Technofixing the Future: Ethical Side Effects of Using AI and Big Data to meet the SDGs[1]

[1]:Sustainable Development Goals

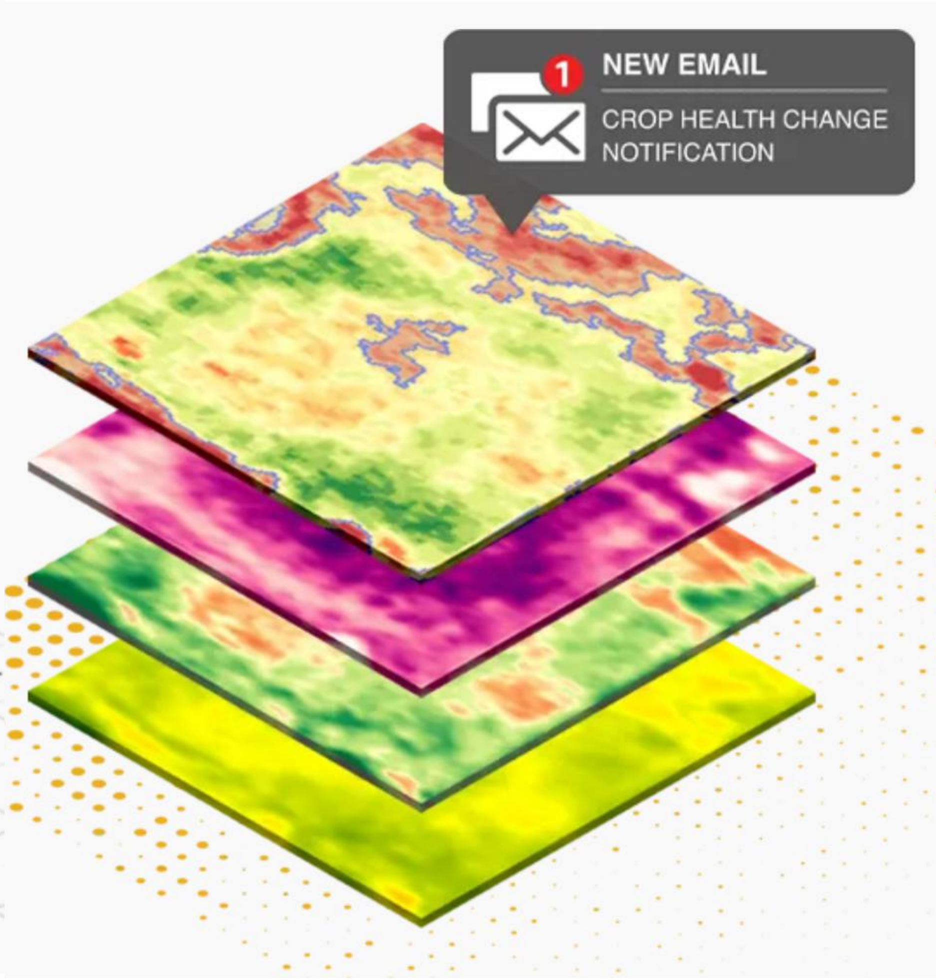
分享人：杜宜林

AI IN AGRICULTURE

Weather Data

Soil Data

Satellite Image Data



AI in Healthcare

Patient-Facing

AI Chatbots
buoy, Your.MD, med what, GYANT, Dr.AI

Wearables & Devices
physIQ, QARDIO, QUANTTUS, sano, DietSensor, kiwi.ai, BIOSERENITY, Jodytrak, aira, Magnea, kinsa, SCANADU, AliveCor, 19 Labs, CliniCloud

Personalized Genetics
iCarbonX 碳云智能, rthm, CG, DAY TWO, Positive Bioscience, LIFENOME, MEDICUS, 23andMe, VIIAGENE, Arivale, habit FOOD, PERSONALIZED, Suggestic PRECISION EATING, VIOME

Mental Health
Woebot, TOUCHKIN, BIOBEATS, Ginger.io, AURA, koko, NeuroLex

Women's Health
CYRCADIA HEALTH, bloomlife, celmatix, flutter, eva, Clue, elyse, 28, Natural Cycles

Skin
SkinVision, skinIO, skin analytics, SCANZ, skin, DOCTOR HAZEL, FIRST DERM

Telehealth

Telemedicine
ada, SENSELY how are you feeling today?, babylon, anamnèse, Remedy, pager, MDLIVE, spruce, Z

Lifestyle Management
AiCure, MedAware, ovia health, next IT Healthcare, Catalia Health, lucina, Health Next Gen, Doctella, NarrativeDx, cardiogram, Wellframe, Welltok, CareAngel

Disease Management
prognos, INFOBIONIC, HealthReveal, HEALTH FIDELITY, DIABNEXT, SENTRIAN, TEMPUS, dm dreamed DIABETES, Healint, WIKAYA, FLASHBACK TECHNOLOGIES, RxPREDiCT, BELONG Beating Cancer Together

Research

Drug Discovery
Mind the Byte, SaaS for Drug Discovery, DATAVANT, QUANTITATIVE MEDICINE, Standigm, ENVISAGENICS, BERG, RECURSION, BenevolentAI, Athelas, Exscientia, NuMedii, healx, transcriptic, Numerate, twoXAR, Atomwise Better medicines faster., INSILICO MEDICINE, CloudPharmaceuticals, Globavir

Information & Clinical Trials
trials.ai, PathAI, BenchSci, CYCLICA, Mendel.ai, Qrativ, peptone, Brite Health, iPLEXUS, antidote, Plex Research, nference, RESULTCARE

Genetic Research
MolecularMatch, Blueprint Genetics, VERGE GENOMICS, NURITAS, diploid, HUMAN LONGEVITY, INC., SOPHIA GENETICS, freenome, FDNA, PATHWAY GENOMICS, deep genomics, DESKTOP GENETICS

Doctor-Facing

Medical Records
APIXIO, AUGMEDIX, clinithink, MODERNIZING MEDICINE, saykara, nimblr, emr.ai, WellSheet, hCue, patientory, SANSORO HEALTH, PROTENUS, BloomAPI

Data Analytics
ZEPHYR HEALTH, burstIQ, MD.ai, Infermedica, CLOUDMEDX, jvion, CareSkore, flatiron, PROFILITY, lumiata, ROAM, Hindsait, Oncora MEDICAL, GRAIL, MEDALOGIX, ensodata, pulseData, healthcare.ai, KenSci, COMBINOSTICS, impacthealth, LYTICS

Medical Imaging
VIZ, AMARA HEALTH ANALYTICS, enlitic, zebra MEDICAL VISION, 推想科技 infervision, CureMetrix, BAYLABS, VUNO, ADVENIO, koios, BUTTERFLY Network, Inc., SIGTUPLE, Behold.ai, imagia, ARTERYS, Mindshare MEDICAL, aidence, IMAGEN, aidoc, Lunit, Cardiologs, Genetesis, HubMedX, DIA IMAGING ANALYSIS, SemanticMD, Blackford analysis

Hospital
FORWARD, Qventus, MedyMatch, gauss, orgo.ai, MEDASENSE, P3d, AURIS, VERB SURGICAL, Qualaris

北京大學
PEKING UNIVERSITY

Enslaving the Algorithm:From a “Right to an Explanation” to a “Right to Better Decisions”?

分享人：魏日升

机器学习算法（不透明）
招聘、贷款歧视

欧盟：90年代 数据保护指示（DPD）
2016 通用数据保护法规（GDPR）

解释权
自动化
个人数据

“

个人责任太重
逻辑复杂

认证和信托
第三方机构

资金
执行

”

Keeping Authorities "Honest or Bust" with Decentralized Witness Cosigning

Ewa Syta, Iulia Tamas, Dylan Visher, David Isaac Wolinsky

– **Yale University**

Philipp Jovanovic, Linus Gasser, Nicolas Gailly, Ismail Khoff, **Bryan Ford**

– **Swiss Federal Institute of Technology Lausanne** (**EPFL**)

背景

- 我们依赖很多权威机构(Authorities)
  - 授时服务(NTP)
  - 证书颁发机构
  - 软件更新服务

- 但权威机构是否真的值得信任？
  - Google发现了多个未经授权的证书
  - 证书来源于MCS Holdings公司
  - 2015年，Google宣布不再信任CNNIC颁布的证书

Authorities也可能不那么靠谱！

Google Security Blog

The latest news and insights from Google on security and safety on the Internet

Maintaining digital certificate security

March 23, 2015

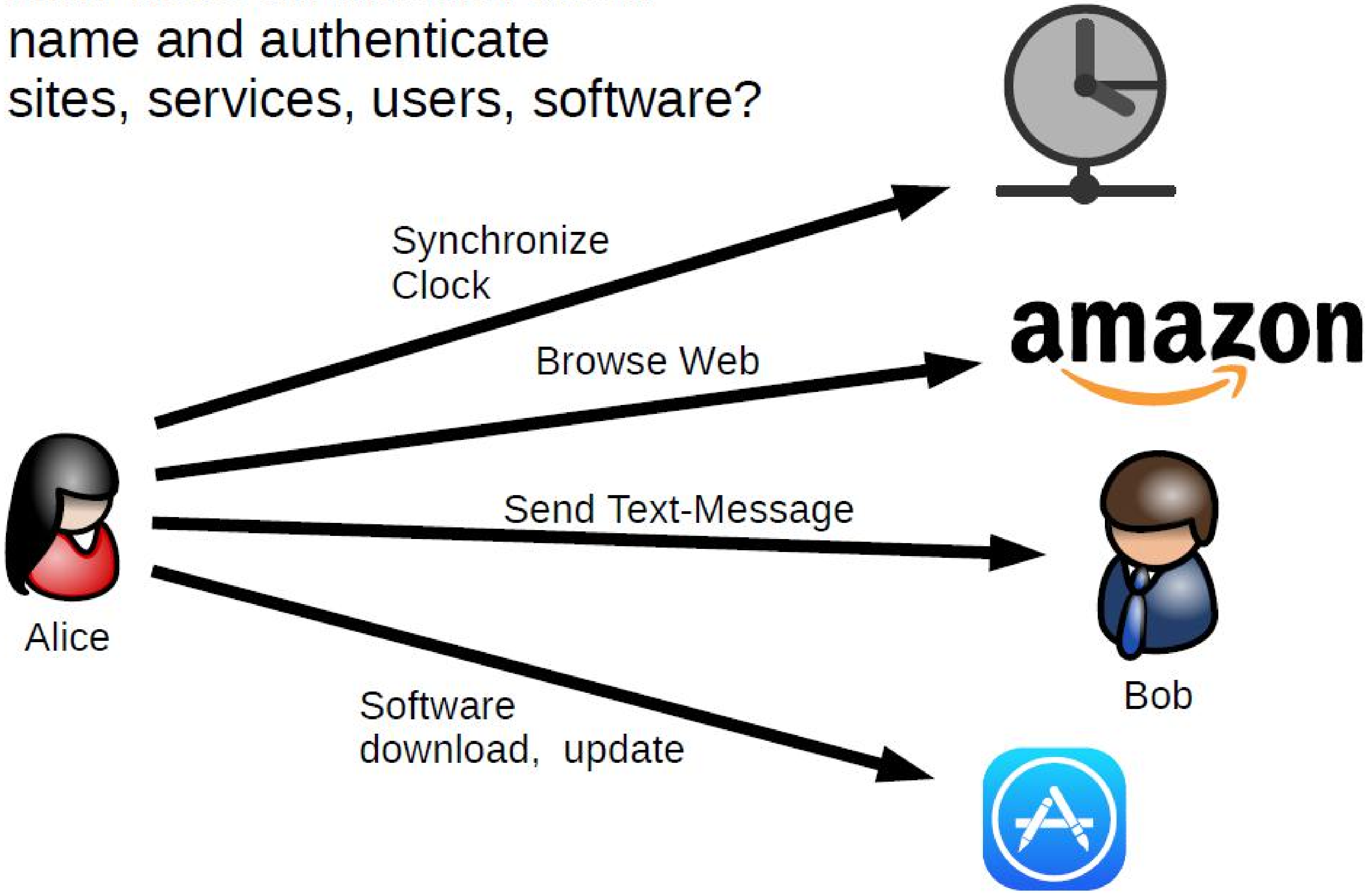
Posted by Adam Langley, Security Engineer

On Friday, March 20th, we became aware of unauthorized digital certificates for several Google domains. The certificates were issued by an intermediate certificate authority apparently held by a company called MCS Holdings. This intermediate certificate was issued by CNNIC.

CNNIC is included in all major root stores and so the misissued certificates would be trusted by almost all browsers and operating systems. Chrome on Windows, OS X, and Linux, ChromeOS, and Firefox 33 and greater would have rejected these certificates because of public-key pinning, although misissued certificates for other sites likely exist.

用户依赖Authorities

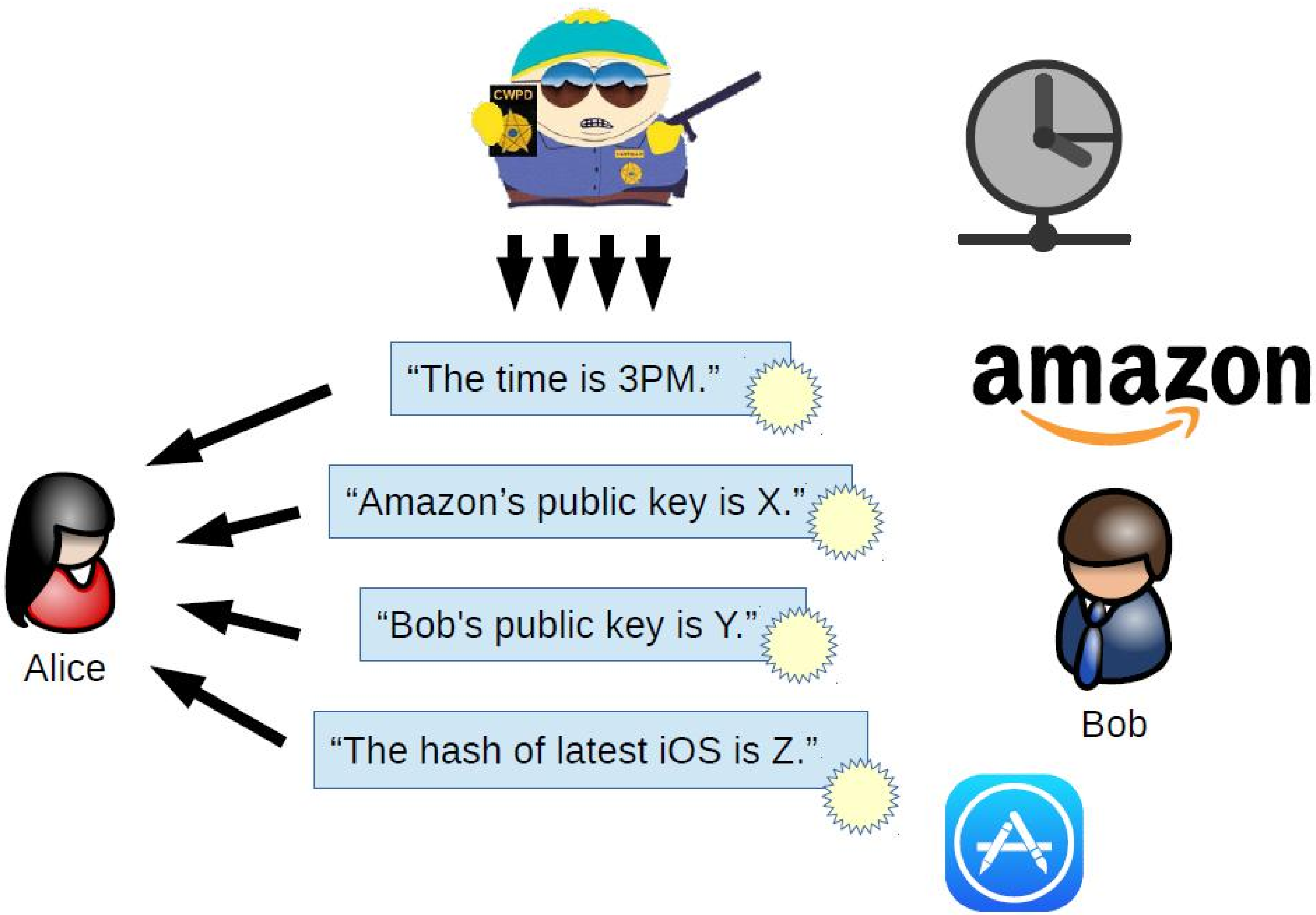
How does an Internet client
name and authenticate
sites, services, users, software?



发出请求

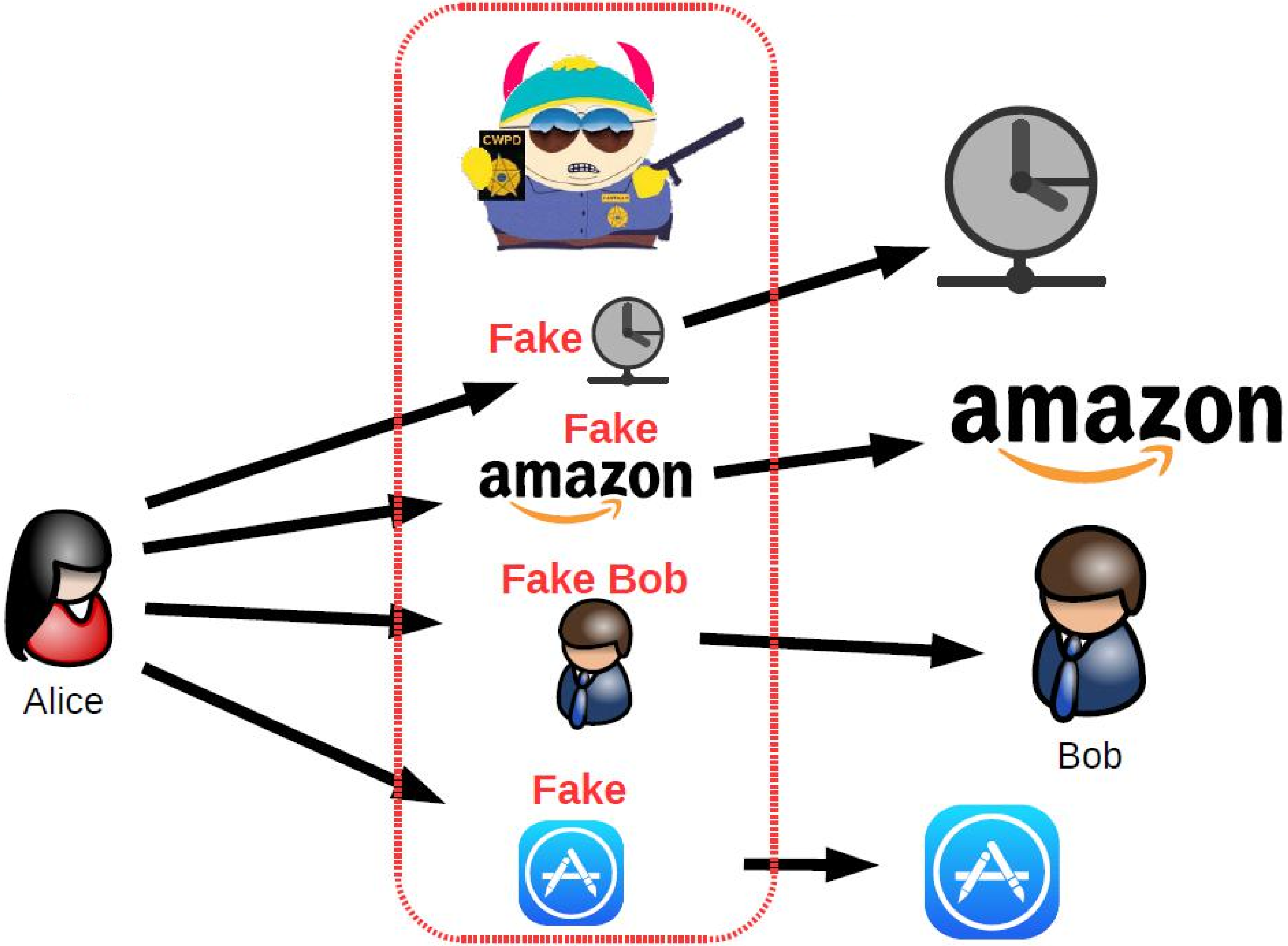

Respect my
Authoritah!
CWPD
?
Alice
What is:
• The current time?
• Amazon's SSL public key?
• Bob's IM public key?
• Latest version of App?
amazon
Bob

Authorities认证



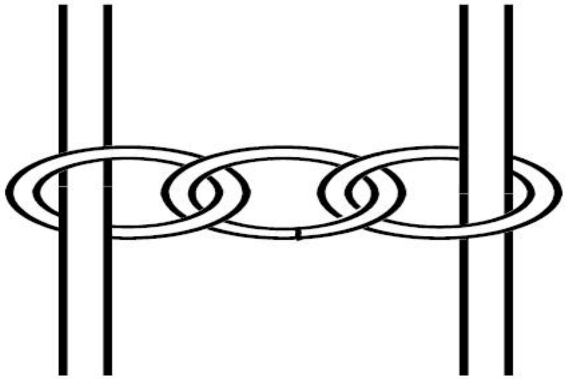
问题1：Authorities不可信

- 中间人攻击
- 冒充用户
- 虚假更新
- ……



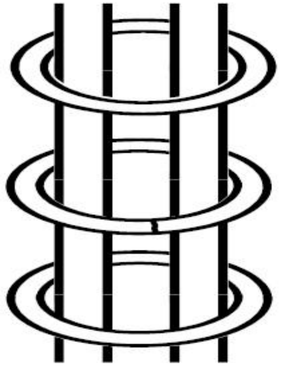
问题2：弱的安全链接

- 弱链接：一个环节攻破，整体不安全



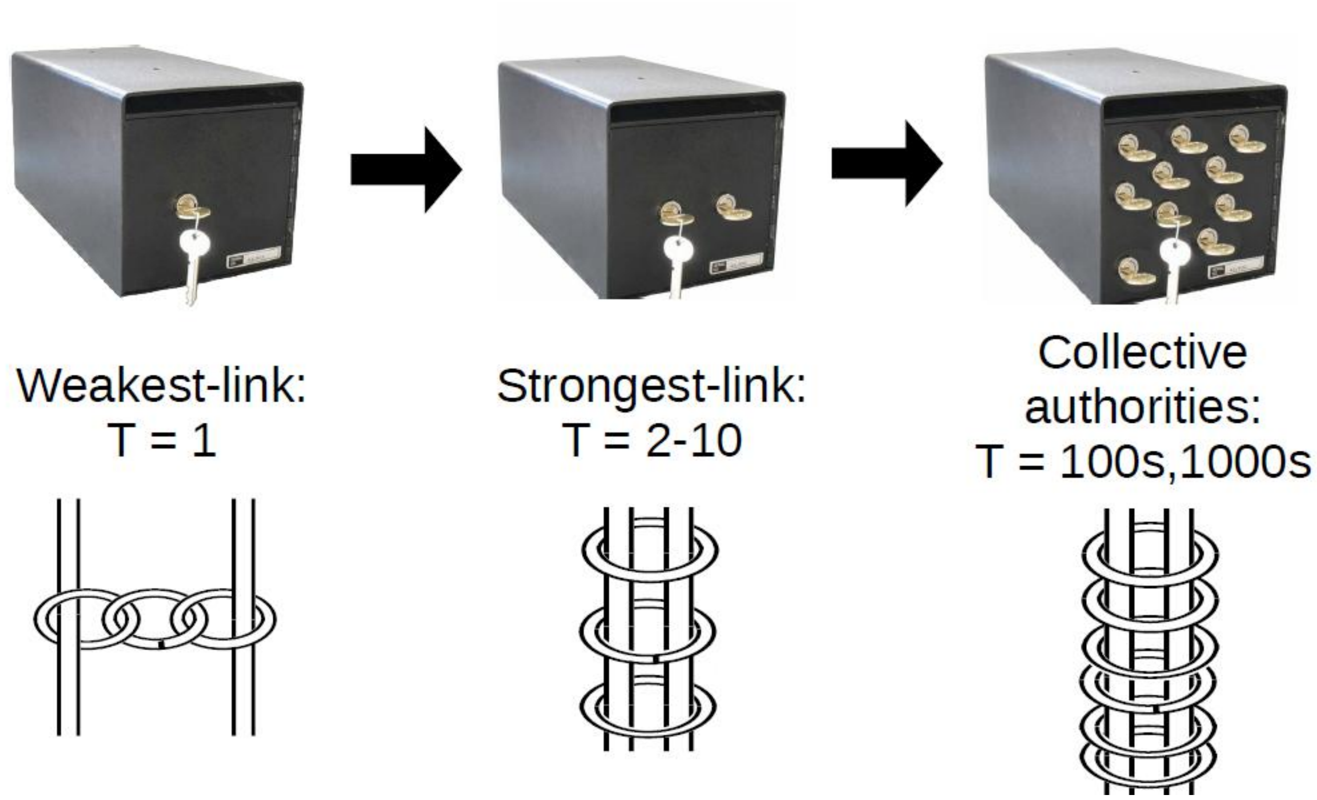
CA被攻击，整个Web访问都不安全

- 强链接：一个环节攻破，不影响整体安全性



解决思路：分散权威

- 将一个权威分散到不同的独立实体



例子

- Tor网络的目录服务器

Ref : https://jordan-wright.com/blog/2015/05/14/how-tor-works-part-three-the-consensus/

技术细节：Schnorr签名

- Generator *g* of prime order *q* group
- Public/private key pair: ($K=g^k$, k)

	Signer		Verifier
Commitment	$V=g^v$	$\longrightarrow$	V
Challenge	c	$\longleftarrow$	c = H(M|V)
Response	r = (v – kc)	$\longrightarrow$	r

Signature on M: (c, r)

Commitment recovery	$V' = g^r K^c \; = g^{v-kc} g^{kc} = g^v = V$
Challenge recovery	c’ = H(M|V’)
Decision	c’ = c ?

群签名

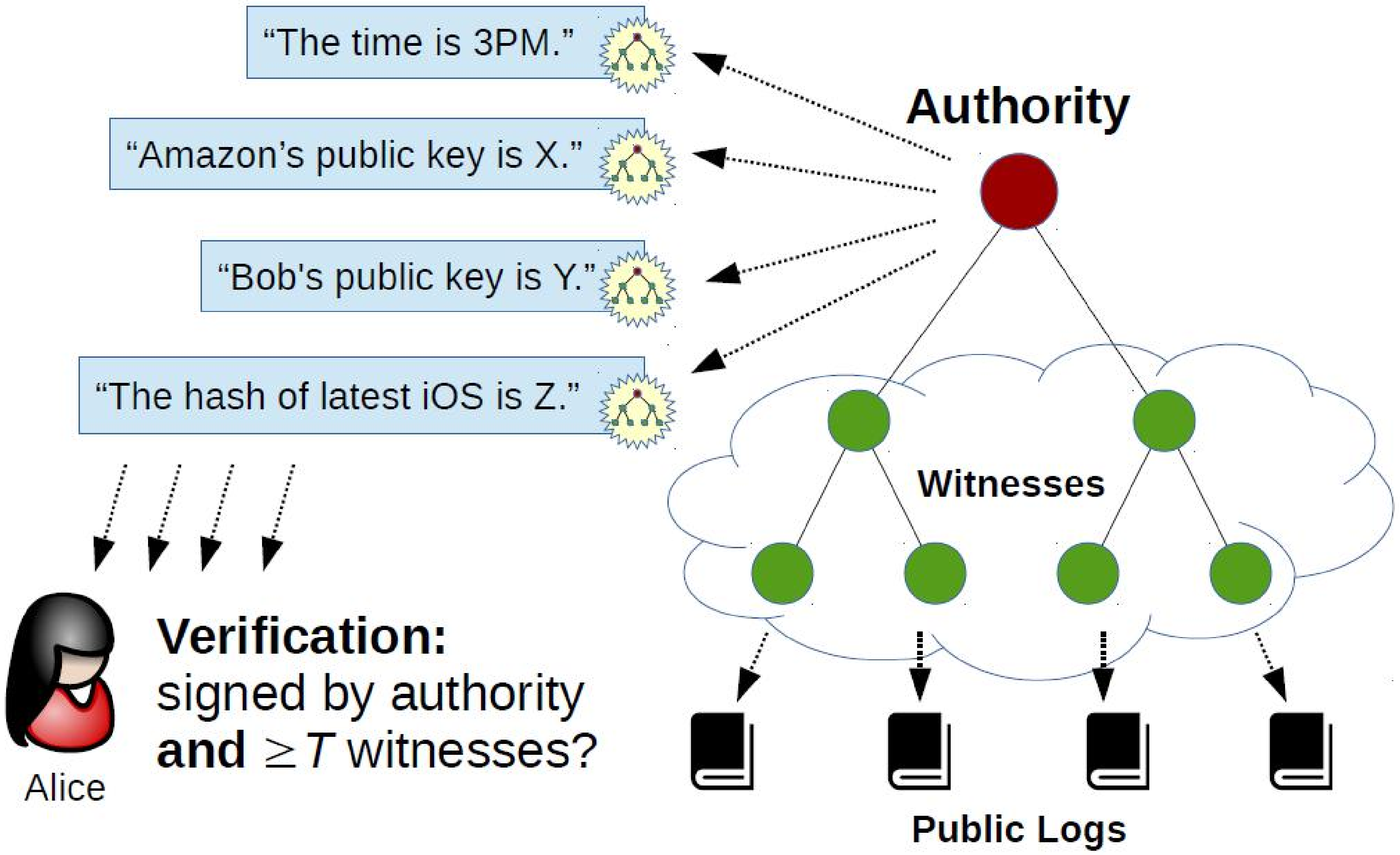
- Key pairs: $(K_1=g^{k_1}, k_1)$ and $(K_2=g^{k_2}, k_2)$

	Signer 1	Signer 2		Verifier
Commitment	$V_1=g^{v_1}$	$V_2=g^{v_2}$	$\longrightarrow$	$V_1 \quad V_2 \quad V=V_1{}^*V_2$
Challenge	c		$\longleftarrow$	$c = H(M\|V)$
Response	$r_1 = (v_1 - k_1c)$	$r_2 = (v_2 - k_2c)$	$\longrightarrow$	$r_1 \quad r_2 \quad r=r_1+r_2$

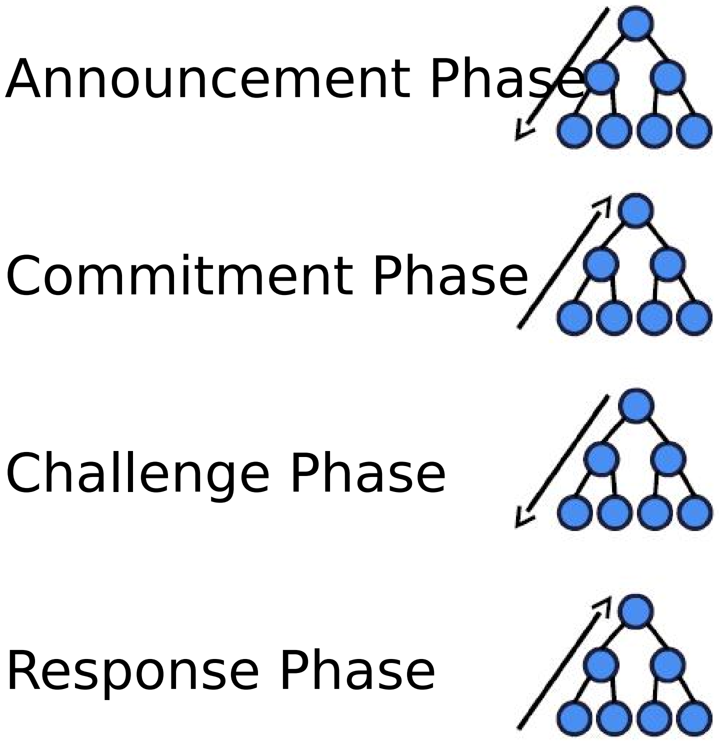
Signature on M: (c, r) Same signature!

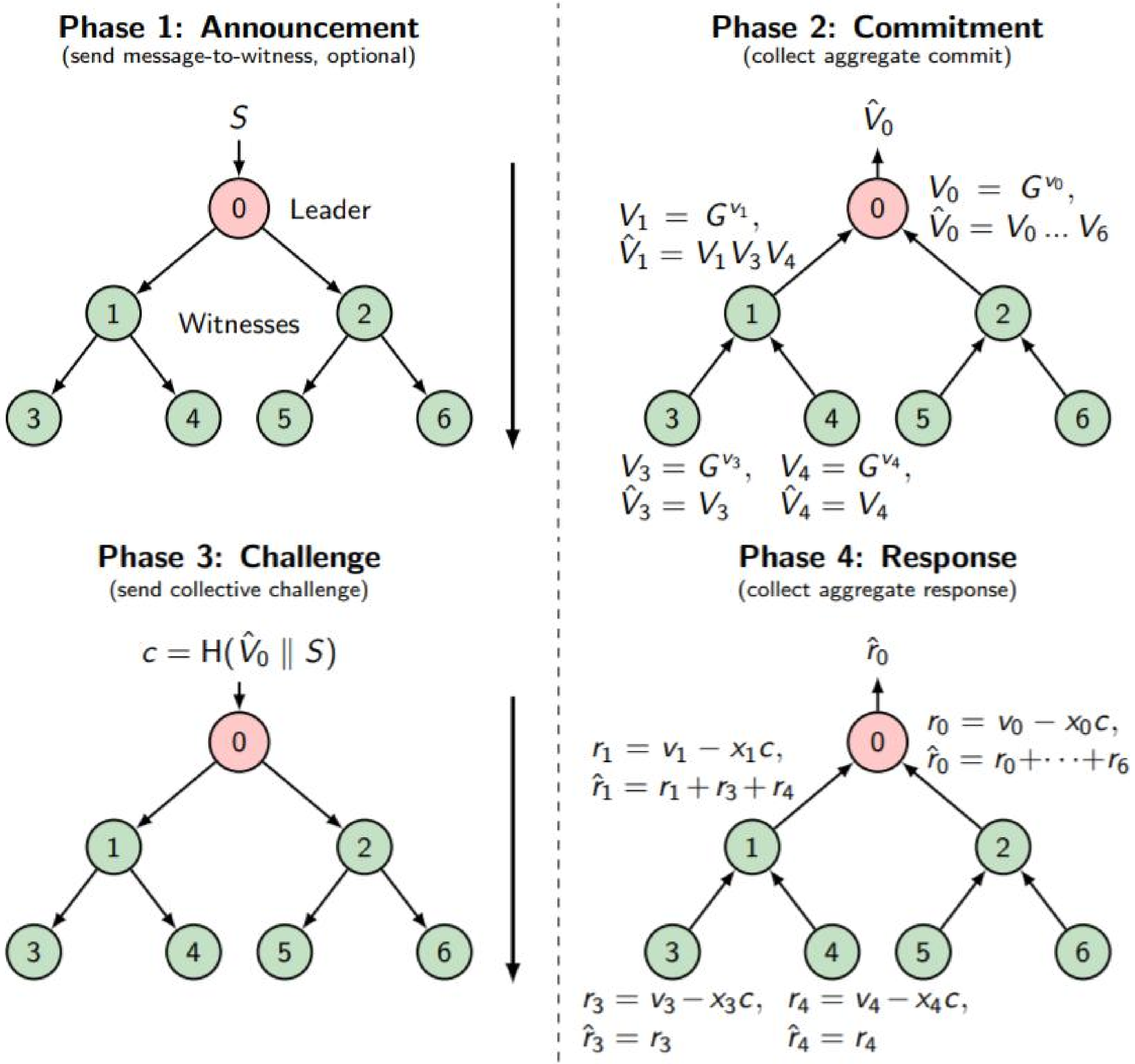
Commitment recovery	Same verification!	$V' = g^rK^c$	$K=K_1{}^*K_2$
Challenge recovery	Done once!	$c' = H(M\|V')$	
Decision		$c' = c$?	

关键贡献：树形结构



协议流程



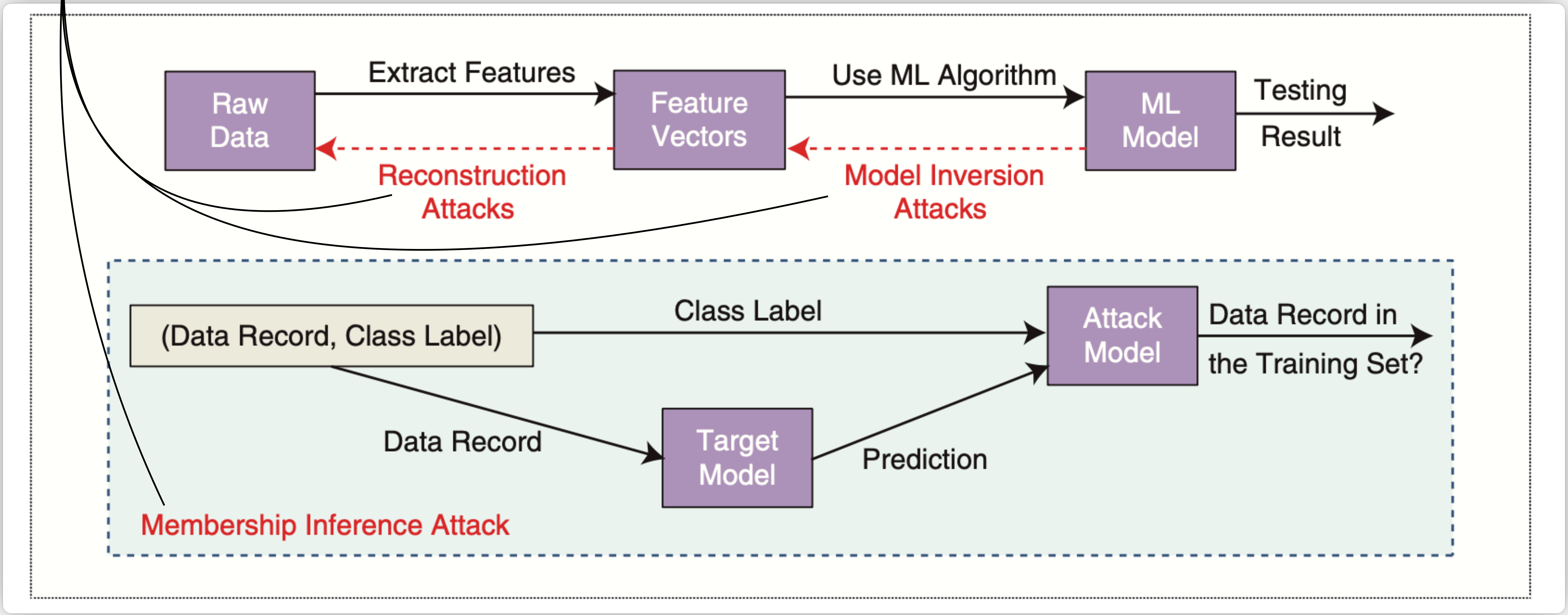


Phase 1: Announcement
(send message-to-witness, optional)
S
Leader
Witnesses
0
1
2
3
4
5
6
Phase 2: Commitment
(collect aggregate commit)
$\hat{V}_0$
$V_0 = G^{v_0}$, $\hat{V}_0 = V_0 \dots V_6$
$V_1 = G^{v_1}$, $\hat{V}_1 = V_1 V_3 V_4$
$V_3 = G^{v_3}$, $\hat{V}_3 = V_3$
$V_4 = G^{v_4}$, $\hat{V}_4 = V_4$
Phase 3: Challenge
(send collective challenge)
$c = \mathrm{H}(\hat{V}_0 \parallel S)$
Phase 4: Response
(collect aggregate response)
$\hat{r}_0$
$r_0 = v_0 - x_0 c$, $\hat{r}_0 = r_0 + \cdots + r_6$
$r_1 = v_1 - x_1 c$, $\hat{r}_1 = r_1 + r_3 + r_4$
$r_3 = v_3 - x_3 c$, $\hat{r}_3 = r_3$
$r_4 = v_4 - x_4 c$, $\hat{r}_4 = r_4$

谢谢！

论文标题：**Privacy-Preserving Machine Learning : Threats and Solutions**

威胁



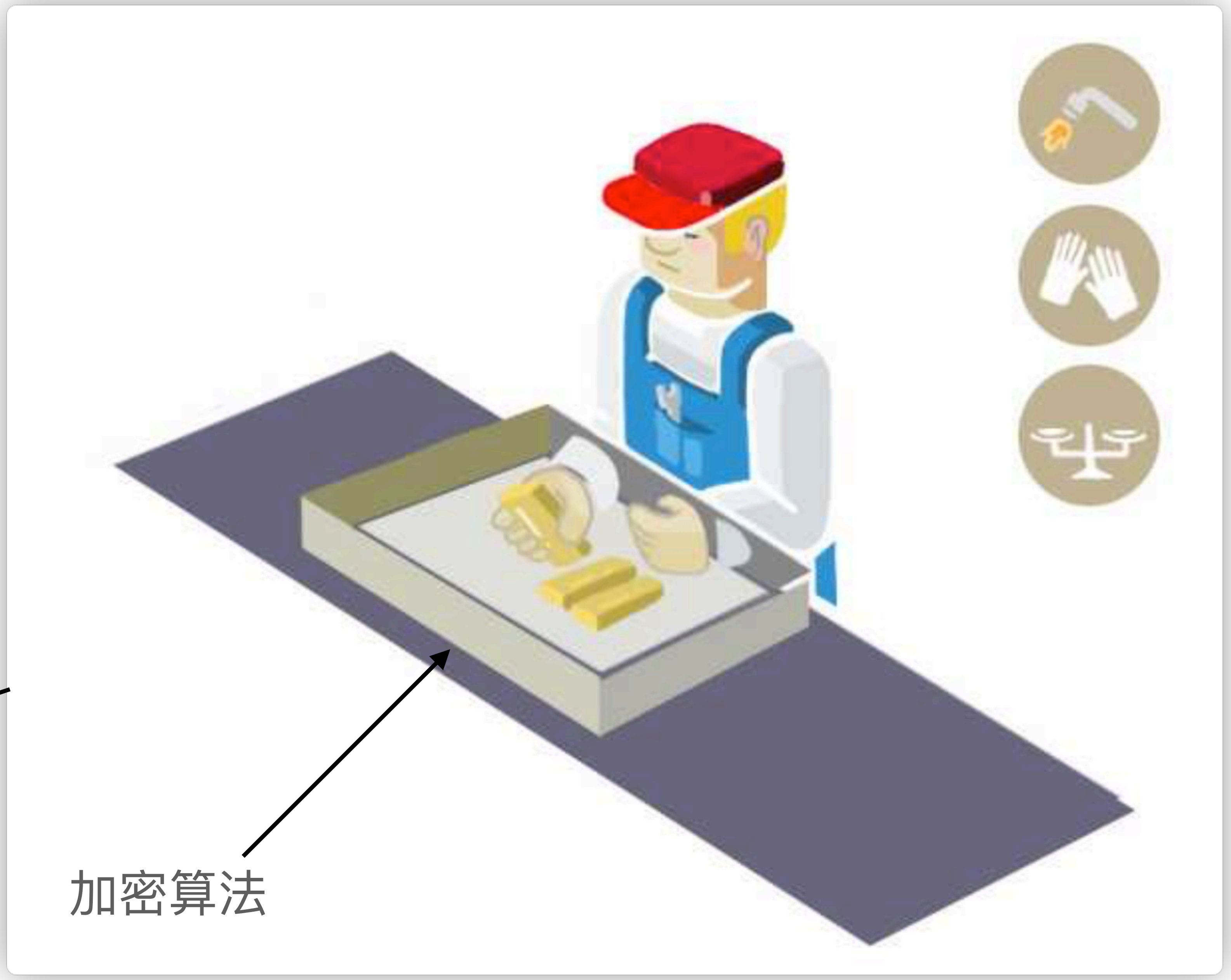
论文标题：**Privacy-Preserving Machine Learning : Threats and Solutions**

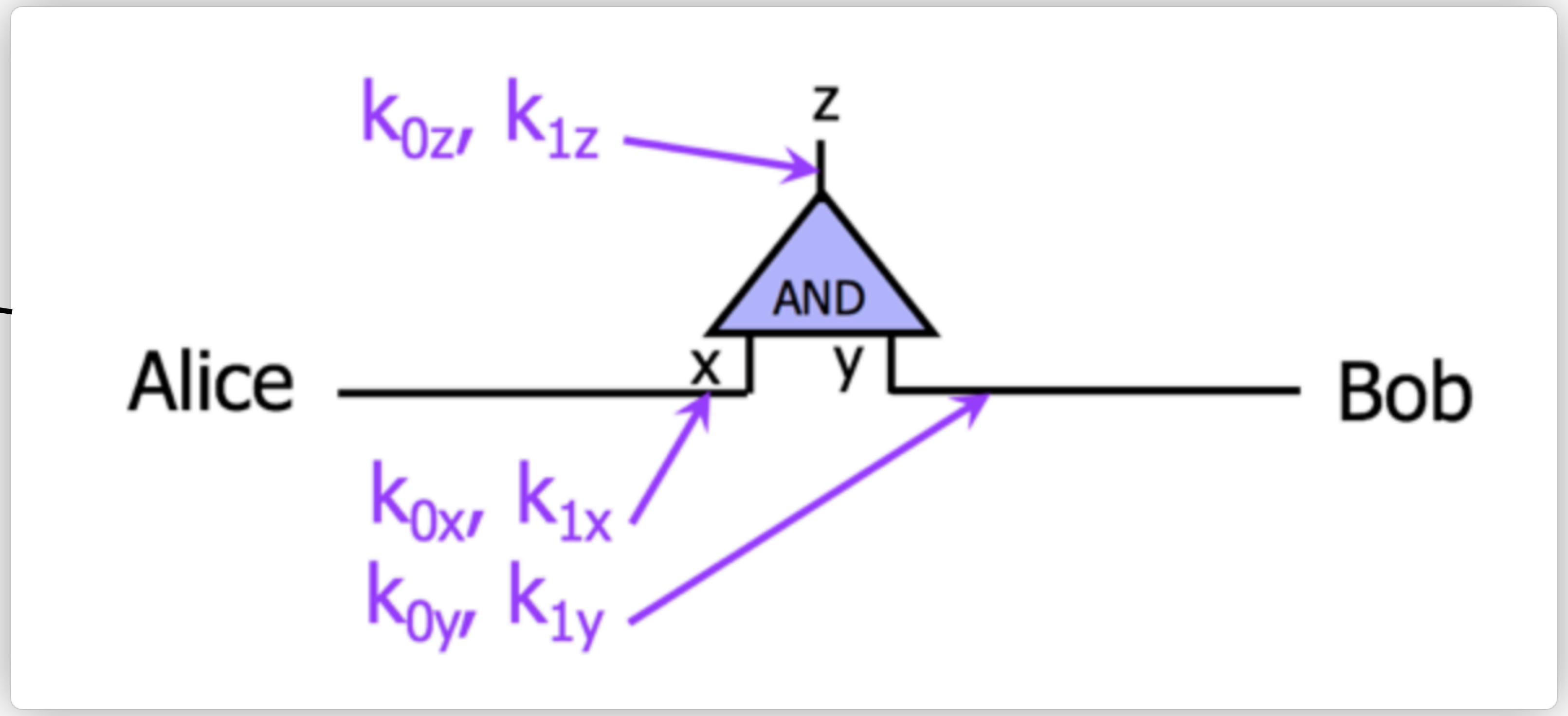
$f(A) + f(B) = f(A + B)$

$f(A) \times f(B) = f(A \times B)$

机器学习中用于隐私保护的密码学技术

- 密钥共享 Secret sharing
- 安全处理器 Secure processors
- 同态加密 Homomorphic encryption
- 乱码电路 Garbled circuits





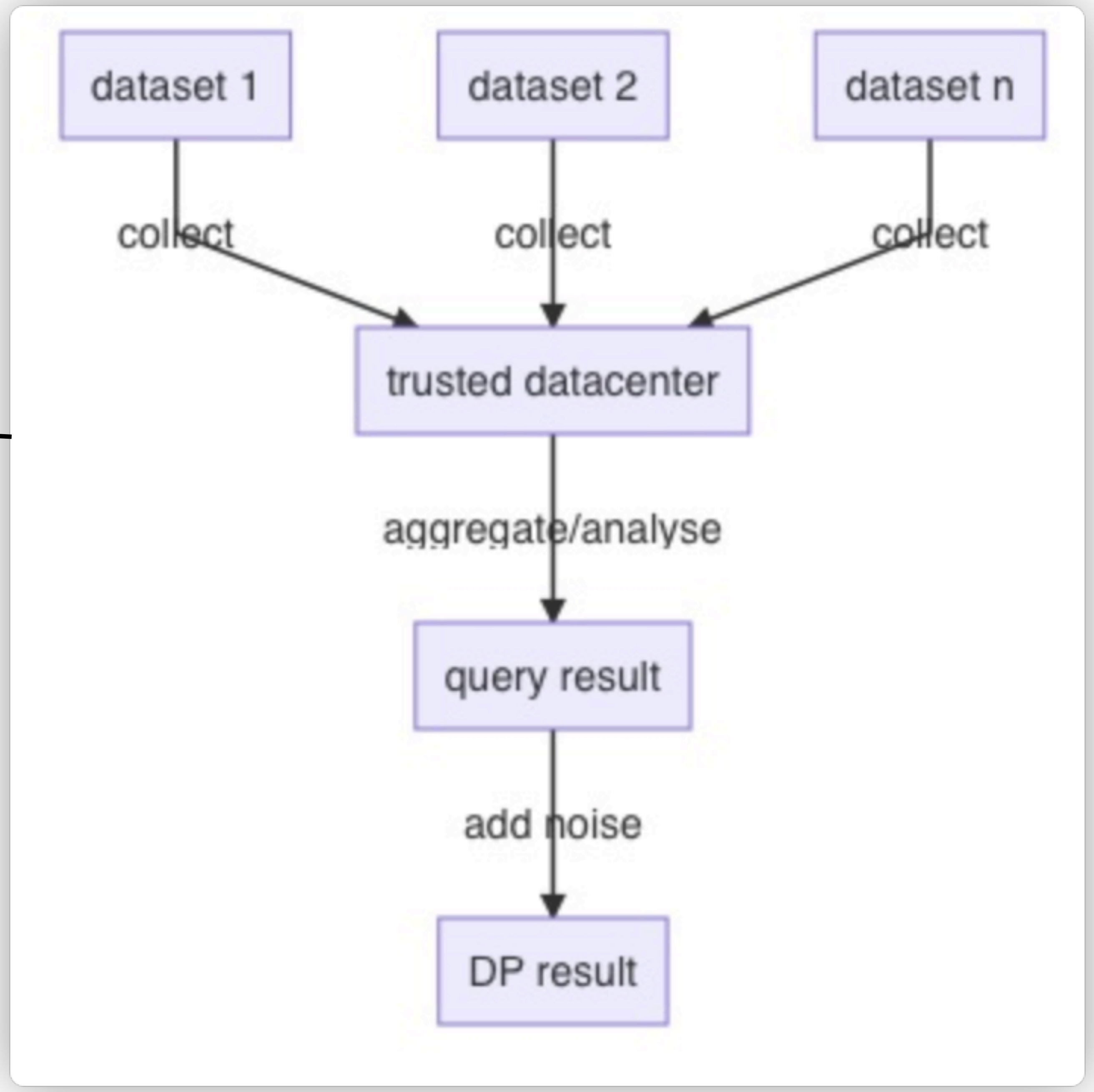
参考链接：

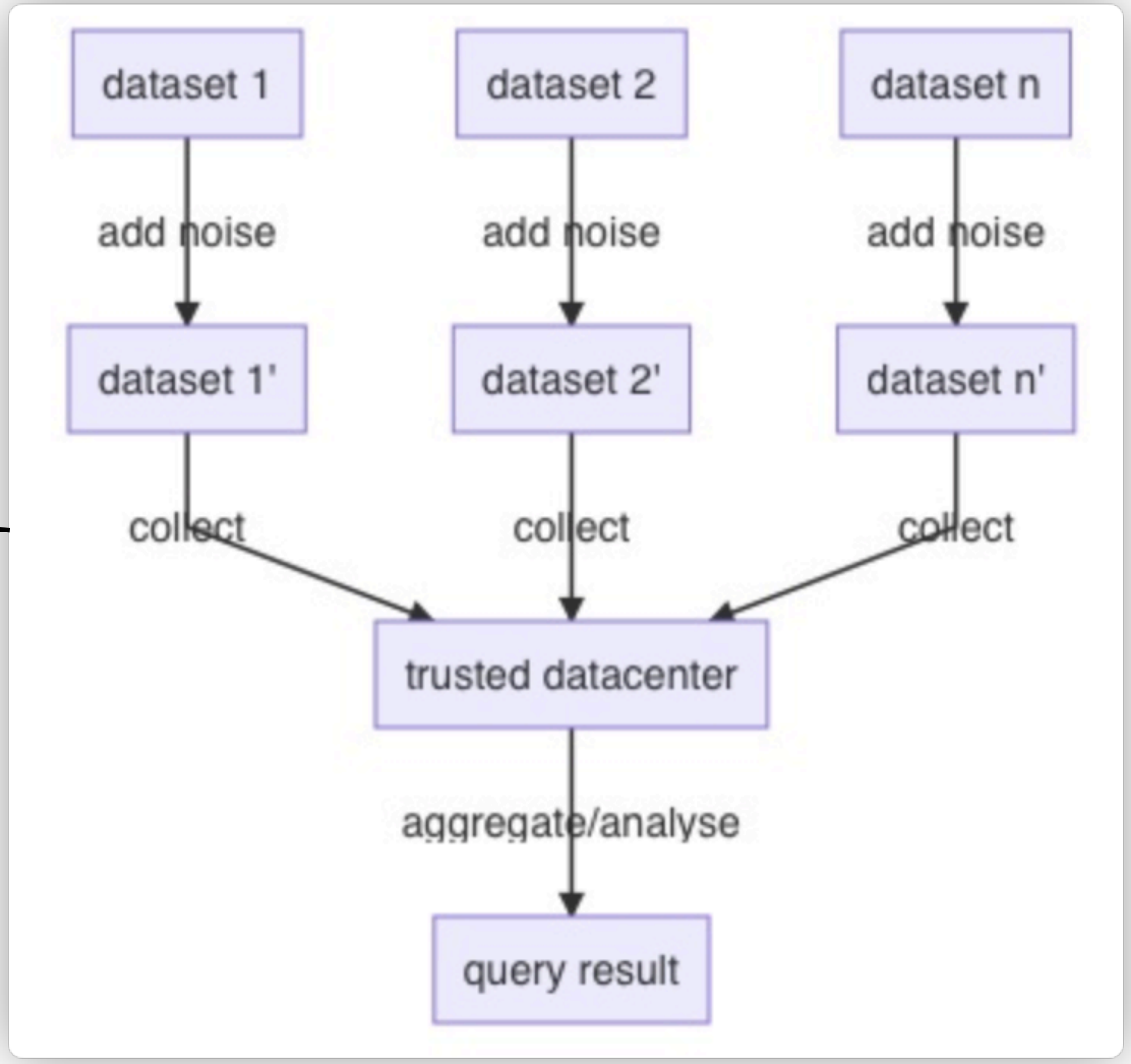
密钥共享 https://zhuanlan.zhihu.com/p/44999983

同态加密 https://www.zhihu.com/question/27645858/answer/37598506
https://zhuanlan.zhihu.com/p/31822335

乱码电路 https://zhuanlan.zhihu.com/p/41172002
https://zhuanlan.zhihu.com/p/25986848

论文标题：Privacy-Preserving Machine Learning : Threats and Solutions





参考链接：

差分隐私 https://zhuanlan.zhihu.com/p/266587002

https://zhuanlan.zhihu.com/p/139114240