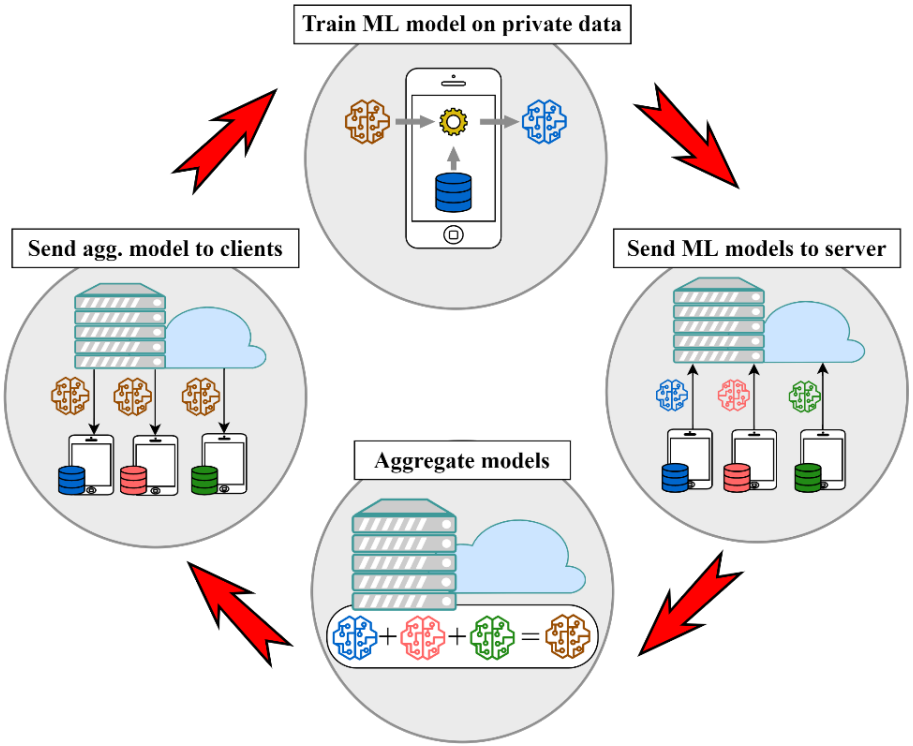
Kripto-Tabanlı Federe Öğrenme

Melek Önen

6 Mayıs 2023

EURECOM
Sophia Antipolis

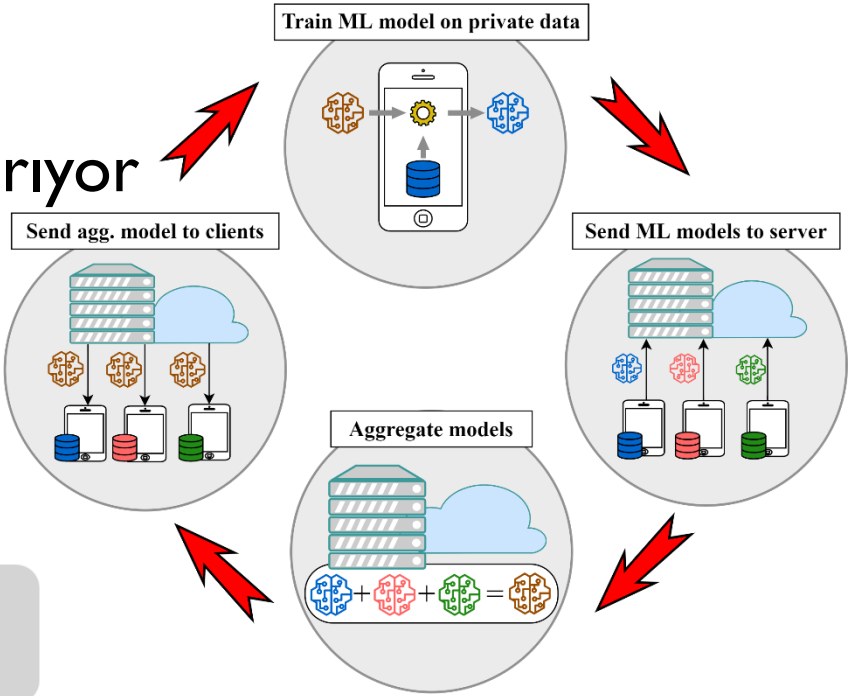
Federe Öğrenme (FÖ)

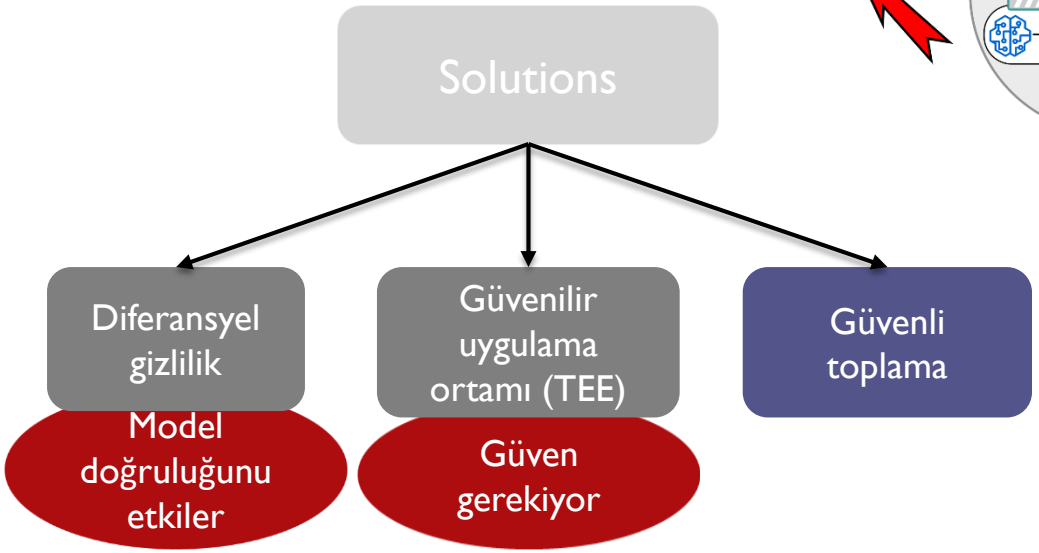


- Amaç: Makine öğrenme modeli eğitmek
  - birçok gizli veri kümesi
    - n FÖ istemci
    - 1 FÖ sunucusu (aggregator=

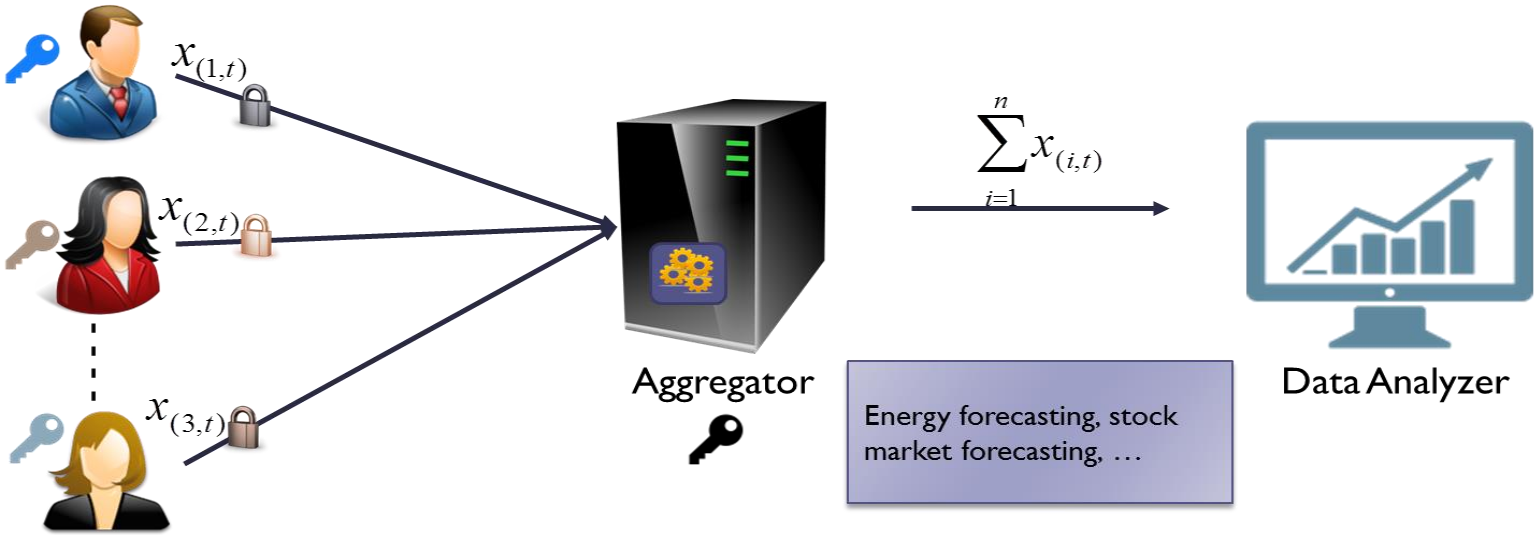
FÖ ve üyelik çıkarım saldırıları

- Dürüst-ama-meraklı sunucu
- Lokal eğitilmiş modeller bilgi sızdırıyor





Güvenli toplama (Secure aggregation)

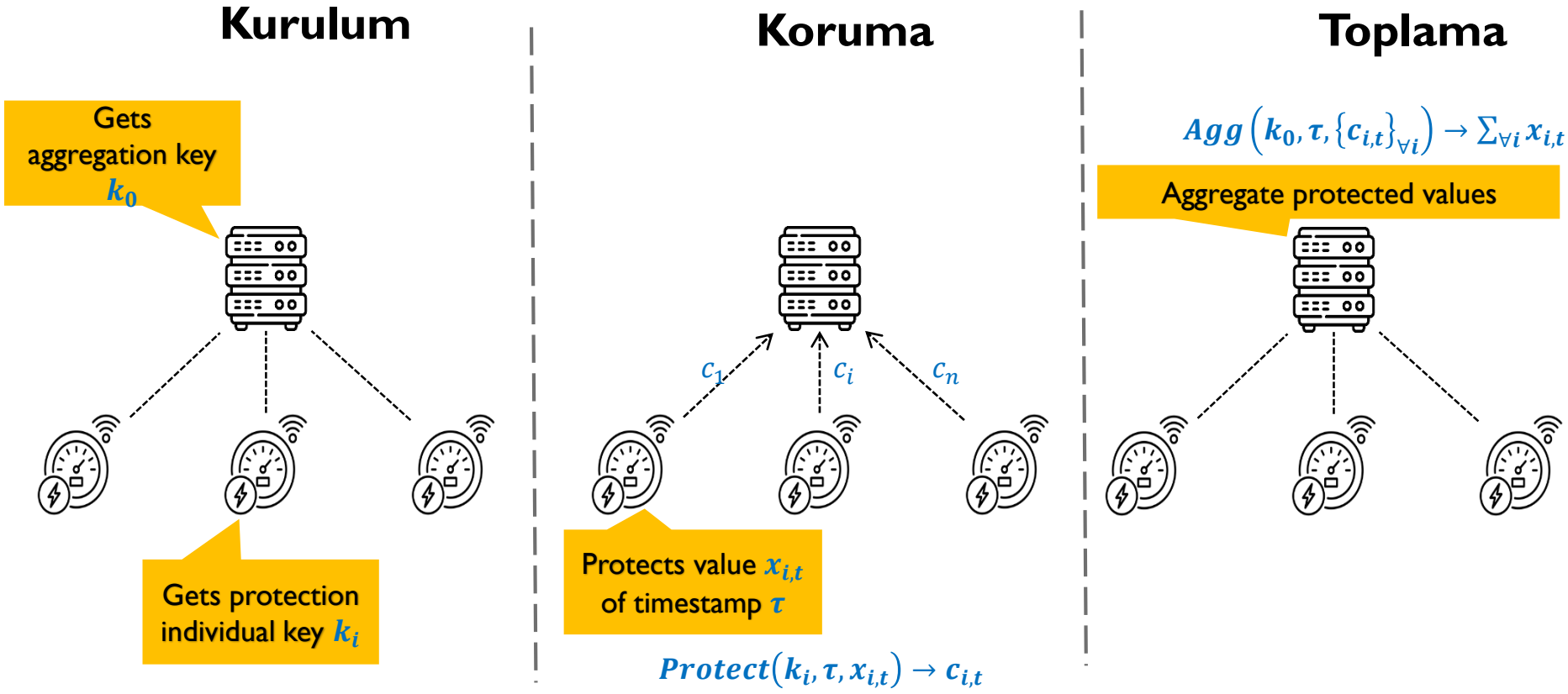


- Tehdit modelleri & Güvenlik gereksinimleri
  - Dürüst-ama-Meraklı (honest-but-curious)
    - Toplayıcı kayıtsızlığı (Aggregator Obliviousness)

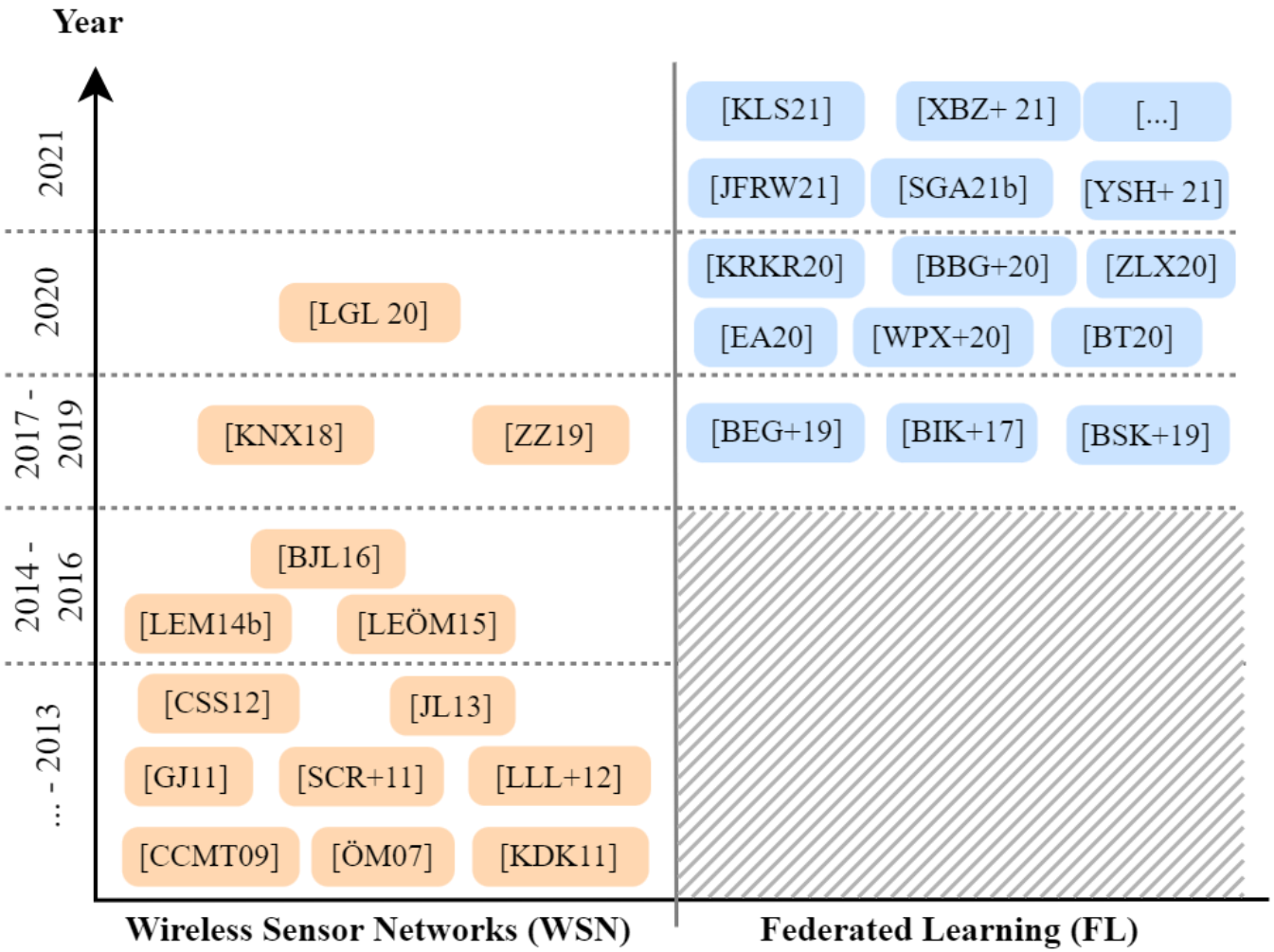
      Toplayıcı bireysel girdilerden hiç bir şey öğrenemez
  - Kötü niyetli model (Malicious model)
    - Sonuç değiştirilemezlik (Result unforgeability)

      Toplayıcı toplam sonucunu değiştiremez

Güvenli toplama aşamaları

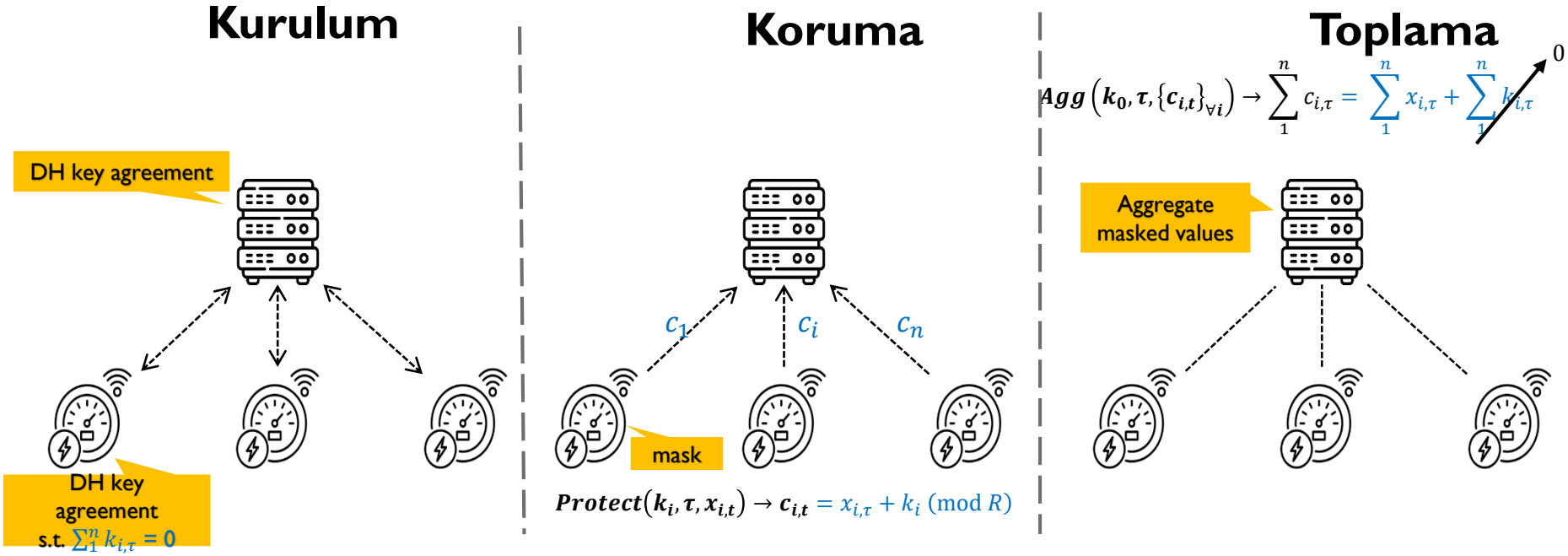


Varolan araştırmalar



Kripto tabanlı güvenli toplama

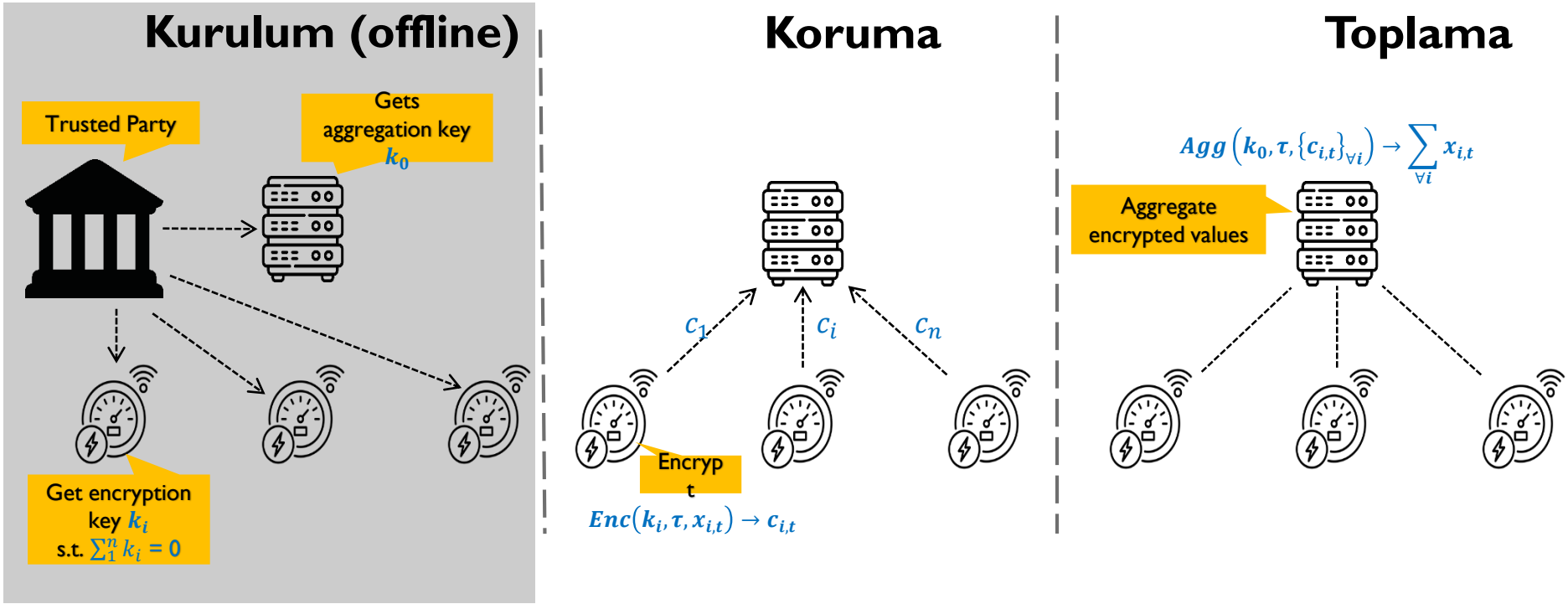
Maskeleme-tabanlı güvenli toplama



Artıları: basit işlemler ☺

Eksileri: kurulum aşaması karmaşık (quadratic) ☹

Homomorfik şifreleme tabanlı güvenli toplama



Artıları: uzun ömürlü anahtarlar ☺

Eksileri: Dinamik istemcilere uygun değil ☹

Konu başlıkları

- Homomorfik şifreleme tabanlı güvenli toplama
- Kötü niyetli istemcilerle federe öğrenme
- İstemci terkleri altında federe öğrenme

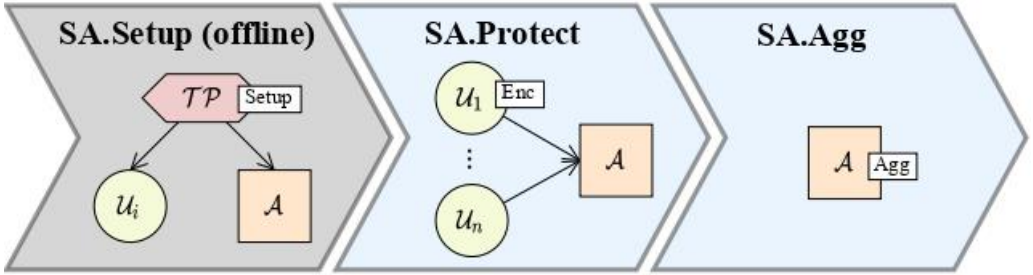
Homomorfik şifreleme tabanlı güvenli toplama

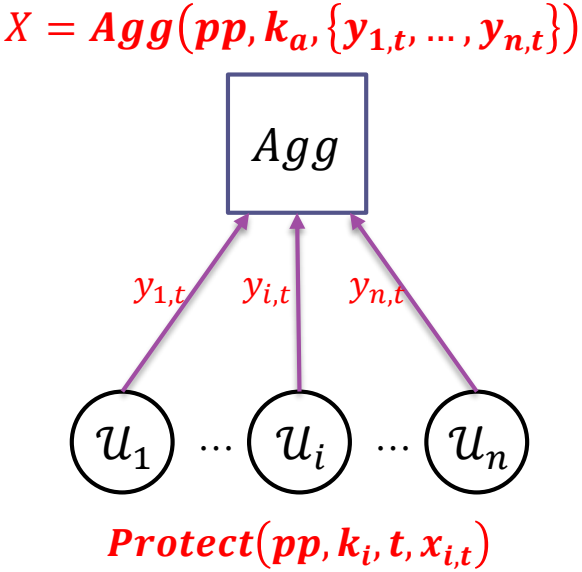
AHE

- Birçok istemci ile homomorfik şifreleme
  - $(k_0, \{k_i\}_{\forall i \in U}, pp) \leftarrow \boldsymbol{AHE.Setup}(\lambda)$: Anahtarların ve genel parametrelerin kurulumu
  - $y_{i,\tau} \leftarrow \boldsymbol{AHE.Enc}(pp, k_i, \tau, x_{i,\tau})$: $x_{i,\tau}$ mesajı k_i anahatarıyla τ birim zamanında şifrelenir
  - $X_\tau \leftarrow \boldsymbol{AHE.Agg}\left(pp, k_0, \{y_{i,\tau}\}_{\forall i \in U}\right)$: şifreli toplam hesaplanır ve deşifre anahtarı k_0 'la deşifrelenir

Homomorfik şifreleme tabanlı güvenli toplama

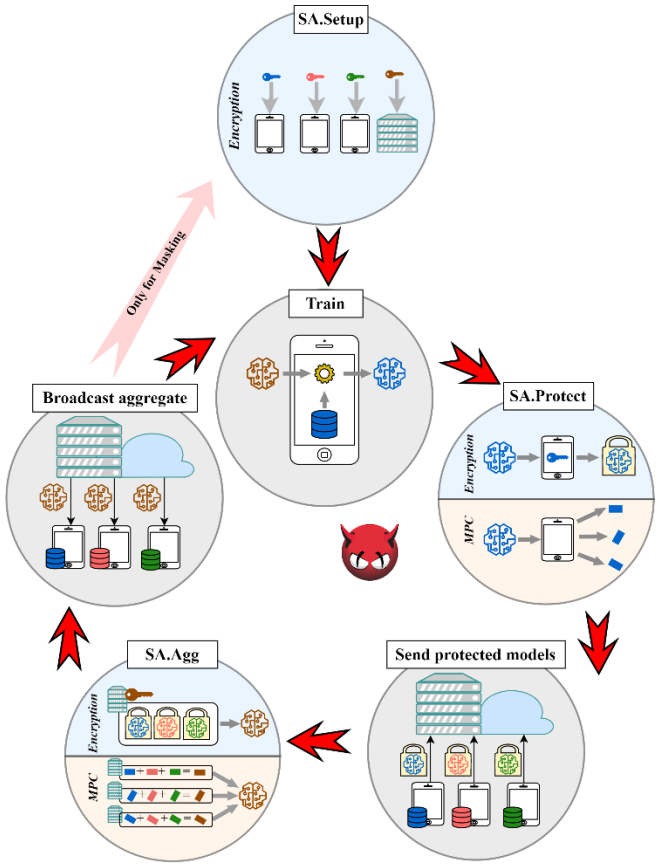
- Joye-Libert şeması
  - $\boldsymbol{Setup}(\boldsymbol{\lambda})$: N = pq, hash H, anahtar k_i, $\sum_i k_i = -k_a$
  - $\boldsymbol{Protect}(\boldsymbol{pp}, \boldsymbol{k_i}, \boldsymbol{t}, \boldsymbol{x_{i,t}})$: $y_{i,t} = (1 + x_{i,t}N)H(t)^{k_i} \bmod N^2$
  - $\boldsymbol{Agg}\left(\boldsymbol{pp}, \boldsymbol{k_a}, \{\boldsymbol{y_{i,t}}\}_{\forall \boldsymbol{i}}\right)$: $X = \frac{H(t)^{k_a} \prod_i y_{i,t} - 1}{N}$



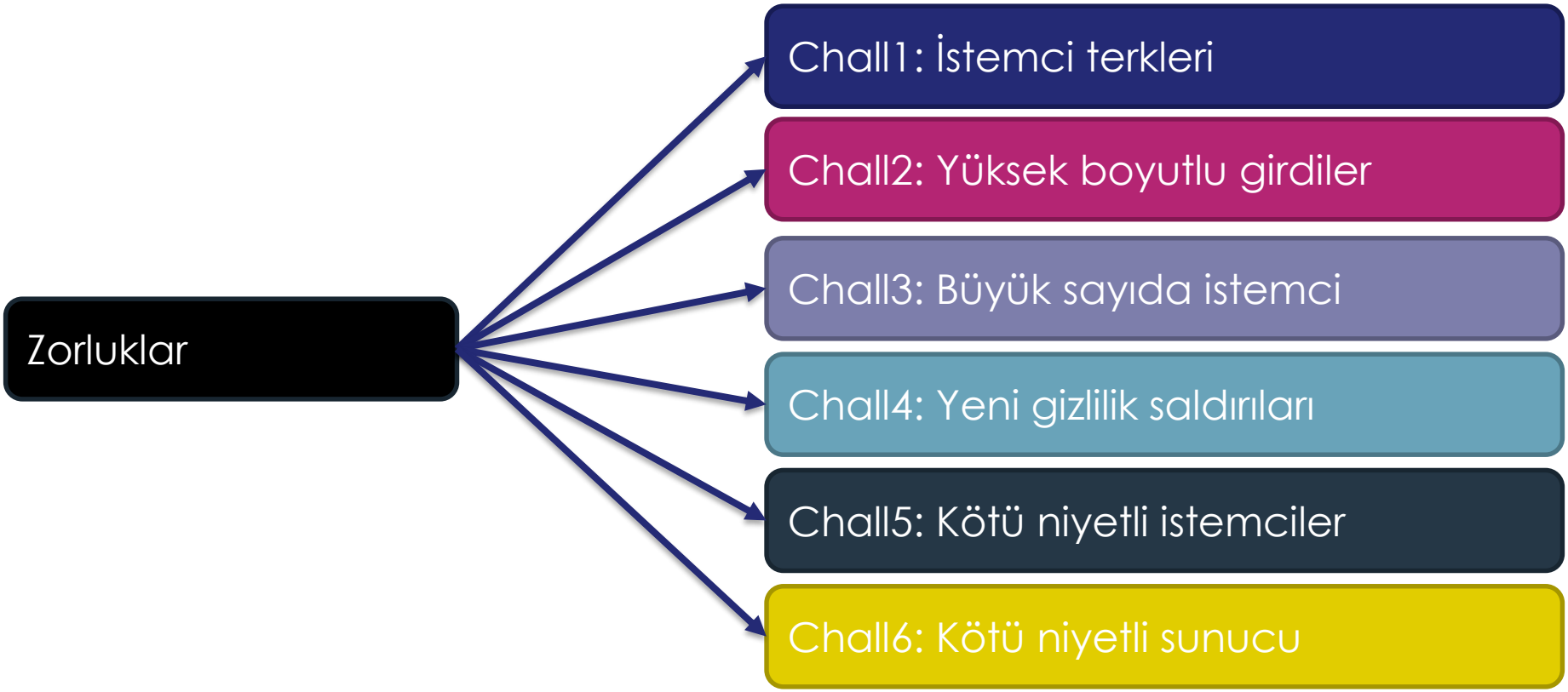


SA to Mitigate Inference Attack on FL

- FL clients: users
- FL server : Aggregator
- Can SA be used, directly?



Federe öğrenme ve güvenli toplama



Yüksek boyutlu girdiler

Chall2: Yüksek boyutlu girdiler

- Girdiler model parametreleri
  - Bunlar yüksek boyutlu vektörler
  - Büyük iletişim ek yükü
- Varolan araştırmalar:
  - Homomorfik şifreleme tabanlı ve grup şifreli [PAH+18], [LCV19], [ZLX+20]
  - Maskeleme tabanlı : quantization bazlı [BSK+19], [EA20]

Chall3: Yüksek sayıda istemci

Chall3: Yüksek sayıda istemci

- Binlerce istemci (cross-device FL)
- Senkronizasyon problemleri

- Varolan araştırmalar:
  - İstemciler grup halinde birleştirme [BEG+19], [BBG+20], [SGA21b], [SMH21], [JNMALC22]
  - Asenkron FÖ [SAGA21]

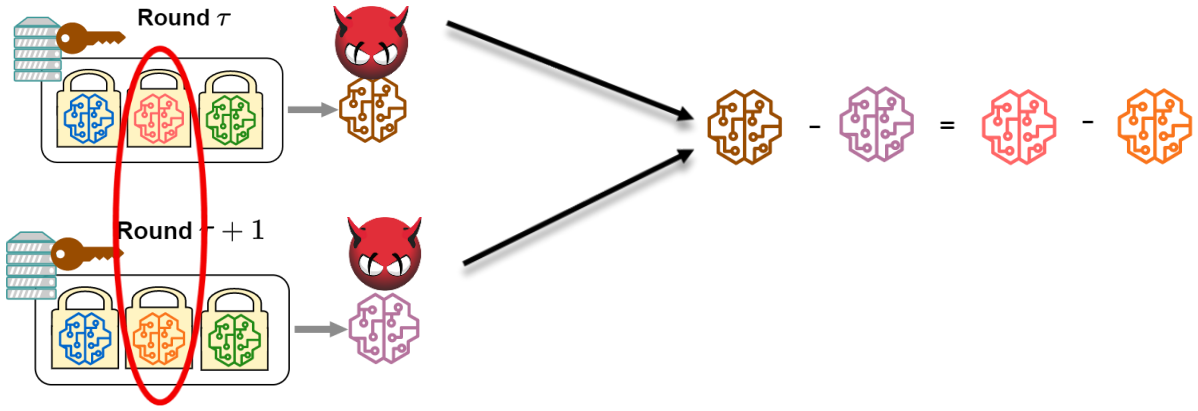
Chall4: Mahremiyet saldırıları

Chall4: Mahremiyet saldırıları

- Geliştirilmiş model public bir model
- Toplam sonucu çoğu zaman korunmuyor
- Bu toplamdan bilgi çıkarımı yapılabilir
  - Üyelik çıkarımları

Varolan araştırmalar

- Dağıtılmış Differential Privacy (DDP) [TBA+19], [SSV+21]
- Toplu bölünme [SAG+21]



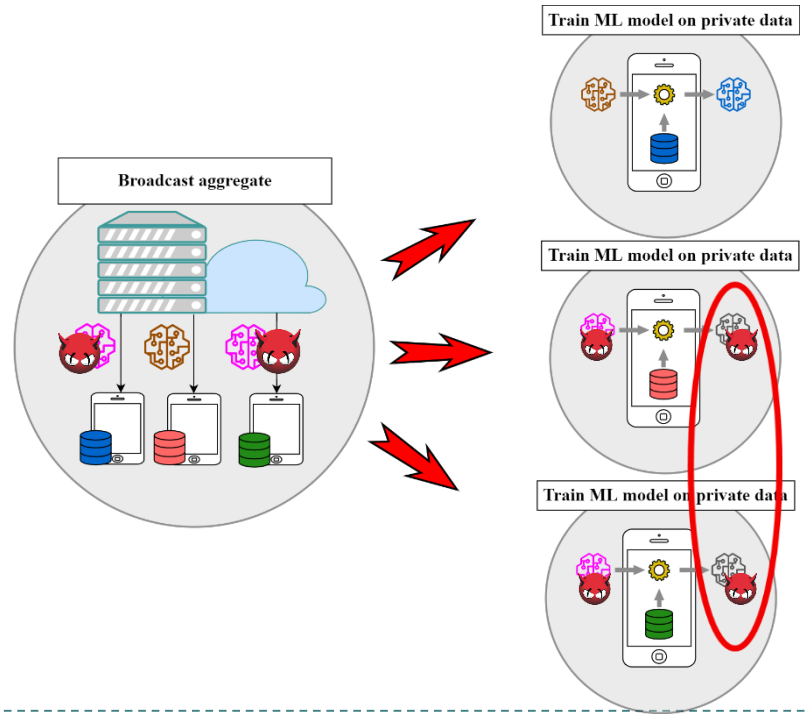
Chall6: Kötü niyetli toplayıcı

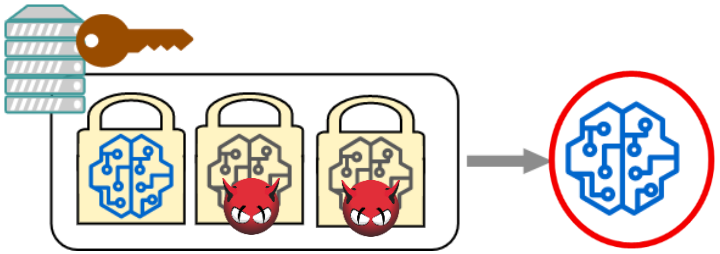
Chall6: Malicious Aggregator

- Kötü niyetli toplayıcı yanlış toplam gönderebilir
  - İstemciler backdoor modeller eğitir

Varolan sonuçlar:

- Homomorfik hash fonksyonlar (HHF) [ZFW+20], [XLL+20]
- Taahhüt şemaları [GLL+21]

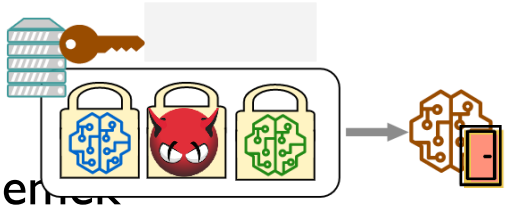


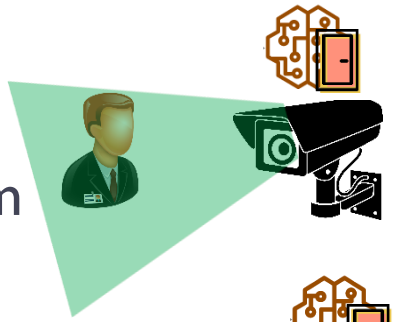


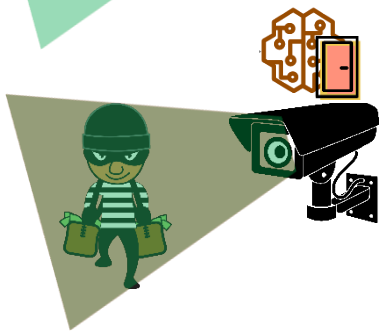
Konu başlıkları

- Homomorfik şifreleme tabanlı güvenli toplama
- Kötü niyetli istemcilerle federe öğrenme
- İstemci terkleri altında federe öğrenme

- Zehirleme saldırıları (poisoning attacks)
  - Saladırı amacı
    - Girdileri değiştirerek model performansını etkilemek
  - Çözüm
    - İstemci girdisini önceden doğrulamak
  - Homomorfik şifreli güvenli toplama ile çözüm
    - OPPRF ile eşik ile karşılaştırma [SACMAT21]

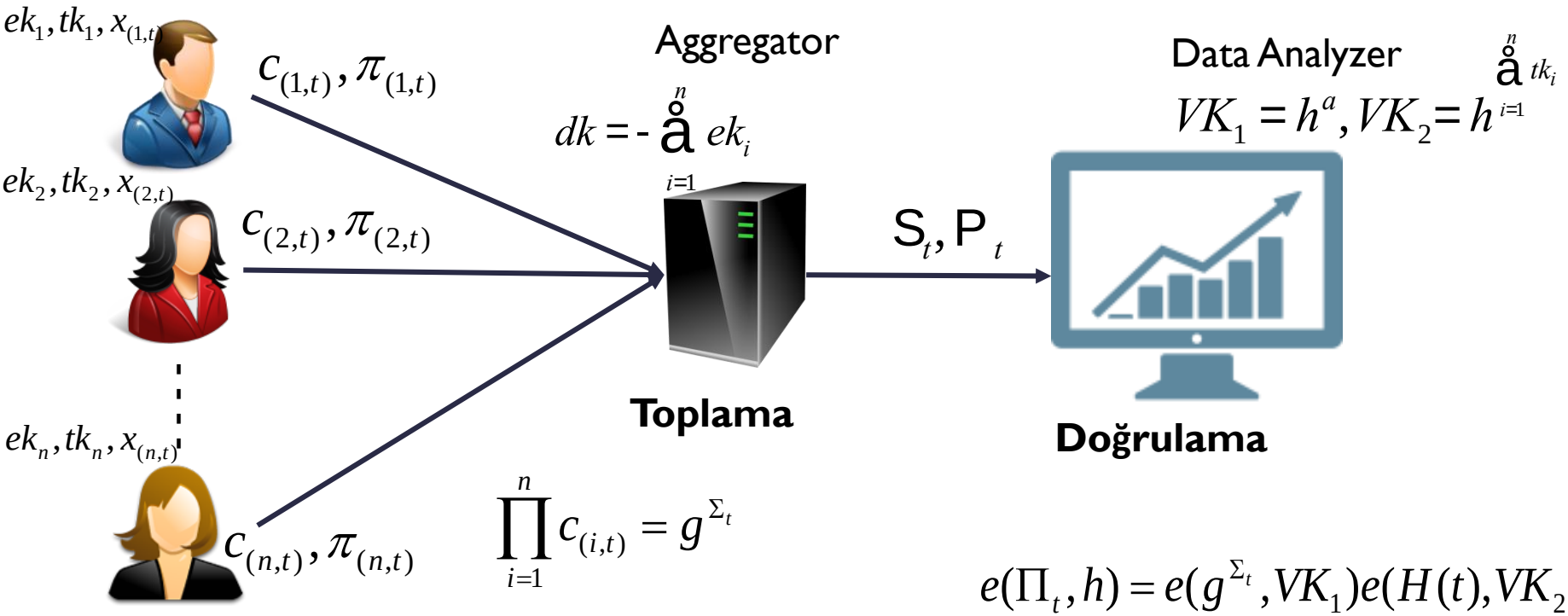






Background - PUDA

[CANS'15]



$$\prod_{i=1}^{n} c_{(i,t)} = g^{\Sigma_t}$$

$$\prod_{i=1}^{n} \pi_{(i,t)} = g^{a\Sigma_t} H(t)^{\Sigma_i tki}$$

$$e(\Pi_t, h) = e(g^{\Sigma_t}, VK_1) e(H(t), VK_2)$$

$$c_{(i,t)} = g^{x_{(i,t)}} H(t)^{ek_i}$$
$$\rho_{(i,t)} = g^{ax_{(i,t)}} H(t)^{tk_i}$$

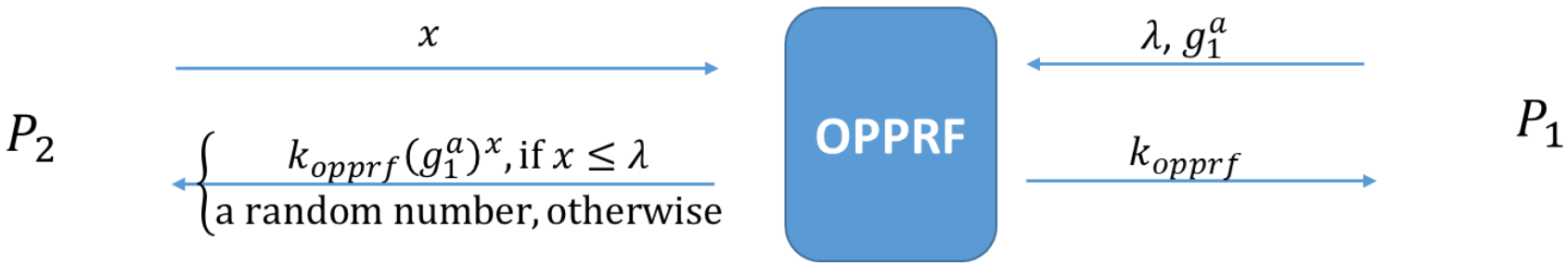
⇒ **Toplamı deşifreleme**

⇒ **Genel doğrulama**

⇒ **Sabit zamanda doğrulama**

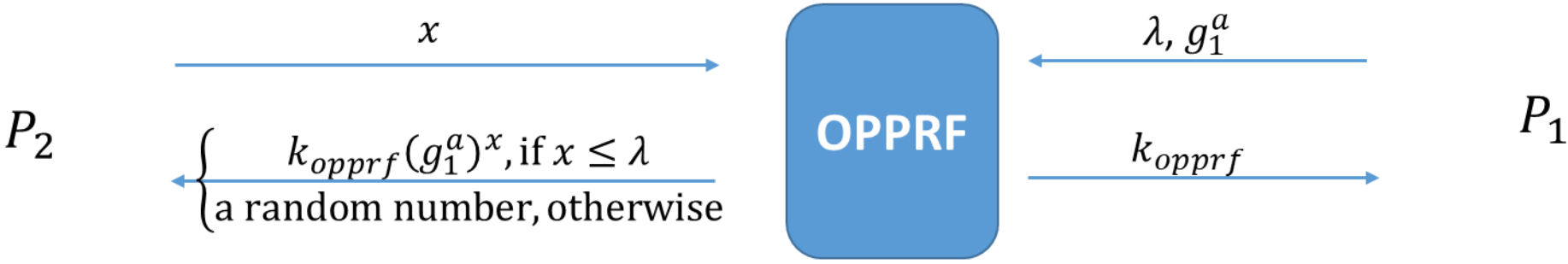
Background - OPPRF

Parameters: cyclic groups $\mathbb{G}_1$ with a generator g_1

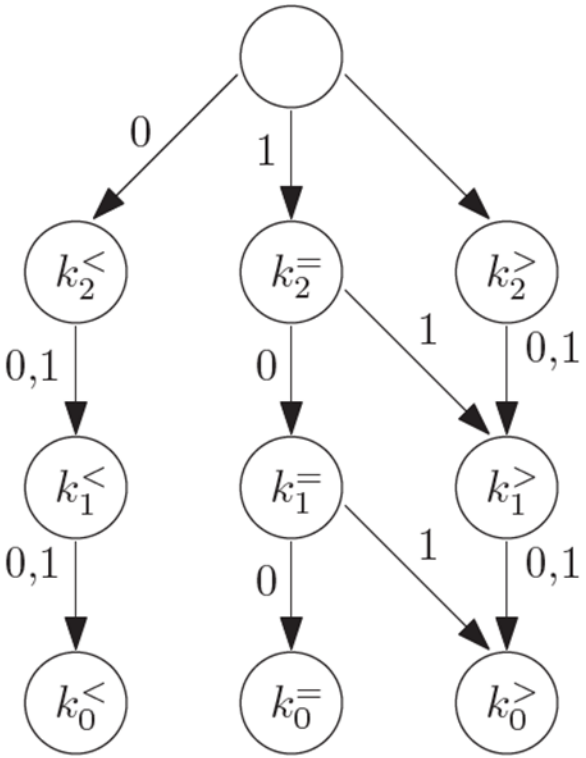


Çözüm

[SACMAT'21]

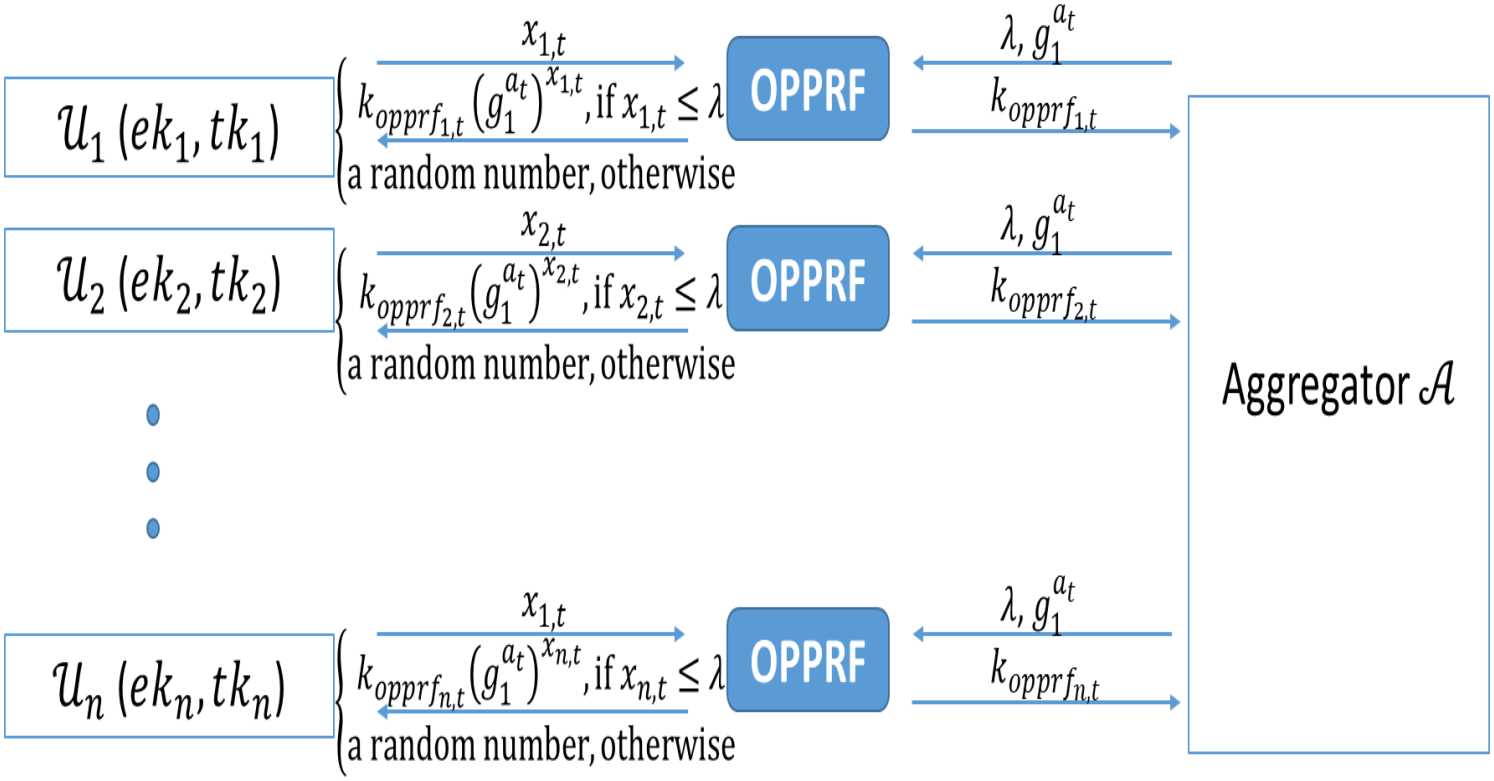


- We modify the Ciampi and Orlandi PSM protocol to adapt it to our case
  - Example oblivious graph tracing for $\lambda = 100$
  - In the end of graph tracing, $k_0^<$, $k_0^=$ or $k_0^>$ is learned by P_2 if $x < \lambda$, $x = \lambda$ or $x > \lambda$, respectively.
  - P_1 sends two encryptions of $k_{opprf}(g_1^a)^x$ under the keys $k_0^<$ and $k_0^=$. P_1 also sends encryption of a random number under the key $k_0^>$.
  - P_2 will be able to get $k_{opprf}(g_1^a)^x$ only when $x \leq \lambda$ that means can compute the integrity tag when when $x \leq \lambda$.



EURECOM
Sophia Antipolis

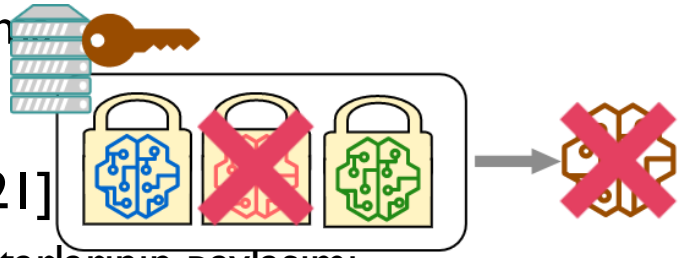
OPPRF ile güvenli toplama



Konu başlıkları

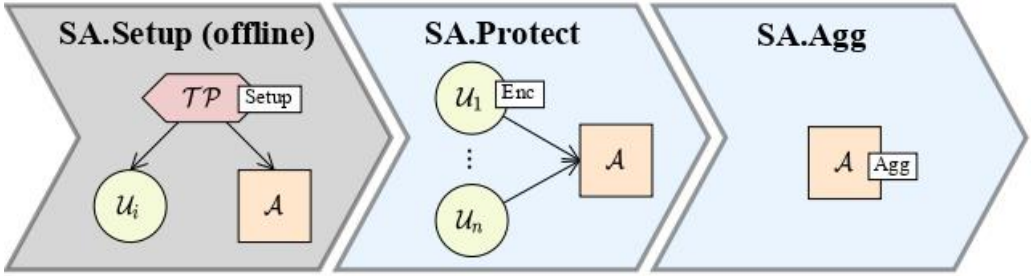
- Homomorfik şifreleme tabanlı güvenli toplama
- Kötü niyetli istemcilerle federe öğrenme
- İstemci terkleri altında federe öğrenme

- İstemci terklerinde toplama gerçekleşemez
  - Sunucu anahtarı istemci anahtarlarına bağlı
    - Maskeleme-tabanlı güvenli toplama
    - Homomorfik şifreleme tabanlı güvenli toplam
  - Varolan araştırmalar
    - Maskeleme ile : [BIK+17], [SSV+21], [YSH+21]
      - Sır paylaşma tabanlı (secret sharing): DH anahtarlarının paylaşımı
      - Bir istemci terkinde, o istemcinin maskesi hesaplanır

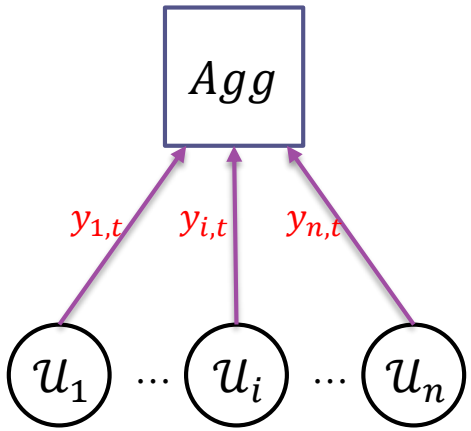


Homomorfik şifreleme tabanlı güvenli toplama

- Joye-Libert şeması
  - $\boldsymbol{Setup}(\boldsymbol{\lambda})$: N = pq, hash H, anahtar k_i, $\sum_i k_i = -k_a$
  - $\boldsymbol{Protect}(\boldsymbol{pp}, \boldsymbol{k_i}, \boldsymbol{t}, \boldsymbol{x_{i,t}})$: $y_{i,t} = (1 + x_{i,t}N)H(t)^{k_i} \bmod N^2$
  - $\boldsymbol{Agg}\left(\boldsymbol{pp}, \boldsymbol{k_a}, \{\boldsymbol{y_{i,t}}\}_{\forall \boldsymbol{i}}\right)$: $X = \frac{H(t)^{k_a} \prod_i y_{i,t} - 1}{N}$



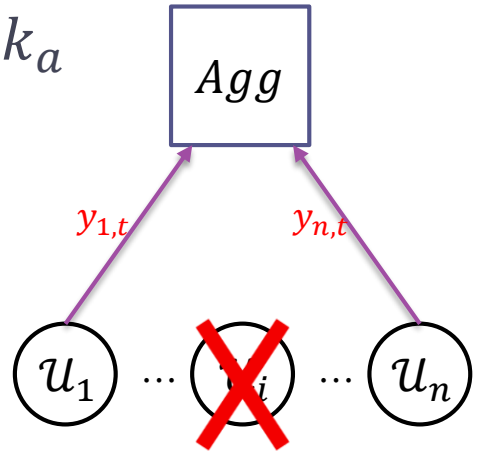
$X = \boldsymbol{Agg}(\boldsymbol{pp}, \boldsymbol{k_a}, \{\boldsymbol{y_{1,t}}, \dots, \boldsymbol{y_{n,t}}\})$



$\boldsymbol{Protect}(\boldsymbol{pp}, \boldsymbol{k_i}, \boldsymbol{t}, \boldsymbol{x_{i,t}})$

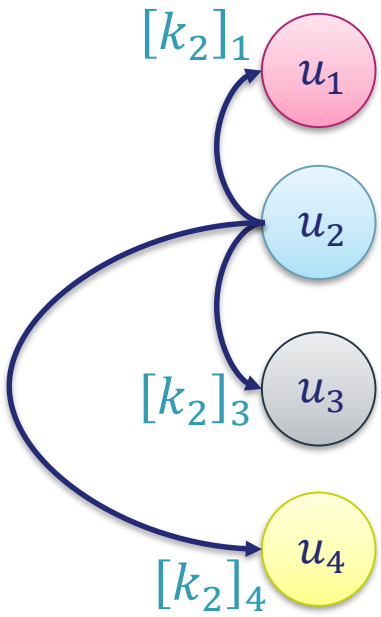
Threshold Joye-Libert Secure Aggregation [ACSAC'22]

- JL hata toleranslı değil
  - Bir veya birçok istemci terkeder ise: $\sum_i k_i \neq -k_a$
  - Toplam hesaplanamaz

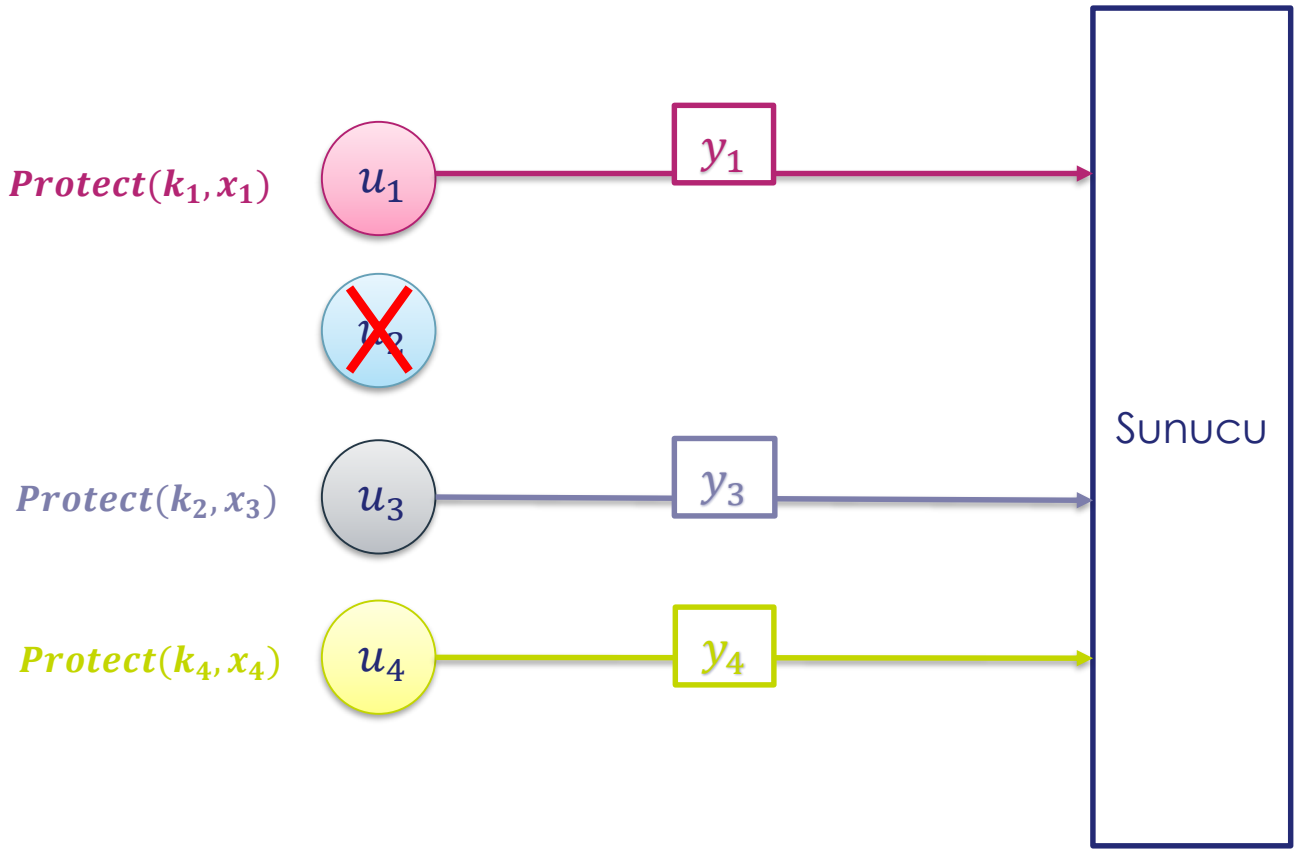


- Threshold Joye-Libert (TJL):
  - Shamir sır paylaşımı (secret sharing),
    - Her bir istemci anahtarını paylaşır
  - $\mathcal{U}_i$ terki durumunda:
    - n istemciden t istemci terk etmiş $\mathcal{U}_i$ için 0'ı şifreler
  - Sunucu toplama işlemini başarıyla tamamlar

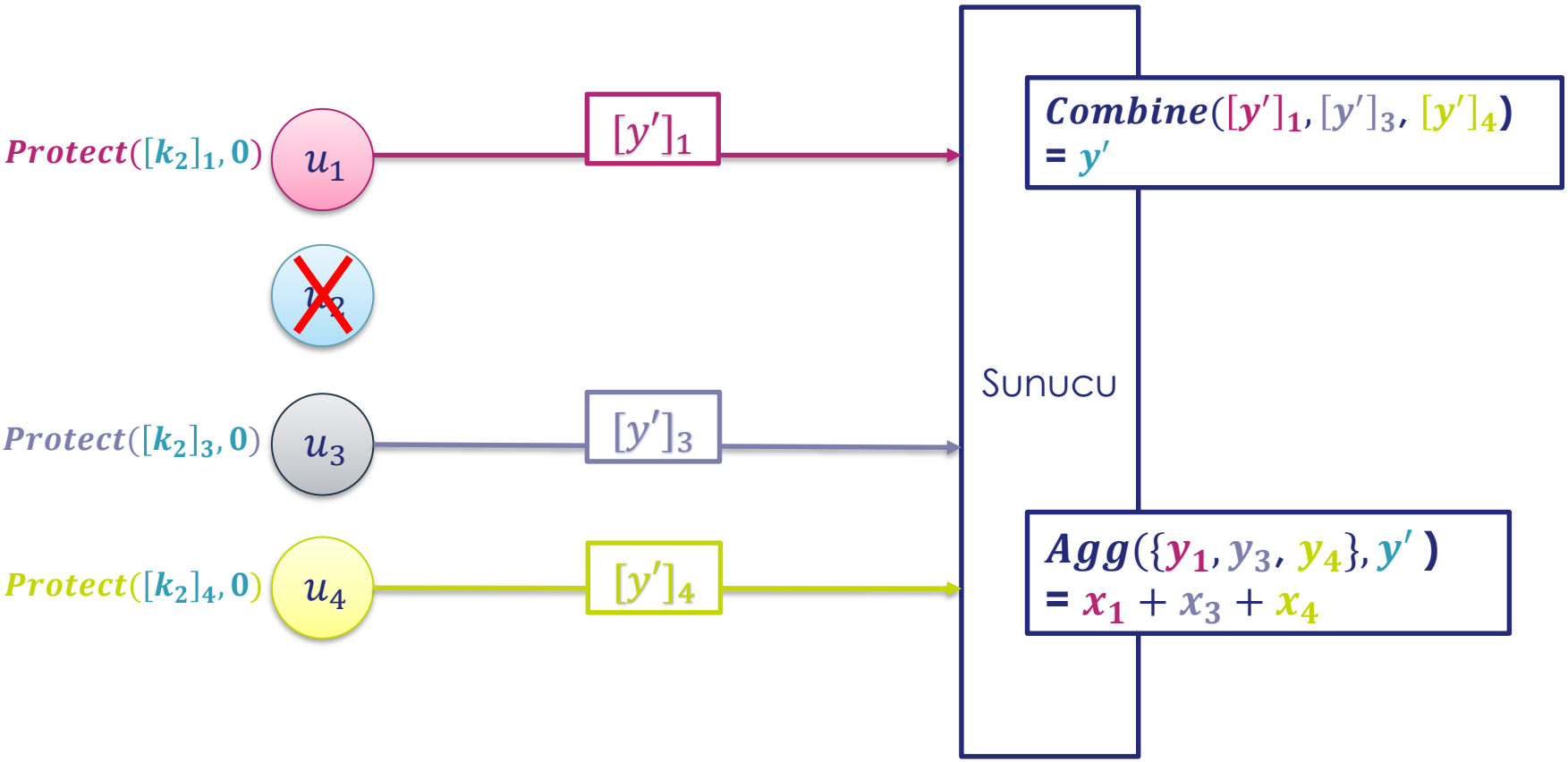
Hataya toleranslı güvenli toplama - Kurulum



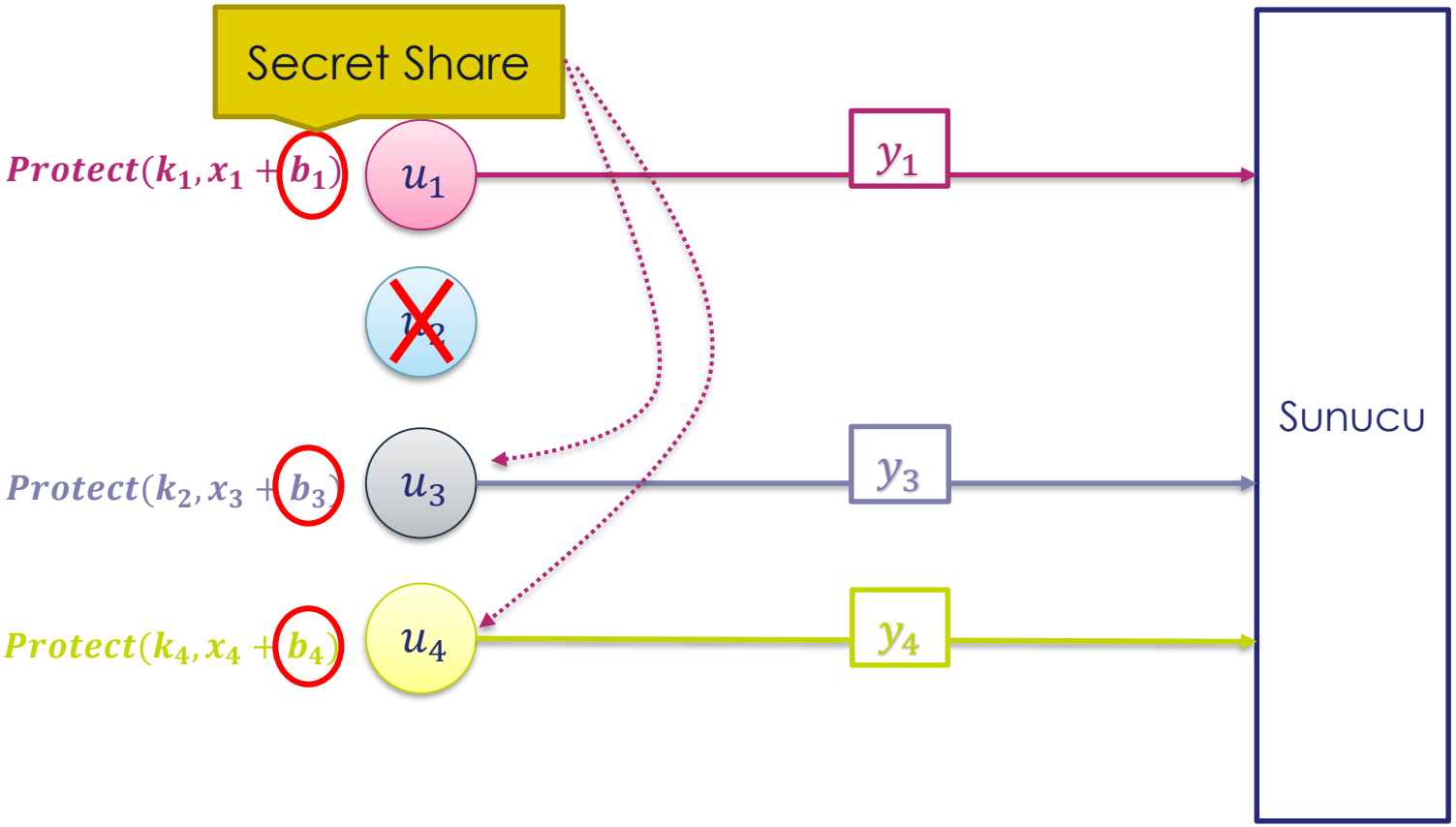
Hataya toleranslı güvenli toplama - Koruma



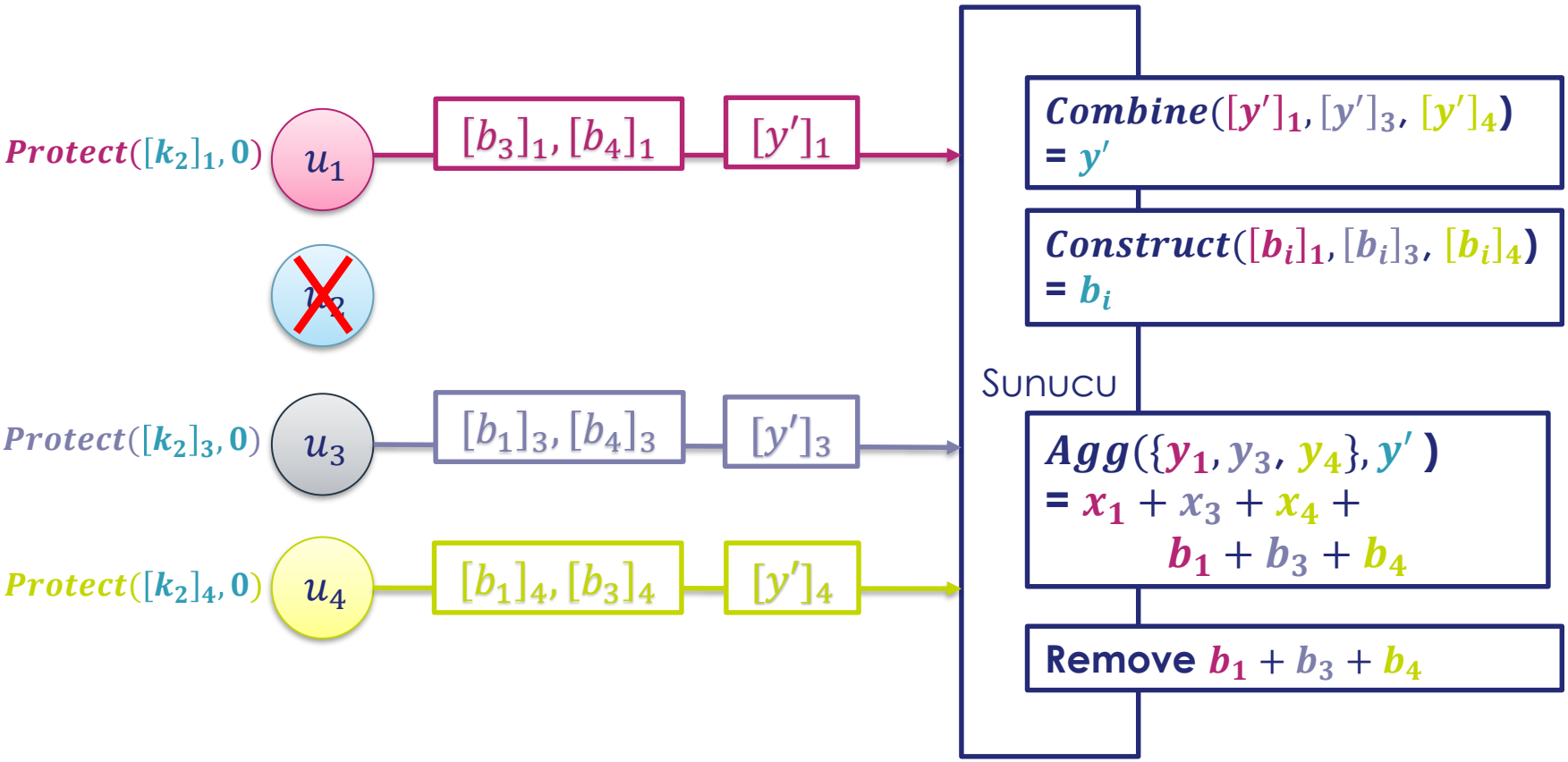
Hataya tolerans güvenli toplama – yeniden oluşturmak



Hataya toleranslı güvenli toplama – yeniden oluşturmak



Hataya toleranslı güvenli toplama – yeniden oluşturmak



Sonuçlar

- Güvenli toplama federe öğrenme için doğrudan kullanılamaz
- Bütün zorlukları ele alan bir sonuç henüz bulunmuyor
- Gelecekteki çalışmalar: kötü niyetli istemci ve sunucularla federe öğrenmede güvenli toplama

Araştırma konuları

Melek Önen

EURECOM
Sophia Antipolis

Çalışma arkadaşlarım

Mohamad Mansouri

THALES

Alberto Ibarrondo

IDEMIA augmented identity

Oubaida Chouchane

Riccardo Taiello

Oualid Zari

Ayse Ünsal

Jakub Klemsa

Kripto-tabanlı

- **Privacy preserving Image registration**
  with Riccardo Taiello, Olivier Humbert, Marco Lorenzi
- **Multi-key homomorphic encryption**
  with Jakub Klemsa, Yavuz Akin
- **Privacy-preserving biometrics**
  with Alberto Ibarrondo, Herve Chabanne, Oubaida Chouchane

Mahremiyet saldırıları ve DP-tabanlı güvenlik çözümleri

- **Membership inference Attacks against Principal Component Analysis**
  with Oualid Zari, Ayşe Ünsal, Javier Parra-Arnau, Thorsten Strufe

UPCARE

THALES

Teşekkürler

Melek Önen

In collaboration with:W. Ben Jaballah, Z. Bilgin, M. Conti, F. Karakoç, M. Mansouri

EURECOM
Sophia Antipolis