Privacy-friendly Forecasting for the Smart Grid using Homomorphic Encryption

J. Bos[1], W. Castryck[2,3], I. Iliashenko[2] and F. Vercauteren[2,4]

[1]NXP Semiconductors [2]COSIC, KU Leuven and imec

[3]Université de Lille-1 [4]Open Security Research

AfricaCrypt 2017

Dakar, Senegal

The Smart-Grid

The Smart-Grid

load consumption
weather conditions
bills
structure of a local utility grid . . .

The Smart-Grid

Benefits

- control of consumption
- optimization of utility production
- improved logistics
- source of research data

[1] European Commission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity. Technical report 365, June 2014.

The Smart-Grid

Benefits

- control of consumption
- optimization of utility production
- improved logistics
- source of research data

"The Third Energy Package requires Member States to ensure implementation of intelligent metering systems for the long-term benefit of consumers. [...] For electricity, there is a target of rolling out at least 80% by 2020, of the positively assessed cases.[1]*"*

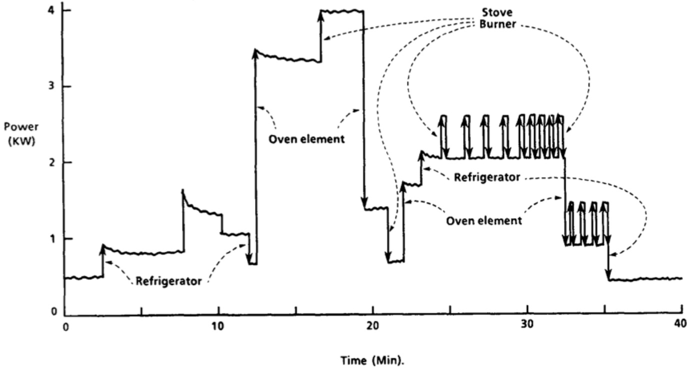
[1] European Commission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity. Technical report 365, June 2014.

Privacy Concerns in the Smart-Grid

Update rate of smart-meters: $\leq$ 15 min (EU recommendation)

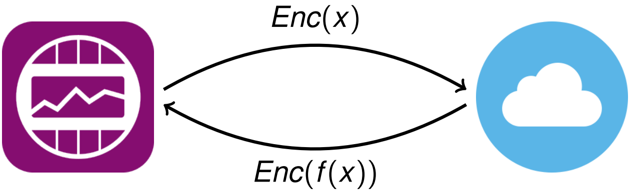
Privacy Concerns in the Smart-Grid

Update rate of smart-meters: $\leq$ 15 min (EU recommendation)

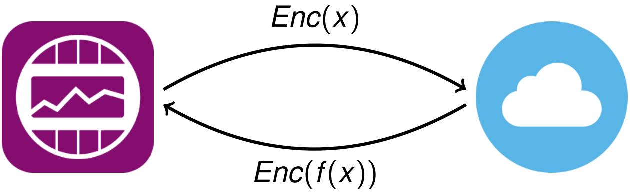


Power step changes due to individual appliance events.
G.W. Hart. Nonintrusive appliance load monitoring.
Proceedings of the IEEE, 80(12):1870-1891, 1992

Homomorphic Encryption



Homomorphic Encryption

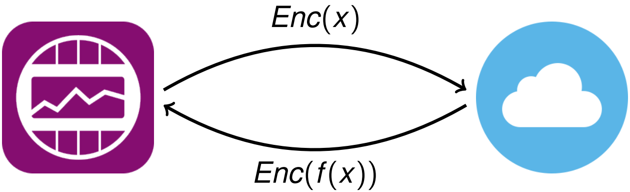


Partially HE: $+$ or $\times$

Somewhat HE: $+$ and up to some consecutive $\times$

Fully HE: $+$ and $\times$

Homomorphic Encryption



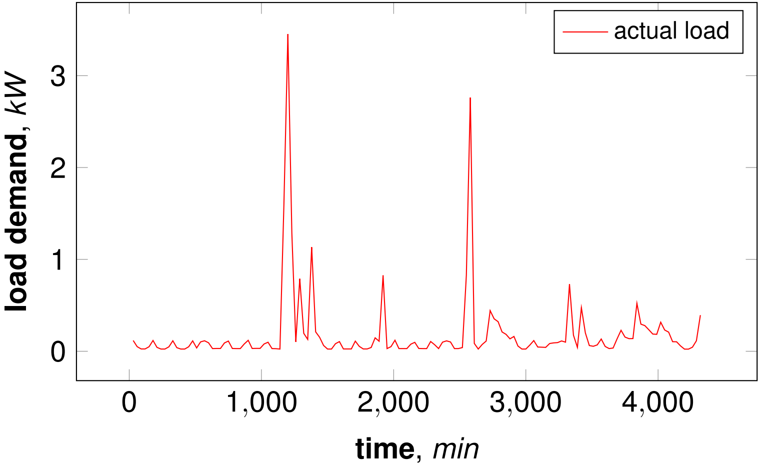
Partially HE: $+$ or $\times$

Somewhat HE: $+$ and up to some consecutive $\times$

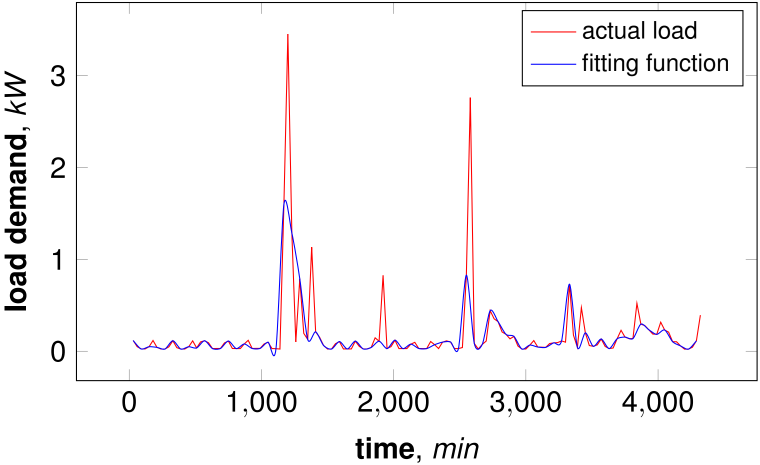
Fully HE: $+$ and $\times$

Complexity ↓

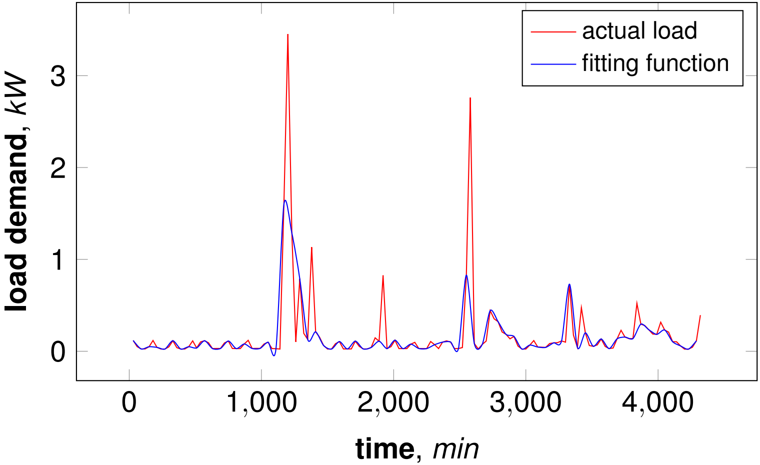
Prediction for the Smart-Grid



Prediction for the Smart-Grid



Prediction for the Smart-Grid



$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i^{forecast} - y_i^{actual})^2, \qquad \mathbf{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i^{forecast} - y_i^{actual}}{y_i^{actual}}\right|$$

Prediction for the Smart-Grid

Time period	ARIMA	BATS	NNET	PERSIST	TBATS
30 min	91	57	49	75	72
60 min	51	59	52	60	63

MAPE for varying periods and algorithms

Source: Veit et al. Household electricity demand forecasting: benchmarking state-of-the-art methods. In Proceedings of e-Energy ’14. 2014.

ANNs are the best choice, but . . .

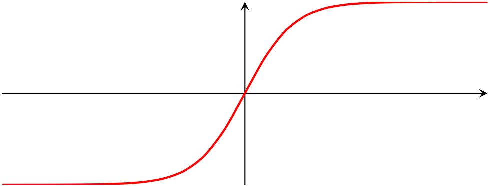
Prediction for the Smart-Grid

Time period	ARIMA	BATS	NNET	PERSIST	TBATS
30 min	91	57	49	75	72
60 min	51	59	52	60	63

MAPE for varying periods and algorithms

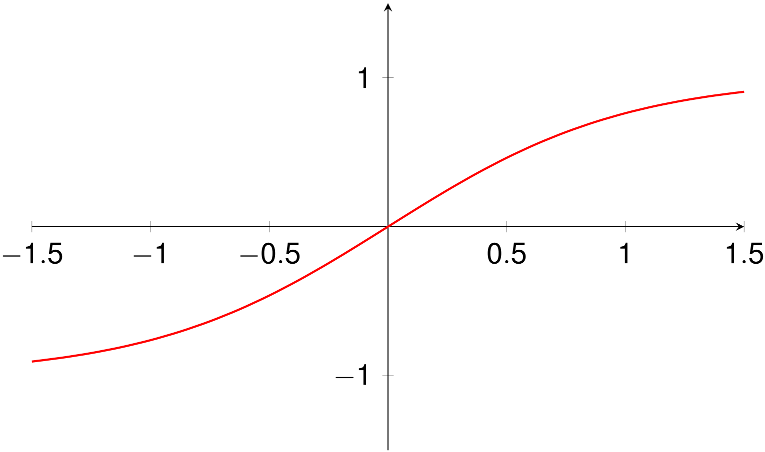
Source: Veit et al. Household electricity demand forecasting: benchmarking state-of-the-art methods. In Proceedings of e-Energy ’14. 2014.

ANNs are the best choice, but . . .

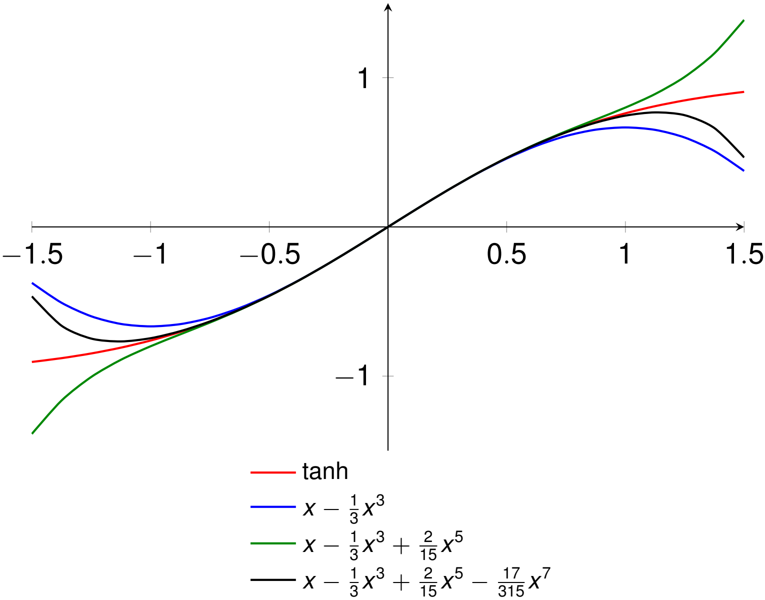


. . . they contain highly non-linear sigmoids as activation functions.

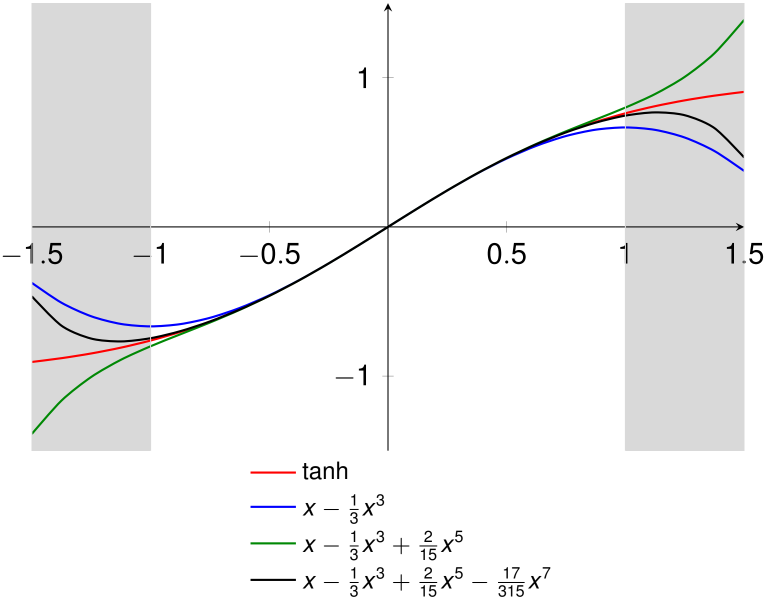
Prediction for the Smart-Grid



Prediction for the Smart-Grid



Prediction for the Smart-Grid



Polynomial Neural Networks

ANNs with polynomial activation functions:

$\mathbf{x^2}$	pattern recognition (Microsoft's CryptoNets)
GMDH	forecasting

Polynomial Neural Networks

ANNs with polynomial activation functions:

x^2 pattern recognition (Microsoft's CryptoNets)

GMDH forecasting

- Published by Alexei Ivakhnenko in 1970.
- Originally used for wheat harvest prediction in Ukraine.

Polynomial Neural Networks

ANNs with polynomial activation functions:

$\mathbf{x^2}$	pattern recognition (Microsoft's CryptoNets)
GMDH	forecasting

- Published by Alexei Ivakhnenko in 1970.
- Originally used for wheat harvest prediction in Ukraine.
- Applied for load forecasting:
  - comparable with conventional ANNs
  - MAPE $\approx$ 2%

Polynomial Neural Networks

ANNs with polynomial activation functions:

x^2 pattern recognition (Microsoft's CryptoNets)

GMDH forecasting

- Published by Alexei Ivakhnenko in 1970.
- Originally used for wheat harvest prediction in Ukraine.
- Applied for load forecasting:
  - comparable with conventional ANNs
  - MAPE $\approx$ 2% over a town, a big city district or a region

Group Method of Data Handling

Approximation by truncated **Wiener series**

$$a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} a_{ijk} x_i x_j x_k + \dots$$

n is the number of previous states of a system.

Group Method of Data Handling

Approximation by truncated **Wiener series**

$$a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} a_{ijk} x_i x_j x_k + \dots$$

n is the number of previous states of a system.

Find coefficients $\{a_i\}, \{a_{ij}\}, \{a_{ijk}\}, \dots$

Group Method of Data Handling

Approximation by truncated **Wiener series**

$$a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} a_{ijk} x_i x_j x_k + \dots$$

n is the number of previous states of a system.

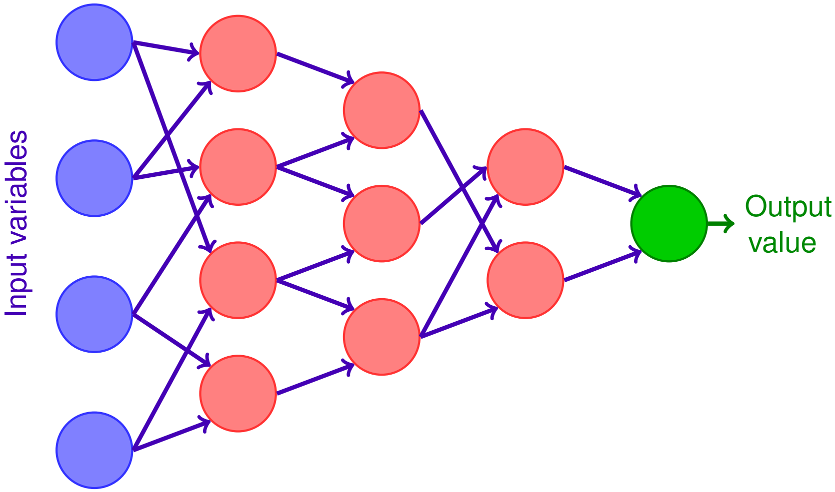
Find coefficients $\{a_i\}, \{a_{ij}\}, \{a_{ijk}\}, \dots$

Can be replaced by a composition of quadratic polynomials

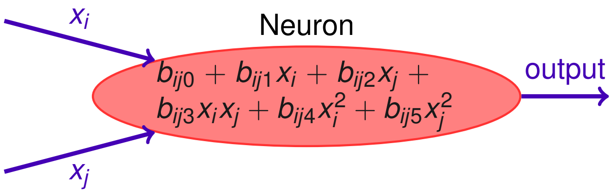
$$\{b_{ij0} + b_{ij1} x_i + b_{ij2} x_j + b_{ij3} x_i x_j + b_{ij4} x_i^2 + b_{ij5} x_j^2\}.$$

Group Method of Data Handling

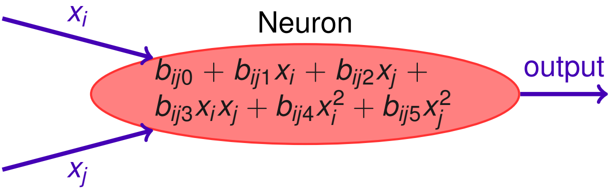
A neural network with the structure 4 – 4 – 3 – 2 – 1.



Group Method of Data Handling



Group Method of Data Handling



Learning

Define coefficients of a quadratic polynomial from

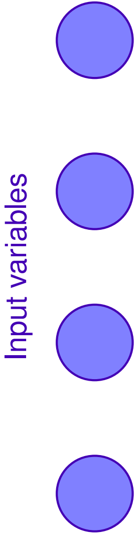
$$Y = bX + e,$$

where

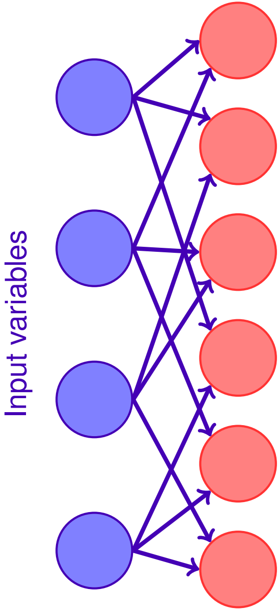
$\mathbf{X} = (1, x_i, x_j, x_ix_j, x_i^2, x_j^2)^\intercal$, $\mathbf{b} = (b_{ij0}, b_{ij1}, b_{ij2}, b_{ij3}, b_{ij4}, b_{ij5})$,
$\mathbf{Y}$ is the expected output, $\mathbf{e}$ is a random noise.

Can be done by the least-squares method.

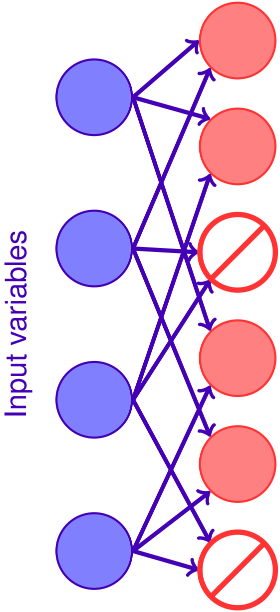
Group Method of Data Handling



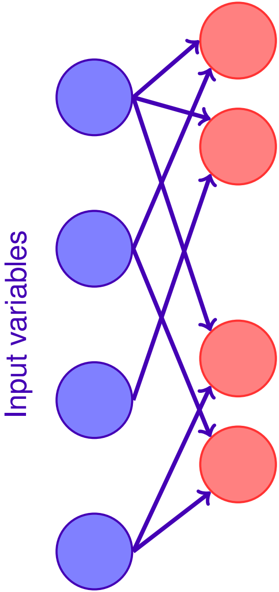
Group Method of Data Handling



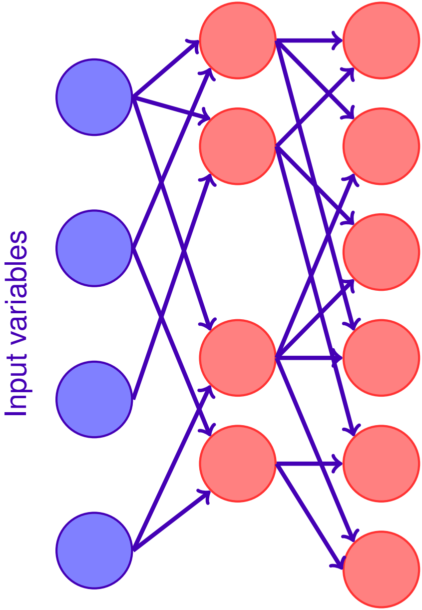
Group Method of Data Handling



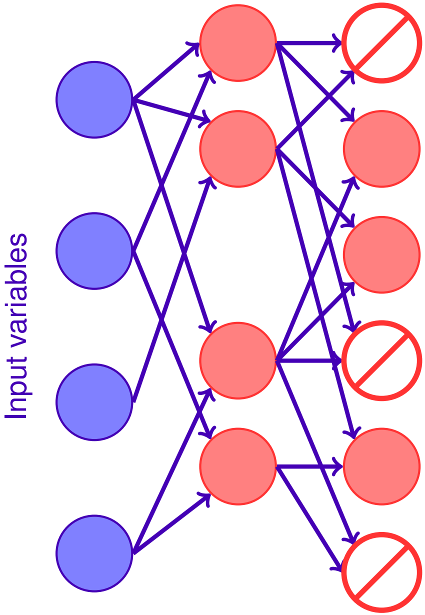
Group Method of Data Handling

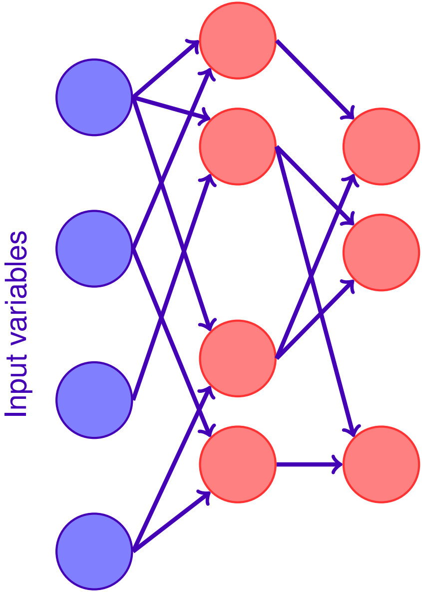


Group Method of Data Handling

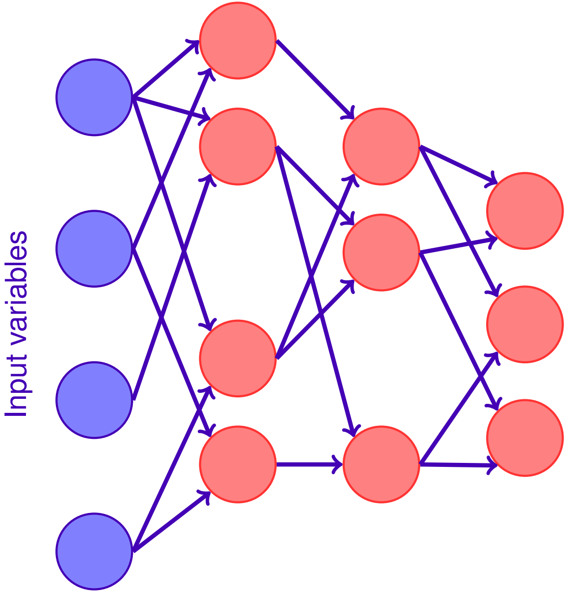


Group Method of Data Handling

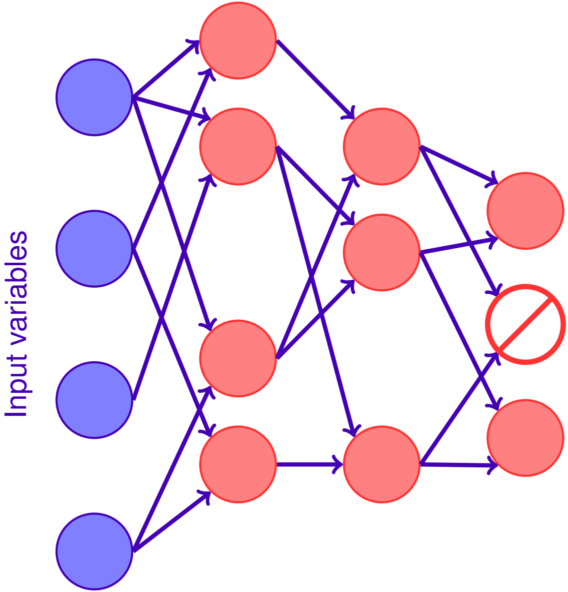



Input variables

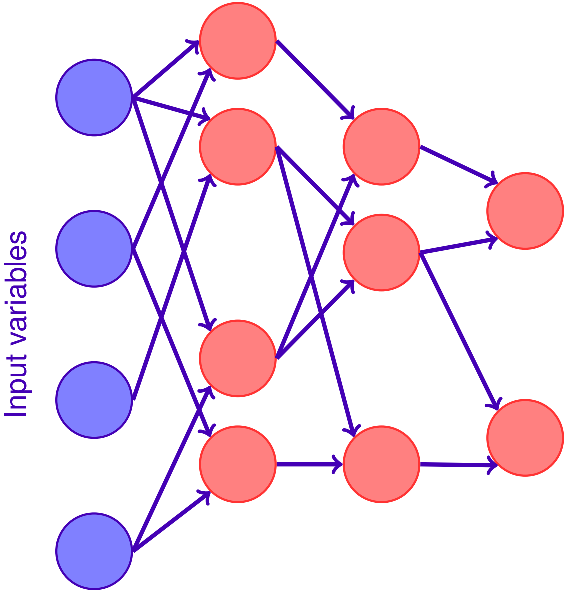
Group Method of Data Handling



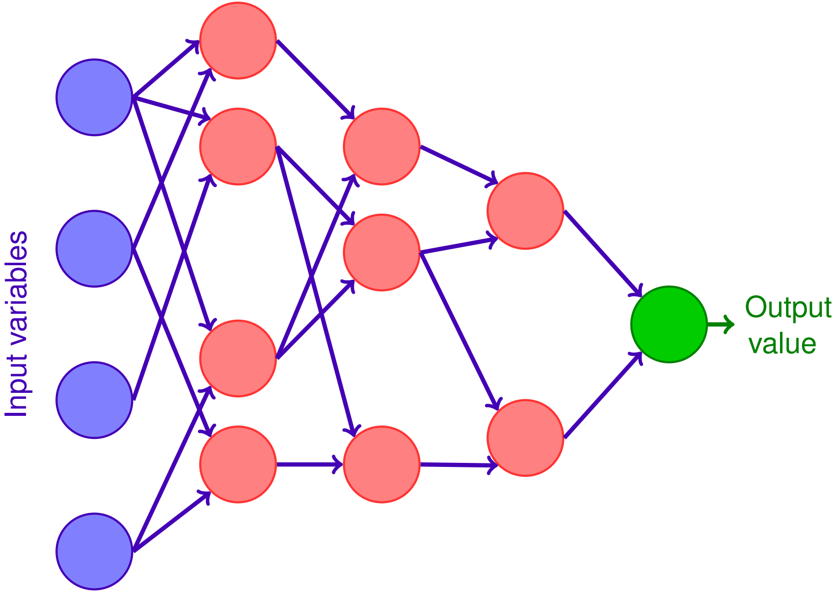
Group Method of Data Handling



Group Method of Data Handling



Group Method of Data Handling



Sample data

Source: Commission for Energy Regulation (CER) is the regulator for the electricity and natural gas sectors in Ireland.

Over 5,000 Irish homes and businesses observed.

Electricity consumed during 30 minutes intervals.

Time frame: July 14 2009 to Dec. 31 2010.

Data splits:

- training set (1 year),
- test set (half a year).

Structure of the GMDH-network

Input layer (51 nodes):

- previous 48 half-hour load measures,
- day of the week,
- month,
- temperature.

Hidden layers:

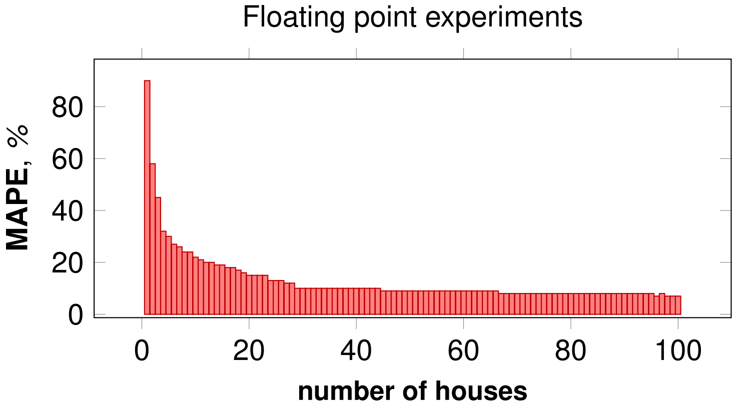
- 3 hidden layers of sizes $8, 4, 2$.

Output node:

- predicted consumption during the next 30 minutes.

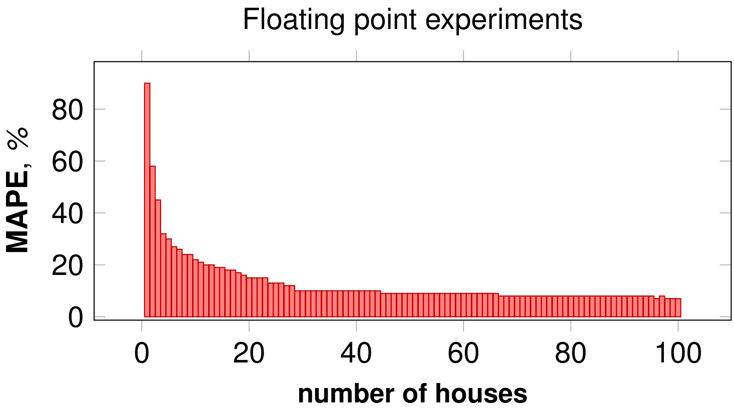
Resulting polynomial is of degree $2^4 = 16$.

Implementation in the Plain Mode: Floating Point



# houses	1	2	5	10	20	50	100
avg. MAPE(%)	90	55	33	23	15	9	7

Implementation in the Plain Mode: Floating Point



# houses	1	2	5	**10**	20	50	100
avg. MAPE(%)	90	55	33	**23**	15	9	7

Fan-Vercauteren SHE

Ring-LWE based SHE (2012).

Parameters:

- ring $\mathbf{R} = \mathbb{Z}[X]/(f(X))$ with $f(X) = X^d + 1$, $d = 2^n$
- moduli $\mathbf{t}, \mathbf{q} \in \mathbb{Z} : q \gg t$,
- plaintext and ciphertext spaces $\mathbf{R_t} = R/(t)$, $\mathbf{R_q} = R/(q)$,
- key and error distributions over R: $\chi_{\text{key}}, \chi_{\text{err}}$
  - discrete Gaussian with small σ.

Fan-Vercauteren SHE

Ring-LWE based SHE (2012).

Parameters:

- ring $\mathbf{R} = \mathbb{Z}[X]/(f(X))$ with $f(X) = X^d + 1$, $d = 2^n$
- moduli $\mathbf{t}, \mathbf{q} \in \mathbb{Z} : q \gg t$,
- plaintext and ciphertext spaces $\mathbf{R_t} = R/(t)$, $\mathbf{R_q} = R/(q)$,
- key and error distributions over R: $\boldsymbol{\chi}_{\text{key}}, \boldsymbol{\chi}_{\text{err}}$
  - discrete Gaussian with small $\boldsymbol{\sigma}$.

Key generation

- $\mathbf{s} \leftarrow \chi_{\text{key}}, \quad \mathbf{a} \leftarrow \mathcal{U}(R_q), \quad \mathbf{e} \leftarrow \chi_{\text{err}}$
- $\mathbf{b} = [-(a \cdot s + e)]_q$

$$\mathbf{pk} = (a, b)$$

Fan-Vercauteren SHE

Encryption

- $\mathbf{e_1}, \mathbf{e_2} \leftarrow \chi_{\text{err}}, \quad \mathbf{u} \leftarrow \chi_{\text{key}}$
- $\mathbf{c_0} = \lfloor q/t \rfloor m + b \cdot u + e_1, \quad \mathbf{c_1} = a \cdot u + e_2$

$$\mathbf{c} = (c_0, c_1)$$

Ciphertexts allow homomorphic addition and multiplication.

Fan-Vercauteren SHE

Encryption

- $\mathbf{e_1}, \mathbf{e_2} \leftarrow \chi_{\text{err}}, \quad \mathbf{u} \leftarrow \chi_{\text{key}}$
- $\mathbf{c_0} = \lfloor q/t \rfloor m + b \cdot u + e_1, \quad \mathbf{c_1} = a \cdot u + e_2$

$$\mathbf{c} = (c_0, c_1)$$

Ciphertexts allow homomorphic addition and multiplication.

Decryption

$$\mathbf{m} = \left[\left\lfloor \frac{t[c_0 + s \cdot c_1]_q}{q} \right\rceil \right]_t$$

Fan-Vercauteren SHE

$$[c_0 + c_1 s]_q = \Delta m + e$$

with $\Delta = \left\lfloor \frac{q}{t} \right\rfloor$.

$$||e||_\infty < \Delta/2 \rightarrow \text{correct decryption.}$$

Fan-Vercauteren SHE

$$[c_0 + c_1 s]_q = \Delta m + e$$

with $\Delta = \left\lfloor \frac{q}{t} \right\rfloor$.

$$||e||_\infty < \Delta/2 \rightarrow \text{correct decryption.}$$

Mult(c_1, c_2):

$\Delta m_{\text{mult}} + e_{\text{mult}}$.

$$||e_{\text{mult}}||_\infty > \delta \cdot \max\{||e_1||_\infty, ||e_2||_\infty\}$$

Fan-Vercauteren SHE

$$[c_0 + c_1 s]_q = \Delta m + e$$

with $\Delta = \left\lfloor \frac{q}{t} \right\rfloor$.

$$||e||_\infty < \Delta/2 \rightarrow \text{correct decryption.}$$

Mult(c_1, c_2):

$\Delta m_{\text{mult}} + e_{\text{mult}}$.

$$||e_{\text{mult}}||_\infty > \delta \cdot \max\{||e_1||_\infty, ||e_2||_\infty\}$$

4 network layers
Security level **80** bits

$$\rightarrow \begin{cases} d = 4096 \\ q \approx 2^{186} \\ \sigma = 102 \\ t \leq 396 \end{cases}$$

Fan-Vercauteren SHE

$$[c_0 + c_1 s]_q = \Delta m + e$$

with $\Delta = \left\lfloor \frac{q}{t} \right\rfloor$.

$$||e||_\infty < \Delta/2 \to \text{correct decryption.}$$

Mult(c_1, c_2):

$\Delta m_{\text{mult}} + e_{\text{mult}}$.

$$||e_{\text{mult}}||_\infty > \delta \cdot \max\{||e_1||_\infty, ||e_2||_\infty\}$$

4 network layers
Security level **80** bits

$$\to \begin{cases} \mathbf{d} = \mathbf{4096} \\ \mathbf{q} \approx \mathbf{2^{186}} \\ \sigma = 102 \\ t \leq 396 \end{cases} \to \begin{matrix} (\mathbf{c_0}, \mathbf{c_1}) \\ \text{186 kB} \end{matrix}$$

Fixed-point Representation [DGLLNW15, CSVW16]

FV plaintext space is R_t.

Fixed-point Representation [DGLLNW15, CSVW16]

FV plaintext space is R_t.

Balanced ternary expansion of $y \in \mathbb{R}$.

$$b_{\ell_1-1} b_{\ell_1-2} \ldots b_0 \,.\, b_{-1} b_{-2} \ldots b_{-\ell_2}$$

with $b_i \in \{-1, 0, 1\}$.

Back to the floating point

$$y = b_{\ell_1-1} 3^{\ell_1-1} + \ldots + b_0 3^0 + b_{-1} 3^{-1} + \ldots + b_{-\ell_2} 3^{-\ell_2}.$$

Fixed-point Representation [DGLLNW15, CSVW16]

FV plaintext space is R_t.

Balanced ternary expansion of $y \in \mathbb{R}$.

$$b_{\ell_1-1} b_{\ell_1-2} \ldots b_0 \,.\, b_{-1} b_{-2} \ldots b_{-\ell_2}$$

with $b_i \in \{-1, 0, 1\}$.

Back to the floating point

$$y = b_{\ell_1-1} 3^{\ell_1-1} + \ldots + b_0 3^0 + b_{-1} 3^{-1} + \ldots + b_{-\ell_2} 3^{-\ell_2}.$$

Replace $3 \to X$ and use $X^d \equiv -1$

$$b_{\ell_1-1} X^{\ell_1-1} + \ldots + b_0 X^0 - b_{-1} X^{d-1} - \ldots - b_{-\ell_2} X^{d-\ell_2} \in R_t$$

Fixed-point Representation [DGLLNW15, CSVW16]

Convert $g(X) \in R_t$ to $\mathbb{R}$.

	X^{d-1}	X^{d-2}	$\cdots$			$\cdots$	X^2	X	1
$g(X) =$	$-b_{-1}$	$-b_{-2}$	0	0	0	0	0	b_1	b_0

Fixed-point Representation [DGLLNW15, CSVW16]

Convert $g(X) \in R_t$ to $\mathbb{R}$.

	X^{d-1}	X^{d-2}	$\cdots$			$\cdots$	X^2	X	1
$g(X) =$	$-b_{-1}$	$-b_{-2}$	0	0	0	0	0	b_1	b_0
	fractional part				integer part				

Fixed-point Representation [DGLLNW15, CSVW16]

Convert $g(X) \in R_t$ to $\mathbb{R}$.

	X^{d-1}	X^{d-2}	$\cdots$			$\cdots$	X^2	X	1
$g(X) =$	$-b_{-1}$	$-b_{-2}$	0	0	0	0	0	b_1	b_0

fractional part | integer part

Replace $X^i \to -X^{i-d}$ for the fractional part.

	X^{-1}	X^{-2}	$\cdots$			$\cdots$	X^2	X	1
$h(X) =$	b_{-1}	b_{-2}	0	0	0	0	0	b_1	b_0

Fixed-point Representation [DGLLNW15, CSVW16]

Convert $g(X) \in R_t$ to $\mathbb{R}$.

	X^{d-1}	X^{d-2}	$\cdots$			$\cdots$	X^2	X	1
$g(X) =$	$-b_{-1}$	$-b_{-2}$	0	0	0	0	0	b_1	b_0

fractional part | integer part

Replace $X^i \to -X^{i-d}$ for the fractional part.

	3^{-1}	3^{-2}	$\cdots$			$\cdots$	3^2	3	1
$h(3) =$	b_{-1}	b_{-2}	0	0	0	0	0	b_1	b_0

Fixed-point Representation [DGLLNW15, CSVW16]

Convert $g(X) \in R_t$ to $\mathbb{R}$.

	X^{d-1}	X^{d-2}	$\cdots$			$\cdots$	X^2	X	1
$g(X) =$	$-b_{-1}$	$-b_{-2}$	0	0	0	0	0	b_1	b_0

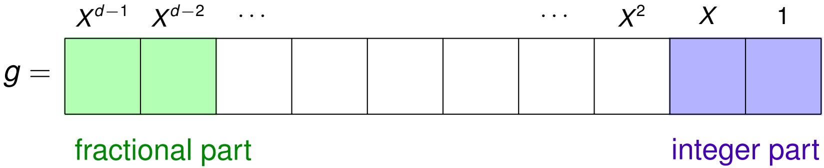
fractional part | integer part

Replace $X^i \to -X^{i-d}$ for the fractional part.

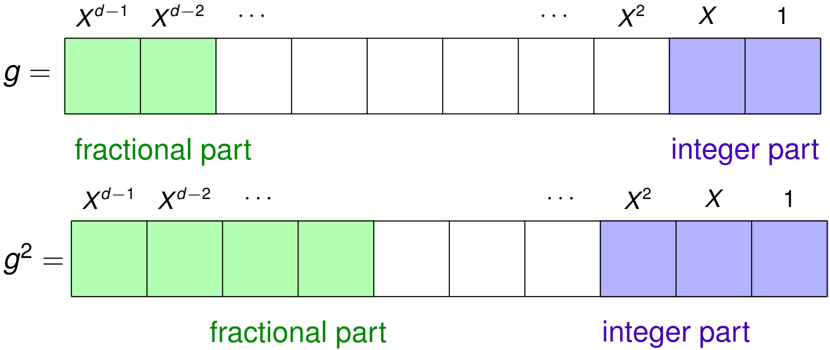
	3^{-1}	3^{-2}	$\cdots$			$\cdots$	3^2	3	1
$h(3) =$	b_{-1}	b_{-2}	0	0	0	0	0	b_1	b_0

$h(3) \in \mathbb{R}$ corresponds to $\mathbf{y} = b_1 b_0 . b_{-1} b_{-2}$.

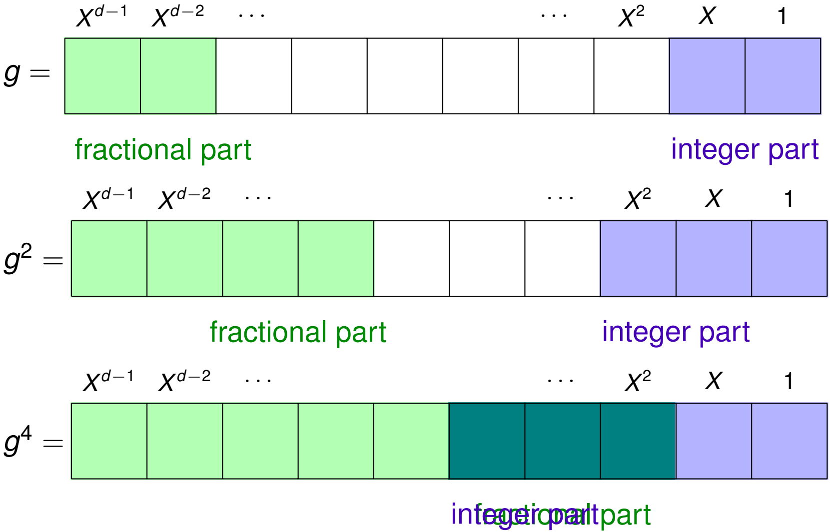
Fixed-point Representation [DGLLNW15, CSVW16]



Fixed-point Representation [DGLLNW15, CSVW16]



Fixed-point Representation [DGLLNW15, CSVW16]



Fixed-point Representation for the GMDH network

$$\begin{array}{l}\textbf{4}\text{ network layers}\\ \textbf{9}\text{ ternary bits}\end{array} \rightarrow \begin{cases} t \gtrsim 2^{106} \\ d \geq 368 \end{cases}$$

Fixed-point Representation for the GMDH network

4 network layers
9 ternary bits

$$\rightarrow \begin{cases} t \gtrsim 2^{106} \gg 396 \\ d \geq 368 < 4096 \end{cases}$$

Fixed-point Representation for the GMDH network

4 network layers
9 ternary bits

$$\rightarrow \begin{cases} t \gtrsim 2^{106} \gg 396 \\ d \geq 368 < 4096 \end{cases}$$

Use CRT with $t = t_1 \cdot t_2 \ldots t_m$ where $\forall t_i < 396$

$$R_t \rightarrow \begin{cases} R_{t_1} \\ R_{t_2} \\ \ldots \\ R_{t_m} \end{cases}$$

Fixed-point Representation for the GMDH network

$$\begin{matrix}\textbf{4}\text{ network layers}\\ \textbf{9}\text{ ternary bits}\end{matrix} \rightarrow \begin{cases} t \gtrsim 2^{106} \gg 396 \\ d \geq 368 < 4096 \end{cases}$$

Use CRT with $t = t_1 \cdot t_2 \ldots t_m$ where $\forall t_i < 396$

$$R_t \rightarrow \begin{cases} R_{t_1} \\ R_{t_2} \\ \ldots \\ R_{t_m} \end{cases} \xrightarrow{Enc} R_q \xrightarrow{GMDH} R_q \xrightarrow{Dec}$$

Fixed-point Representation for the GMDH network

$$\begin{matrix}\textbf{4}\text{ network layers}\\ \textbf{9}\text{ ternary bits}\end{matrix} \rightarrow \begin{cases} t \gtrsim 2^{106} \gg 396 \\ d \geq 368 < 4096 \end{cases}$$

Use CRT with $t = t_1 \cdot t_2 \ldots t_m$ where $\forall t_i < 396$

$$R_t \rightarrow \left\{ \begin{matrix} R_{t_1} \\ R_{t_2} \\ \ldots \\ R_{t_m} \end{matrix} \right. \xrightarrow{Enc} R_q \xrightarrow{GMDH} R_q \xrightarrow{Dec} \left. \begin{matrix} R_{t_1} \\ R_{t_2} \\ \ldots \\ R_{t_m} \end{matrix} \right\} \rightarrow R_t$$

Fixed-point Representation for the GMDH network

$$\begin{array}{l}\textbf{4} \text{ network layers} \\ \textbf{9} \text{ ternary bits}\end{array} \rightarrow \begin{cases} t \gtrsim 2^{106} \gg 396 \\ d \geq 368 < 4096 \end{cases}$$

Use CRT with $t = t_1 \cdot t_2 \ldots t_m$ where $\forall t_i < 396$

$$R_t \rightarrow \left\{ \begin{array}{l} R_{t_1} \\ R_{t_2} \\ \ldots \\ R_{t_m} \end{array} \xrightarrow{Enc} R_q \xrightarrow{GMDH} R_q \xrightarrow{Dec} \begin{array}{l} R_{t_1} \\ R_{t_2} \\ \ldots \\ R_{t_m} \end{array} \right\} \rightarrow R_t$$

Combine 13 co-prime factors to get

$$t = 95059483533087812461171515276210 \approx 2^{106.229}$$

Results in the Encrypted Mode

Test platform
Intel Core i5-3427U CPU, 1.8GHz,
FV-NFLlib

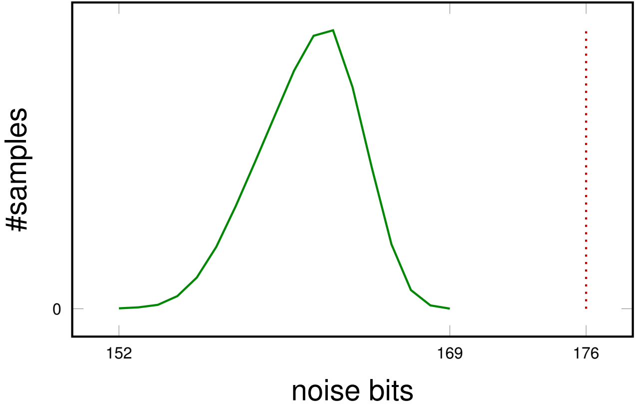
Time
One modulus t_i: **2.5** sec
One prediction: **32** sec

Results in the Encrypted Mode

Test platform
Intel Core i5-3427U CPU, 1.8GHz, FV-NFLlib

Time
One modulus t_i: **2.5** sec
One prediction: **32** sec

Output noise distribution



Conclusion

- First homomorphic prediction algorithm.
- Reasonable timings.
- High accuracy comparable with the best forecasting methods.
- Other possible applications:
  - financial data,
  - biometric data.

Conclusion

- First homomorphic prediction algorithm.
- Reasonable timings.
- High accuracy comparable with the best forecasting methods.
- Other possible applications:
  - financial data,
  - biometric data.

Thank you!