Combined Network Coding and Paillier Homomorphic Encryption for ensuring Consumer Data Privacy in Smart Grid Networks

Presented by David Murray, PhD candidate on behalf of the team:

D. Murray, B. Zhao, G. Elafoudi, J. Liao, L. Stankovic, V. Stankovic Centre for Intelligent and Dynamic Communications University of Strathclyde, Glasgow, UK
Monitoring domestic energy usage within the smart grid

- How will smart meters be used?
  - Automatic and accurate billing, improve energy usage
  - Enhance energy distribution and efficiency

- How can signal information processing/data analytics turn smart meter data into ‘useful’ information?
  - Inform and enhance current energy information
  - Provide itemised billing down to individual appliances and activities
  - Provide advice on retrofit advice

- How much and what kind of data do we need for effective data analytics?
  - Electricity, temperature, light, humidity, occupancy
  - Data collection 15mins, 1min, 30 sec, 1sec…

- Can we draw better conclusions from individual/household detailed energy consumption?

- How is data collected, where does it go?
- How do we ensure privacy of data and detect tampering?
How we implemented it at Strathclyde to gather smart meter readings
From data acquisition to analytics

Energy, Environmental & Occupancy Sensors → Real-Time Display → Gateway → Server Database → Data Checking & Cleaning → Statistical & Predictive Analysis
Analytics to provide intuitive feedback

- We developed a scalable database that effectively manages incoming smart meter data, and provides an easy-to-query design.

- Designing and developing robust and low-complexity non-intrusive load monitoring disaggregation algorithms for low sampling rate load data (< 1 Hz).

- Appliance characterisation and modelling to provide appliance upgrade recommendations.
Data acquisition, management and repository

- An automated platform for remote data collection and real-time monitoring
  - Including remote and non-broadband customers
  - Keeping communications between home and server to a bare minimum, without compromising on data quality

- Scalable repository of energy and environmental measurements
  - Data checking
  - Easy to add/remove sensors and houses
  - Query for correlations amongst sensor readings in the database
  - Query for correlations across houses (e.g., compare refrigerator consumption across selected houses)
  - Collecting/processing data in real time from a large number of houses (currently 40 houses)
Power Disaggregation

- Non-intrusive appliance load monitoring (NALM): Algorithmic solutions to disaggregate overall household’s power readings to individual appliances
  
- Find the consumption of individual appliances without using separate individual appliances power meters (IAMs)

- Activity modelling allows for more relative feedback that people can understand and adjust habits
  - Improve: You used X amount of power yesterday on the kettle
  - If you didn’t over fill the kettle you could save.

- Very active research topic
  - Currently, only few commercial solutions that operate well, work only at extremely high sampling rates ~kHz
Smart Meter readings – UK Dept of Energy and Climate Change infrastructure

Block diagram from REFIT project team
Smart grid communications architecture

- Smart metering (SM) system located in houses comprises of a hub and sensor network
  - HAN based on 2.4GHz and 433MHz

- From the houses, data transmitted to Communications Service Provider (CSP) via wireless NAN
  - Communication range 0.25miles
  - At least two CSPs in the range of each house
  - CSPs – routers that forward packets possibly multi-hops

- Data Communication Company (DCC) receives packets from CSP and performs processing

Smart grid communications architecture
Problem statement

- Lots of sensitive personal household data being gathered within the HAN and transmitted via WAN/NAN

- Data contains billing information and energy consumption from which domestic routines can be inferred, inc. when the household is away on holiday…
Security in the smart grid

- Smart grid prone to cyber attacks and tampering due to sensitivity of information in the network

- Key concern is against tampering, eavesdropping and traffic analysis

- Types of attacks expected:
  - Entropy Attacks
  - Packet Erasures
  - Packet Modification / Corruption
  - Eavesdropping / Analysis
Proposed Solution

- Pairing network coding (NC) with Paillier homomorphic encryption (PHE)

- Due to its homomorphic additive and multiplicative properties, PHE simplifies computation of cipher texts and incorporation of linear NC

- Intermediate nodes can still perform NC in the conventional manner without needing access to the systems private key
Random Linear Network Coding (RLNC)

\[ X = G \ast M \]

\[ x_k = \sum_{i=1}^{S} g_{ik} m_i \]

- \( k \)-th NC symbol/packet
- \( m_i \) – \( i \)-th source message
- \( g_{ik} \) random local encoded coefficient

Benefits compared against traditional routing
- Throughput
- Efficiency
- Scalability
- Resilience to attacks and eavesdropping

Issues
- Larger transmission overhead
- Linear dependency of coefficient vectors
Paillier Homomorphic Encryption (PHE)

\( p, q \) are two \( k \)-bit primes where \( k \) is the security parameter, are random and independent such that:
\[
gcd(pq, (p - 1)(q - 1)) = 1
\]

\[
n = pq
\]

\[
\lambda = \text{lcm}((p - 1)(q - 1))
\]

Choose random integer \( g \) where:
\[
g \in \mathbb{Z}_{n^2}^*
\]

Check modular multiplicative inverse:
\[
\mu = \left( L\left( g^\lambda \mod n^2 \right) \right)^{-1} \mod n \quad \text{where} \quad L(\mu) = \frac{\mu - 1}{n}
\]

Public Key : \((n, g)\)
Private Key : \((\lambda, \mu)\)
PHE: Encryption and Decryption

Encryption:
For a message $m \in \mathbb{Z}_n$
Select random $r \in \mathbb{Z}_n^*$
Calculate ciphertext using public key:
$$E(m) = c = g^m r^n \mod n^2$$

Decryption:
Given received ciphertext $c < n^2$,
recover the message using private key:
$$m = D(c) = L(c^\lambda \mod n^2) \mu \mod n^2$$

Homomorphic Properties

$$E(m) = (g^m r^n) \mod n^2$$

$$E(m_1)E(m_2) = g^{m_1m_2} (r_1r_2)^n \mod n^2 = E(m_1 + m_2)$$

$$D(E(m_1)^k \mod n^2) = km_1 \mod n$$
Threat model

We consider the threat posed by an attacker with the following characteristics:

- The attacker can eavesdrop all network links
- Has knowledge of scheme used
- Is computationally bounded
- Can inject or erase packets in the network
Proposed scheme
Proposed scheme – Combined RLNC and PHE encryption

- Each house SM system performs NC on its HAN dataset
- The local encoding vectors used are then encrypted using the public key

\[ c_i(e) = E_{ek}(g_i(e)), \quad (1 \leq i \leq h) \]
\[ c(e) = \left[ c_1(e), c_e(e), \ldots, c_h(e) \right] \]
- CSP performs conventional NC on all incoming encoded and encrypted packets from the houses

\[
g(e) = \sum_{i=1}^{h} \beta_i(e) g(e'_i) \]
\[
E_{ek}(g(e)) = E_{ek}\left( \sum_{i=1}^{h} \beta_i(e) g(e'_i) \right) \\
= \prod_{i=1}^{h} E_{ek}(\beta_i(e) g(e'_i)) \\
= \prod_{i=1}^{h} E_{ek}^{\beta_i(e)} (g(e'_i))
\]
$M_1 = [m_1^1 m_2^1]$

$X_1 = G_1 M_1$

$E(g_{11}^1), E(g_{12}^1)$  $X_{11}$

$E(g_{21}^1), E(g_{22}^1)$  $X_{12}$

House1

$M_2 = [m_1^2 m_2^2]$

$X_2 = G_2 M_2$

$E(g_{11}^2), E(g_{12}^2)$  $X_{21}$

$E(g_{21}^2), E(g_{22}^2)$  $X_{22}$

House2

$X_{CSP} = G_{CSP} X$

$(E(g_{11}^1))^{g_{CSP}^{11}}, (E(g_{12}^1))^{g_{CSP}^{12}}$  $X_{11}^{CSP}$

CSP
Proposed Scheme – Decoding & Decryption at Sink

- DCC/sink will first decrypt NC coefficients
- Resulting global encoding vectors are used for RLNC decoding in the conventional way, e.g., using Gaussian elimination
- Note that all houses have knowledge of the public key and only the DCC has knowledge of the private key for decryption
- At the sink decryption can be carried out once a sufficient number of packets from the same generation has been received

\[
\begin{bmatrix}
x_1 \\
\vdots \\
x_h
\end{bmatrix} = G^{-1} \begin{bmatrix}
x^{csp}(e_1) \\
\vdots \\
x^{csp}(e_h)
\end{bmatrix}
\]
Decrypt encoding coefficients

\[ D(\text{E}(g^1_{11})^\text{CSP}^{1}_{11}) = g^{\text{CSP}_1}_{11} g^1_{11} \]

Gaussian Elimination for NC decoding

\[ \overline{M} \quad \text{Recovered message} \]

DCC
Analysis

- Is efficient and does not incur a significantly high overhead
- Features privacy against packet analysis
- Encryption prevents against earliest decoding by packet analysis
- Each message is of the same size helping prevent size correlation and buffering reduces effect of time order correlation attacks
- Inefficient against Pollution and Entropy attacks
Resilience to packet drops

- CSPs dropping incoming packets
- RLNC provides erasure protection, and additional PHE does not affect erasure protection of RLNC
- Lemma 1: *If only one CSP is dropping all packets it receives, then all messages from any house can be recovered at the DCC*

Pollution and entropy attacks

- Attacker injects dummy packets
- CSP can generate malicious content that is mixed with incoming packets
- Lemma 2: *If only one CSP is injecting malicious content and mixing it with incoming packets, then all messages from any house can be recovered at the DCC*
before decryption

after decryption =>
Clear separation between corrupted and useful packets
Adaptive threshold to separate corrupted/injected packets from the original ones.
Computational Overhead

- Communications Hub (SM) encoding:
  - PHE: $O(N^2)$
  - Multiplications & modulus operations: $O(N^2 \log n)$

- CSP encoding:
  - Per packet $O(N^2 \log n)$
  - Total: $O(N^3 \log n)$

- DCC decoding:
  - $O(N^2 \log n)$
Conclusions

- With smart meter rollouts being deployed worldwide, it is critical to ensure secure transmission of sensitive personal data that is subject to eavesdropping and malicious attacks.

- We propose a combined homomorphic encryption algorithm with network coding that provides security and privacy while maintaining RLNC allowing for a high chance of recovering packets at the sink.

- Open work:
  - Implementation of lightweight verification scheme to prevent pollution attacks propagating throughout the network
  - Assessing robustness to a large scale attacks
  - Practical implementation
Acknowledgements

- All presented work is being addressed as part of two UK Research Council projects by our team, within multi-disciplinary and multi-institutional (academic + industrial) consortia:
  - REFIT: Personalised Retrofit Decision Support Tools for UK Homes using Smart Home Technology
  - APAtSCHE: Aging Population Attitudes to Sensor Controlled Home Energy