都是有的时候有些开发者也会给你提供一些线索，所以你是所有这些仿病毒的技术你也要搜索一些。这些都是大量的信息，比如说恶意网络攻击的汇报都搜集起来，当然会有一些时间，我们所有的这些都是有价值的，我们通过这个分析之后就了解这些威胁背后的一个逻辑，以及互联网背后的一些问题，与这个不同威胁原因是什么等会儿介绍一下。

我们还做了一些其他工作包含，在针对这些威胁增加一些额外的数据收集来源，我们做的工作就是把路径源进行收集，我们收集了AAS，就是我们IP主机的位置进行场景开发，另外我们还收集国家的一些相关的信息。举个例子，刚刚是举例说明，差不多就是数据收集，我们进行收集之后我们还有一个域，我们通过这个域可以知道产生这些恶意活动的情况，然后我们把这些恶意的行为相关数据进行编码处理，然后登记在某一个地方，帮助我们了解发生的威胁，以及威胁的类型。最后我们的成果就是，我们收集了总计大概要5100多万恶意攻击报告，其中有一些有标签，有一些没有标签。其中有超过64的IP地址，其中我们在收集数据的时候发现一个问题，大概15%的数据都贴了标签。涉及到我们网络上的恶意的一些攻击行为，其中也涉及到我们专用IP和主机，所以这个讲的就是关于这个标签。这个标签流程也是非常重要的，可以帮助我们了解在发生恶意攻击行为的时候这样IP的情况。

大家可以通过这个图表可以看到，我们有（英文）占比是10.37，然后（英文）是64%，关于这个效率方面大家可能对这方面感兴趣，如果好奇想要看相关的内容，了解哪一个效率是最高的，我们需要采取一个非常有效的方法帮助我们更好的设置整个流程。大家可以看到这里有一些数据的显示，这里面关于主机受攻击的收集情况。我们可以看到通过有些报告，我们主机受到的攻击。大家可以看到在恶意攻击当中我们有多少台主机受到了影响，这个所讲的就是我们需要加强主机的专业化，我们需要专用的主机来抵御网络上的攻击行为。除此之外我们还有一个举证图可以部署的，基本上来说我们可以设定这样一个域值，机遇这个域值进行部署。

第二个方面就是通过机器学习的方法，其追有一些机器张贴了标签，可以帮我们对恶意攻击行为进行分类，另外我们这个学习的工具，我们也想拓展到其他的方面，大家可以看到它包含一些特性包括天、月、年，以及这些类型所对应不同数据是什么。大家可以看到这里列出了还有一些国家，哪些国家设定在IP地址受到了攻击，还有通过这样方式我们需要了解我们需要了解AAS，稍候我会给大家展示一些国家具体数据收集，以及恶意攻击情况。

前面也提到关于效率是如何，这个情况我就不在这里介绍了，我们总的来说会把这样的知识进行这样一个拓展，进行评价和验证。我们90%的报告都进行建模。我们构建了这样一个1%的跨度，我们通过这个筛选也发现了一些独一无二的IP地址，也发现了一些独特的现象、特性，大家可以看到这里针对IP的独特特性，其中有一些它的数据是非常高的，是超过60%的一个占比。另外我们IPS占的比例超过80%，其中有些IP地址在过去几年当中也受到了攻击。我们是AWS，这个是AWS所构建的一个体系，我们叫（英文）这是其中一个例子，实际上在有一些恶意攻击情况下，经常性反复性被攻击的对象。我们针对不同的国家，也进行了专项的分析，我们发现不同国家分布情况也是不一样，有一些国家受攻击比例是比较高的，是超过一万份恶意攻击报告的收集。这一方面美国和俄罗斯都做的比较好。

这一页所显示的就是我们叫做ASES系统，就是我们自动化系统我们发现有些是涉及到亚马逊，其中进程是具有侵略性的，还有（英文）这里面都是一些自动的系统。这个图表上都有显示，包括（英文）的一个数据。关于这个数据化的影响我也想跟大家提到一下。这是时间的一个演变过程，这个是基于我们所收集到恶意攻击报告所得来的，这个时间跨度是从2007-2017年，大家可以看到这个容量这个报告容量实际上是处于上升的过程，这个驱动因素是有一些，比较高水平FS还有exploit kit。在2008、2009年有学生跟我反映这个系统是有很好的工具键，关于这个（英文）还有一些恶意攻击活动过程都有在这个图表上显示。2009、2013年这两年是比较特别比较有意思的，这样一个情况实际上有由，大家可以看到在这个跨度上我们处理方式这个效率更高。我们更多接入了移动设备，移动网络，大部分的受害者大部分的IP，根据我们的报告就是在这两年的时候受到了网络攻击的影响，主要是因为我们接入大量移动网络、移动设备。近期大家可以看到，29%在2017年的时候这个（英文）所代表的数据达到了29%这就是我们事实的情况。

我们也按照不同的年份对我们受到攻击情况进行了分类，（英文）实际上在追赶，（英文）实际上也是。我们针对这一方面做了大量的工作，我们也开展了很多的工作，（英文）有一些实际上我们检测系统、探测系统探测不到的，实际上有些垃圾邮件没有纳入到我们黑名单当中，基本上来说我们对这个进行建模，构建一个数据链，构建这样一个流程。我们是可以针对不同量级的威胁构建一个模型，不好意思这个显示的可能不太正常。实际上我们可以测量这样一个高寿命，可以帮助我们了解有一些IP还有一些恶意的域，可以安装一些探测设备来避免病毒的攻击。另一方面对这个密度比较低的地方，我们提高这样一个弹性、灵活性，基本上来说我们可以在受到攻击以后进行快速的恢复。

根据IP我们对它的分布进行了总结，IP从国家的角度来看有一些特定的国家，无论是出于什么原因，可能他们会参加到恶意攻击活动当中。在这个方面我们基础设施就需要构建它一个弹性，保证我们网络的安全，我们也对其中一些信息进行了一些建模，这对我这些方面的看法。实际上我们通过数据的收集也可以发现有一些目标群体，实际上是受到了反复的攻击，比方说美国，平均受攻击时间是少于1.5周的，另外有一些特定群体目标群体是我们发现受到了反复的攻击，这是在美国比较常见的情况。这个对我们来说也是比较担心的，当然这里关于活动频率我要跳过，我们确实要给大家看一看，这个细节我当然也要跳过，我们是对这种恶意的攻击是进行了分析，我们主要把不同的活动进行分组，我们有的时候也有人会有一些硬核的这些。有的时候比如说钓鱼活动或者什么活动，这些都是恶中之恶的，他们会催生其余子图。有些具体的IP它会存在，哪怕你把它子的病毒或者是活动消除，那些顽固不化母的占点还是会存在，还是像僵尸一样存在，这也许是我们对这种调研活动我们要最终连根拔除彻底摧毁，我们就是要找到它的根源知道它的巢穴。

我们要找到母巢穴，把母巢穴这些罪魁祸首全部连跟铲除这是非常重要的。这是关于威胁量级的分析，这我跳过。我再跟大家介绍一下隐含的信任，这些代码或数据集在这里我们都有，是有一个链接的，大家可以其实经过这样一个链接进去看一看，比如说有的时候也会有些移动的活动，我们也在这个领域我们也有所涉及，大家够可以扫二维码用这个链接看一下，我们对这个研究数据集还有代码，还有更多一些洞察。这个也就是说我们是通过这些数据的分析来研究这个具体的一个威胁，

接下去我跟大家说，在网络当中一个或者是暗示性的信任，或者是盲从的信任，在这个方面很多人都发表了论文。有说在我们各个大会当中人们都进行了这方面的研究，这是是BBC.COM，这是大家所喜欢的一个网页，你如果是载入这个网页。当然有的时候，如果我访问这样BBC.COM网页最后确实我们访问这样的数据，是4个不同的国家还有9个网络，我们发出了172个这样的请求是进入4个国家9个外网，仅有14个是来自于BBC.COM域的，有35个是外部域的，有19个我们叫做这种外项型，有的是你比较喜欢的网页，有时候你当然可以说这是一个非常明确的信任，当然除了明确信任之外其他都是涵饰的信任。我们发现在BBC.COM下面会有好几个分支，有一个（英文）还有一个（英文），（英文）下面还有（英文），还有GST.COM，还有（英文），这样一层一级我们对各种不同依赖级都仔细分析，特别是分析这样含蓄性的信任，我们在20个网站当中，其实80%这样的网站都会创造这样一个信任链。80%他们都是会包含这样的一种含蓄的一个信任，比如说这种模糊性或者是那种灰色区域，你真的不知道到底出现什么样的情况。有一些极端的情况当中我们也会找到一些网站，是在20万个当中，有的时候它甚至达到38个层级，各种各样的一个资源一个请求，反反复复等等。所有的这些网站它其实都会从外面，比如说在三级或者是三级以下，他们都会引用外部的资源这种载入。从安全角度来说，有的时候外部进行载入往往会有敏感的信息，比如说（英文）这些脚本或者是有的时候，往往是针对你这种自信方面，有的时候有些含蓄性的这样的信任，可能会值得怀疑。我们要找到到底这些含蓄性的这种信任有多少是恶意的，当然我在这里说的是这个论文当中非常细节的东西，有6%这样一个份额，2.4%其实是比如说请求一些怀疑是可疑（英文）的脚本，我们有时候甚至很有把握的说这是一个恶意这样一个域传来的东西。当然相关这些数据代码在这个ULL，都有大家可以阅读这篇论文。

非常感谢大家的聆听，如果有时间我会接受提问。

主持人：谢谢，我们有时间接受一个提问。

提问：做的非常棒，我在这里有一个问题。就是建立你的一个数据集，这个问题就是并不是有很多这样标签的数据，并不仅仅是我们域的安全。我们的社区如何变的更好，是不是有一些教育可以变得更好？

DaliKaafar：可能有几点我能分享，谋划从一个点你要经历非常痛苦的过程，你要做一些人工的一些处理，比如说你希望要花一些时间，花一些这样的力气，来得到一些比如说方法，比如说这个是真正去做一些，当然有的时候比如说众包的数据并不一定真实，你要自己验证。有的时候机器学习也是一个工具，但并不是最好的工具。我们确实注意到了有的时候从IP的角度在某种威胁参与过程当中，也是非常复杂，有的时候很难捕捉到这样一个标签。这种特性当然你可以知道它有这样特性，但不一定保证它就能威胁你。

主持人：我们比预期稍微晚了10分钟，现在10：40，我们10：50准时开始，接下来会有来自日本国立情报研究所季老师，还有亓老师分享区块链方面，还有最后一位关于网络测量，我们10：50准时开始。

（中场休息）

主持人：我们继续开始，我们有请季教授给我们做报告。

Yusheng Ji：很高兴有这样机会来到这里，今天我跟大家分享的就是关于SINET信息，这是一个日本学术一个平台网络。希望我的演讲内容不会让大家感到失望，我的演讲主题是人工智能驱动网链，简称SINET5，我是来自日本国立情报所的，我的名字叫Yusheng Ji。

我今天的汇报内容包含以下几个方面：首先我简要介绍一下NII，也就是我们组织机构情况。另外我还会跟大家介绍一下SINET演进过程，另外分享一下SINET使用案例，以及SINET网络架构情况，我会简要跟大家分享一下，然后还会介绍一下SINET服务，其中包含我们移动平台，这个移动平台也是近期项目SINET用户推出的。还有最后一个方面就是下一代SINET6时间安排和计划。

首先介绍一下NII我们的NII全称是日本国立情报所，这个是由教育文化体育科技部投资组建一个机构，它位于日本。它承担两个主要的任务，第一个方面就是研究任务，研究和教育任务。

第二个主要的任务就是服务，提供服务。关于研究方面，实际上NII开展的工作包含情报学方面的一些研究工作，这个情报学从最基本的情报学延伸到我们实际问题的研究解决方面，另外我们还有一个项目是针对这个模式教育的一些学术方面的方案，大家可以在下面圆圈当中可以看到，左侧Reserch研究我们分成了不同的部分，首先第一个我们主导的原则，我们第一个部门是关于情报学原则研究，主要是解决情报学的一些基本性的问题包含计算、运算，还有一些理论的一些分析。这些都是针对情报学方面的，第二个就是我们情报学研究部门其中也涉及到计算机还有通讯还有硬件以及软件。

第三个部门我们称之为数字化内容和媒体研究科技部门，有点像我们的信息归档，以及涉及到应用方面内容，还有AI应用内容。

第四个部门涉及到信息科学，它的名字叫信息社会研究部。最开始我们在实验室构建了一些网络，我们也有人参与到这个网络研究当中，涉及到安全问题还有一些政策的制定。相关的研究工作就是这个部门来完成的，除此以外还有这个叫我们的研究生或者硕士研究生，博士研究生一个教育项目。或者就是硕士研究生毕业之后将会接受三年博士研究生受教育过程。第二个方面就是我们所说的服务右侧，我们为大学还有研究机构、学术机构，所有日本这些机构我们都会向它提供服务。在右侧这个圆圈当中可以看到我们SINET整个信息网络构建，除此以外我们也还提供这个云服务还有网络安全，还有ID方面的服务，还有信息流动传播等等方面的服务。

今天我主要跟大家介绍的就是SINET。这一页上面所解释实际上更浅易圆圈内容相似，只是解释方式有所不同，在NII内部，大家可以看到在下面我们有这个网络，就是我们SINET，实际上我们这个SINET在全国范围内提供100个Gbps（音）骨干内容。我们还有国际线路，有通往国内、国外、亚洲，此外我们还有教育服务方面的服务。

下面这一页主要介绍一下SINET的历史，在最早1987年这个网络和大学进行连接，当时我们构建了一个叫数据包交换网络协议，大家可能也有所闻。这个就是用来帮助我们和大学之间构建这样一个连通。1992年的时候SINET采用了互联网的协议，然后从此以后就成为了因特网的骨干组成部分。那个时候是它的速度是6-50兆比特一秒，后来又是要更快。在2002年这是我们的SINET第二个版本，我们把它叫做（英文），它和SINET是平行建构。这个提供是10级BPS这样一秒这样一个速度，我们当时认为这个带宽也应该是非常宽的了。SINET它缩写的意思就是学术信息网的意思。这个SINET不断一个演进，2007年是SINET3，这个SINET3速度达到40GBPS速度，它成为混合型骨干的网络。2011年是SINET4，它是涵盖了日本的所有的县，当然日本的县其实就相当于中国省这样一个行政单位。所有日本省都受到了SINET4覆盖，现在已经是SINET5，它是大带宽达到100Gbps的带宽，Gbps就参照比特每秒。现在我们也将会上升到400Gbps这样一个带宽。

另外我们除了我们SINET5，我们还有一个移动的平台，这是在去年低的时候我们进行推出的。还有一些我们安全服务等等，关于SINET5这当然是我们一些服务。我们想超过900个大学和研究机构提供服务，包括日本国立大学还有各个省这样的省级这样一个大学，另外把这个服务提供给一些私立的大学，当然还有其他一些研究机构。其实我们来自单位也是其中一个大学间一个研究单位。我们总共把服务提供910个机构，总共有3万个用户使用我们的SINET。

这是SINET5国际线路，这个是从日本一直连接到了欧洲，然后再到纽约通过洛杉矶到纽约，甚至连接到了新加坡。现在我们是在2019年三月份它升级到100Gbps这样的速度，现在我们也有非常好的韧性回路。这个韧性回路就是一个备用的一个，这当然也是美国国家信息架构，和其他一些网络进行合作建立这样的备用回路。这是它的回路，这是包括了我们南半球、北半球整个速度是100Gpbs网络速度。

这是SINET5的一些节点，在我们的右边这是一些节点，所有节点都是在数据中心，在左边这是其他一些组织，比如说大学组织等等，这些大学组织其实它就是我们SINET服务一个定业方。这些用户他们当然自己准备他们一个接入线，然后接入到数据中心。因为现在我们SINET已经升级它一个网速，已经是100Gbps了，很多大学还有研究机构他们也有一些大量的数据，他们也把他们的一个线路速度上升到了100Gbps。

像这个服务SINET其实也和云服务直接相连，包括和亚马逊、微软、NTD等等有29个这样云服务中心，所以这些都是和SINET连接起来，提供高性能、安全服务连接，服务于大学的用户。这是云服务，所有云服务平台都和SINET相连接，然后再给大家提供服务。这是一些使用的案例，一个使用的案例就是高能源物理，比如说加速器也产生大量的数据，当然它需要高速的网络来互相使这些加速器互相连接，这当然也是一个国际合作的项目。他们用第二层或者是第三层VPM的服务来互相连接来传输大量的数据。

第二个使用案例是高性能的运算，有九个日本国立大学，他们都有超级计算机，他们也用SINET使彼此互相连接，传输大量的数据。在这个图片当中你可以看计算机叫K，这当然是好几年之前叫K计算机，几年之前还是全球第一的计算机。

第三个使用案例叫做融合的一个科学，这也是一个国际协作的项目。这个是从日本到欧洲还有法国，通过SINET使这些机构连接在一起。你知道在日本很多地震，为了更好感觉到这样一个地震，所以在日本有很多的传感器的一个布置，很多都是在靠近海岸线上面，有一千个这样的点，其实都是通过SINET连接在一起。通过和我们SINET相连，像这样地震学方面的数据，各个科学家之间都可以分享，一千个地震感应电的数据都可以连接在一起，科学家可以通过这些数据连接来精确这些地震的情况。

还有更多的例子，另外一个例子就是天文学，在天文学方面我们是有天线来对空间进行观测，为了有更为准确的观测，我们希望很遥远的地方也架起这样的天线。在各个天线之间还有非常大的距离，所以通过这样我们有一个虚拟这样的非常巨大的一个天线。通过这些天线相距非常远构成一个大天线来观测。在这些空间我们也有非常快的传送速度，基本上可以8.4Gbps传输速度。

另外一个我们叫测地学，测地学我并是很熟知，这是国际协作的项目，这个项目主要是来监测地球的一个情况，就是地球怎样进行它的旋转，还有这些大陆地区它是怎么样移动。其实在地球上不同的地点都可以对地球一个转动进行测量。

第七个使用的案例叫Alma Telescope，在南美沙漠里面有超过60个这样高精度的天线，他们是建在一起来形成一个非常大的一个天线，确保有非常好的一个观测。

第八个使用案例，关于卫星的，这些卫星它们都发射空间。（英文）现在是离开地球有两千英里，它们是收集数据，是要60分钟才要让光能够传输，这当然是非常远的一个距离。这个是和德国一个合作项目，从日本把数据送到德国，再由德国送到空间去。还有空间观测项目还有另外其他一些项目。这个比如说像医疗科学方面的，像医学图象数据的分析，像这些医学图象，各个不同点可以分享特别是在日本很多的大学，他们都有自己的附属医院，像医学科室他们都可以来分享这些图象，他们使用我们的SINET通过SINETL2VP7的服务来进行分析。还有远程诊断使用8K的视频，这个也是使用SINET传输8K的视频，把这个视频从一个点发到另外一个点来进行疾病的诊断。另外一个是进行远程医学教育，还有医学资料备份，大学当中也是这样做，还有其他一些医院。一般来说病人他是去不同的医院，这样的情况下面其实不同的医院，其实他们应该有很好信息共享，万一出现紧急情况，另外这个数据也是隐私性，也比较敏感，另外我们也要确保信息保密，这也是非常重要的项目。

然后再跟大家介绍一下SINET5的一个网络架构，这个当然更加技术性，一般来说是有三个层次在底部这是一个光纤传输层，还有数据包传输层。在SINET其实有三个层面，这三个层面其实是全网格的IPMMS。这是一个传输层这是ROIDM，也就是从光学，从光学纤维转化成电信号，还有IPOS它是互相直接的连接，是没有IP路由，在里面没有IP路由，减少延时，数据包并不是到路由器。那么其实在路由C当中会绕过这个路由直接到达路由B，这个是更好一个延时的表现。在这层再使用我们切片服务网络，是有五个这样一个切片在网络当中。这个是跟我们一般的网络的层级是不一样的，我们有LV3是一般用户都能使用到。第二层级SINET是有专门项目当中才能使用到的，在这个层级我们学校的校园也可以进行实现共享，除此以外还有我们第二层安需服务。在第二个层级上我们用户使用者可以进行，可以自定义，以自定义的方式对这个服务进行配置，除此以外还有SDN服务，这就是我刚刚讲到五个切片，当然这个当中都需要容量非常大的路由器。在大的路由器内部当然会设置不同逻辑路由器，除此以外还需在联网的时候，路由器进行联网的时候还需要提供针对性的互联网服务，可以确保我们这个网络是独立的。所以为了实现这一目标我们想增加新的切片，可以实现对网络服务的一个重新配置。

另外一个例子是关于大容量数据传输，我前面也提到了，整个大的科学范畴底下，我们需要传输大容量的数据，比方说在这个高能阶段，高能领域我们需要，我们在Backup个城市构建了一个高能研究机构。每秒钟这个数据传输量都达到了30Gbps，都是通过SINET从日本传输到美国的，大家可以看到中间数据量有多大，大家在右侧可以看到我们高性能运算数据备份。这个超级计算及需要对数据进行备份，它的速率可以达到90个Gbps。就是考虑到这些方面的需求，我们想对整个网络系统进行升级，把这个系统升级到400个Gbps。

除此以外SINET5还有一些新的服务内容，比方说第二级的安需服务刚刚我也提到了，另外还有一个虚拟的校园，除此以外还有DDosmitigatian、（英文）缓解，还有（英文）实验台。前面我也提到了我们的目标是，把光纤连接提高到400个Gbps每秒，实际上我们在东京和大阪之间数据流量超过了400个Gbps，我们计划在年底对整个系统进行一个升级改造。这个项目实际上是新开展一个项目，这个是SINET5一个移动平台，截止到目前为止现在SINET是固定电的连接，现在的连接是一个固定式的与用户实现互联。所以我们计划构建这样一个移动式的平台，我们通过移动运营商，大家可能知道日本有三大运营商，通过这样移动运营商提高他们容量，通过VPN再连接到SINETI，通过这样一个平台可以搜集物联网相关一些数据，这些就是我们正在开展的一些活动。

下面这一页讲的就是，我们这些服务实际上是从2018年12月份开始落地的，我们已经开始了38个试连续项目，这些服务覆盖面包括农业，还有医疗行业、养殖行业以及其他社会各个体系、各个部门，这些都是我们提供一些相关的服务内容。除此以外我们还有厂家，也就是提供数据分析工具这些厂家，提供这种数据分析平台和数据分析的工具ITT、KDDKI、还有亚马逊还有微软等等，这些都是我们合作的厂家。

这是具体情况介绍，下面一页针对移动领域大家都知道，我们即将进入5G时代，我们的SINET系统，也计划为我们用户提供5G服务。实际上5G也会在明年构成我们服务范围，也会我们奥运会提供一些服务。接下来的计划就是针对SINET6的，目前我们使用的是SINET5。我们规划是在三年之后推出SINET6，在SINET6当中我们目前只是粗略构建这样一个模式，我们目标可能是达到100个Gbps，以及提供5G服务，提供无线的通信连接，以及有线连接，谢谢大家。

主持人：现在有时间提一个问题。我有个问题这个问题是关于构建这样一个商业模型，完全是由政府出资建立的吗？

Yusheng Ji：是的是的，是由政府出资建立，这个服务是免费的，这个服务是针对所有大学免费提供的，我也发现这个技术也用在我们医疗行业数据传输等等。这些都是我们商业机遇，现在也是免费使用的吗？目前为止这些使用都是免费提供的，实际上来说我们也希望把这个拓展到高校，拓展劳动学生用户当中，其中有一件我们需要了解的，因为我们是隶属于教育部门的，你刚刚提到这个是商用，以后也是有这方面的商机的。

主持人：我对刚刚您介绍整个流程和进展印象真的是非常深刻，谢谢。

我们下面演讲嘉宾是来自北京邮电大学的亓峰老师，今天他给我们报告内容是基于区块链去中心化可信命名服务技术及应用，有请亓峰老师。

亓峰：很抱歉我用中文讲，我PPT都是用英文，我本来翻译成英文好了，只能麻烦咱们同声传译努力了。

今天给大家汇报五个方面，一个是区块链，这个是最大话题之一，也是最激烈问题之一。我们并不研究区块链技术本身，包括区块链在业务工作上面应用，主要研究区块链如何提升安全和可靠性方面的工作。

基于区块链做了四个中心，第一一个是去中心化互联网命名服务平台，这个是在2013年就成立了互联网服务社区，研究基于区块链命名去中心化，发现很多有价值出现出来我们把它应用到下面三个领域：一个是互联网，互联网一个很重要的就是域名的解析，大家从前几天新闻也知道俄罗斯已经在做这件事情。第二个方面是未来网络领域。未来网络学术他的比较多，但落地比较很困难，我们认为其中一个重要的问题不是数据层面上的问题，寻址的问题没有解决。第三个物联网应用。

第一个简单介绍一下区块链它的价值，区块链的价值刚才讲了区块链是最后争议的信息之一，无论是在学术界、商界、政治界议论都比较多，很多评价也走向两个极端，有的认为它很好，有的认为它一无是处。但是区块链里面有一个共识是大家共同认可的，区块链是通过技术而不是管理方式来实现了可信，它本身就有很多安全的问题，它也没有解决安全的问题，但是它是基于不可信、不安全网络上实现的可信，而是是很警惕实现的这种可信。

现在区块链很多可能很多是应用领域噱头，我们个人感觉它真正发挥价值地方至少有这么四个必要条件，第一个是有新的数据需求，就是我自己数据不够或者我用别人的数据，或者我自己的数据还想跟别人共享。

第二个需求，多业务主体。再一个是彼此不互信，还有一个业务是强相关，我既要给别人业务交互又不信任它，这四个条件用区块链才有意义。另外现在很多争议最根本的问题，对于区块链根本就不明确。

区块链有三个理解：一个是区块链技术，第二个是区块链跟互联网一样，大家知道互联网最多应用其实不是在互联网是在企业网，或者是在国家网络层面上，所以技术和网本身是两件事情。还有一个就是区块链的思维，我今天仅仅谈第二个，我只谈区块链它的应用。

区块链应用必然要涉及到区块链这个生态，这个区块链生态，我现在是画了简单一个图，是有两个边界。第一个边界是内部的边界，这紫颜色的线代表共识，有三个：一个是所有的币。第二个是共识方案法。第三个智能合约。另外币要想用必须要跟外面世界产生交互才可以，基于这个生态下其实最典型应用有三类，第一个叫数字货币，这就是比特币这就所谓的币圈，第二个是数据存证，大量应用系统都是在这个层面上在做。

我今天讲一个问题就是数据的命名，就是数据在外面我是把名字放到链上，为什么这样做？是因为区块链本身事实性或者它的容量是达不到，现在网络需求。如果把名字放在链上可能达到数据性和要求。

基于这个思维我们在互联上构建了一个PPK社区，P2P来（英文），基于区块链做了一个去中心化命名平台。这个平台比较简单，我们同时基于供应链联盟链。供应链它可以在整个互联网领域里面来做到标识唯一性，而联盟链可以作为一个组织内部来做到标识唯一性，这样就给咱们互联网所谓的公网或者是内网就可以相映射了。第二个层面是构建了根标识对接网络和扩展标识对接网络，当然他们两个分别基于公有链或者是私有链。另外做了一个ODNI标识协议层，对外提供两类服务，就是标识注册管理服务和标识解析服务。基于这个平台上面后面可以做很多很多应用，像浏览器、物联网、数字资产、数据共享等等等等。大家有兴趣可以看下面这个网站，相关成果都在这个网站里面，就总而言之这个平台是提供了一种旺多环境下自主命名标识和交换数据内容所有开发数据的一个系统。

这个是ODIN标，上面是静态的命名方式，下面是通过安全可以命名标识的结构。基于这个命名服务我们可以把标识做到唯一、可信，映射到现有不同的服务，再一个可以把标识比如说PPK映射到外部网站，可以做到唯一的映射。也可以映射到我们自己的一个内容服务，也可以映射到（英文）服务，都是可以唯一映射过去的，这样保证了它的可信性和唯一性。另外这个服务是基于区块链的，如果大家还认为区块链可信也可以把标识映射到本身上甚至合约上都是可以的。

目前这个社区注册数量在2018年从100，现在已经注册数量已经跃升到三万了，看来大家对这个比较感兴趣。它解决了两个问题：第一大家认为比特币可信你服务它就是可信。另外一个它免费，如果把我域名地址放在我服务上之后，我不用购买所谓域名服务商每年租金了，因为这个能保证我是唯一可信的，而且不会被别人盗用的。然后ODIN有这么几个特点，最根本就是唯一和可信性。

下面我来讲ODIN在其他领域的应用，第一个是互联网，互联网目前无论是除了我们网络大国之外其他都想做一进事情，就是去中心化。因为现在的互联网域名从技术上讲，它构建了三个中心，一个是命名中心、发布中心、解析中心，运行技术是可靠没有任何问题的，现在问题是网络大国在控制这一件事情。从技术上来讲，我们用户看到域名解析服务是通过递归解析来做。首先做一个解析首先要到根上到中间层再顶层来做的。

这样存在三个方面：第一个是解析中心化带来的，如果解析无法通达就带来孤立式的风险，它根本的问题就由于每一次解析先找根，如果网络问题断了会导致你国内域名都访问不了了。第二类风险也是由于解析中心化带来的，如果解析服务给你停止了它就会造到致盲性的风险，这个风险历史上也发生过，然后根本原因是跟前面一种风险有点相类似，但是不是由于网络的问题，是由于就是这个服务所导致的。还有一个是由发布中心化带来的叫消失性风险，如果就是说美国它把你的数据给你改了，那你肯定就访问不到了，而这种事情也发生过了，在伊拉克战争就发生过这样它就把数据给你改掉了，现在已经不是在谈理论了，实践当中已经出现了。

第四类风险是由于发布中心带来的，如果它把它的数据改成让你感觉不到，你能访问到最终目的地，但是中间经过一个转接，转接中间代理再给你访问四级网站，这中间所有访问网站是一览无余，这是一个致命的风险，我们交劫持性的风险。最根本的问题是四类风险怎么解决掉，用中国的话就叫“大土豪分田地。”

用区块链怎么解决问题呢？先说在根域层面上，让各国共同参与构建一个可信的共识链，让根都放到这个链上，这样一个任何国家独自改了无效。第二个是构建权威，权威链是企业在注册自己域名的时候能不能建全球一个链，这样我自己管我自己的，就是说别人改了是不行的，用这么一个办法。这么一个办法你看既修改了域名体系中的运营和服务这两个环节，其他环节是不需要动的，甚至未来的链和现有的系统可以并行来运行的，相互访问都是可以的，他们之间的接口没有必要改变。最终构建层基于区块链域名可能是这样一个情况，就是右面对于根有一个共识链，对于权威有一个更大的共识链。对于权威来讲数据大概要到百T级，所以需要大的一个存储系统。这个项目不是我们自己在想象，今年国家科技支撑计划立项已经开始在做了，据说俄罗斯已经变成现实了。

再一个是具体的一个技术实现的方案，因为主要是在权威域名方面，它的数据链太大，如果你把数据放到链上是不可行的，所以就需要我们命名服务，然后把代表域名的命名放到链上，而域名数据本身是要放到下面网络存储上，比如说用RPIF都可以，这样可以包括数据的融云和效率。

第二个是未来网络，未来网络是很好的一个概念，而且它最根本的一点，我先把未来网络狭隘成RSN，RSN是以数据的名字来寻址如何避免互联网导致今天的状况，就是你的域名可能是不信的，如果这个解决不了的话，RFN推广是很困难的。既然RSN核心就是数据寻址那数据命名服务安全性、可靠性是最重要的。关键是它的可信性，而区块链本身天然具有可信这样一个特点，所以说用区块链来解决RSN寻址的问题这可能是好的办法。刚好我们做了SINET这一套体系就很容易筛到RSN这一套来了。

它跟RSN想结合好处有粮店，第一点RSN能解决实时数据问题，区块链刚好实时性又不足，但是如果把数据的名字放到区块链上，数据名字的变化是很平凡的甚至不会变化的，这样就用数据传输这样就解决了，区块链不能存储太多的问题。同时通过对加密技术的应用，特别是是非对称加密技术应用，能在RSN网络上能构建出咱们现在（英文），这样可以解决数据分发加密性的问题，这样就构成切片式的数据共享的一个服务。

再一个是具体构建方式，就是ICN为数据提供了高效、快捷的共享网络体系。数据及设备命名与寻地址是ICN网络的核心服务、区块链可制成安全。

再一个是具体一个实现方式，这个实现方式它面对三种应用场景，一种应用场景是在局部范围内，RSN如何安全可信、甚至切片化来实现数据共享。第二个场景是在全网范围内，跨业务主体如何实现数据可信共享。第三个应用场景是如何解决数据安全来进行共享。这里面核心是密钥分发，如果这个解决了可以使网络安全可信。

第三个就是物联网，物联网本身核心要解决也是数据共享的问题，如果把数据都到系统来再进行共享有几个问题：第一个数据处理之后你的数据不是原始数据，你可信性就出现问题了，如何处理原始性数据就能被不同设备共享，这样也解决了这个问题，刚好基于区块链命名就能解决这个问题了，这样是大概是基于RSN和命名一个功能结构。核心是网络结构，中心是解决通讯问题，外皮是解决命名问题。

这个是基于具体区块链的实现，这是它物理的结构，如果要是说在咱们RP网络上加上一层RSN需要这些设备，首先需要有网络节点，它基于区块链和RSN。所以区块链可以终结在终端上面。目前已经在几个领域做了一些实验，这些不是我们的实验，是欧洲（英文）地带实验，它每用区块链，但是它用RNS已经做到了电网的实时性，我们现在正在跟国家电网做这个实验。第二个是跟国家电网一起做的，就是解决太阳能发电上网补贴的，现在补贴要经过电网经过政府、经过银行等等几个环节大概要半年多补贴才能发到太阳能发电业主上面去，而用这个数据直接上链之后可以在原始数据共享，所以任何一个部门都可以不用跟其他部门协同自己就能保证数据的可信，开展工作，发放可能就几天，或者实时发放下来。

另外还有一个应用场景就是家庭用电数据的采集，采集完之后可以可信共享，比如说电网可以用，我的电器设备生产商也可以用，用它长期观测它设备的运行状况，以及对设备性能的规划，我汇报的就这些谢谢大家。

主持人：谢谢亓峰教授精彩的报告，我们可以有一个问题。那我问一下，您前面讲到域名的时候，注册那个都是可以那样，就是在解析的时候怎么来在区块链这样情况。因为现在ODIN解析不是简单的映射，因为它还有布到各个地方我怎么能域名，同一个域名不同的地方，您在解析的时候用区块链怎么实现，类似于负载均衡。

亓峰：说到现在国家研发计划里面比较重要的一块。有两层，在根域名，根域名这个问题不存在，主要是在权威层。权威层本质是它域名数据记录集里面，就已经像你说根据各地或者根据它在什么情况下应该解析成不同的地址，已经把这个数据放到里面去了。解析是很简单的，核心是域名数据的你怎么来制作，或者怎么来存放，核心困难在这个地方。由于数据存储是基于网络的存储，如何来保证你的数据存储在，你的域名数据存储在就近的地方，这个也是现有数据存储里面没有彻底解决的问题。你像RSN不管你，你直接上网就可以了，但是在域名解析不行的，你要根据几个维度存量来。另外要确保数据同步性，这个是想办法在解决的，就是可控采用比如存储编码的方式，来解决域名数据的存储，的确是一个问题。

主持人：谢谢，谢谢再次感亓峰教授。我们最后一个报告是湖南大学的谢鲲教授，我们下面有请谢教授分享基于稀疏感之的大规模网络测量和异常检测。

谢鲲：非常感谢邀请，我来给大家介绍一下我们近期的一些工作。我主要是围绕着大规模网络性能测量和异常检测做了一些工作。我的汇报主要围绕三个方面第一件事情就是咱再生大规模网络情况下如何进行网络部署，如何进行低开销测量方法研究。第二个是当我们采到数据有数据缺失如何进行准确数据检测，第三个就是数据高精度检测。

首先讲第一个，如何保证网络测量准确性的同时，降低网络代价提高网络效率成为一个突出的难题，当网络越来越大有很多节点的时候，尤其是面向大规模的网络，比如说SDN我们需要有全网状态信息，才能够做比较优化路由策略和功能能量策略。我任何一个节点对之间进行测量，如何说我测量都需要这样的性能，这样的代价测一次SN（音）平方级。当网络需要持续不断测量，这样测量代价就更高。我们知道主动测量，测端到端的延时它通过主动注入数据包来实现。对于这个吞吐量可能需要压路数据包，压路数据包我们发现不光是对，对网络造成很多的影响，本身这种测量的方式就会使网络性能变差，因此我们就在想如何来降低测量，我如何拿到全网网络性能数据同时，来降低网络测量代价呢？如何进行网络部署呢？这就我们所关注的问题。

现在我们发现在数据处理领域还有信号处理领有有稀疏感知技术，稀疏感知技术比如说压缩传感，比如说这个数据本身是稀疏可以通过少量测量来找出少量的数据，在举证填充还是扩展到（英文）。张量填充同样的，传统的应用都应用于异向数据处理。我们想是不是可以通过这个数据来做低开销网络来做一种手段。在上面网络图如果我要测端到端的延时，这里有六个节点，我建立举证，这个举证行就是原节点，列就是目标节点，是不是可以通过测量少部分的原节点，对性能来推测出剩下的性能呢？如果可能的话，就对我们低开销网络部署带来了可能。我们在想这些技术的可行本身是低质是稀疏的，我们网络测量数据、网络性能数据是否有这个特点呢？确确实实有这个特点，我们拿着现在网络性能数据，这个是端到端的延时数据，我们建成类似这样子的举证。我们来看举证是否具有低质属性呢？确确实实，我们发现前面少数举证能量就占所有数据能量，就说明这个数据具有低质性，低质性就给我们使用稀疏感值提供测量，提供了前提。当然这种挑战是在于，虽然我们稀疏感知技术可以通过小量测量推，但是网络测量是在线的过程。我在拿到数据之前我是不知道这个数据本身的分布的，所以我们举证填充有很多的理论，给到一个举证然后给出数据分布或者给出质的表征，我告诉你罪与预值测多少就可以恢复了。我在数据没到来的时候我不知道数据是怎么分布的，我很难知道测量到哪做的这种推测。

虽然我们知道可以用稀疏感知来做这个东西，但是在技术和理论应用上都需要突破。基于网络测量提升填充来进行网络测量的话，首先第一件事情我们知道举证填充，当知道它一部分信息比如说举证质，我们可以推测储它需要最低测量数量是多少。在测量过程中我们知道，我们不知道，我们只知道是已知测量数据不知道数据的分布，因此很难通过举证的质来推测出质量。

第二件事情，我们知道当这个测量的位置不同的位置它对数据最终恢复的对策是不一样的，我们在线测量的时候我们并不知道数据分布，也不知道哪个位置是有效的位置，所以这个都给我们带来了挑战。我们需要确定它准确更好的测量位置和最少的测量数量，来降低整体的测量开效。

我们在这方面做了一些前期的工作，首先在2015年的（英文）我们第一次提出来通过性能填充来做网络测量。当测量点部署随着测量（英文），随着测量点的增加它会，这个误差会越来越低。当超过了一定的测量数量的时候，再增加对于测量恢复结果精度影响不大，因此我们在想我们能不能找到这一个点呢？我们就发现提出来一个定律，我们通过前后两个虚拟测量，来估测两次虚拟偏差，如果两次偏差推测这个差不多的话，我们用理论定理，这个时候第二次后面这一轮就可以停到。我们通过提前测量停止，我们来降低测量代价。

右边也是关于我们网络测量，数量确定一个相当一个工作，我们发现了，我们如果说过于历史的数据，我们在做推测的的时候事实上觉得没有什么效果。因此我们采用滑动窗口来测，我们发现前后两个滑动窗口它的质，它的举证质具有强相关性，也就是说前后两个滑动窗口我们发现它的质就是增加和减少1，我们从定理上证明了这个，我们就基于这个发现我们提出了基于全向学习的方法。如果在前面窗口能够确定我测量数量，我后面可以基于全向学习来学习和初始确定我测量情况，然后我们再通过一些学习的方式来确定这个时候的测量的最终数据，数量来减少网络测量，减少整个测量的部署。

关于测量位置的确定，我们也提出了基于信息量测量位置测定方法。通过在动态过程中来比较前后两次恢复结果的偏差，来选择偏差大的地方，作为它的测量下一部分测量位置，还提出一个相对简单的测量柵格划分的测量位置调度算法，我们利用举证填充，举证填充性格就是当某一行或者没有一列如果没有数据，那这个很难推算的。我们就将整个举证面划成了很多个区域，通过不同颜色区域之间进行简单调度的方法，来减少这个每一行每一列没有这个数据采集到的可能的这种情况，从而大大来提高恢复的精度。

再近期我们又提出一个，前面讲的工作是在举证内部，我是决定哪个位置进行调度，哪个位置进行测量。在最近我们的工作，好，就是说当我们前面已经拿了一些历史数据，在网络层两中我们很可能对未来一个短期进行规划。我在未来短期我如何来确定，未来短期我测什么地方，来保证我推测准确性呢？这个就是我们最近提出基于主动采样测量调度算法，也是举证的方式，但是我们这一次可以做到了不是在举证内部，我们可以做到举证外部。如何来短期进行调度，通过历史的学习，来保证我调度测的位置，可以能够准确恢复我全部的数据，来减少整个测量的开销。这一些是前面这些工作是我们关于测量调度发放的论文，然后在测量数据分析的时候，我们发现当我们把数据采集到了之后，第一来是我们根据采样来获得数据的，第二件事情在数据传输过程中，可能是用UDP传输，传输很可能出现数据会丢失，不可避免数据丢失在很多网络应用中我们需要完整数据怎么办？我们就发展把数据推测出来，填充出来。数据如何建模？如果建模成一阶的销量的话，那么可以知道它只包含了一维的信息，如果建模成二维的举证的话，它可以包含了行和列之间相关的信息。我们知道网络流量天和天、周和周有相关性，因此我们将这个网络性能数据监测一个高阶的张量。我们可以想到基于这种建模本身对于信息的挖掘，对于信息的表达是更准确的。

因此我们在这个基础上做填充算法精度会更高，或者对我测量的要求就会更低。这个时候我做测量分析的时候，我们确确实实就是围绕这一块儿来做，我们首先，这是我们在2016年第一次将张量填充用于做网络填充数据分析上来。我们把网络数据建模成张量，当我们建成张量之后，一定相比于之前举证精度更高，但是相比于居轴集成代价会更大。我们在想我们肯定要求它得到更高的精度，但是集成代价高我们希望相应的减少。我们对比前后两个时段张量填充和分解的过程，我们发现在前面一个时间这个数据到来之后，当后面数据到来之后我们整个张量分解是要求出用于张量分解的参数。如果我在后面时刻张量分解这个参数能够通过前面时刻的结果直接计算，而不是通过完全重新来分解和迭代计算来进行，这样我们代价会大大降低。因此我们这篇文章就是将前一论测量结果、数据分析张量结果，到后一段时间到来新的数据，直接通过演算得出新的分解的结果，我们在持续恢复过程中我们计算的代价降低。

但是我们在很多时候，我们就是第一个在前面用2016年发现用张量来处理网络之后得到精度提高的情况下，我们在想张量一定会提高网络会得到更多的精度。我们想第二个问题难道每一天的测量它的模式是一样的吗？也许我们在实际网络测量开展过程中可能是说比如说我前半天五分钟测一次，后半天八分钟测一次，这个时候就会导致我这每一个举证面举证的列数是不一样的，它们就那中间那个图是测量的轴，每一个点就表示我开展一次测量，它的点数在每个轴上发现是不一样多的，就到证据证面大小不一样，我们知道用张量来建模数据分析精度会更高，但是是不规则张量，虽然我们知道是不规则的，但是我们知道网络特征是这样的，比如说我上午九点钟和第二天上午九点钟，我们测量数据本身规则还存在，我们讲数据在物理时间轴上进行划片，划成多个内部是这个均一分布的测量，但是每一个物理轴上进行划片，表示每一个小块儿物理时间和上链是对应的，通过这样子划分之后，采用数学方法将一个不规则张量重新变形，变成地下规则张量。基于规则的张量进行填充，获得推测的数据然后再反变型回去得到原始的数据，这样子我们就提供一个在不规则张量情况下，变换说明情况下，一个张量填充算法，我们可以让使得在我们在实际测量中能够用上这个张量填充。这个是我们2017年的工作。

在前面东西我们都考虑在王者数据情况下推测全部的数据，在一些真实环境中我们可能更关心的是，TOPk的流量，也许我们拿到部分测量数据然后我们现在想知道的就是，在每一个时间链上拓不开的流量是多少？难道我们在想，难道我是要全部推出参数，将所有数据全部出来，然后再举证排序这样，如果这样子有一些数据是我不关心的，有些数据是小的数据，我也帮他去算了一遍，这个时候我们做法是提出了离散张量填充算法，我们跟原来算法不同的时候原来所有表征和过程都是用实数型，到我们这儿用离散型，我们在这样子框架下面，我们对于本身不是大的数字，我们不做推测，不做计算。我们能够快速玻璃掉不是大的数据，然后来实现快速推测。事实上我们快小和计算存储都大大降低。这个是我们放在2017年（英文）的工作，通过我们拿到测量数据之后，这个时候我们就希望在测量数据在这个数值上面来分析和监测网络是否出现异常。网络是否出现异常，当我们拿到数据的时候，以这个数据为例，我们一开始是基于举证，在举证框架下做异常检测，我们这里的异长检测是是是无监督检测，我只要数据情况下我来分析，我现在收到数据它的传统的模式，和数据上跟传统模式当中的差异，通过这种差异我来解释异常。这个是我们一开始做的网络异常检测，一开始我们建成举证，这个是一个（英文）的举证，行是（英文）面是时间轴。在正常情况下这个数据是低质举证，我如果能把收集到的观测举证已经被异常污染的观测举证能够分离出来，分离成低质举证加稀疏举证，我就可以借于稀疏来进行检测。但是那些东西来做的时候异常检测精度不高，那个框架是在图象处理用来分前倍几的工作，它是通过直接来约束低质和稀疏来检测异常，那个工作我们想这个比全面（英文）PC，比PC检测方法精度肯定要高，这个是框架，这个框架在做的过程中是迭代进行的，它每一个迭代分为两个字问题，第一个子问题和第二个子问题，第一个子问题是通过（英文）分解来做的，（英文）分解代价比较高，在这篇文当中我们讲整个过程在中间把这一块儿问题把它归结成一个寻找机的过程，我们快速借助于聚敏感函数，将整个数据进行快速重组，然后来借助一个聚敏感函数，我们搞一个（英文）缩放，来进行数据分割，实现来快速找到它的低质块，我们达到最好的精度的同时，我们速度比传统方法要提高几倍。

在前面做的过程中，我们事实上已经发现基于张量做异常检测一定会比基于举证检测精度高。这个是收到观测数据，我们如果能把观测数据分成低质张量和稀疏张量这个时候我们确确实实可以比举证来说可以提高精度，如果在做的时候，这里面依然存在它在整个过程这种张量复杂度一定会提高，因此我们就提出一些相应的方法，来降低在张量分解过程中的复杂度，这个在我看来是用张量模型来做网络异常检测第一篇论文，然后这一个效果甚至我们想象得到，比举证来做分解的效果大大提高了。

在做的过程中，整个举证过程中都是基于现性代数，在做的过程中只用了现性特征，这个数据可能会有一些非现性特征，如果说这个数据分布流性体上，我们再算两点之间距离，我们将非现性特征在刚刚我国中我们只考虑现性特征，如何来表示这个非现性特征，如果将非现性特征引入到张量分解当中，我是把正常模型拿到我如果也能把正常模型学到，这样子我精度一定能提高。因为它用张量的同时也学习到了现性和非现性。

我们有的时候有些场景需要我们快速进行网络异常检测定位，比如说数据到了时候我就告诉他这个数据是否存在异常，我们就研究了一些在线检测方法，需要快速定位当前到来数据是否存在异常，我们这个方法是提出一个双向二维PC的分解方法，通过检测新到的数据对于PC投影，在二维PC投影的方向，来快速定位网络是否存在异常。这个是我们刚刚今年在（英文）发的就是，刚刚那个方法我们在快速投影的时候有两面投影，实际上在数据面表达虽然建成张量。这里我需要快速进行异常检测，但是这个地方我们把它做成三维，这个地方是用滑动窗口来做，这个东西跟前面框架有类似，但是在这一个过程中我们发现整个滑动窗口在滑动过程中后一个阶段分解跟前一个分解是有关系的，而且在算法在执行迭代过程中，迭代下一论和它的前一论是有关系的，因此我们利用这个关系不再在每一个轮做重新的分解，而是通过迭代关系进行更新计算，使得这个计算比前面时间提高六千种。

这个是我们做的大规模测量技术相关成果，谢谢大家。

主持人：谢谢谢教授，谢老师精彩的报告，我们有时间有一个问题。

提问：谢主任问一下，你用几型张量分解？你现在不是张量分解和张量填充吗？问一下张量用的是几型张量？

谢鲲：我们做的还是比较简单的，我把这个网络数据建成三阶张量，当然还可以建成更高阶的张量。

提问：我不你用的是几比几？

谢鲲：我有两种，一个维度是ODIN，一个是（英文）一个是（英文）。

提问：我们现在实验环境还是在模拟实验环境。

谢鲲：我们这个过程我们拿的现网运行跑的数据，跑的数据比如说端到端的延时，我来看哪个地方测哪个地方不测，来推出你数据的精度。

主持人：谢谢，再次感谢谢教授。所以我们这个报告就都结束了，非常感谢大家上午在这儿一上午非常专注的聆听，我们明年开会应该还是会有这个，希望明年大家再能聚，再到一起探讨这方面的一些工作，谢谢大家，再次感谢！

—结束—