Error detection in data storage systems and distributed voting protocols

Alexei Uteshev

St.Petersburg State University

09.07.2021

Overview

1. Distributed voting protocol (Shamir scheme)

2. Reed–Solomon codes (Berlekamp & Welch protocol).

3. Polynomial interpolation (redundant data set with systematic erroneous values).

Voting protocol: attitude to confidentiality

1. All the decisions are assumed to be open.
2. All the decisions are assumed to be secret.

Backgrounds: interpolation

$$\{(x_j, y_j)\}_{j=1}^{K} \subset \mathbb{Q}^2$$

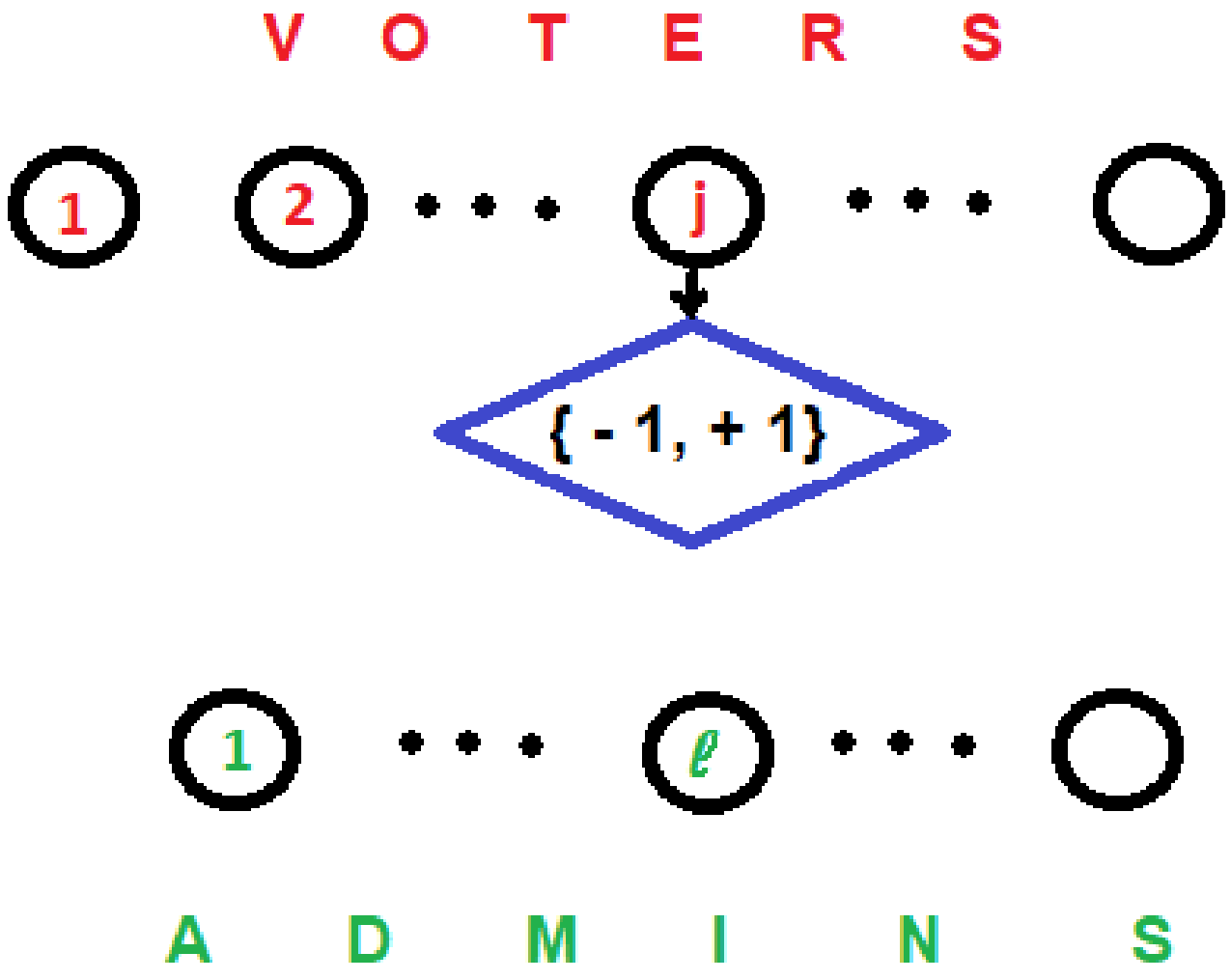
Find a polynomial $f(x)$ such that $\{f(x_j) = y_j\}_{j=1}^{K}$.
If $\deg f \leq K - 1$ then Lagrange representation is valid:

$$f(x) \equiv \sum_{j=1}^{K} y_j \frac{W_j(x)}{W_j(x_j)} \equiv \sum_{j=1}^{K} y_j \frac{W_j(x)}{W'(x_j)} \,.$$

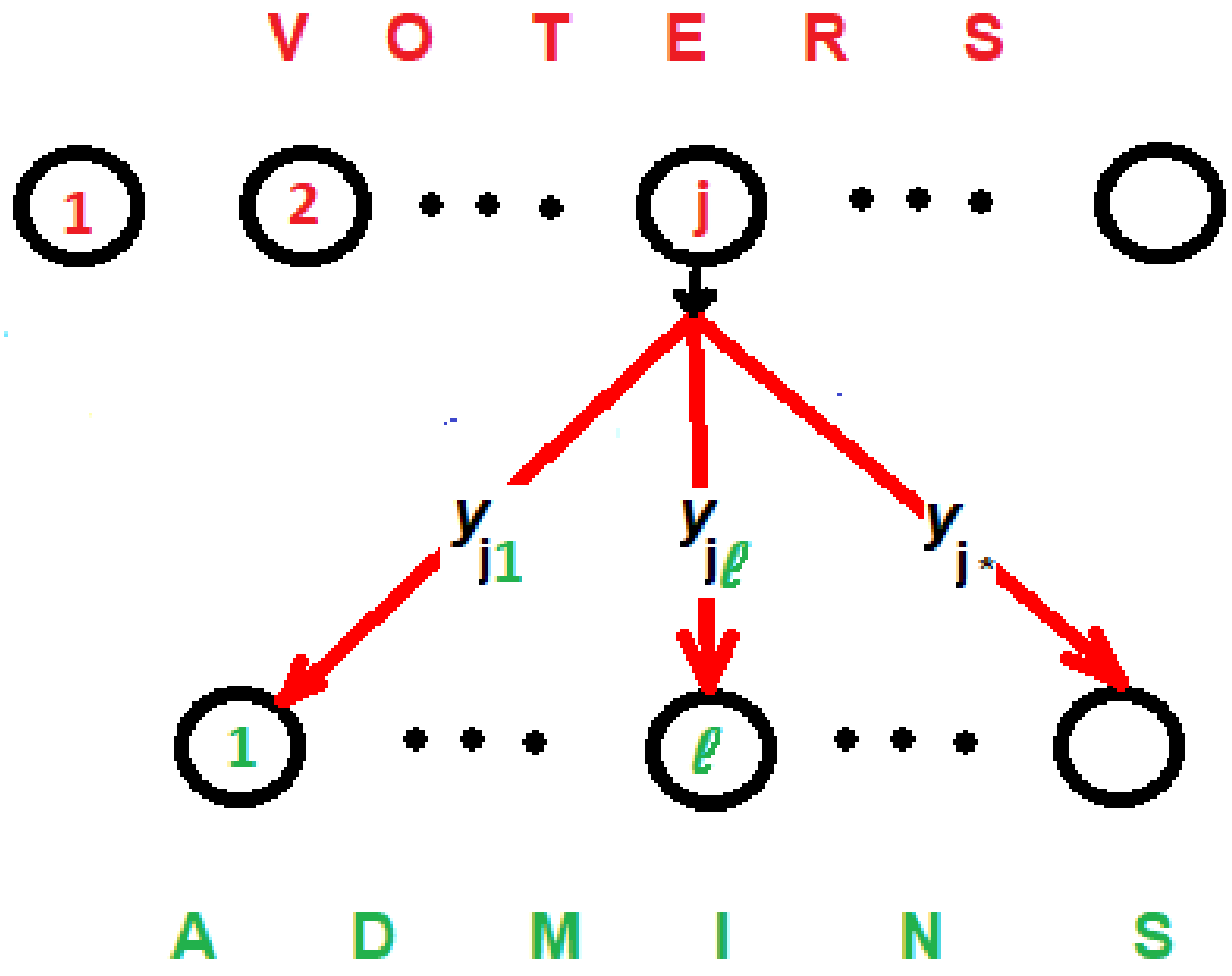
Here

$$W(x) := \prod_{j=1}^{K} (x - x_j) \;, W_j(x) := \frac{W(x)}{x - x_j} \quad \text{for} \;\; j \in \{1, \ldots, K\} \,.$$

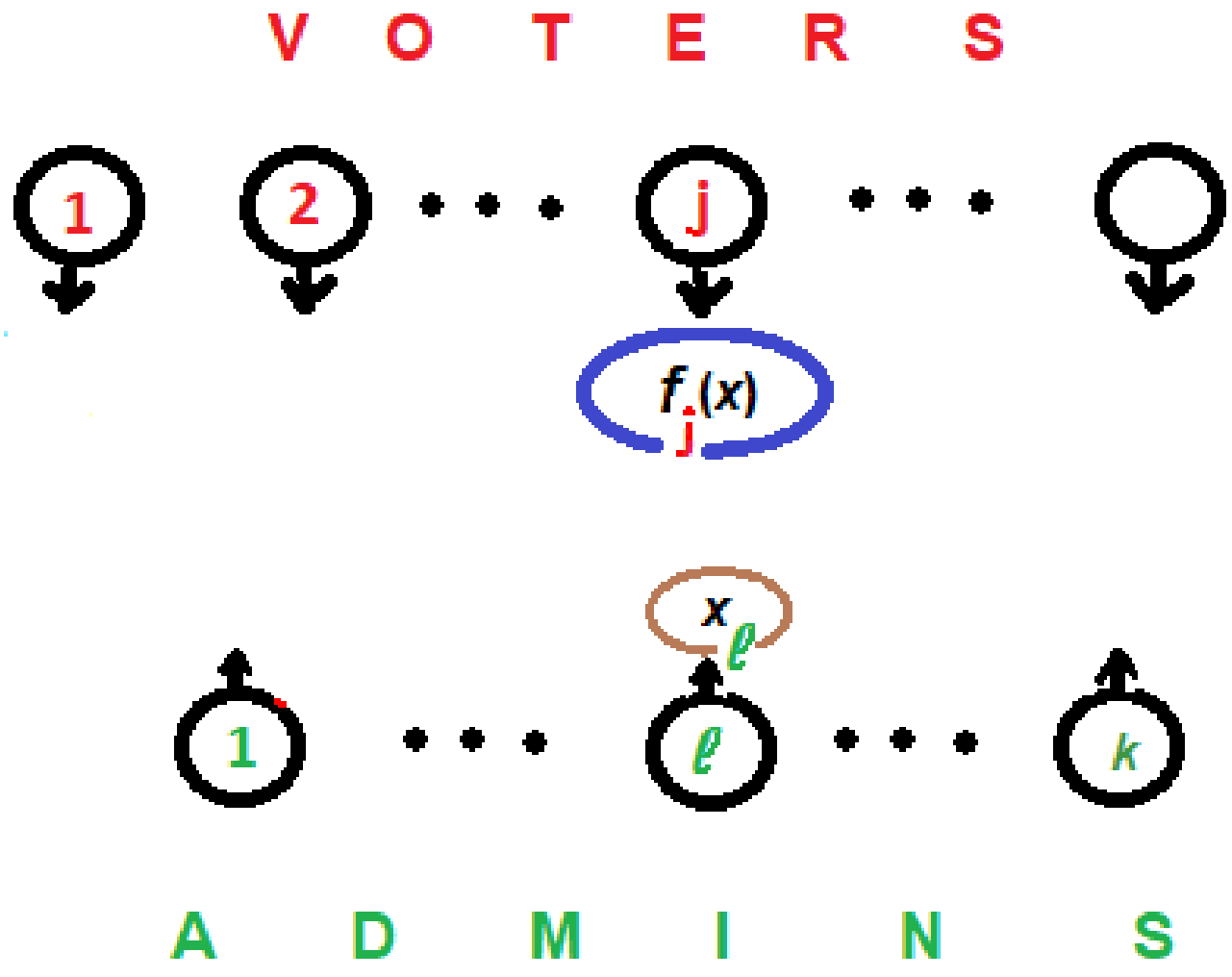
Decentralized voting



Decentralized voting



Shamir's secret sharing



Shamir's secret sharing

jth voter generates a polynomial over $\mathbb{Z}$:

$$f_j(x) := \underbrace{A_{j0}x^{k-1} + \cdots + A_{j,k-2}x}_{\textit{arbitrary}} + \left\{ \begin{array}{c} +1 \\ -1 \end{array} \right\}$$

$f_j(0) =$ his secret vote

Shamir's secret sharing

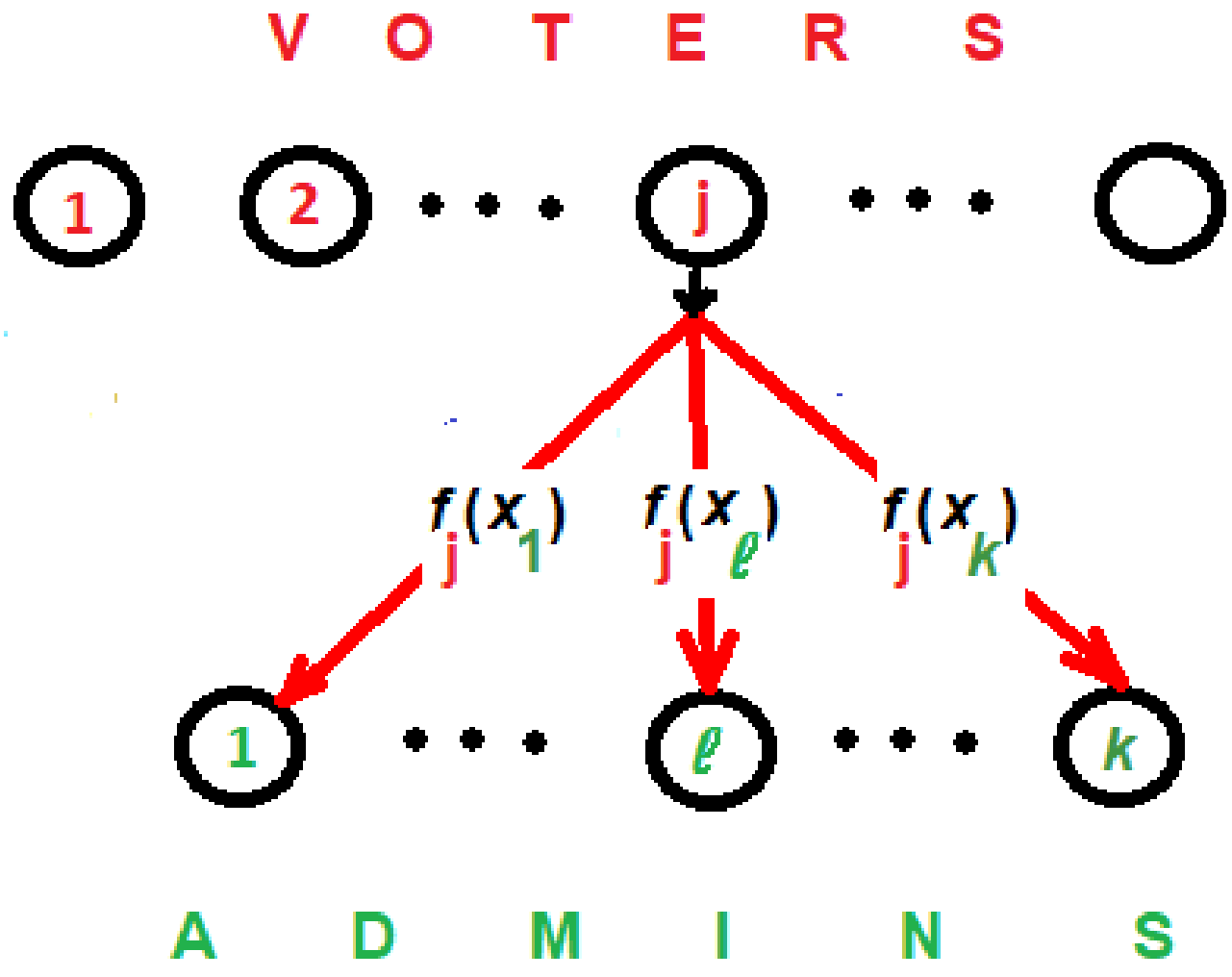
jth voter generates a polynomial over $\mathbb{Z}$:

$$f_j(x) := \underbrace{A_{j0}x^{k-1} + \cdots + A_{j,k-2}x}_{\textit{arbitrary}} + \left\{ \begin{array}{c} +1 \\ -1 \end{array} \right\}$$

$f_j(0) =$ his secret vote

$f_j(x_1), \ldots, f_j(x_\ell), \ldots$ — shares of the secret

Shamir's secret sharing



Decentralized voting

$$F(x) := \sum_j f_j(x)$$

$\deg F \leq k-1, F(0) = \sum_j f_j(0) =$ result of voting

ℓth admin collects all the delivered shares $Y_\ell := \sum_j f_j(x_\ell) = F(x_\ell)$

$F(x)$ can be found as an interpolation polynomial for

x	x_1	x_2	$\ldots$	x_k
y	Y_1	Y_2	$\ldots$	Y_k

Decentralized voting: redundancy

What if the number K of admins is greater than $k = 1 + \deg f_j$?

Any subset of k admins are able to restore the result of voting.

Decentralized voting: redundancy

What if some of admins' data are compromised?

$$\{\widehat{Y_1}, \ldots, \widehat{Y_K}\} \neq \{Y_1, \ldots, Y_K\}$$

x	x_1	x_2	$\ldots$	x_K
y	$\widehat{Y_1}$	$\widehat{Y_2}$	$\ldots$	$\widehat{Y_K}$

Assumption. Let we have an owerhelming redundancy of the correct values.

Error detection: some algebra

The data set

x	-3	-2	-1	0	1	2	3	4	5
y	19	-2	-7	-8	3	14	37	35	107

is generated by a second order polynomial

Error detection: some algebra

The data set

x	-3	-2	-1	0	1	2	3	4	5
y	19	-2	-7	-8	3	14	37	35	107

is generated by a second order polynomial with the exception of some erroneous values.

$$\tau_\ell := \sum_{j=1}^{K=9} y_j \frac{x_j^\ell}{W'(x_j)};\ W(x) := \prod_{j=-2}^{5} (x - x_j)$$

$$\tau_0 = \frac{1}{70},\ \tau_1 = \frac{53}{1680},\ \tau_2 = \frac{193}{1680}, \ldots$$

$$\mathcal{H}_1(x) := \begin{vmatrix} \tau_0 & \tau_1 \\ 1 & x \end{vmatrix},\ \mathcal{H}_2(x) := \begin{vmatrix} \tau_0 & \tau_1 & \tau_2 \\ \tau_1 & \tau_2 & \tau_3 \\ 1 & x & x^2 \end{vmatrix}$$

Error detection: some algebra

$$\mathcal{H}_3(x) := \begin{vmatrix} \tau_0 & \tau_1 & \tau_2 & \tau_3 \\ \tau_1 & \tau_2 & \tau_3 & \tau_4 \\ \tau_2 & \tau_3 & \tau_4 & \tau_5 \\ 1 & x & x^2 & x^3 \end{vmatrix}$$

$$\equiv \frac{33}{313600}(x+2)(x-1)(x-4)\,.$$

x	-3	-2	-1	0	1	2	3	4	5
y	19	-2	-7	-8	3	14	37	35	107

$$f(x) = 4\,x^2 + 3\,x - 8.$$

Error locator polynomial

Th. 1. $E :=$ number of errors. If

$$E < (\text{number of points} - \text{degree of a polynomial})/2$$

then the polynomial

$$\mathcal{H}_E(x) := \begin{vmatrix} \tau_0 & \tau_1 & \tau_2 & \cdots & \tau_E \\ \tau_1 & \tau_2 & \tau_3 & \cdots & \tau_{E+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \tau_{E-1} & \tau_E & \tau_{E+1} & \cdots & \tau_{2E-1} \\ 1 & x & x^2 & \cdots & x^E \end{vmatrix}_{(E+1)\times(E+1)}$$

possesses the zero set coinciding with the erroneous values of x in the data set.

Backgrounds

Carl Jacobi (1846) Rational interpolation

Berlekamp & Welch (1986) Error correction codes

$$\underbrace{D_0, D_1, \ldots, D_{k-1}}_{\textit{information}}, \overbrace{D_k, \ldots, D_{n-1}}^{\textit{redundant}} \subset \mathbf{GF}(2^N)$$

$$f(\mathfrak{x}): \quad \begin{array}{r|cccc} \mathfrak{x} & \mathfrak{a}^0 & \mathfrak{a}^1 & \ldots & \mathfrak{a}^{k-1} \\ \hline f(\mathfrak{x}) & D_0 & D_1 & \ldots & D_{k-1} \end{array}$$

$$D_j := f(\mathfrak{a}^j) \quad \text{for} \quad j = k, \ldots, n-1$$

Systematic vs. nonsystematic errors

Infinite fields: problem of outliers

x	-3	-2	-1	0	1	2	3	4	5
y	19.2	-1.8	-7.1	-8.2	3.2	14.1	37.0	34.9	106.5

Is it possible to distinguish systematic and nonsystematic errors?

Thank you for attention!