两种常见的技术：

1.使用中间人方法

弊端：允许弃用密码套件，一个异构型的网络需要很多经验丰富的管理员管理

2.key的共享和委托

介绍：企业方向证书或者私钥。客户端共享会话密钥。

以上这两种技术，都违背了端到端加密的初衷，并且实现困难。在多个MB的方法中，对于正则表达式的检测都需要解密数据包。MB需要得到完全的信任

-------------------------------------------------------------------------

本文主要讨论一些加密方案，并研究它们在真实世界的使用部署

将这些归类为4种基本技术，中间人方式作为第5个类别

定义一个通用的框架，标记它们的使用情况和限制。定义三个系统模型用于描述通过MB的流量（面向客户，面向服务器，客户和服务器的）。提供了一种个细化的标准。

定义了信任模型，确认威胁和安全需要

将隐私保护分为两类（主动和被动的）并进行对比

被动的无需解密或改变底层协议

主动的需要解密或者修改底层协议（例如：访问控制，可搜索加密，机器学习和可信任硬件）

分析了现有的方法，使作者更了解以后产业化的需求

解密流量是最好的办法，但是什么样的流量需要被解密，这是需要进一步研究的问题

-------------------------------------------------------------------------

实验模型和用例

blindbox blindIdS Bhargavan

图123 展示了高级的框架内。

MB被设置在client和service端之间时反应最位迅速。

第二部分提到的框架这里有详细的分析

面向客户：MB接收多个来自client的通向serv的流量，MB执行检查（广告拦截，防火墙等）。

面向服务器：例如（web服务器）：加密流量从服务器端点路由到服务器的MB，进行流量审查。ps为企业员工不在公司网络下访问公司的资源。

与面向客户的不同之处在于连接一般有客户端发起。MB主要用于保护服务器：数据库服务器等

----------------------------------------------------------------------------

信任模型

与以往的论文相同，都认为至少有一个端点时诚实的

当两个端点都是恶意的时候：Goltzsche David Goltzsche, Signe Rüsch, Manuel Nieke, Sébastien Vaucher, Nico Weichbrodt, Valerio Schiavoni, Pierre-Louis Aublin, Paolo Costa, Christof Fetzer, Pascal

Felber, Peter Pietzuch, and Rüudiger Kapitza. 2018. ENDBOX: Scalable Middlebox

Functions Using Client-Side Trusted Execution. In 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2018, Luxembourg,

June 25-28, 2018. IEEE Computer Society, 1–12 ENDBOX试图解决这个问题

本文的假设：一个实体可以是诚实的半诚实的或者不诚实的

表1有list（共同之处：MB都是诚实或半诚实的）只有endbox考虑了两端都不是诚实的情况

SafeBrick [46], ShieldBox [56] and SGX-Box [31].被部署在云端，可能是恶意的

主流假设1：MB不是恶意的，以合端点必须是诚实的

主流假设2：云端的MB是恶意的。或者两端是恶意的MB是诚实的

安全要求：对加密流量进行检测

------------------------------------------------------------------------------

技术

被动的无需解密或改变底层协议

主动的需要解密或者修改底层协议（例如：访问控制，可搜索加密，机器学习和可信任硬件）

部分检测：基于可搜索加密的技术仅仅在精确匹配和正则表达式匹配上解密

全面检测：检测所有的payload

中间人可以解密流量进行主动检测

中间人检测的

优点：可以不改变底层的tls结构

功能性：流量可以被解密，也可以对流量再加密（可以处理加密和不加密的两种流量）

性能：根据一些公司的调查，setup工作需要一定的网络流量和计算开销

局限：安全性问题，对于外包到云的MB不能得到完全的信任。

可搜素加密

不解密底层流量进行token匹配的工作

blindbox使用可搜索加密。主要想法：MB的规则即受用tls的会话密钥进行加密，mb只需要比较client发来的是否一致即可。但是MB不能使用key解密流量

减轻运算，blindIDS产生了，使用公钥对用于对token的匹配。这也在造成了tls需要被跟换的问题

spabox：使用的Diffie-Hellman 基于不经意 伪随机 函数 用于 加密 规则 的准备。并且还用到了机器学习

【 Xingliang Yuan, Xinyu Wang, Jianxiong Lin, and Cong Wang. 2016. Privacypreserving deep packet inspection in outsourced middleboxes. In 35th Annual

IEEE International Conference on Computer Communications, INFOCOM 2016, San

Francisco, CA, USA, April 10-14, 2016. IEEE, 1–9.】：提出比blindbox更高效的过滤器，扩展了令牌机制，虽然高效但是需要托管。

privDPI：改进了blindbox的握手机制，减少setup时间，并且使用可以重复使用的中间规则

优点：TLS不需要被修改，提供了不解密加密流量的检测办法（可疑时依然解密）

局限性：规则和token的rule的映射。serv、client都需要安装协议。需要独立的信道

性能：client因为要处理token计算过于密集，mb只执行匹配

访问控制技术：client和service都知道有个MB。并且它们同意赋予MB决定加密流量的访问控制权限。（主要特点是MB可见）

现有的解决方案种，需要对tls等底层协议进行修改才能实现（mcTLS）mbTLS可以做到不修改底层协议

优点：安全性上，提供了一种问责机制，serv端可以对部署的MB进行验证（独一无二）

性能：不需要SE（可搜索加密）之类的，效率更高。不需要客户端安装整数，不需要MB解密任何流量

功能：全面

局限性：tls协议需要改变，安全上MB如果被托管可能会泄露隐私

机器学习

机器学习可以不改变当前setup的情况下进行检测。主要思路是分析协议的报头，从加密载荷中提取信息

优点：不需要改变tls，从而保证了协议设计之初的安全性

局限性：功能性，针对不同的情况需要改造迎合不同类型的中间件，性能上，需要高逼真度的训练，很难获得

可信任硬件

解密流量不在mb，而在一个可以信任的地方。主要需要把会话密钥安全的运送到可信任硬件

SGXBOX

------------------------------------------------------------------------------------------------------------------

比较

中间人技术依然是最受业界欢迎的技术

当前的主要问题是保护tls的端到端安全，尽量少的泄露信息给MB

安全性上：主要体现在token可能会泄露信息，MB可能会与恶意的client或者serv合谋，ML技术如何获得准确的训练集

性能上：SE方案可以保护隐私，但是缠上的开销巨大（仅仅信道就要求两个）。机器学习的解决方案理想，但是需要进一步的探索

新的方案正在向安全硬件（安全飞地）方面转移

看看ENDBOX SGXBOX

翻译：

概论：

中间件在一个计算机网络系统的检查和分析网络流量检测恶意通信，监控系统性能，并提供运营服务。然而，加密的流量，这已成为越来越普遍，阻碍了中间件提供上述服务的能力。 一个常见的在实践中解决这个问题是由采用中间人（MITM） 的方法，其特征在于， 一个 加密的业务流之间两个端点被中断，中间人解密并分析流量。中间人的做法被很多组织所使用，但是带来了很多实际应用和隐私方面的担忧。实际上，由于成本的原因在 MITM的延迟发生由于到的加密，企业开始寻找一种消耗更少的解决方案。目前已经还了讨论上的许多 努力需要来配置MITM。 此外， MITM 违反终端到终端的隐私，对于用户隐私的忧虑也在提升。 此外，一些中的MITM实现被发现到被有缺陷的。 总的来说，新的方法和隐私保护技术使对加密的流量检测被提出。

我们系统地检查了这些技术，以比较它们的优点，局限性和挑战。通过定义一个由系统架构，用例，信任和威胁模型组成的框架，我们将它们分为四个主要类别。这些是可搜索的加密，访问控制，机器学习和受信任的硬件。我们首先讨论中间人方法作为一个基线，然后讨论在它们的每个细节，并提供了一个深入比较它们的优点和局限性。通过这样做 ，我们描述了采用该技术的实际限制，优势和陷阱。继此，我们给出现有方法之间的区别和在业界的表现， 这引导我们现在面临的挑战和研究的方向。

1引言：

包检查和分析已被用于对检测，减轻和阻止在家庭和企业网络可疑的活动。这个实现通过 实时检查的报头和有效载荷的网络流量。为这个目的部署的设备被称为中间件。一个中间设备（MB）提供各种服务 ，在当今的计算机网络基础架构中必不可少。一个主要的服务包括部署MB提供给系统和用户安全，用于例如，作为个人和组织的防火墙，入侵检测和预防系统，家长过滤，数据泄露检测系统，取证分析工具，如良好的恶意软件分析系统。在另外在 安全方面，它也常见到部署MB提供性能和运营服务。这些包括服务于代理/高速缓存作为在内容分发网络（CDN），广域网优化，协议加速，访问控制，计费和使用监控和网络地址 翻译。合规服务也是部署了一个 MB，来履行义务：支持合法拦截和控制的非法内容和隐私。

传统上， M被部署在的企业或客户网络的一侧。被称为“预置”的解决方案。随着网络功能 虚拟化（NFV）中的出现，依赖于专门的和昂贵的硬件部署的MB被也被质疑，并且存在一个明显转变朝部署基于软件的中间件功能。换句话说，我们已经开始向看到外包到的云基础架构中的MB 。 虽然云计算基础架构提供了更大的灵活性和动态可扩展性相对于基于硬件的设备来说，但是他们也面临新的挑战， 尤其是在这些系统确保安全和隐私。

为了使服务提到上面在一个有效行为，中间件经常执行网络检查和分析使用一个公认的技术称为为深包检查（DPI）。DPI工具深入检测网络数据包的报头和有效载荷，对比以传统包检测，传统方法只检测数据包报头。DPI可以是有状态的，与不同状态存储期间的分组的处理， 例如流量特点，应用状态。然而，目前 87％ 90％的网络通信量使用TLS加密。根据谷歌 transparancy报告，2020年11月不同平台使用chrome的流量中有81%到98%使用https。 现有 DPI 技术即是仅能够的检查纯分组属性，这个有限的使用将严重影响各个网络服务。这以为着MB需要具有分析加密流量的能力，以平衡隐私和提供服务。

1.1业界实践和新方法

两个常见的技术广泛地用于在该行业来检查加密的流量是：1.分裂TL 技术，MITM的方法。2.key的共享和委托。

MitM。在中间人，不同于建立p2p的tls会话，而是建立一个client和MB之间的会话。通 过这样做的话，MB可以检测client的数据，解密并分析。该 MB 重新加密和转发数据到server，新TLS 会话在MB和服务器之间。这提供了一个切实可行的解决方案是可以被部署，而不需要任何修改TLS协议， 但它需要一个客户端安装MB根证书。该根证书使MB假装自己是所述服务器（即目的地端点）。在实际上，是MB假冒的服务器。这种部署是安全的根证书被安全地存储，更新的TLS的使用。不幸的是，一些的部署已经被示出是不安全的，由于对的弱点在其执行的所述底层协议，例如作为允许弃用密码套件。此外，基于sherry等人的研究，具有异构网络设备的大型网络将需要许多经验丰富的管理员来管理它们。这带来了另一个隐私担忧：MITM 方法被使用时，因为许多的这些设备可以具有解密的数据权限。 确定哪些设备具有权限跟踪流量是一件困难的事。

密钥共享和委派。在这种方法中，企业分享他们的证书，或私钥，以及客户共享会话密钥用的中间件。他们都支持通过行业实际使用。对于例如，一个服务器共享其私有密钥与CDN，或者在服务器共享静态密钥与中间件，如去尾中的企业运输安全（ETS）标准化由ETSI [ 19 ]。 问题包括：不支持完美正向保密（PFS），TLS 1.3不兼容。无论是中间人方法还是密钥共享的方法都违背了端到端加密和数据数据隐私并且难以实现。MB存储解密的数据还带了工具和数据泄露的问题。这意味着MB的管理员需要仔细并诚实的进行设置。US-Cret已经发出警告，https可能以为这种设置导致变得薄弱。

隐私保护的方法：一个调查关于用户接受对加密的流量检查（即TLS）通过中间盒进行审查。 1976年参加者接受了调查，75.8％表示对隐私有顾虑， 而 70.9％的受调查者担心政府在监控。此外，使用中间人密钥共享作为一个工具来检查网络流量可能还违反保密规定。例如，83.2%的调查者认为第三方检测应该被事先通知。

1.2MB的外包

由于管理管理设备的困难也助长了新的中间件服务的发展。当前的一个趋势时MB外包到云。它减轻了企业的负担：购买相关硬件，安装（包括软件和硬件），配置，操作和维护中间盒。从而使该企业以专注于它的关键业务。相比较于企业内部的MB，部署在云上的MB更容易带来关于数据泄露的担心，因为数据需要被路由到处理者所在的地方。因此许多关于在云端的MB的研究，包括包括隐私保护检查的加密流量。在另外，基于云计算的MB必须提供低延迟的操作，因为流量被重新路由到了云中。一个例子是：embark。低延迟的实现是通过一个架构称作为：APLOMB。一般的解决办法是创建一个token的tls加密数据流。可搜索加密的方法被用于和MB上的规则集进行匹配。

1.3本论文的研究概述

本质上，在寻找一个在两个极端之间的解决方案，1.能够对加密流量执行深入检测的方法 2.对解密流量使用中间人检测和密钥共享方法。近几年被提出了一些解决方案取解决这些问题。本文把这些方法分为四个大类，本文同时还把中间人方法作为第五个类别进行比较。不同的方法需要被清楚的描述自己的优势和不足。（a）定义一个通用的架构，标记它们的用例和约束。我们定义了三个不用的模型，这些类型的流量会作为网络流量通过MB。我们把它们归类为面向客户端的，面向服务器的，服务器客户端的。（b）我们定义了一个信任模型，包含了MB系统中不同的参与者。（c）将隐私保护技术分为主动的和被动的。被动的检测技术不对流量解密进行检测，或者不改变底层的协议。主动检测协议则会解密流量或者改变底层的tls协议内容。这些技术还被分类为访问控制，可搜索加密，机器学习和可信硬件。（d）分析当前研究和潜在的研究方向。我们检测当前模型和存在的攻击。我们相信有进一步提升的可能。

为实用和高效的部署与最优检测能力，它似乎解密再加密有效载荷是一个更好的方法。一个问题是多少加密数据需要被解密，且需要保护隐私。我们相信，基于这一事实，一个新的MB的安全协议（MSP）需要被提出并提供标准化的解决方案。它可能是可信硬件的方法，例如 如使用英特SGX，正在被开发，加密流量的解密在一个MB无权访问的地方进行解密和检测。

【59】它研究的各种机制，的MB中为安全外包的一个普遍的方式。我们的工作在其基础上进行了进一步的补充，我们定义了不同的系统模型和用例。 在欧盟机构的网络安全 （ENISA） 也公布一个调查加密的流量分析。该调查描述6个关键要素（即应用程序识别，网络分析， 用户信息识别，加密的恶意软件，文件/设备/网站/位置手指甲的检测印刷和DNS隧道检测）和技术基于对机器学习。

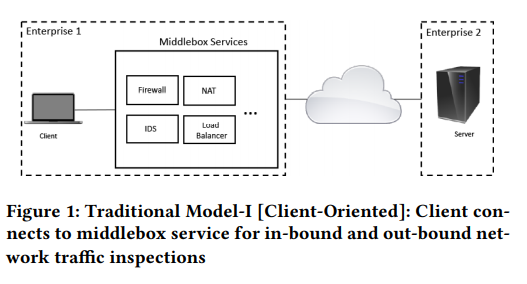
2 系统模型和用例

blindbox blindIdS Bhargavan

图 1， 2 和 3 示出一个一般的高级别体系结构的MB的操作在一个计算机网络环境。基于对这些架构我们描述了现有的和潜在的使用情况。我们注意到在内置的中间盒设置中，MB通常直接作用域一个通过的流量，其中一个或多个中间盒被放在client盒service直接的线路上。MB常见的使用包括：评估控制，计费和使用情况监视，资产跟踪，名称或标签解析和操作控制。 MB也可能会存储这些流量并进行分析。 网络安全用例包括网络防火墙， 应用程序防火墙，入侵检测系统（IDS）和入侵防御系统（IPS）。

2.1 面向客户端的MB

面向客户：MB接收多个来自client的通向serv的流量，MB执行检查（广告拦截，防火墙等）。

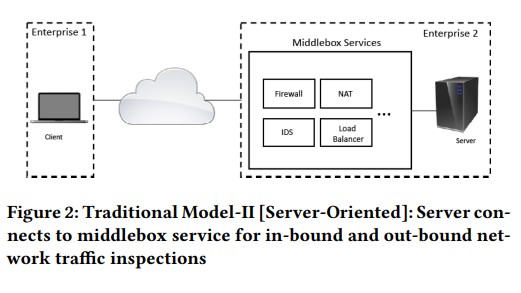


无论是入站还是出站的流量都能进行检测。它通常包括安装特定软件，或修改在连接在所述 客户端设备中以便于使所述的MB来执行该检查。

用例：常见的使用案例包括防火墙，广告拦截，个人数据保护和匿名化的信息。Sherry描述这样一个例子，学生使用大学网络安装提出的解决方案，以他们数据的保护隐私，以检查加密流量。 另一个使用的情况下，过滤器。将最相关的部署方案为企业 是 在 面向客户的 中间件。

2.2面向服务器

这种MB常见于企业网络。例如（web服务器）：加密流量从服务器端点路由到服务器的MB，进行流量审查。ps为企业员工不在公司网络下访问公司的资源。与面向客户的不同之处在于连接一般有客户端发起。MB主要用于保护服务器：数据库服务器等。



3 信任模型

两个端点至少一个是诚实的，两个端点都是恶意的时候：ENDBOX: Scalable Middlebox

Functions Using Client-Side Trusted Execution.

【endbox的概述

MB的分散式部署，利用客户端的限制资源处理客户端的流量

挑战：

1客户端必须被信任

2admin需要能控制MB的功能

endbox在客户端可以收信任的执行MB

本文：可以在源头或者目的地都识别流量

INTEL SGX：intel提供的。安全区内存存储在称为安全区页面缓存（EPC）的系统保留的内存范围内，该范围已进行透明加密

可以支持本地和远程验证

endbox：在客户机上执行MB（基于SGX---TEE），数据包的加解密都发生在TEE中

唯一入口：仅接受使用正确的endboxclient端加密的流量。确保没有流量可以绕过endbox。绕过的流量也不可读

endbox取消了MB的验证，而是把对流量的验证工作前置到客户端，具体方式是使用SGX这种安全飞地的方式。由intel的背书

-------------------------------------------------------------------------------------------------------------------

endbox的客户端组成

数据的分装等功能在安全区外执行，加密解密在安全区内执行

为了确保流量都被检测，用户不能直接连接到网络，只能通过vpn连接到网络。

vpn客户端对报文进行进一步的处理（在安全区内）

1.可以根据设置由一个或多个MB进行处理

2.根据特定的功能，包头或者payload可能会被更改。甚至整个数据包有可能被丢弃

2.数据包被签名并加密到TEE外面，被传递会vpn客户端的不可信任空间

4.传输数据包

serv端进行反向处理。

vpn客户端分为两部分：可信的和不可信的

--------------------------------------------------------------------------------------------------------------------

证书和密钥管理

私钥从不离开安全区

--------------------------------------------------------------------------------------------------------------------

处理网络加密流量

对网络加密流量从处理

假设：用户使用不受信任的tls库，且这些会话密钥被发送给诚实的Clieck接口，并且在endbox的客户端运行了。这些密钥被用于解密一些特殊的数据包

通过蟹盖openssl中的一个自定义函数，这个自定义函数通过openvpn调用会话密钥。

使用这种方法，endbox可以解密流量（客户并不知道），client不需要信任ca挥着其他提供接入服务的机构。通过endbox转移会话密钥并不会造成风险

--------------------------------------------------------------------------------------------------------------------

安全保证：用户不适用vpn和endbox连接的话，流量会被防火墙丢弃------------------【这点可以借鉴，当流量不符合时MB直接丢弃，省去serv端的种种操作（包括不限于对数据包头的处理）】

client和serv周期性的交换信息，客户端可以使中间盒一直有效，保持最新状态

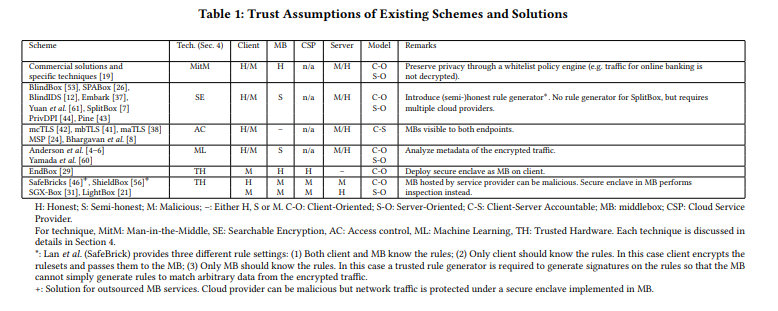
由于并非集种部署MB，对于拒绝服务攻击由很好的阻拦效果，得益于分布式部署，当发生故障时，影响不至于扩散致整个网络

】

3.1 假设

主要参与者有三 client MB serv，在某些方案中甚至有四个参与者。对于每个实体可以认为是诚实的，办诚实的，恶意的。

信任：下表中列出了信任方案



第二行中都是用可搜索加密技术，MB可以被部署在CS端，符合下面描述的信任模型1

第三行的mbTLS mcTLS等使用访问控制技术。对MB无要求

机器学习方法属于第一种安全模型

可信任硬件中，endbox和其他使用可信任硬件的相比减少了一种适用情况

Endbox是个列外，认为两边都可能是恶意的。使用安全飞地的MB有SafeBrick [ 46 ]， ShieldBox [ 56 ] 和 SGX-Box [ 31 ] 在这里，该服务提供商托管的宏块可以是恶意的，因为执行的MB功能，包括数据包检查的执行中的安全飞地。这意味着一个恶意的服务提供商应该不能获取或分析安全飞地中

总而言之，有两种不同的信任模型。信任模型1：MB是半诚实的，两个端点有一个是诚实的。（这种假设一般部署在 中间人 可搜索加密 访问控制 机器学习等技术提供的MB上） 信任模型2：MB可以是恶意的，两个端点都可以是恶意的，如果两个端点都是恶意的，则MB需要不是恶意的（一般使用可信任硬件技术）

3.2安全要求

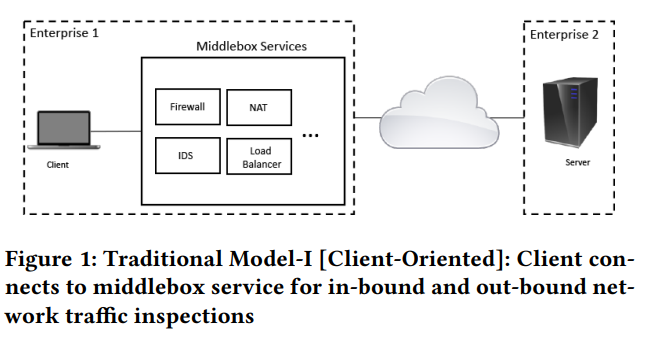
主要目标是：在加密的数据上进行检测，且能保护隐私。只有最终的端点能知道信息，其他设备不能知晓任何不被允许知道的信息。

4使用的技术

被动的无需解密或改变底层协议（b）主动的需要解密，或者说解密部分信息或者修改底层协议（例如：访问控制，可搜索加密，机器学习和可信任硬件）。主动检测技术的部分检测：基于可搜索加密的技术仅仅在精确匹配和正则表达式匹配上解密。全面检测：检测所有的payload。

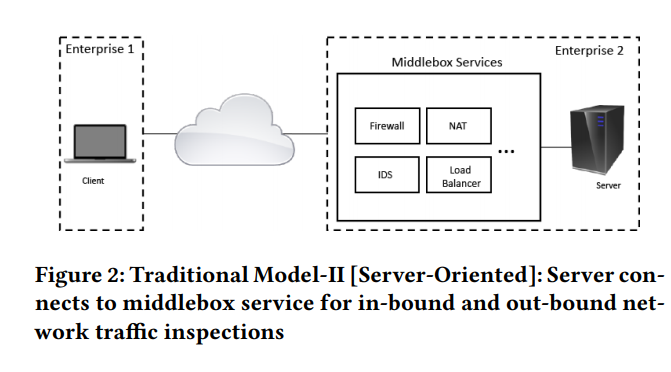
4.1中间人

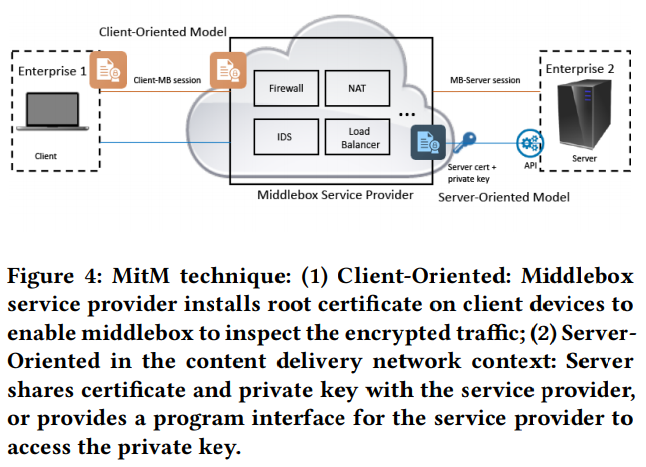
中间人的用途是主动检测，主要思路是是MB向service一样，这样依赖MB可以解密信息并重新加密信息转发到网络上。对于一个面向客户端的模型如下图



实现一般是由一个安装了证书的MB安装在客户端。（受信任的CA存储在浏览器）。客户端发起一个对话的时候，MB会截获流量并伪装成服务器

在面向服务器的模型上也可以进行类似的部署，如下图





上图对面向客户端和面向服务器的两种情况，如上图所示。面向客户的签名已经做了说明，下面解释面向服务器的形况：服务器与服务提供者共享证书和密钥，或者提供一个接口为服务的提供者提供一个获得私钥的机会。

中间人检测评价

优点：可以不改变底层的tls结构

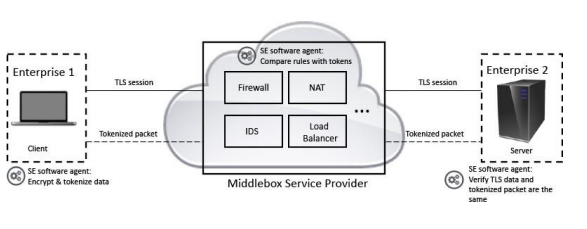
功能性：流量可以被解密，也可以对流量再加密（可以处理加密和不加密的两种流量）

性能：根据一些公司的调查，setup工作需要一定的网络流量和计算开销

局限：安全性问题，MB可能会学习到流量中的信息，对于外包到云的MB不能得到完全的信任。

4.2可搜索加密

主要思想是通过使用映射方案对应加密关键字和关键词。SE（可搜索加密使用主动检测方案）。sherry引入了blindbox，这是第一个是哟个SE技术进行DPI的方案。下图给出一个基于SE的方案



blindbox先通过handshake使用OT和GC建立一个aes加密。使用OT和GC的目的是：client端的密钥K和Mb端的rule都不希望对方知道。Handshake之后会生成一个aes加密，其中的key由client端提供，其中的payload由MB提供。Blindbox使用华东2窗口对token进行切分。对每个切分的token使用handshake的key进行加密。具体实现时还需要考虑针对词频的攻击，使用salt进行加密改为使用，只要client和service两端同时对出现的关键字进行salt+1即可解决词频泄露的问题。这里的token是单独使用一个ssl/tls会话传送的。对blindbox减轻计算量提出了blindIDS【blindids和blindbox的不同之处在于RG:不同之处：可能会在网络上监听有自己签名的数据包。论文设定：MB和RG不会联手的原因：市场声誉的限制。

使用的模型：

（未加密的）规则 R 和一个检测算法，表示为 Detect，将（未加密的）流量 T 和该集合 R 作为输入。在后文中，我们说该流量是恶意的 iff Detect （T， R）=0。否则，流量是安全的，并且 Detect（T，R）=1。在两种情况下，也可以提供一些辅助信息 aux作为检测算法的输出。

过程：

setup：

每个人都有一个公钥（可以没有私钥），有公共安全参数生成的一些潜在key

RULEGEN：输入生成的一组公共安全参数，RG使用自己的私钥对rule进行加密，生成Brule，发送给MB

Send：使用公共参数，用自己的私钥和R的公钥发送Tmsg，Tmsg，加密后为Emsg

detect：检测;MB使用自己的公钥，加密Emsg，加密Brule，输出0、1 1match 0safe

recv：使用自己的私钥，Emsg作为输入。输出Tmsg

解决的blindbox的不足：1.大小 2.明文规则 3.不使用sslKey

主要特点：

开销不取决于规则的数量，使用的不是ssl加密（使用可解密的搜索加密Decryptable Searchable Encryption (DSE)一个工具），需要握手。setup的开销转移到了MB的检测部分，DSE检测协议使用的检测时间比MB的检测时间更长，但是减少了内存的开销

blindbox的内存开销大小主要是为每个sender准备一个乱码电路，大小0.5M

】spabox也使用使用了SE技术【

解决了恶意程序的检测机器学习问题

构架一个trie的数据结构加快搜素速度

机器学习模型可以加入检测

分析结果只有客户端能看见---比blindbox进步之处，保证了分析结果的安全性

所有在中间间的流量都保持加密状态（与设计的协议相比，设计的协议也是token保持加密状态，但是token的加密并不是aes加密，而是异或加密）

特点：

ssl流量不在通过middlebox转发，而是直接发送给另一个client

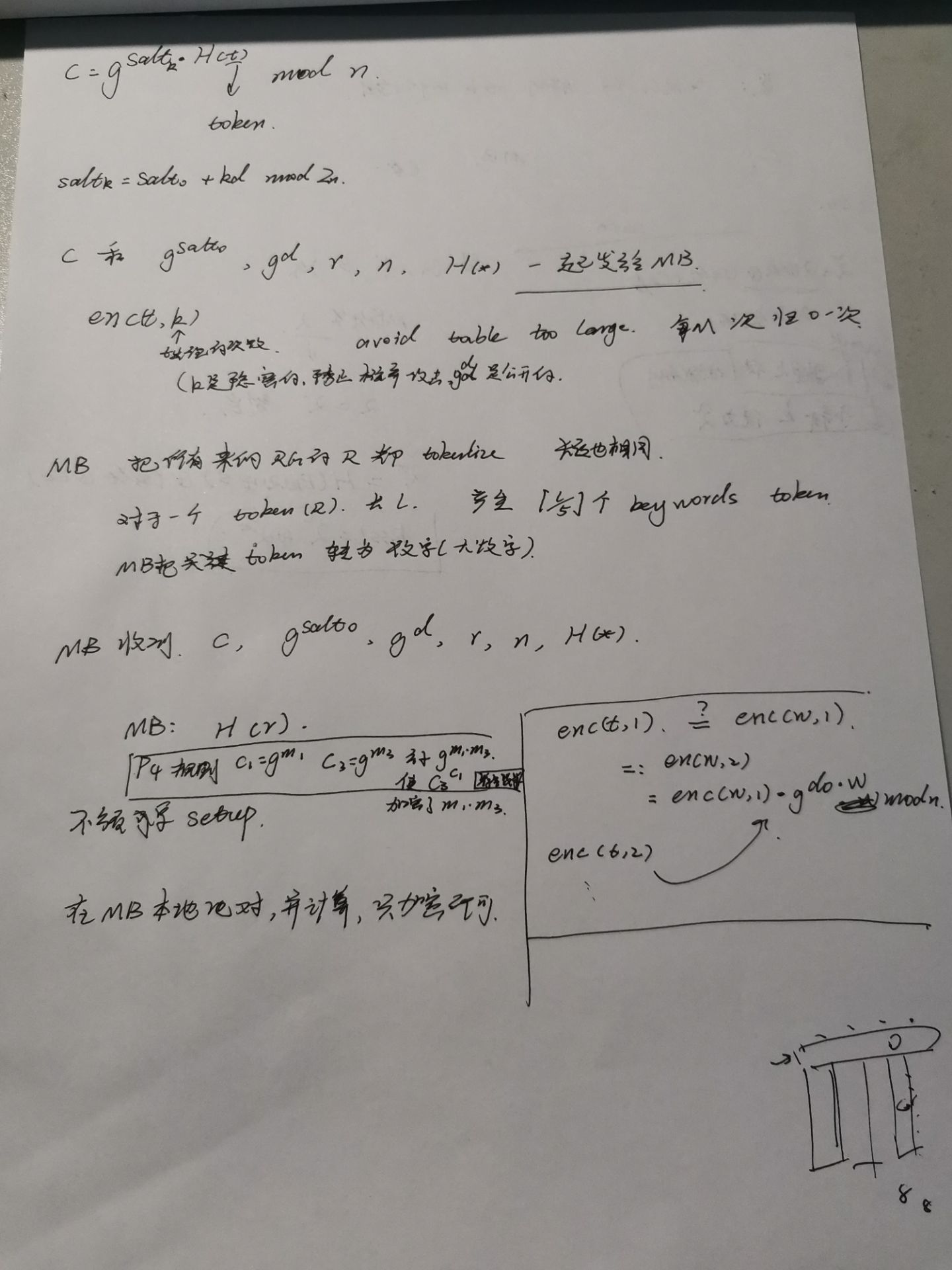
对攻击的定义不在是只有关键字的匹配，而是增加了程序符合已知的训练模型，机器学习的匹配

sap连接也采用令牌化的token，源连接不做任何处理。与其他协议的不同在于，为了适应机器学习，token需要进行特殊的预处理。（其他协议仅仅进行分割）

MB进行两次匹配：关键字的匹配和机器学习模型的匹配。机器学习的分析结果mb也不应该看到，将其发送给接收者。mb可以断开关键字匹配的规则。

bilndbox使用的setup工作，这里使用公钥和私钥完成（效率问题需要解决）

匹配方案：



】privdpi【

仅进行一次setup，产生的模糊规则集可以进行复用

】pine是对privdpi的改进【pine对于privdpi：减少了预处理步骤，支持规则隐藏，可以动态添加规则】

不解密底层流量进行token匹配的工作

blindbox使用可搜索加密。主要想法：MB的规则即受用tls的会话密钥进行加密，mb只需要比较client发来的是否一致即可。但是MB不能使用key解密流量减轻运算，blindIDS产生了，使用公钥对用于对token的匹配。这也在造成了tls需要被跟换的问题

spabox：使用的Diffie-Hellman 基于不经意 伪随机 函数 用于 加密 规则 的准备。并且还用到了机器学习

privDPI：改进了blindbox的握手机制，减少setup时间，并且使用可以重复使用的中间规则

SE技术方案的特点：

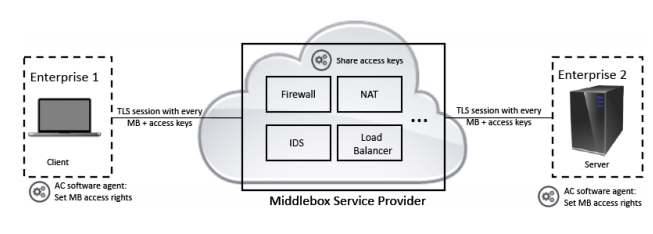
优点：TLS不需要被修改，提供了不解密加密流量的检测办法（可疑时依然解密）

局限性：规则和token的rule的映射。serv、client都需要安装协议。需要独立的信道

性能：client因为要处理token计算过于密集，mb只执行匹配

4.3访问控制技术

AC提倡客户端和service之间的准入client和service都知道MB部署在它们之间。此外，他们被赋予的能力，以验证和分配访问权限给这些MB。AC使用主动检查。下图示出了一个典型的设置为基于AC-方案。



Naylor等。引入了一个方案称为mcTLS。它会修改现有的TLS 协议 ，以允许在客户端，在MB和服务器以建立安全和验证通道，并交换在会话密钥中添加read、write密钥。mcTLS协议根据最小特权MB访问策略对网络流量划分。每个MB被分配一个读或一个写密钥。这涉及到提供不同级别的访问控制，以读和/或写上的加密通信。该加密的流量是这样区分的：即一些有效载荷可以被解密。这种技术的好处是对于哪些数据是不可以被解密的问题上可以细粒度的控制并灵活的保护隐私。换句话说，mcTLS仅部分地保留隐私：MB被授权解密的流量（有解密密钥）。McTLS改变了底层的tls协议，这个问题使器不能被广泛应用。

Bhargavan。 证实了mcTLS的攻击，并分析了协议。简单地说，一个的攻击利用的证据是 mcTLS，对于一个client和MB，MB和service之间并没有对会话的建立使用授权和在handshake完成的消息。Bhargavan提出了一个协议，不需要修改tls底层。

AC的技术特点：

优点：安全性上，提供了一种问责机制，serv端可以对部署的MB进行验证（独一无二）

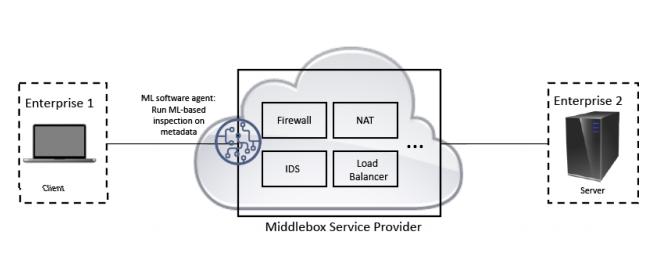
性能：不需要SE（可搜索加密）之类的，效率更高。不需要客户端安装整数，不需要MB解密任何流量

功能：全面

局限性：tls协议需要改变，安全上MB如果被托管可能会泄露隐私

4.4机器学习的方案：

主要思想是：分析协议的报头协议和提取加密流量的流量特征（例如可分析在一个时间周期内数据包的发送），观察附加信息，观察握手协议，观察有效载荷。ML是唯一使用被动检测的技术。下图展示一个ML的示意图

目前有许多协议使用ML进行流量的识别。Yamada提出了一个方案进行异常检测使用：数据包的大小和时间对加密流量进行检测。安德森提出的技术为恶意软件检测还采用各种TLS 报头信息和DNS 数据。安德森证明企业网络流量相比，某些基于恶意软件的加密流量具有不同的特征。这些特性时结合观测的数据，允许精确归类恶意软件的流量。他们都还表明该随机森林法优于其他方法分类恶意流量。然而，通过仔细设计的的功能，线性回归更优于随机森林法。

有的技术基于指纹识别恶意流量的。对于例如， JA3 / JA3S， 基于TLS的元数据 （例如 握手消息）生成的恶意软件指纹。 Anton提出了一个技术：通过观察数据流创建指纹。上述技术都需要实现了解恶意软件。

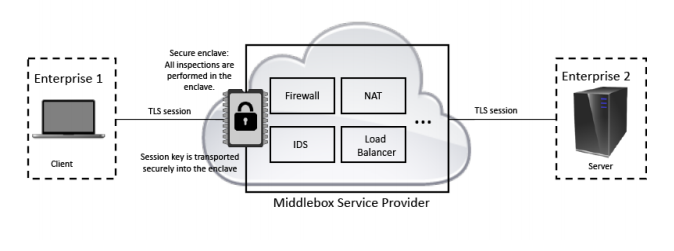
技术特点：

优点：不需要改变tls，从而保证了协议设计之初的安全性

局限性：功能性，针对不同的情况需要改造迎合不同类型的中间件，性能上，需要高逼真度的训练，很难获得

4.5可信硬件

解密流量不在mb，而在一个可以信任的地方。主要需要把会话密钥安全的运送到可信任硬件。可信硬件（TH）已也已部署了隐私保护的数据包检查。利用intel提供的安全飞地实现。 主要思想是：client和service在安全飞地中的MB共享会话密钥。检测和再加密都是再安全飞抵中进行。MB的提供者需要保证绝对安全。TH也是一种主动安全策略。



主要特征是解密流量和检测都在安全飞地执行。主要要求是：安全的运输会话密钥到安全飞地（不是客户端发送就是service发送）

SGXBOX，ENDBOX【

MB的分散式部署，利用客户端的限制资源处理客户端的流量，减轻服务器端的压力

endbox在客户端可以收信任的执行MB

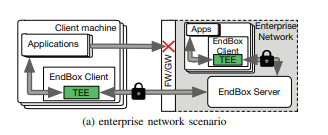
INTEL SGX：intel提供的。安全区内存存储在称为安全区页面缓存（EPC）的系统保留的内存范围内，该范围已进行透明加密

可以支持本地和远程验证

endbox：在客户机上执行MB（基于SGX---TEE），数据包的加解密都发生在TEE中

唯一入口：仅接受使用正确的endboxclient端加密的流量。确保没有流量可以绕过endbox。绕过的流量也不可读

endbox取消了MB的验证，而是把对流量的验证工作前置到客户端，具体方式是使用SGX这种安全飞地的方式。由intel的背书



endbox的客户端组成

数据的分装等功能在安全区外执行，加密解密在安全区内执行

为了确保流量都被检测，用户不能直接连接到网络，只能通过vpn连接到网络。

vpn客户端对报文进行进一步的处理（在安全区内）

1.可以根据设置由一个或多个MB进行处理

2.根据特定的功能，包头或者payload可能会被更改。甚至整个数据包有可能被丢弃

2.数据包被签名并加密到TEE外面，被传递会vpn客户端的不可信任空间

4.传输数据包

serv端进行反向处理。

vpn客户端分为两部分：可信的和不可信的

安全保证：用户不适用vpn和endbox连接的话，流量会被防火墙丢弃------------------【这点可以借鉴，当流量不符合时MB直接丢弃，省去serv端的种种操作（包括不限于对数据包头的处理）】

独有优势：

由于并非集种部署MB，对于拒绝服务攻击由很好的阻拦效果，得益于分布式部署，当发生故障时，影响不至于扩散致整个网络

】

特点：

优点：不改变tls底层协议。流量的检测再一个屏蔽的环境中进行

局限性：SGX尚再研究阶段，已有成功侧面攻击SGX的例子。功能性上，硬件的成本

4.6比较

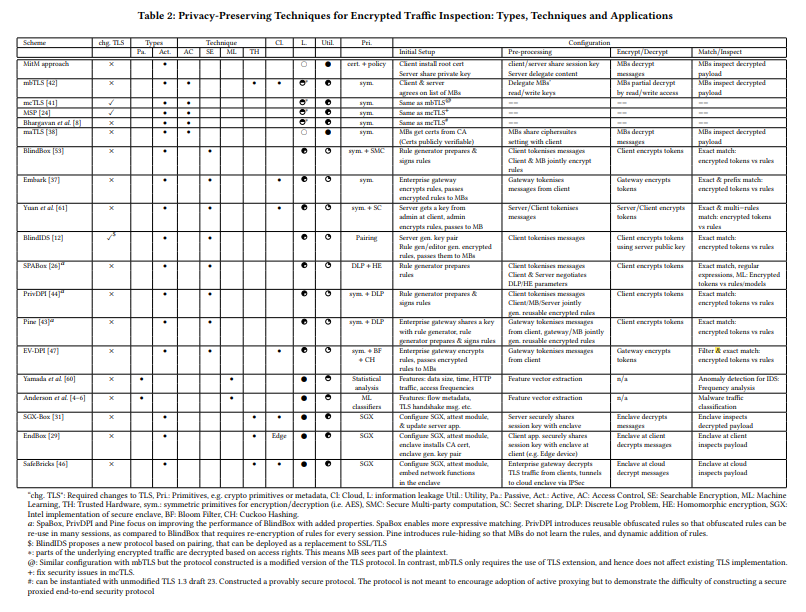
安全性：多少数据被泄露出去

1. 完全被reveal，网络数据中的流量对于MB是可见的
2. 部分被reveal只有部分网络流量被解密。解决方案基于SE和AC的对数据访问都在这个范围。之间也有细小的差别SE只在token与规则集匹配的时候才会解密数据包。对于AC方法来说，数据通过时都会被解密。所以相对于AC方法，SE泄露的数据更少。
3. 数据完全隐藏，MB不能看到数据包的任何内容，ML和TH技术在此范畴

效果

效果

1. 完全的功能方案提供的检查方式相当于直接在明文上进行检查
2. 部分功能，方案仅能在部分数据上进行匹配，例如：SE对token的匹配和对前缀的匹配。因此对SE的功能性。ML的功能性为，AC和TH使用的数据最多，起在安全飞地中直接解密，功能性在除了中间人方案外最强



第一列表示是否对TLS进行变更

CL表示是否在云端

L：表示数据泄露

Util：表示功能性

Type表示使用的技术是什么类型的：pa表示被动的，act表示主动的

Pri：原语（加密元或元数据）

Sym：对称加密（AES）

SMC：多方安全计算

SC：秘密共享

DLP：离散对数问题

HE：同态加密

SGX：intel的安全硬件系统

BF：bloom filter（过滤器）‘

CH：cuckoo哈希

Spabox，privdpi，pine聚焦于通过增加属性的方式提升blindbox的性能，SpaBox可实现更具表现力的匹配。PrivDPI推出可重复使用的模糊规则，所以是模糊的规则可以被重复使用。Pine可以动态的增加规则。

5讨论

许多方案已被提出，但是业界还是更愿意使用基于中间人方案。主要目标依然是保证端到端安全性的前提下减少对MB的信息泄露。

5.1安全性

信息泄露，基于SE的方案中令牌是可能泄露信息的，进一步的研究中认为TH可以避免这类问题。外包的MB中可能存在和谋，恶意客户端或服务器可以串通服务提供商通过提供 所述会话密钥以便为提供商底层流量，从而绕过例如SE AC TH ML的已有方案。一个新的安全模型提出用于解决这一问题。对于例如， 在对其中的场景的企业外包其内容管理到一个内容分发网络（CDN）提 供商，该企业股票的 证书 私有密钥，或委托内容到的CDN提供商。 假 设 说 的 CDN 供应商 使用 其 自己 的云 基础架构，也 就是 没有 接应 在 这种 情况下。 在 此相反， 如果 一个 企业的外包网络威胁检测到一个 托管服务提供商，其中的供应商使用一个第三方云提供商来托管其 各种 MB 服务， 然后 在 可能 的 勾结 可能 需要到 被 考虑 到 考虑。（待完善） ML技术的挑战，ML使得对流量的检测不在需要加密或者解密流量数据包，存在的问题在于：现有技术是否能覆盖MB所需要的功能，以及，所提供的训练集是否足够准确。

5.2性能

基于SE的方案何以保护隐私，但是会产生巨大的开销。SE一般都需要至少两个信道。但是因为其不需要专用硬件，使其依然是一个技术路线。同样也不会像AC方案一样解密底层的数据。因此如何提升效率就是一个挑战。在TH硬件广泛使用后，SE方法和TH的结合依然可以是TH的效率得到提升。

机器学习的方法提出了理想的解决方案，但是效率低下

基于AC的方案可以作为中间人的一个替代。为了避免中间人方案的安全为题，基于中间人提出了AC的技术，要求所有MB在客户端和服务器之间负责访问控制的协议。如果部署了这个功能，client和service可以决定哪些MB查看哪些会话。它的性能将类似于一个普通的中间人方案，因为MB像之前一样，解密-检测-加密。但是在AC技术方面，真正的挑战不在性能方向，而是安全性，配置和实用性

TH方案，最近的研究在向TH方向倾斜，由intel提供的SGX技术可以在大多数新的intel cpu上实现。但是这样同样也会泄露出消息的最小产生日期是什么（根据加密的方法推断）。因为安全飞地可以在内存上，所以是高效的

5.3效果

AC和TH具备提供完整数据解密检测的能力。

无需客户端服务器实现完整功能。AC技术面临的挑战是：能否在不事先决定client和service之间设计的MB的前提下实现这种效用。不清楚什么时候引入新对MB或删除现有的MB，client和service如何更新它们的通信。此外AC引入了上下文的概念，client和service可以灵活的对其数据设置访问MB的策略，这以为着MB只能读取一个特定的信息。

SE和ML的扩展，部署硬件，对于老旧的系统不应该成为一个选择。可以寻求基于SE和ML的方法

6结论

我们提出一个全面的调查加密流量隐私保护检查。我们定义一个信任模型，并根据现有的最新方案对不同的网络设置进行分类。从我们的编辑，我们进一步归类的当前计划为 4种主要技术，对展示的优势和局限性都有建议。主要的困难是要填补实际应用中的缺口，现在还有很多依然是基于中间人的，并不包含隐私。