POLITECNICO DI MILANO





A Security Framework for Smart Metering with Multiple Data Consumers

Cristina Rottondi, Giacomo Verticale, and Antonio Capone

Department of Electronics and Information

Politecnico di Milano

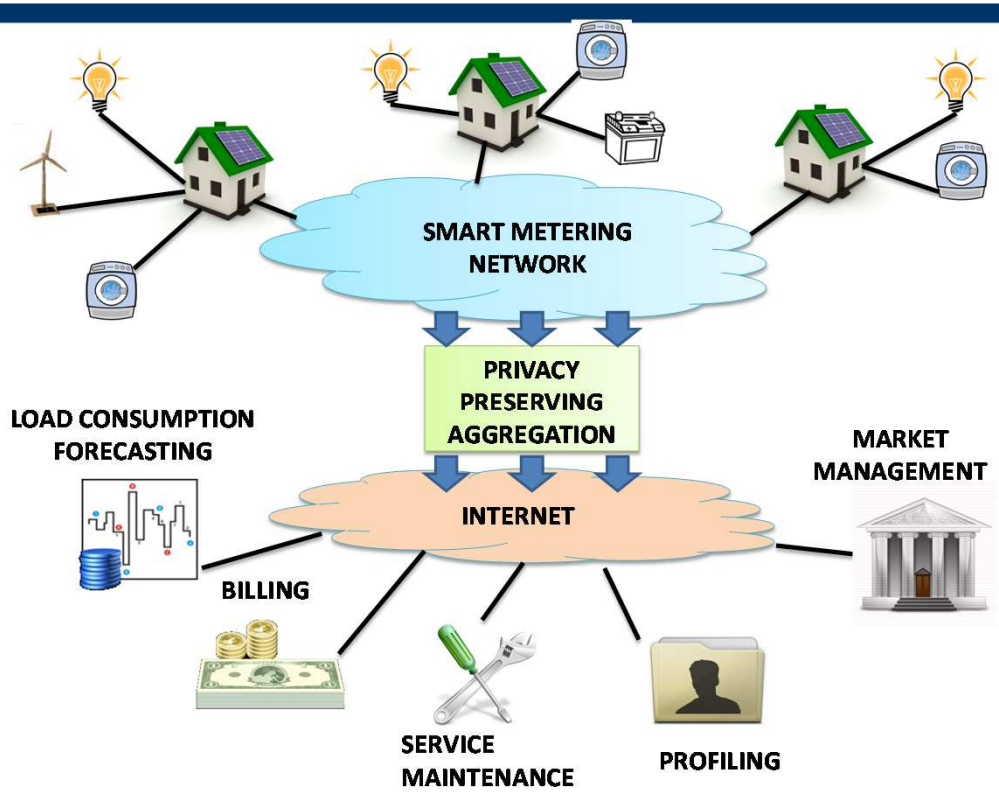
Introduction

- **Privacy concerns** about the Automatic Metering Infrastructure (AMI)
  - Non-Intrusive Load Monitoring leak information about user habits
  - meter service providers (“data consumers”) should have access to aggregated data in time and in space
- This work proposes an **aggregation architecture** based on multi-party computation preventing the collection of individual user measurements.
  - Focus is on low complexity meter even with multiple data consumers
  - We discuss scalability issues in error-free and errored channels

Outline

- Privacy requirements in Smart Grids
- The proposed privacy framework
- Allocation of information flows and scalability issues
- Control of unreliable communications
- Conclusion

Scenario and Requirements



- Configurable levels of spatial and temporal aggregation
- Prevent information leakages derived by collusions of nodes
- Scalability to millions of meters (“data producers”)
- Operability even in case of unreliable communication network
- Meter complexity should be insensitive to the number of consumers

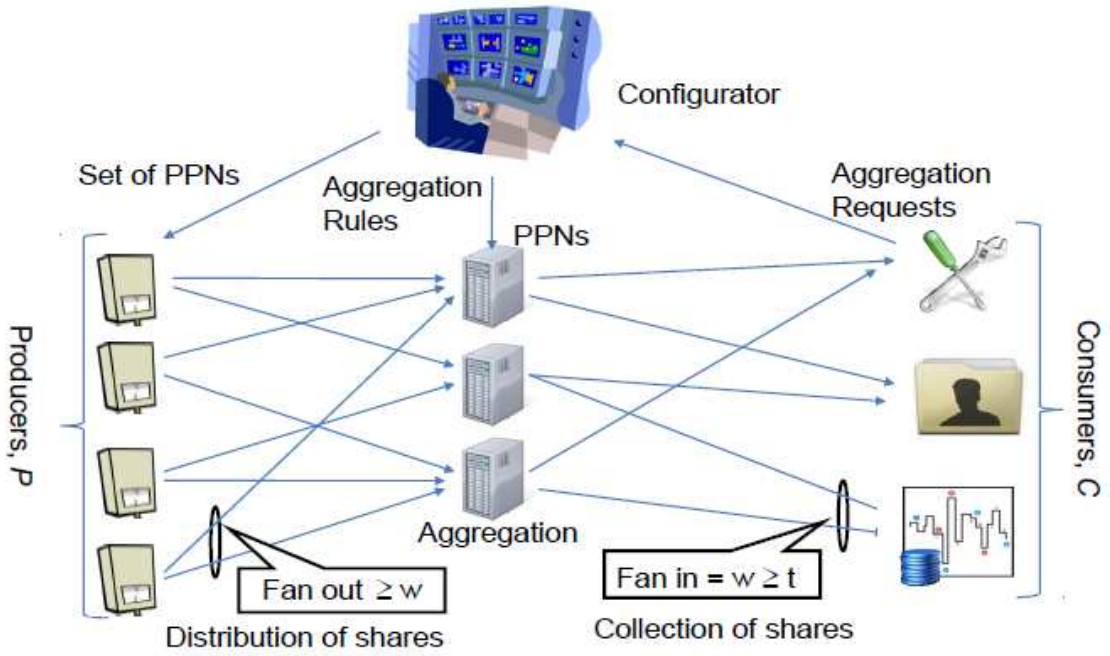
State of the Art on Privacy-Preserving Aggregation

- Already a few proposals exploiting various:
  - Cryptographic Building Blocks (homomorphic encryption, multiparty computation, zero knowledge, perturbation)
  - Attacker Models (honest-but-curious vs dishonest-but-non-intrusive)
  - Architectures (dedicated aggregation gateways vs aggregation at the meters)

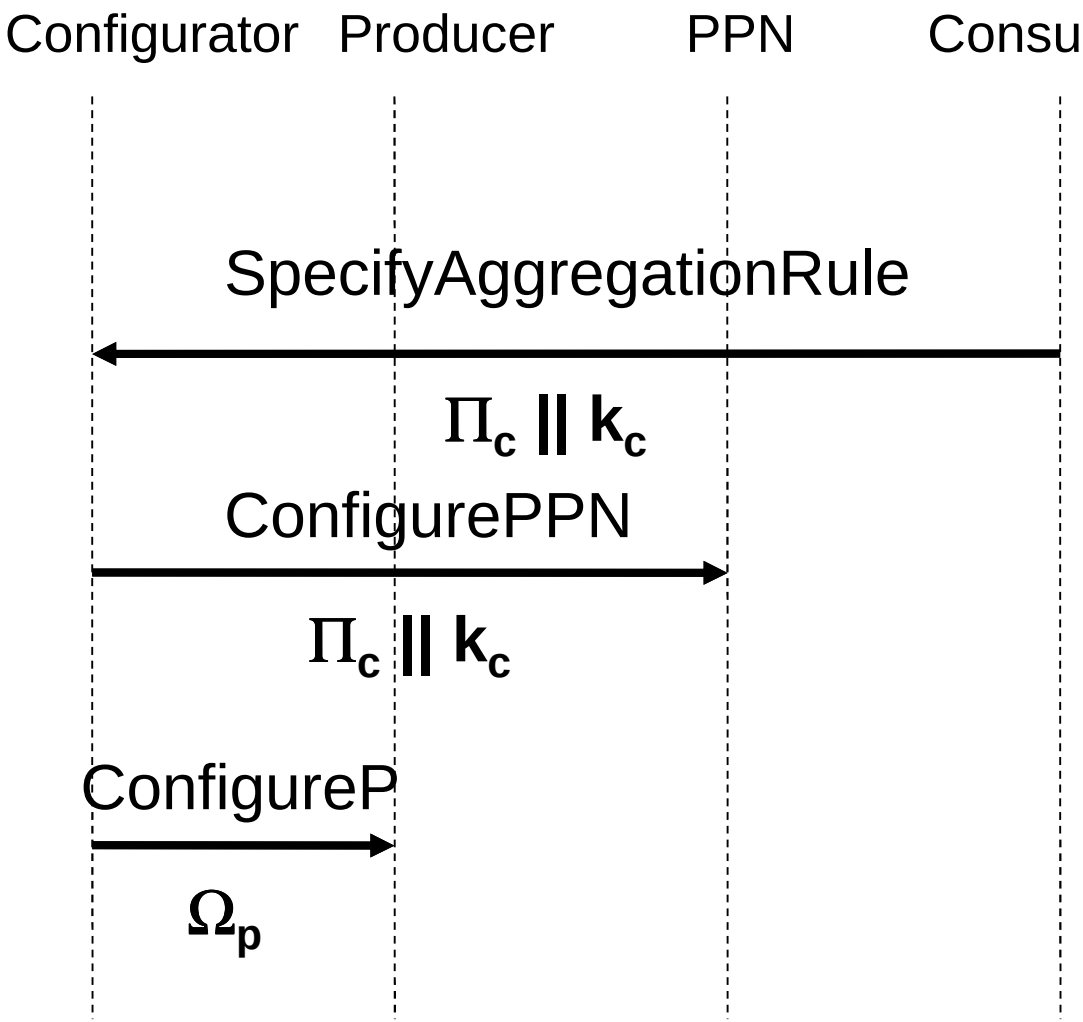
- No one shows **composability gains** when there are multiple data consumers. If the same meter participates to multiple aggregations, the meter must send the data multiple times.
  - Higher meter cost
  - Higher communications cost

The Proposed Privacy Infrastructure

- Includes a set of nodes called Privacy Preserving Nodes (PPNs)
- Data producers split the measurements using Shamir's secret sharing scheme with ***w*** shares and threshold ***t***
- The PPNs perform aggregation of the users' data by exploiting the homomorphic properties of Shamir's scheme
- The data consumers recover the aggregated data by collecting multiple shares from the PPNs



Communication Protocol (I)
The Configuration Phase

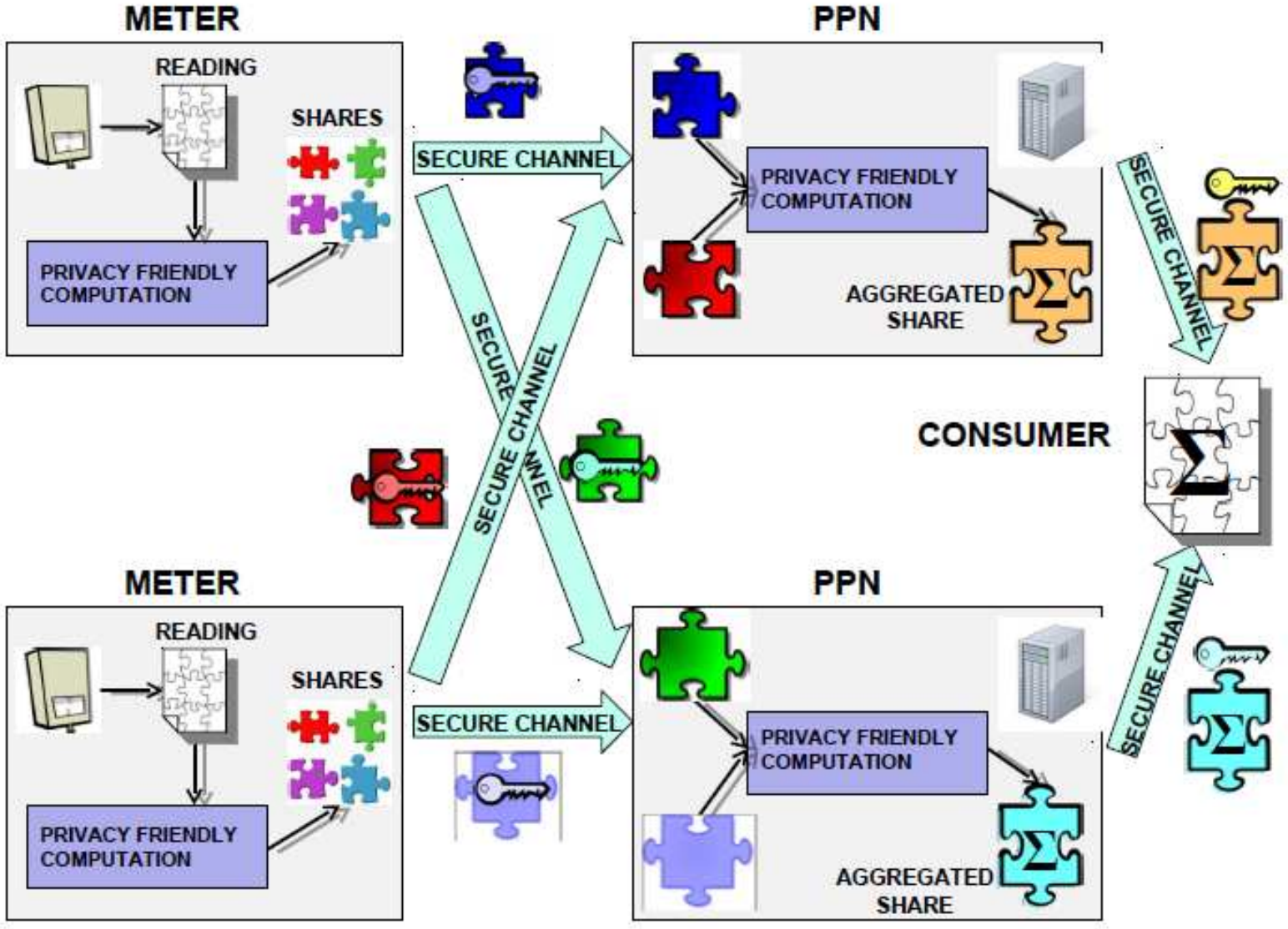


Π_c = set of Producers monitored by Consumer c

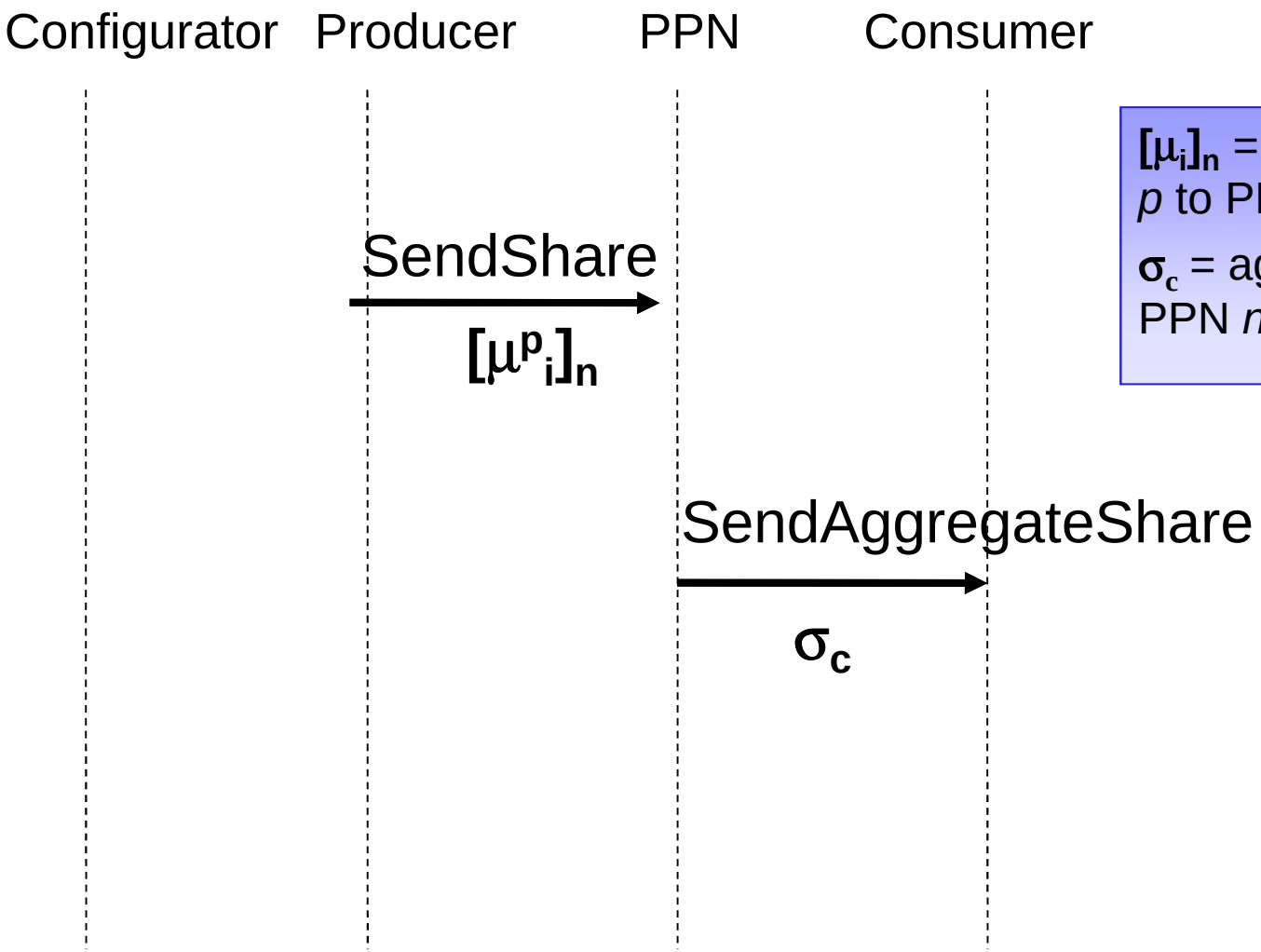
k_c = time aggregation requested by Consumer c

Ω_p = set of PPNs to which Producer *p* sends a share

Communication Protocol (II)
Sketch of the Data Phase



Communication Protocol (III)
The Data Phase



$[\mu_i]_n$ = share sent by Producer *p* to PPN *n* at round *i*

σ_c = aggregated share sent by PPN *n* to Consumer c

Communication Protocol (IV)
Calculation of Shares and Aggregate Shares

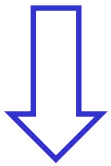
- Parameters:
  - *t, w*
  - A large enough prime number *q*
- Steps at each round:
  - Each Producer generates a measurement μ_i
  - Each Producer chooses *t-1* integer random numbers $r_1, r_2, \ldots, r_{t-1}$ uniformly distributed in [0,*q-1*], computes w_p shares:

$$[\mu_i^p]_n = \mu_i + r_1 n + r_2 n^2 + \ldots + r_{t-1} n^{t-1} \bmod q \qquad i = 1,2,\ldots,w_p$$

    and securely communicates the shares to the Wp PPNs
  - Each PPN sums the k_c consecutive shares of all $p \in \Pi_c$ for every Consumer which is connected to the PPN and sends it the aggregated share
  - Each Consumer obtains the aggregated shares by *t* PPNs and recovers the aggregated measurements

Deployment of Information Flows

- Design issues
  - The PPNs represent a costly resource
  - The number of messages exchanged between the nodes increases with the number of installed PPNs
  - Trade-off between the complexity of the aggregation and the number of PPNs in the system

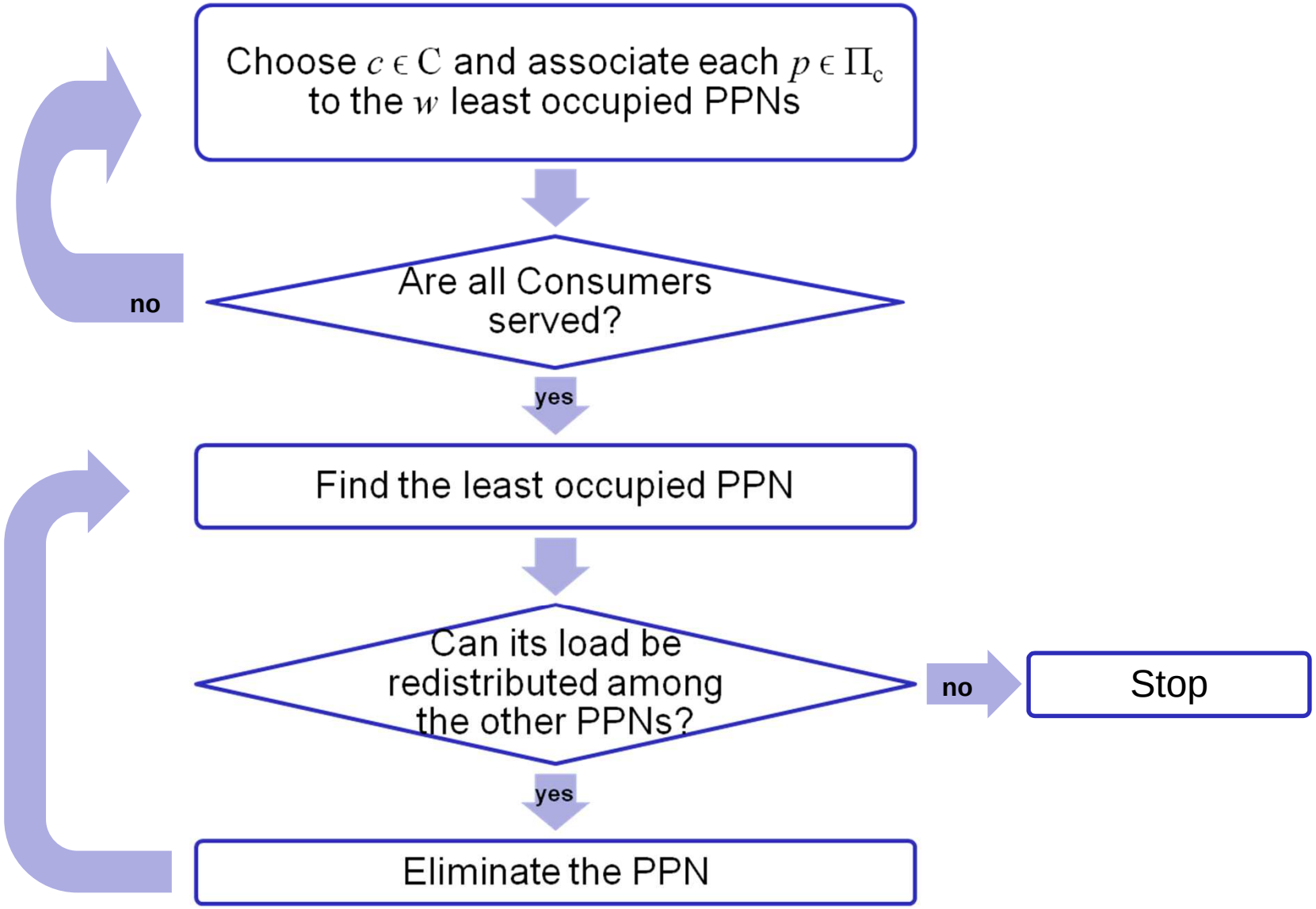


- The number of PPNs should be minimized
- We introduce an ILP formulation and a greedy algorithm to optimize the deployment of the information flows between the Producers, the PPNs, and the Consumers

Integer Linear Programming formulation

- **Objective function:** minimization of the number of PPNs
- **Inputs:**
  - Number of shares used in SSS
  - Maximum computational load at each PPN (expressed in number of sums)
  - The sets of Producers monitored by each PPN
- **Outputs:**
  - Number of installed PPNs
  - Communications flows between Producers, PPNs and Consumers
- We have proved that the problem is NP-hard

Heuristic approach



Performance of greedy vs ILP

		Greedy		ILP		
$\|C\|$	$\|P\|$	Average Result	Time	Average Result	Time	Average Gap
10	100	4	19.9 ms	4	2.1 s	0%
10	1000	4	96.7 ms	4	49 s	0%
10	10000	4	997.6 ms	4	45 min	0%
50	100	13.4	29.8 ms	13	294.7 s	3.08 %
50	1000	13.7	227.7 ms	13	44 h	5.38%
50	10000	14.5	2.7 s	N/A	N/A	N/A

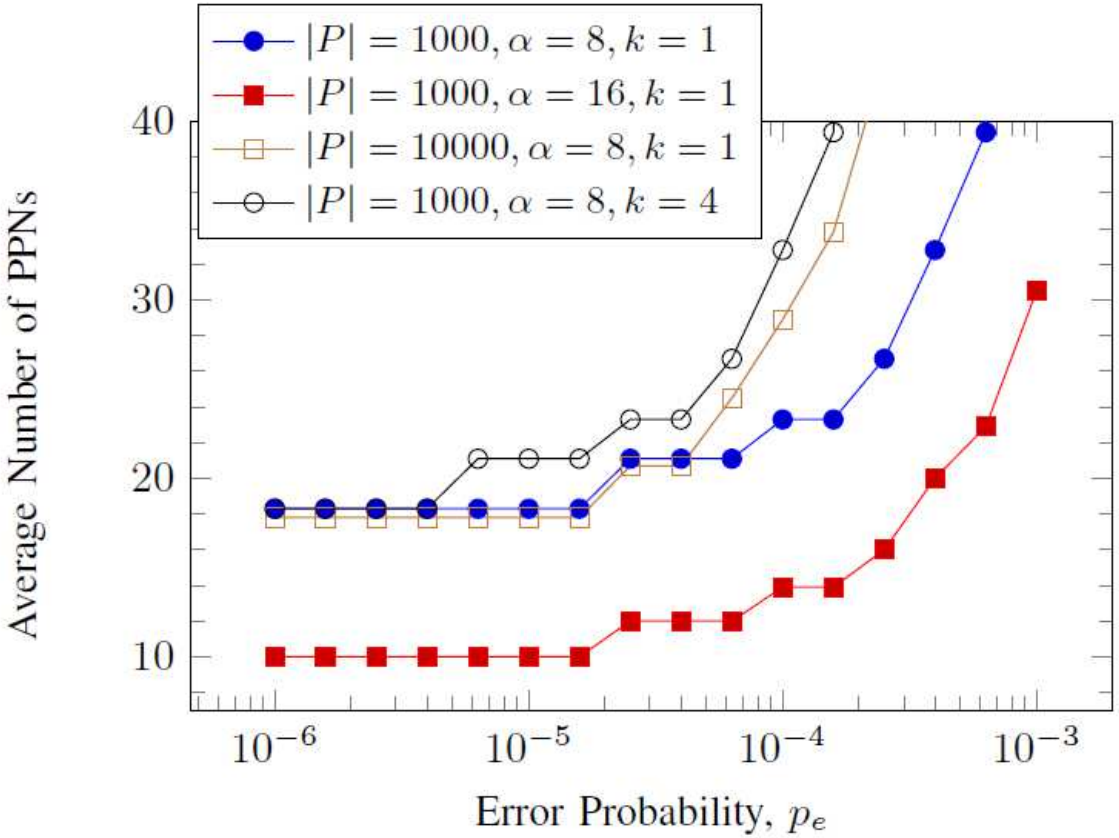
- Error free scenario assumed
- Results obtained by the greedy algorithm close to the optimum
- Short computational time even for large instances

Scalability Evaluation with Communications Errors (I)

- Shares can be missing at the PPN because of delays or losses
- If any shares are missing, the aggregated share cannot be calculated
- If less than ***t*** aggregated shares are available at the Consumer, the aggregated measurement cannot be recovered

- Therefore, the scalability of the privacy preserving infrastructure is limited by the network reliability

Scalability Evaluation with Communications Errors (II)



α = sums at the PPN, per producer

k = time aggregation

All simulations with |C|=50, t = 4

- The number of installed PPNs grows with the number of Producers, the transmission error probability and the reduction of the maximum number of sums per Producer
- Therefore, transmission errors limit the scalability of the system

Conclusion and Future Work

- A novel protocol for the AMI which handles customers' measurements in a privacy-friendly way
- The new functional nodes called Privacy Preserving Nodes are able to perform multiple aggregations of the customers' data with different spatial and temporal granularities without having access to the data
- The scalability of the architecture has been evaluated using an ILP formulation and a greedy algorithm
- Results show that in an error-free scenario the architecture is scalable to millions of meters.
- Dealing with communication errors increases the number of PPNs, limiting the scalability of the system