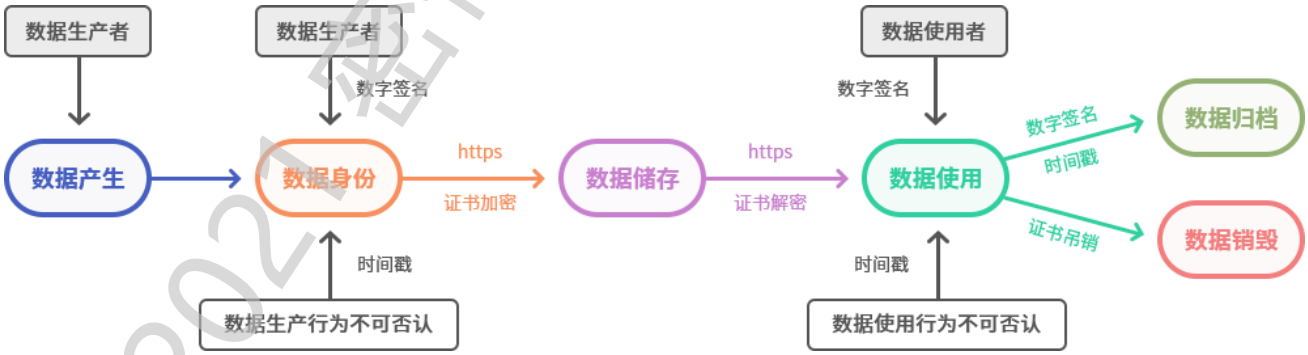
大数据应用与隐私保护，二者不可得兼？

（2021 年 2 月 25 日）

我国已经开始实施国家大数据战略，主要包括大力推动大数据技术产业创新发展、构建以数据为关键要素的数字经济、运用大数据提升国家治理现代化水平、运用大数据促进保障和改善民生、切实保障国家数据安全等。在这个大战略的指引下，我国的大数据应用发展取得了一定成绩，但是由此带来的大数据隐私保护问题也不少，并且已经成为一个能否健康稳定发展的一个重要关键指标，大数据应用和隐私保护必须二者得兼才能健康发展。笔者此文意在大数据采集、使用和存储等方面提出一些个人见解。

先说说数据的生命周期吧，所有数据都会经过五个时期，从数据产生开始到数据有身份，再到数据存储，再就是数据使用，最后可能是数据归档不再使用或者是数据销毁不复存在。而要保障数据的全生命周期安全，当然离不开 PKI 技术。PKI 技术（Public Key Infrastructure，公钥基础设施）是保障大数据安全的唯一可靠技术，彻底解决了（1）数据的机密性（Privacy);（2）数据生产和使用方的身份真实性（Authentication);（3）数据的完整性（Integrity);（4）数据生成行为和使用行为的不可否认性（Non-repudiation）等四大令人头痛的数据安全问题（**PAIN**）。

PKI 技术的重要应用就是各种数字证书的数字签名和加密应用，如下图所示，看看 PKI 技术是如何保护大数据安全的。



1. **数据产生**

数据生产者可以是人或物(数据采集终端)，由生产者产生数据。

2. **数据身份**

数据产生后应该用数据生产者的身份证书给数据数字签名来证明数据生产者的真实身

份，当然是数字签名加时间戳，来证明数据产生行为和生产时间的可信和不可否认。同时也可以证明这个数据的所有权和数据身份，让每份数据都有可信身份，用于数据版权保护、数据溯源和数据处理之前的合法性认证，防止非法数据入库污染大数据。这一步非常重要，肩负多重责任，要从源头保护数据和保证数据质量。但是目前的操作这一步基本上都是缺失的。

3. **数据存储**

数据生产出来后也有了身份，就需要提交到云端服务器上存储，数据提交必须使用https加密传输，以防止数据在传输过程中被非法窃取和非法篡改。这一安全保障措施非常重要，但现实情况是，大量的数据采集都是采用明文的 http 方式提交到云端服务器，这一点急需改进。如果这个数据是机密数据，还可以用有权阅读此数据的人员的公钥证书加密此数据后再提交到云端服务器，以确保此数据只能是有权阅读者才能解密阅读，防止机密数据被非法使用。

4. **数据使用**

数据使用者当然必须采用 https 浏览数据或者下载数据，否则无法保证数据下发过程中被非法窃取和非法篡改。如果数据是加密的，则数据使用者需要用其证书私钥解密才能正常阅读。最重要的是：数据使用不仅仅要有权限控制，而且使用者必须用数字签名来证明其合法身份，并附署时间戳签名来证明数据使用时间，数字签名加时间戳不仅能确认使用者的身份，而且能有效保证数据使用行为的不可否认和可信使用时间，用于后续审计和追追溯。

5. **数据归档和数据销毁**

这一步就是数据的生命周期的结束，可以是把数据用归档证书数字签名加时间戳归档封存数据，其目的是不仅可以确保数据不能再被篡改，以便以后审计和溯源需要，同时时间戳签名则能证明可信归档时间。如果决定要销毁数据，可以是在服务器上的物理删除。但是，为了彻底销毁已经下发给用户使用的数据，可以采用吊销加密证书的方式来彻底销毁数据，以确保该数据无法再使用。这个销毁操作要求数据在下发给用户之前必须是采用了证书加密的。

大家从上面的 PKI 数据保障措施可以看出，只要我们在数据生命周期内全程采用数字签名和加密技术就能保障数据的全生命周期的安全，大数据应用和隐私保护是二者可以得兼的！

为了让大家能理解以上概念，我还是拿邮件数据来说明吧。电子邮件是互联网的第二大数据源，每天邮件数量高达 2690 亿封，这些数据中含有大量的个人隐私信息和商业秘密，甚至

有许多金融付款和信用卡等高度机密信息，但是目前都是明文传输和明文存储的，非常不安全。

下面，我就基于上述大数据生命周期的五个不同时期，来讲一下密信技术是如何将 PKI 技术应用于保护电子邮件数据安全和保护电子邮件隐私信息安全的。

(1)**数据产生**：用户使用密信 App 写好了邮件，就完成了数据产生，邮箱主人就是数据生产者。

(2)**数据身份**：用户可以使用密信 App 的数字签名功能为每封邮件加上数字签名来证明邮件的身份，证明这个数据的身份，并确保这个数据不可能被非法篡改，一旦被篡改，则收件人在收到邮件后密信 App 或 Outlook 等会提示数字签名无效，数据已经被篡改。同时，自动配套附署的密信邮件时间戳服务也能证明这个数据的生产时间可信。

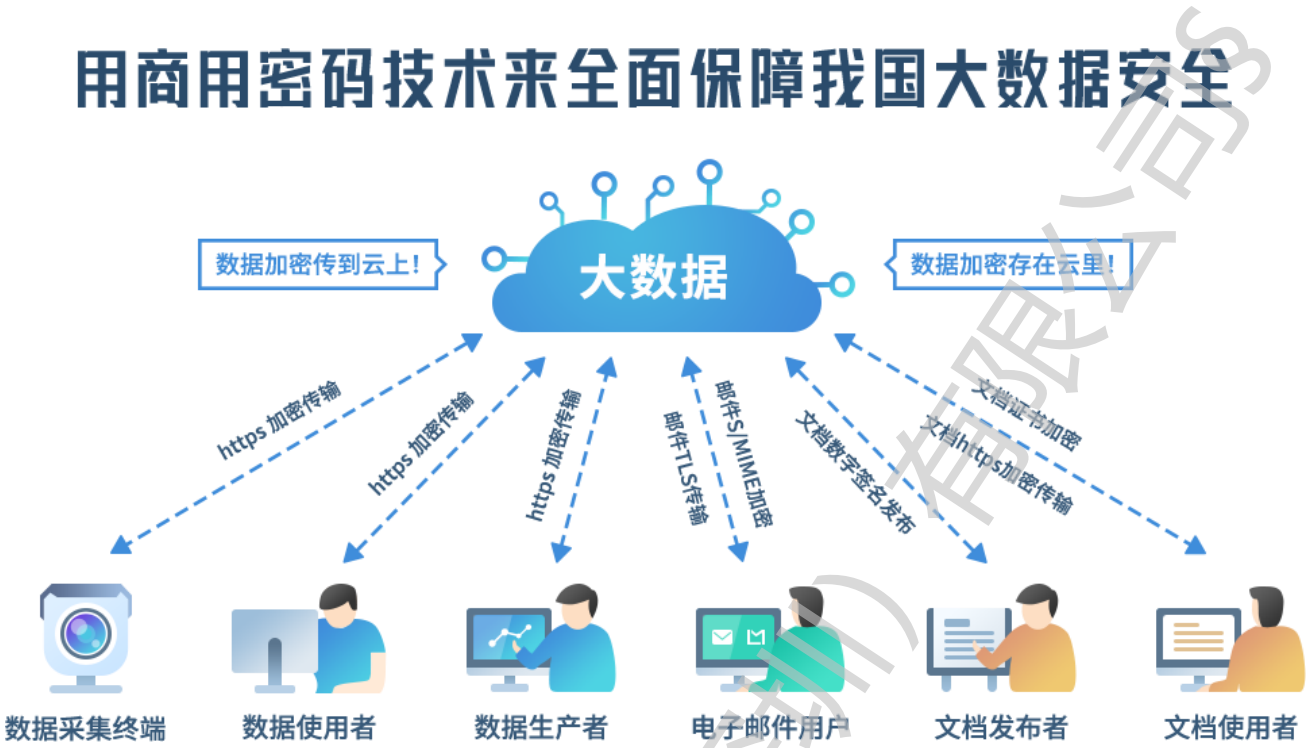
(3)**数据存储**：密信 App 发送的邮件默认是有数字签名、加密和时间戳的，邮件服务器也有 TLS 实现类似 https 加密的机制保障邮件加密传输到邮件服务器上。而邮件本身的证书加密则保证了电子邮件是以密文方式存储在邮件服务器中，保证了机密信息的存储安全，特别是使用云邮件服务的用户，能保障邮件机密信息不会被非常窃取和非法篡改和不会被泄密。

(4)**数据使用**：收件人收到邮件后，密信 App 会自动用收件人的私钥解密此加密邮件，并验证数字签名，展示发件人的真实身份，有效预防邮件欺诈。由于收件人也有数字证书，也就是明确了数据使用者的真实身份，能解密此邮件也就是证明了数据使用者的使用行为是不可否认的。

(5)**数据归档和数据销毁**：如果邮件本身已经数字签名和加密，则可直接用于归档而无需再签名。如果要销毁数据，则直接在邮件服务器上彻底删除邮件即可。

从上面的例子可以看出，密信 App 实现了全自动邮件加密和数字签名加时间戳，能有效地帮助用户实现邮件数据的安全使用和隐私保护。需要特别指出的是：密信技术全面采用了国密 SM2 算法和国密证书实现了邮件数据加密和数字签名加时间戳，符合我国《密码法》对关键信息基础数据保护的合规要求。

笔者认为：不管是什么数据，特别是各种数据采集终端采集的数据，都应该全面应用 PKI 技术来保障其全生命周期安全，用国密算法来保障我国大数据安全。从数据产生源头开始标识数据身份，加密传输到云端，并加密存储在云端，也应该加密实现数据交换和数据使用，并采取有效措施真正销毁应该销毁的数据，采用数字签名加时间戳来固化归档数据。只要这样，才能确保大数据应用和隐私保护二者可得兼也。



----------END----------

想联系我讨论此话题？请使用 密信 App（ - ）扫码发加密邮件给我，我一定会回复您的加密邮件。