

UNIVERSITATEA BUCUREȘTI
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ

Protocoale Criptografice de Grup

Lucrare de Licență

Coordonator:

Lect.dr. Ruxandra F. Olimid

Student:

Dragoș Alin Rotaru

BUCUREȘTI
2015

Abstract

Criptografia modernă se găsește pretutindeni în viața de zi cu zi de la ATM-uri, cartele telefonice la protocoale de autentificare, licitații electronice sau bani digitali, luând amploare o dată cu apariția sistemelor cu cheie publică. O definiție potrivită este *studiul științific al tehnicilor pentru a securiza informația digitală, tranzacțiile și calculul distribuit* [1]. În cadrul criptografiei există primitive multiple care stau la baza tuturor protocoalelor precum: semnături digitale, vot electronic, scheme de partajare, schimbul sigur de chei. Un rol important îl deține criptografia de grup, care permite definirea și analiza securității unor protocoale la care iau parte mai mult de 2 participanți (vot electronic, semnături de grup, partajarea de secrete).

Păstrarea confidențialității și redundanței datelor pe perioade lungi de timp reprezintă o sarcină dificilă din cauza multiplelor probleme care pot apărea precum atacuri, erori umane, defecțiunea unor componente hardware sau dezastre naturale [2]. Problemele sunt cu atât mai actuale astăzi, când organizațiile își păstrează datele în cloud pentru scăderea costurilor necesare întreținerii unui centru propriu de date (datacenter), deci informațiile sunt stocate și procesate în afara unui control fizic al proprietarului.

Lucrarea de față analizează câteva dintre primitivele criptografice prezentând protocoale construite pe baza lor precum și contribuții personale în dezvoltarea lor [3,4]. În final introducem o arhitectură implementată cu ajutorul schemelor de partajare, eliminând astfel câteva vulnerabilități prezente la sistemele menționate anterior.

Cuprins

Structură	5
0.1 Introducere + Motivatie - deocamdata text reciclabil	5
1 Funcții Pseudoaleatoare	7
1.1 Introducere	7
1.2 Definiții de Securitate	7
1.3 Funcții Pseudoaleatoare Constrânse	8
2 NIKE	10
2.1 ID-NIKE	10
2.1.1 Construcția Boneh-Waters (2013)	11
2.2 Multi-Party ID-NIKE	12
2.2.1 Proprietăți de Securitate	12
2.2.2 Construcție Multi-Party ID-NIKE bazată pe funcții pseudoaleatoare constrânse	12
2.3 Contribuții	12
3 Scheme de Partajare	13
3.1 Introducere	13
3.2 Schema unanimă	14
3.3 Schema Shamir	14
3.4 Schema Ito, Saito și Nishizeki	15
3.5 Schema Feldman	17
4 Sisteme de stocare de lungă durată	19
4.1 Introducere	19
4.2 Criptare vs. scheme de partajare	19
4.3 Sisteme de stocare de lungă durată bazate pe scheme de partajare	20
4.3.1 PASIS	20
4.3.2 GridSharing	21
4.3.3 POTSHARDS	23
4.4 Sistemul de stocare Alouneh et al. (2013)	24
4.5 Contribuții	26
4.5.1 Vulnerabilități evidențiate	27

4.5.2	Implementare și rezultate practice	29
5	Implementare alternativă a sistemelor de stocare	36
5.1	Arhitectură bazată pe scheme de partajare	36
5.2	Instanțe Amazon EC2	36
5.3	Reconstituirea fișierelor	36

Introducere și organizarea lucrării

- TODO Completeaza descrierea
- Secțiunea ?? oferă o scurtă introducere în domeniul schemelor de partajare și descrie în detaliu construcțiile care vor fi folosite în secțiunile următoare.
- Secțiunea 4.3 conține cerințele pe care trebuie să le îndeplinească un sistem de stocare de lungă durată și compară sistemele de stocare de lungă durată care utilizează criptarea cu sistemele de stocare bazate pe scheme de partajare a secretelor. De asemenea, sunt descrise câteva sisteme considerate relevante pentru înțelegerea contribuțiilor până în prezent.
- Secțiunea 4.4 introduce un sistem de stocare criptanalizat.
- Secțiunea 4.5 conține contribuțiile de cercetare, punând în evidență problemele cauzate de determinismul sistemului introdus în Secțiunea 4.4.

0.1 Introducere + Motivatie - deocamdata text reciclabil

Termenul de criptografie este definit în dicționarul Oxford ca fiind *arta de a scrie și a rezolva coduri*. Criptografia modernă s-a desprins de cea clasică în jurul anilor '80, motivând implementarea rigurozității matematice pentru definirea construcțiilor criptografice. Aceasta pentru că în anii anteriori, experiența a dovedit nesiguranța metodelor de criptare, criptanaliza lor fiind uneori trivială (cifrul lui Cezar, Vigenere [5], [6]) sau uneori atinsă cu ceva mai mult efort precum Enigma și alte metode din cel de-al doilea război mondial. [7]

Schemele de partajare sunt utilizate în aplicații precum: protejarea și recuperarea cheilor criptografice, vot electronic, certificate distribuite unor autorități, licitații on-line sau sisteme de stocare de lungă durată. [8]

O schemă de partajare reprezintă o metodă de a distribui un secret unor participanți, oferind fiecărui participant o componentă (share) astfel încât doar o submulțime de participanți pot recupera secretul inițial.

Cum volumul de datelor electronice este în continuă creștere, necesitatea stocării sigure a devenit extrem de necesară.

Date sensibile trebuie menținute secrete, câteodată pentru lungi perioade de timp (zeci de ani). Exemple includ informații militare, documente legale sau dosare medicale: informațiile militare pot fi clasificate zeci de ani, un testament trebuie menținut secret până la deschiderea sa oficială, un dosar medical necesită păstrare sigură (cel puțin) pe toată durata vieții pacientului.

Capitolul 1

Funcții Pseudoaleatoare

1.1 Introducere

O funcție pseudoaleatoare (PRF) $F : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$ este un algoritm polinomial determinist definit pe mulțimea cheilor \mathcal{K} unde \mathcal{X} , \mathcal{Y} reprezintă domeniul, respectiv codomeniul lui F . Primele construcții au fost definite în 1986 pornind de la un PRG (generator pseudoaleator) sigur [9]. Această primitivă este folosită în special pentru generarea de chei în protocoale actuale precum TLS (fostul SSL) [10].

Un caz particular de PRF în care $\mathcal{Y} = \mathcal{X}$ și F este bijectivă, sunt permutările pseudoaleatoare (PRP): AES, 3DES, având proprietatea utilă de a admite inversă.

Intuitiv, funcția F poate fi evaluată eficient (în timp polinomial) în toate punctele $x \in \mathcal{X}$ din momentul în care o cheie $k \in \mathcal{K}$ este fixată. Din punct de vedere al securității, un adversar PPT (timp probabilistic polinomial) \mathcal{A} nu poate face diferența între o instanță aleatoare $F(k, \cdot)$, $k \in \mathcal{K}$ și o funcție aleasă aleator din toate funcțiile posibile având domeniul \mathcal{X} , codomeniul \mathcal{Y} .

1.2 Definiții de Securitate

În procesul formalizării noțiunii de securitate *adaptivă* a unei funcții PRF F , vom avea în vedere un adversar PPT \mathcal{A} . În definiție sunt prezentate 2 experimente cu adversarul \mathcal{A} : $EXP(b)$, $b \leftarrow^R \{0, 1\}$, iar avantajul lui va consta în abilitatea de a le distinge. Notăm cu $s \leftarrow^R S$ faptul că s este ales în mod aleator uniform din mulțimea S .

La începutul experimentului este apelată funcția $Setup(1^\lambda)$ care primește la intrare un parametru de securitate λ și întoarce o cheie secretă $k \leftarrow^R \mathcal{K}$. Fie $V, C \subseteq X$ 2 mulțimi de puncte (inițial vide) pentru care \mathcal{A} cunoaște valorile $F(k, \cdot)$ trimise la **F.eval**, respectiv la **F.challenge**. De asemenea, \mathcal{A} are acces la 2 tipuri de oracole:

- **F.eval**: pentru un $x \in \mathcal{X}$ întoarce $F(k, x)$ dacă $x \notin C$ și actualizează mulțimea $V \leftarrow V \cup x$, altfel \perp .
- **F.challenge**: pentru un $x \notin V$ întoarce $F(k, x)$ dacă $b = 0$ și $y \leftarrow^R \mathcal{Y}$ dacă $b = 1$ iar $C \leftarrow C \cup x$.

La finalul experimentului, \mathcal{A} întoarce o valoare $b' \in \{0, 1\}$.

Fie W_b evenimentul ca în urma eperimentului $EXP(b)$, b' să fie 1 la sfârșitul jocului. Notăm cu $AdvPRF_{\mathcal{A}, F}(\lambda) = |Pr[W_0] - Pr[W_1]|$ avantajul lui \mathcal{A} .

Definiția 1.3.1. *O funcție PRF F este sigură dacă pentru orice adversar PPT \mathcal{A} , $AdvPRF_{\mathcal{A}, F}(\lambda)$ este o funcție neglijabilă.*

1.3 Funcții Pseudoaleatoare Constrânse

O funcție cPRF (pseudoaleatoare constrânsă) $F : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$ este o funcție PRF care deține *chei de constrângere* ce permit evaluarea funcției F în subdomenii $S \subseteq \mathcal{X}$ menținând proprietățile de securitate ale funcțiilor PRF.

Definiția 1.3.1. *O funcție PRF $F : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$ este constrânsă la o mulțime $S \subseteq 2^{\mathcal{X}}$ dacă există o mulțime în plus de chei \mathcal{K}_c și 2 algoritmi:*

- **F.constrain**(k, S): *un algoritm nedeterminist care pentru o cheie $k \in \mathcal{K}_c$ și o mulțime $S \subseteq \mathcal{S}$ întoarce o cheie $k_s \in \mathcal{K}_c$ pentru care se poate face evaluarea lui $F(k, x)$ pentru orice $x \in S$.*
- **F.eval**(k_s, x): *un algoritm determinist care pentru o cheie $k_s \in \mathcal{K}_c$ și un punct $x \in \mathcal{X}$ întoarce $F(k, x)$ dacă $x \in S$, sau \perp dacă $x \notin S$*

unde $\perp \notin \mathcal{Y}$

Alte construcții pe baza cPRF definite în [11] și care vor fi folosite ulterior sunt: *stânga/dreapta PRF* - funcții cPRF ce admit evaluarea lor având partea stângă a domeniului fixată; *bit-fixate PRF* - funcții cPRF care pot fi evaluate în domenii corespunzătoare unei expresii având unii biți fixați. Autorii [11] menționează că aceste construcții sunt posibile în cadrul funcțiilor multiliniare iar problema devine mai interesantă (deocamdată nerezolvată) în contextul LWE [12].

Definiția 1.3.2. *Fie o funcție $F : \mathcal{K} \times \mathcal{X}^2 \rightarrow \mathcal{Y}$ PRF. $\forall v \in \mathcal{X}$ o funcție stânga/dreapta admite 2 chei de constrângere k_v^L și k_v^R ce permit evaluarea funcției F în toate punctele $(v, x) \in \mathcal{X}^2$, respectiv $(x, v) \in \mathcal{X}^2$. Notăm această funcție cPRF stânga/dreapta cu $PRF^{L/R}$.*

Definiția 1.3.3. *Fie o funcție $F : \mathcal{K} \times \{0, 1\}^n \rightarrow \mathcal{Y}$. $\forall v \in \{0, 1, ?\}^n$ o funcție bit-fixată admite o cheie de constrângere k_v ce permite evaluarea funcției F în*

toate punctele $x \in \{0, 1\}^n$ ce satisfac modelul lui v . Notăm această funcție $cPRF$ bit-fixată cu PRF^{bf} .

Noțiunea de securitate pentru $cPRF$ [11] este asemănătoare cu cea pentru PRF : 2 experimente $EXP(0), EXP(1)$, doar că în acest caz este necesară existența unui oracol $F.constrain$ pentru a permite adversarului să interogheze F conform definiției 1.3.1.

La începutul fiecărui experiment este apelată funcția $Setup(\lambda)$ și întoarce o cheie secretă $k \leftarrow^R \mathcal{K}$. Inițial mulțimile $V, C \subseteq \mathcal{X}$ sunt vide iar adversarul \mathcal{A} are acces la 3 oracole:

- **F.eval**: pentru un $x \in \mathcal{X}$ întoarce $F(k, x)$ dacă $x \notin C$ și actualizează mulțimea $V \leftarrow V \cup x$, altfel \perp ;
- **F.constrain**: pentru o mulțime $S \in \mathcal{S}$ întoarce $F.constrain(k, S)$ dacă $S \cap C = \emptyset$ și actualizează mulțimea $V = V \cup x$, altfel \perp ;
- **F.challenge**: pentru un $x \notin V$ întoarce $F(k, x)$ dacă $b = 0$ și $y \leftarrow^R \mathcal{Y}$ dacă $b = 1$ iar $C \leftarrow C \cup x$;

La finalul experimentului, \mathcal{A} întoarce o valoare $b' \in \{0, 1\}$.

Fie W_b evenimentul ca în urma experimentului $EXP(b)$, b' să fie 1 la sfârșitul jocului. Notăm cu $AdvPRF_{\mathcal{A}, F}(\lambda) = |Pr[W_0] - Pr[W_1]|$ avantajul lui A .

Definiția 1.3.4. O funcție $cPRF$ F este sigură dacă pentru orice adversar PPT \mathcal{A} , $AdvPRF_{\mathcal{A}, F}(\lambda)$ este o funcție neglijabilă.

Capitolul 2

NIKE

NIKE (Non Interactive Key Exchange - schimb de chei non-interactiv) reprezintă o primitivă criptografică importantă în care mai mulți participanți pot stabili de comun acord o cheie comună fără ca aceștia să fie on-line în același timp. Un exemplu des întâlnit în zilele noastre este schimbul de chei Diffie-Hellman [13]. Construcția a fost apoi extinsă la 3 persoane de Joux folosind funcții biliniare [14] iar Boneh și Silverberg au generalizat la n persoane (multi-party) bazându-se pe funcții multi-liniare [15].

2.1 ID-NIKE

Criptografia bazată pe identitatea participanților a fost introdusă de Shamir [16]. Aceasta presupune folosirea identității (email, nume, stradă), informație care îl identifică pe participant într-un mod unic, fără să genereze o pereche de cheie publică, secretă. Principalul dezavantaj este nevoia de o terță autorizată care să administreze generarea de chei secrete, însă avantajul este utilizarea de mai puține resurse din partea participanților. Așadar criptografia bazată pe ID este ideală într-un mediu în care utilizatorii dețin resurse computaționale limitate (rețea de routere).

ID-NIKE combină principiile oferite de NIKE cu ID-urile participanților. Formal, ID-NIKE constă în 3 algoritmi:

- $Setup(\lambda)$: primește ca date de intrare parametrul de securitate λ și întoarce o cheie secretă msk împreună cu parametrii publici $params$;
- $Extract(msk, id_i)$: primește ca date de intrare cheia secretă msk , identitatea $id_i \in \mathcal{ID}$ și întoarce cheia privată sk_{id_i} ;
- $KeyGen(params, sk_{id_i}, id_j)$: primește ca date de intrare parametrii publici $params$, o cheie secretă sk_{id_i} , o identitate $id_j \neq id_i$ și întoarce cheia

comună k_{id_i, id_j} .

Corectitudine

Este garantată dacă pentru orice pereche de participanți, $id_i, id_j \in \mathcal{ID}$, $id_i \neq id_j$ $KeyGen(params, sk_{id_i}, id_j) = KeyGen(params, sk_{id_j}, id_i)$.

Securitate

Amintim definiția de securitate adaptivă pentru ID-NIKE din [17]. Fie $EXP(b)$, $b \leftarrow^R \{0, 1\}$, două experimente cu un adversar $PPT \mathcal{A}$. La începutul fiecărui experiment $EXP(b)$ se execută $Setup(\lambda)$ oferind adversarului \mathcal{A} $params$ și menținând msk privat. \mathcal{A} are acces la 3 oracole:

- $Ext(id_i)$: primește ca date de intrare o identitate id_i , apelează $Extract(msk, id_i)$ și întoarce sk_{id_i} ;
- $Rev(id_i, id_j)$: primește ca date de intrare 2 identități id_i, id_j , $id_i \neq id_j$, apelează funcția $Extract(msk, id_i)$ și întoarce cheia comună k_{id_i, id_j} din $KeyGen(params, sk_{id_i}, id_j)$;
- $Test(id_i, id_j)$: primește ca date de intrare 2 identități id_i, id_j , $id_i \neq id_j$. Dacă $b = 0$ atunci întoarce $Rev(id_i, id_j)$, altfel întoarce $sk_{id_i, id_j} \leftarrow^R \mathcal{SK}$ unde cu \mathcal{SK} am notat spațiul de chei posibile.

Pentru a elimina situația în care adversarul \mathcal{A} câștigă trivial, acestuia nu îi este permis să interogheze Ext pentru o identitate în $Test$, precum și faptul că toate interogările Rev și $Test$ să fie distincte.

La finalul experimentului, \mathcal{A} întoarce o valoare $b' \in \{0, 1\}$.

Fie W_b evenimentul ca în urma experimentului $EXP(b)$, b' să fie 1 la sfârșitul jocului. Notăm cu $AdvKE_{\mathcal{A}}(\lambda) = |Pr[W_0] - Pr[W_1]|$ avantajul lui \mathcal{A} .

Definiția 2.1.1. *ID-NIKE este adaptiv sigură dacă pentru orice adversar PPT , \mathcal{A} funcția $AdvKE_{\mathcal{A}}(\lambda)$ este neglijabilă.*

2.1.1 Construcția Boneh-Waters (2013)

În figura 2.1 prezentăm construcția Boneh-Waters bazată pe funcții constrânse stânga/dreapta [11].

Setup(λ): Fie $F : \mathcal{K} \times \mathcal{X}^2 \rightarrow \mathcal{Y}$ o funcție PRF stânga/dreapta sigură, alege în mod aleator $msk \leftarrow^R \mathcal{K}$ și întoarce parametrii publici $params$ corespunzători funcției F ;
Extract(msk, id_i): calculează $F.constrain(msk, \{(id_i, \cdot)\})$ pentru a obține $k_{id_i}^L$ și $F.constrain(msk, \{(\cdot, id_i)\})$ pentru a obține $k_{id_i}^R$, apoi întoarce $sk_{id_i} = (k_{id_i}^L, k_{id_i}^R)$;
KeyGen($params, sk_{id_i}, id_j$): întoarce $F(msk, (id_i, id_j))$ dacă $id_i < id_j$ și $F(msk, (id_j, id_i))$ dacă $id_i > id_j$ (în ordine lexicografică).

Figure 2.1: Construcția Boneh-Waters [11]

Corectitudine**Securitate****2.2 Multi-Party ID-NIKE**

Multi-Party (mai multe părți) ID-NIKE reprezintă generalizarea protocolului ID-NIKE pentru cazul cu n participanți ($n > 2$). Mai întâi definim modelul de securitate pentru acest protocol după care introducem construcția multi-party bazată pe funcții bit-fixate [3].

2.2.1 Proprietăți de Securitate**2.2.2 Construcție Multi-Party ID-NIKE bazată pe funcții pseudoaleatoare constrânse****2.3 Contribuții**

Capitolul 3

Scheme de Partajare

3.1 Introducere

O schemă de partajare constă în distribuirea unei informații secrete \mathcal{S} la mai mulți participanți $\mathcal{P} = \{P_1, \dots, P_n\}$ astfel încât oricare mulțime de participanți, predefinită ca făcând parte dintr-o structură de acces pe care o vom denumi \mathcal{A} , să poată reconstitui secretul \mathcal{S} . Formal, o schemă de partajare este reprezentată de o pereche de algoritmi (Gen, Rec) :

- $Gen(\mathcal{S}, m)$ este un algoritm care primește la intrare un secret \mathcal{S} și un număr întreg m și întoarce un set de componente s_1, s_2, \dots, s_m .
- $Rec(s_{i_1}, s_{i_2}, \dots, s_{i_q})$ este un algoritm care primește ca parametri de intrare o mulțime de componente și întoarce \mathcal{S} dacă mulțimea $\{P_{i_1}, P_{i_2} \dots P_{i_q}\} \in \mathcal{A}$.

Majoritatea schemelor constau în mai multe etape precum:

- *Inițializare.* Presupune inițializarea variabilelor de mediu necesare.
- *Generare.* O entitate autorizată (numită dealer) \mathcal{D} folosește algoritmul Gen pentru a genera componentele.
- *Distribuție.* Componentele sunt trimise participanților cu ajutorul unui mijloc de comunicare sigur, fără ca acestea să fie vizibile unui atacator.
- *Reconstrucție.* Dându-se o mulțime de componente, se folosește algoritmul Rec pentru a recupera secretul \mathcal{S} .

Schemele de partajare se clasifică în funcție de cantitatea de informație secretă pe care o pot obține persoanele care nu fac parte din \mathcal{A} :

- *Sisteme perfecte de partajare:* componentele nu oferă nici o informație teoretică despre \mathcal{S} indiferent de resursele computaționale.

- *Sisteme statistic sigure*: o fracțiune de informație este dezvaluită despre \mathcal{S} , independent de puterea computațională a adversarului.
- *Sisteme computațional-sigure de partajare*: se bazează pe faptul că reconstituirea lui \mathcal{S} se reduce la o problema *dificilă* (spre exemplu, problema Diffie-Hellman [18]) în lipsa unor informații oferite doar grupului de acces \mathcal{A} . [8]

Primele scheme de partajare au fost dezvoltate independent de Shamir și Blakley în 1979 [19, 20].

Denumite și scheme majoritare (k, n) , acestea rezolvau cazul în care oricare grup de participanți cu un număr mai mare sau egal decât k (mărimea pragului) pot reconstitui secretul \mathcal{S} din componentele primite de la dealerul \mathcal{D} . Dacă schema este perfect *sigură* atunci oricare grup cu un număr de participanți mai mic decât k nu obține nici o informație despre \mathcal{S} .

Schemele majoritare (spre exemplu schema Shamir) sunt insuficiente pentru a permite partajarea pentru anumite structuri de acces. Considerăm cazul în care dorim să partajăm un secret între 4 participanți: P_1, P_2, P_3, P_4 astfel încât $\{P_1, P_2\}$ și $\{P_3, P_4\}$ să fie singurele mulțimi autorizate pentru reconstrucția secretului \mathcal{S} (i.e. $\mathcal{A} = \{\{P_1, P_2\}, \{P_3, P_4\}\}$). În mod evident, problema nu poate fi rezolvată cu o structură de acces de tip prag: anumite mulțimi cu un număr de 2 participanți trebuie să poată reconstrui secretul ($\{P_1, P_2\}, \{P_3, P_4\}$), în timp ce altele nu ($\{P_1, P_3\}, \{P_1, P_4\}, \{P_2, P_3\}, \{P_2, P_4\}$)

Astfel de scheme de partajare pentru structuri de acces generale au fost dezvoltate de Ito, Saito și Nishizeki, realizând o generalizare a schemei Shamir. [21] Benaloh și Leichter au demonstrat că schemele de partajare de tip prag nu pot fi folosite pe structuri general monotone (familie de submulțimi ale lui \mathcal{P} cu proprietatea că dacă $A \in \mathcal{A}$ și $A \subset A'$, atunci $A' \in \mathcal{A}$) și obțin o construcție mai eficientă ca Ito et. al din punct de vedere al numărului de componente distribuite participanților. [22]

3.2 Schema unanimă

Presupunând că împărțim un secret \mathcal{S} la n participanți astfel încât \mathcal{S} să poată fi recuperat doar dacă toți cei n participanți își combină componentele pe care le dețin. Metoda este echivalentă cu o schemă (n, n) majoritară. Un exemplu este schema introdusă de Karin, Greene și Hellman (Fig.3.1). [23]

3.3 Schema Shamir

Schema Shamir oferă mai multă flexibilitate decât schema unanimă prin faptul că oricare k (sau mai mulți) participanți din cei n pot recupera \mathcal{S} , însă mai

Inițializare:

- Fie $S \in Z_q$ unde $q > 1$ și q prim;
- Fie n numărul de participanți;

Generare: Dealerul \mathcal{D} :

- Alege $n - 1$ valori aleatoare $s_i \leftarrow^R Z_p, i \in \{1, 2, \dots, n - 1\}$;
- $s_n = S + \sum_{i=1}^{n-1} s_i \pmod{q}$;

Distribuție: Dealerul \mathcal{D} :

- transmite în mod sigur participantului P_i componenta $s_i, i \in \{1, 2, \dots, n\}$;

Reconstrucție: Cei n participanți:

- Calculează $S = \sum_{i=1}^n s_i \pmod{q}$.

Figure 3.1: Schema unanimă [23]

puțin de k participanți nu obțin nicio informație despre \mathcal{S} . Schema Shamir este deci o schemă (k, n) majoritară.

Intuitiv, având k puncte în plan (x_i, y_i) , $x_i \neq x_j, i, j \in \{1, 2, \dots, k\} \forall i \neq j$, există o curbă polinomială unică de grad $k - 1$ ce trece prin ele. În schimb, pentru a defini o curbă polinomială de grad k care trece prin $k - 1$ puncte date, există o infinitate de soluții. Evident, orice submulțime de valori s_i de mărime egală cu k este suficientă și necesară pentru a reconstrui polinomul f . După interpolarea componentelor deținute de cel puțin k dintre participanți, secretul \mathcal{S} se află în $f(0)$ (Fig. 3.2). [20]

Pentru un atacator care deține chiar și $k - 1$ valori s_i , acesta nu determină nimic despre \mathcal{S} , spațiul de soluții posibile fiind identic față de situația în care deține 0 componente.

3.4 Schema Ito, Saito și Nishizeki

În continuare vom descrie modalitatea de distribuire a componentelor de la care au pornit Ito, Saito și Nishizeki pentru ca schema să aibă o structură de acces $\mathcal{A} \subseteq 2^P$ (submulțime a setului de participanți) monotonă (i.e. $\forall A \in \mathcal{A}, A \subseteq A' \Rightarrow A' \in \mathcal{A}$) [21]. Folosind construcția unei scheme majoritare (k, n) autorii au reușit să descrie elementele din \mathcal{A} folosind rezultatul unei reuniuni de mulțimi de componente cu un număr de elemente mai mare sau egal decât k (Fig. 3.3). [21] Notăția $x : Pr$, înseamnă că x are proprietatea Pr .

Dezavantajul acestei structuri este numărul de componente necesar pentru o structură de acces oarecare \mathcal{A} . Un mod simplu de construire al funcției *Assign*

Inițializare:

- Fie $S \in Z_q$ unde $q > 1$ și q prim;
- Fie n numărul de participanți a.î. $q > n$;
- Fie k numărul minim de componente puse în comun pentru a determina S ;

Generare: Dealerul \mathcal{D} :

- Alege n valori distincte $x_i \leftarrow^R Z_q \setminus \{0\}$, $i = 1, 2, \dots, n$;
- Alege $a_i \leftarrow^R Z_q$, $i \in \{1, 2, \dots, k-1\}$, $a_{k-1} \neq 0$;
- Construiește polinomul $f(x) = a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \dots + a_1x + S$;
- Calculează $s_i = f(x_i)$, $i \in \{1, 2, \dots, n\}$;

Distribuție: Dealerul \mathcal{D} :

- Transmite participantului P_i componenta s_i , $i \in \{1, \dots, n\}$;

Reconstrucție: Orice mulțime cu dimensiunea k (sau mai mare) de participanți distincti P_1, P_2, \dots, P_k :

- Interpolează punctele s_i pentru a obține polinomul f :

$$f(x) = \sum_{i=1}^k s_i \prod_{1 \leq j \leq k, j \neq i} \frac{x - x_j}{x_i - x_j} \quad (3.1)$$

- Află secretul reconstruit $S = f(0)$.

Figure 3.2: Schema Shamir [20]

din figura Fig. 3.3 este următorul: pentru fiecare mulțime minimală $A \in \mathcal{A}$ ($\forall B \in \mathcal{A}$, $B \neq A$, $B \not\subseteq A$) se folosește o schemă de partajare unanimă $(|A|, |A|)$ a lui S pentru participanții din A .

Exemplu 3.4.1. Fie structura de acces $\mathcal{A} = \{\{P_1, P_2\}, \{P_1, P_3, P_4\}\}$.

- Generăm componentele s_1, s_2 a.î. $s_1 \oplus s_2 = S$ cu ajutorul schemei unanime $(2, 2)$,
- Generăm componentele s_3, s_4, s_5 a.î. $s_3 \oplus s_4 \oplus s_5 = S$ cu ajutorul schemei unanime $(3, 3)$

Participanții primesc în felul următor componentele:

- $P_1 : \{s_1, s_3\}$;
- $P_2 : \{s_2\}$;
- $P_3 : \{s_4\}$;
- $P_4 : \{s_5\}$;

Inițializare:

- Fie q un număr prim $q, q > 1, z \in \mathbb{N}$ nenul și $\mathcal{C} = GF(p^z)$;
- Fie $S \in \mathcal{C}$ secretul;
- Fie structura de acces \mathcal{A} ;
- Fie n numărul de participanți;

Generare: Dealerul \mathcal{D} :

- Alege n valori distincte $x_i \leftarrow^R Z_q, i = 1, 2, \dots, n$;
- Alege $a_i \leftarrow^R \mathcal{C} \setminus \{0\}, i \in \{1, 2, \dots, k-1\}, a_{k-1} \neq 0$;
- Construiește polinomul $f(x) = a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + \dots + a_1x + S$;
- Atribuie $s_i = f(x_i) \ i \in \{1, 2, \dots, n\}$; Fie $Shares = \{s_1, \dots, s_n\}$;
- Alege $D_i \subseteq Shares \ 1 \leq i \leq n$;
- Alege funcția $Assign : P \rightarrow 2^Q$:
 - $Assign(P_i) = D_i \ 1 \leq i \leq n$
 - $\mathcal{A} = \left\{ Q \subseteq Shares : \left| \bigcup_{P_i \in Q} Assign(P_i) \right| \geq k \right\}$;

Distribuție: Dealerul \mathcal{D} :

- Transmite participantului P_i componenta $Assign(P_i), i \in 1, 2, \dots, n$;

Reconstrucție: Participanții din structura de acces \mathcal{A} :

- Procedează identic ca în schema Shamir.

Figure 3.3: Schema Ito, Saito, si Nishizeki [21]

3.5 Schema Feldman

Uneori, canalul de comunicație prin intermediul căruia dealerul \mathcal{D} transmite componentele nu este întotdeauna sigur. Din fericire, participanții pot constata cu ușurință integritatea componentelor primite folosind scheme de verificare. Primele direcții către astfel de construcții au fost oferite în anul 1985 de către Chor et al. [24]

În continuare vom exemplifica o schemă de verificare non-interactivă precum cea a lui Feldman, construită pe baza schemei Shamir. (Fig. 3.4). [25] Spre deosebire de schema Shamir, securitatea schemei Feldman este una computațională și deține un model de adversar pentru care problema logaritmului discret este dificilă.

Pentru verificarea componentei $s_i = f(i)$, participantul i probează ecuația:

Inițializare: Folosind notațiile din Fig.3.2, considerăm polinomul f generat în urma Schemei Shamir.

- Se alege un $g \in Z_q$ generator al lui Z_q a.î. este dificil de calculat $\log_g x$

Generare: Dealerul \mathcal{D} :

- Calculează *angajamentele*: $c_0 = g^S$, $c_j = g^{a_j}$, $j \in \{1, \dots, k\}$;

Distribuție: Dealerul \mathcal{D} :

- Distribuie *angajamentele* c_j , $j \in \{0, \dots, k\}$ fiecărui participant $i \in \{1, \dots, n\}$;

Figure 3.4: Schema Feldman [25]

$$g^{s_i} = c_0 c_1^{i^1} c_2^{i^2} \dots c_k^{i^k} = \prod_{j=0}^k c_j^{i^j} = \prod_{j=0}^k g^{a_j i^j} = g^{\sum_{j=0}^k a_j i^j} = g^{f(i)} \quad (3.2)$$

Capitolul 4

Sisteme de stocare de lungă durată

4.1 Introducere

Un sistem de stocare, în general, trebuie să satisfacă cel puțin următoarele 3 condiții:

- Disponibilitatea: Informația trebuie să rămână accesibilă tot timpul, chiar și în prezența erorilor de tip hardware.
- Integritatea: Abilitatea sistemului de a răspunde cererilor într-un mod care garantează corectitudinea lor.
- Confidențialitatea: O persoană care nu face parte din grupul de acces să nu obțină permisiunea de a afla informații de orice fel despre datele existente în sistem.

4.2 Criptare vs. scheme de partajare

Una dintre soluțiile existente pentru a construi acest sistem este criptarea datelor folosind o cheie înainte de inserarea lor în spațiul de stocare. În momentul în care un utilizator autorizat dorește să efectueze o citire a unor date, întrebuițează cheia potrivită pentru a le decripta.

În practică există algoritmi de criptare eficienți precum AES însă aceștia nu garantează confidențialitatea datelor în cazul în care apare un adversar fără o limită computațională.

Un dezavantaj al criptării este administrarea cheilor, standardele de securitate schimbându-se în fiecare an. De fiecare dată când cheile sunt înnoite atunci este necesară recriptarea datelor de pe fiecare bază de date. Cu cât disponibilitatea este mai mare - volumul de date crește - recriptarea informației devine o operație foarte costisitoare.

Majoritatea tehnicilor de criptare se bazează pe dificultatea factorizării unui număr sau cea a calculării logaritmului discret însă o dată cu posibila dezvoltare a calculatoarelor cuantice aceste probleme nu vor mai fi atât de dificile [26]. Pentru schemele de partajare, spre deosebire de criptare, avantajul unui adversar nu depinde de puterea sa computațională, acestea garantând securitatea teoretică precum schema Shamir [20].

Dezavantajul schemelor generale de partajare este dimensiunea componentelor, exponențială în funcție de numărul de participanți. [27]

4.3 Sisteme de stocare de lungă durată bazate pe scheme de partajare

O alternativă la soluția cu criptare care asigură atât confidențialitate cât și redundanța necesară este întrebuințarea sistemelor de stocare de lungă durată bazate pe scheme de partajare [2, 28, 29].

4.3.1 PASIS

PASIS este o soluție pentru un sistem descentralizat care oferă beneficii precum securitate, redundanță a datelor și auto-întreținere [28]. Structurile descentralizate împart informația la mai multe noduri folosind scheme de redundanță precum "Redundant Array of Independent Disks" (RAID) pentru a asigura performanța, scalabilitatea sistemului dar și integritatea datelor [30]. RAID reprezintă o tehnologie ce combină 2 concepte ortogonale precum data striping (aranjarea datelor pe discuri multiple într-o manieră secvențială - Fig 4.1) și redundanță pentru o disponibilitate ridicată [31].

PASIS folosește schemele de partajare pentru a distribui informația nodurilor de stocare dintr-o rețea. Aceasta introduce agenți pe partea clientului pentru a scrie sau șterge date din noduri, dar și agenți pentru mentenanță. Componentele obținute în urma partajării unui fisier sunt stocate în rețea cu ajutorul agenților (Fig. 4.2). Pe lângă conținutul brut al componentelor se adaugă metadate pentru a reține adresa nodului din rețea la care au fost stocate dar și noua denumire cu care este salvată în rețea.

Atunci când un participant inițiază o cerere pentru a citi un fisier, agentul PASIS aflat local procedează după cum urmează:

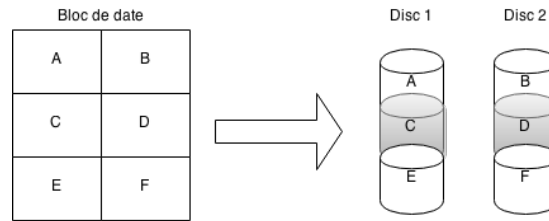


Figure 4.1: Procesul de striping aplicat unui bloc de date

- Caută numele celor n componente care alcătuiesc fișierul într-un serviciu care listează toate datele.
- Inițiază cereri de citire la cel puțin k din cele n noduri.
- în caz ca acesta nu primește cel puțin k răspunsuri se întoarce la pasul anterior încercând interogări la noduri diferite.
- Reconstituie fișierul obținut din cele k componente.

Operația de scriere este similară cu cea de citire, aceasta oprindu-se atunci când în cel puțin $n - k + 1$ noduri s-au stocat cu succes componente.

Autorii specifică soluții pentru auto mentenanța sistemului cu ajutorul resurselor umane prin monitorizarea periodică stării sistemului folosind log-uri sau ajustarea parametrilor din cadrul schemei de partajare.

4.3.2 GridSharing

în 2005, Subbiah și Blough propun o nouă abordare pentru a construi un sistem de stocare securizat și tolerant la erori numit GridSharing [?].

Schema Shamir nu oferă siguranță în ceea ce privește detectarea sau actualizarea unor componente incorecte introduse de un atacator. Din acest motiv, scheme de verificare precum cele prezentate în Secțiunea 3.5 sunt folosite.

Subbiah și Blough folosesc un sistem care înlocuiește schemele de verificare cu o schemă de partajare unanimă XOR (considerăm cazul $q = 2$ în Fig. 3.1) pentru a păstra securitatea construcției. în cazul detectării componentelor incorecte, este adoptată o strategie de tipul replicate-and-voting. Componentele sunt replicate pe un număr mare de servere astfel încât determinarea validității va fi stabilită în funcție de numărul de servere care le conțin.

Se identifică 3 tipuri de defecțiuni care pot apărea pe serverele unde sunt stocate datele:

- Abandonări: un server este *abandonat* dacă nu mai răspunde vreunui mesaj din rețea și s-a oprit din a mai efectua vreo operație.

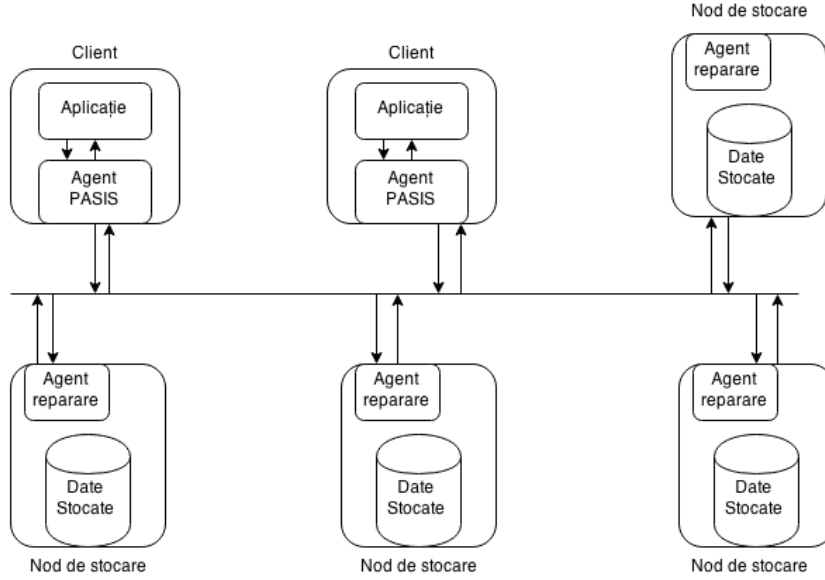


Figure 4.2: Arhitectura PASIS cu 4 noduri și 2 clienți [28].

- Bizantine: atunci când serverul nu respectă întotdeauna protocoalele inițiale iar componentele salvate local au fost compromise.
- Scurgeri de informații: serverul execută protocoalele corect dar e posibil ca un adversar să fi obținut componentele stocate.

Primele 2 modele definite mai sus sunt preluate din calculul cu sisteme distribuite. Cel de-al 3-lea model a fost introdus pentru a defini atacatorul care folosește vulnerabilitățile cu intenția de a *învăța* din informații.

Arhitectura GridSharing constă în N servere unde cel mult c servere pot fi abandonate, b servere bizantine și l cu scurgeri de informații. Cele N pot fi aranjate într-un grid cu r linii și N/r coloane (considerăm pentru simplitate că $N \pmod{r} = 0$). Caracteristicile modelului bizantin și cel specific scurgerilor de informații permit dezvăluirea componentelor unui adversar de pe cel mult $l + b$ servere.

Exemplu 4.3.1. Notăm $\binom{x}{y}$ fiind combinări de x elemente grupate câte y . Considerăm ca împărțim un secret \mathcal{S} la 4 linii (participanți) astfel încât sistemul să permită 2 componente de tip b , 1 componentă de tip l și 20 servere. în cazul acesta vom folosi o schemă majoritară XOR $\left(\binom{4}{3}, \binom{4}{3}\right) = (4, 4)$.

Vom avea 4 componente, (s_1, s_2, s_3, s_4) a.î. $s_1 \oplus s_2 \oplus s_3 \oplus s_4 = \mathcal{S}$. Distribuirea se face în felul următor:

- Serverele situate pe prima linie primesc s_1
- Serverele situate pe a 2-a linie primesc s_2
- Serverele situate pe a 3-a linie primesc s_3
- Serverele situate pe a 4-a linie primesc s_4

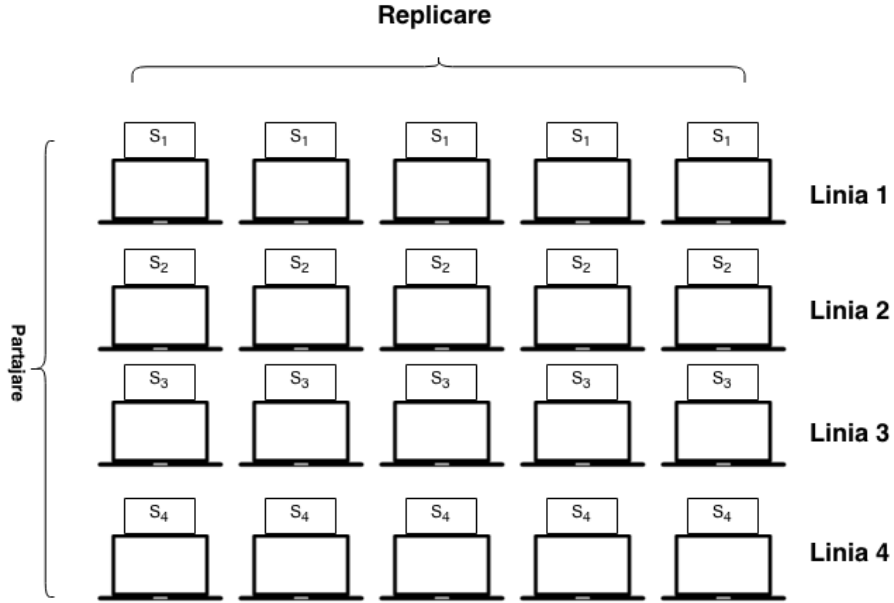


Figure 4.3: GridSharing cu 4 linii, 20 servere dintre care 2 bizantine, 1 cu scurgeri de informații [?].

4.3.3 POTSHARDS

în 2007 este propus un nou sistem care combină caracteristicile PASIS și Grid-Sharing adăugând posibilitatea de migrare a datelor la noduri noi: POTSHARDS (Protection Over Time, Securely Harboring And Reliably Distributing Stuff) [2].

POTSHARDS poate fi gândit ca o aplicație pe partea clientului care comunică cu o mulțime de noduri (arhive) independente, folosind reconstrucția componentelor într-un mod securizat și semnături algebrice pentru a asigura un grad ridicat de păstrare a integrității fișierelor [32].

Confidențialitatea este asigurată prin adoptarea unei scheme de partajare XOR unanimă, la fel ca în GridSharing.

Ca prim pas, POTSHARDS preprocesează fișierul într-un obiect, partajează obiectul în fragmente la care adaugă meta-date, numite *shards* (Fig. 4.4). [2] Acestea sunt trimise apoi arhivelor independente, fiecare având propriul domeniu de securitate, localizate în *regiuni*. Pentru a reconstitui cu succes informația inițială, meta-datele shard-urilor conțin detalii despre structura pointerilor aproximativi, indicând regiunea în care se află următorul shard. Pointerii aproximativi sunt folosiți pentru a reconstitui întreaga arhivă doar din shard-uri.

Procesul de fragmentare a datelor este prezentat în Fig. 4.5.

Pentru ca reconstituirea unui fișier să fie fezabilă unui utilizator, acestuia îi este întoarsă o listă cu locațiile exacte shard-urilor corespunzătoare. Obținerea unui shard de către un atacator nu este folositoare, pentru a detecta următorul shard, un atac brut force constă în cereri multiple în zona indicată de pointerul aproximativ. Un astfel de atac nu va trece neobservat de POTSHARDS deoarece unul dintre scopurile sale este să stocheze datele într-un mod cât mai uniform distribuit [2].

4.4 Sistemul de stocare Alouneh et al. (2013)

Alouneh et al. propun un sistem pentru stocarea sigură a datelor un timp îndelungat folosind schema Shamir cu câteva modificări. Aceste schimbări se vor arăta esențiale mai târziu în menținerea securității [33].

Pentru stocarea unui fișier în sistem (abordând filozofia majorității sistemelor de operare - orice este un fișier), o aplicație client îl împarte în blocuri de octeți

Obiect					
160	128				
Hash	ID	Date			

Fragment					
160	128	128	8	128*Nr	
Hash	ID Obiect	ID Frag.	Nr	Lista Shard-uri	SplitXOR(Obiect)

Shard		
128	128	
Hash	ID	SplitShamir(Fragment)

Figure 4.4: Entități de date în POTSHARDS. *Nr* e numărul de shard-uri produse de un fragment. *SplitXOR* reprezintă o componentă rezultată în urma partajării unanime XOR. Analog *SplitShamir* reprezintă o componentă rezultată în urma partajării folosind schema Shamir [2]

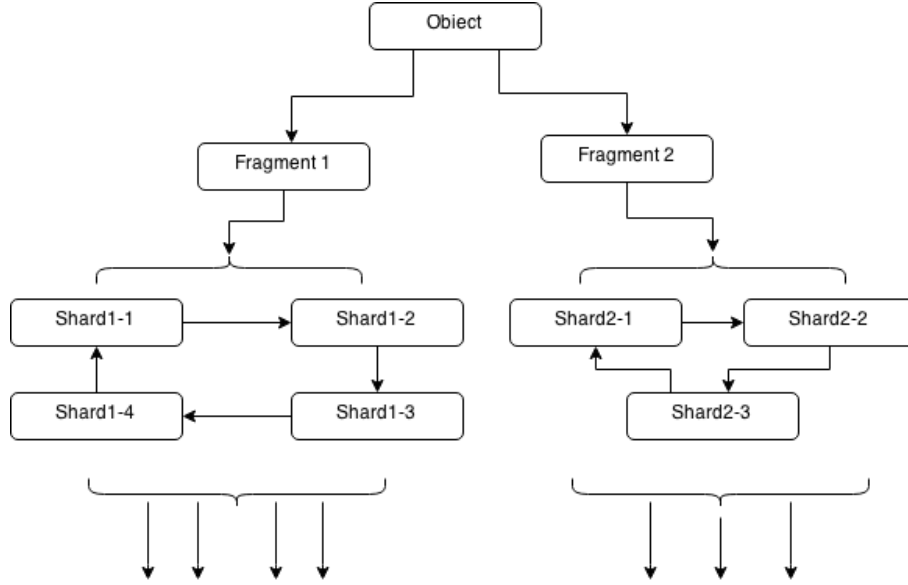


Figure 4.5: Distribuirea unui obiect in POTSHARDS

de lungime k . Pentru fiecare bloc, octeții devin coeficienții unui polinom f de grad k , componenta corespunzătoare participantului i va fi reprezentată de valoarea lui $f(i)$, $i = \{1, 2, \dots, n\}$. Menționăm că toate operațiile se vor efectua în $GF(256)$ modulo un polinom ireductibil (în implementarea sistemului, autorii folosesc $x^8 + x^5 + x^3 + x + 1$). Procedul este descris în detaliu în Fig 4.6.

Exemplu 4.4.1. Vom exemplifica modul de calcul în $GF(256)$ ($\text{mod } g(x)$) unde $g(x) = x^8 + x^4 + x^3 + 1$. Luăm polinomul $f(x) = 10 + 15x$ corespunzător unui fișier format din octeții (în această ordine) 10 15.

$$\begin{aligned}
 f(01) \text{ (mod } g(x)) &= 10 + 15 \text{ (mod } g(x)) \\
 &= (x^4) + (x^4 + x^2 + 1) \text{ (mod } g(x)) \\
 &= x^2 + 1 = 000000101_{(2)} \\
 &= 05_{(16)}
 \end{aligned} \tag{4.1}$$

$$\begin{aligned}
 f(02) \text{ (mod } g(x)) &= 10 + 15 \cdot 02 \text{ (mod } g(x)) \\
 &= (x^4) + (x^5 + x^3 + x) \text{ (mod } g(x)) \\
 &= 00111010_{(2)} \\
 &= 3A_{(16)}
 \end{aligned} \tag{4.2}$$

Reconstituirea unui fișier (Fig. 4.7) se realizează din orice mulțime de componente A de dimensiune minim k folosind interpolarea Lagrange (utilizată și în

Date de intrare: Un fișier binar \mathcal{S} ;

Date de ieșire: n fișiere binare distribuite la noduri din rețea;

Procesarea componentelor: Aplicația existentă pe partea clientului:

- Dacă \mathcal{S} nu are o lungime divizibilă cu k :
 - Concatenează la sfârșitul lui \mathcal{S} octeți până când $len(\mathcal{S}) \pmod k = 0$;
- Împarte \mathcal{S} în blocuri de lungime k ;
- Repetă pentru fiecare bloc B_t de lungime k :
 - Construiește polinomul $f(x) = B_{t_{k-1}}x^{k-1} + B_{t_{k-2}}x^{k-2} + \dots + B_{t_1}x + B_{t_0}$;
 - Calculează $f(i)$ pentru $1 \leq i \leq n$;

Distribuție: Aplicația la nivelul clientului:

- Distribuie componenta $f(i)$ nodului din rețea cu indicele i :

Figure 4.6: Schema Alouneh et al. - Generare [33]

cadrul schemei Shamir):

$$f(x) = \sum_{i \in A} f(i) \prod_{j \in A, j \neq i} \frac{x - j}{i - j} \quad (4.3)$$

Exemplu 4.4.2. Vom exemplifica interpolarea conform ecuației 4.3 folosind componentele (4.1) și (4.2) calculate în exemplul anterior:

$$\begin{aligned}
 f(x) &= 05(x - 02)(01 - 02)^{-1} + 3A(x - 01)(02 - 01)^{-1} \\
 &= 05(x - 02)03^{-1} + 3A(x - 01)03^{-1} \\
 &= F6(05 + 3A)x + F6(05 \cdot 02 + 3A \cdot 01) \\
 &= F6 \cdot 3F \cdot x + F6 \cdot 30 = 15x + 10
 \end{aligned} \quad (4.4)$$

Noutatea arhitecturii constă în diminuarea redundanței componentelor la un factor de k , spre deosebire de sistemele descrise în Secțiunile 4.3.1, 4.3.2 și 4.3.3. Reducerea spațiului ocupat de componente este datorat înlocuirii coeficienților generați aleator din schema Shamir cu octeții din fișierul ce va fi partajat, fiecare componentă având nevoie de $1/k$ din dimensiunea fișierului \mathcal{S} .

Abordarea acestei metode deterministe conduce la insecuritatea sistemului, Alouneh et al. afirmând în mod eronat că securitatea este indusă în mod automat de schema Shamir.

4.5 Contribuții

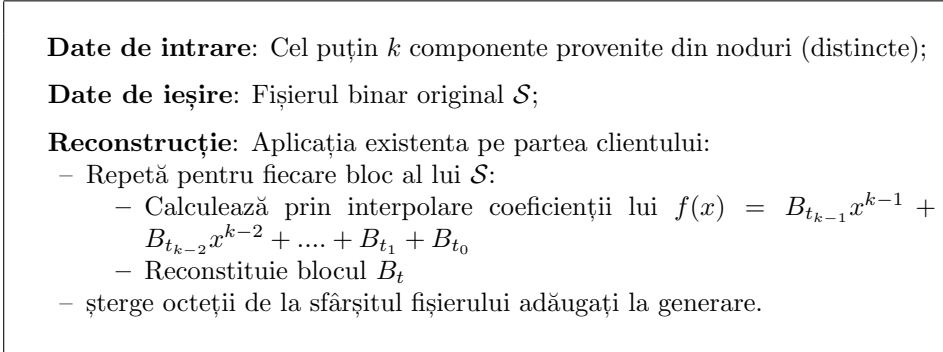


Figure 4.7: Schema Alouneh et al. - Reconstrucție [33]

TODO Reformulare împreună cu mentorul am analizat sistemul Alouneh et al. prezentat în Secțiunea 4.4. [33] Am indentificat erori majore ale acestuia și am implementat sistemul descris de autori pentru a demonstra practic, nu doar teoretic anumite greșeli pe care le vom evidenția în următoarele secțiuni.

4.5.1 Vulnerabilități evidențiate

Spre deosebire de schema Shamir, unde coeficienții (cu excepția termenului liber, care este egal cu secretul) sunt aleși într-un mod aleator uniform, sistemul propus de Alouneh et al. folosește o schemă modificată Shamir în care coeficienții sunt extrași din conținutul fișierelor originale. Alegerea este motivată de faptul ca mulțimea componentelor și efortul computațional depus pentru generarea coeficienților se reduce la un factor de k ori, spre deosebire de schema Shamir.

Datorită determinismului, partajarea unui fișier de mai multe ori implică obținerea de componente identice. Determinismul duce la câteva atacuri simple în momentul în care un atacator obține informațiile stocate într-un nod, indiferent de mărimea pragului folosit în metoda de partajare. în acest sens, am evidențiat 2 atacuri simple în cazul în care componentele sunt calculate în ordine $(f(01), f(02), \dots)$:

- Detectarea tipului unui fișier
- Detectarea conținutului unui fișier

Deoarece Alouneh et al. nu menționează o metoda de padding, am indicat că un atac bazat pe felul în care se realizează completarea fișierului (padding) \mathcal{S} înainte de partajarea sa poate fi fezabil, în condițiile în care s-a demonstrat ca această alegere este esențială în păstrarea securității [34].

Detectarea tipului de fișier

În tehnologia informației, la începutul fiecărui fișier se află o secvență de octeți (denumită semnătură sau antet) cu rolul de a identifica tipul acestuia. Tabelul 4.1 indică 7 din cele mai uzuale antete.

Se consideră cazul în care partajarea unui fișier *pdf* se face cu ajutorul sistemului descris în Secțiunea 4.4, folosind $k \leq 4$. Polinomul corespunzător $f(x)$ va fi întotdeauna același. Presupunând ca numerotarea nodului i este aceeași, putem determina cu ușurință dacă este stocat un fișier *pdf* fără a lua în calcul conținutul fișierului. Acest atac este fezabil deoarece valoarea lui k este publică iar i poate să fie descoperit în momentul distribuirii.

Cu alte cuvinte, dacă un adversar obține controlul unui singur nod, bazându-se doar pe valoarea primei componente poate detecta tipul unui fișier.

Pentru a exemplifica, un adversar poate distinge cu probabilitate ridicată între fișierele *doc*, *gif*, *pdf*, *png*, *rar*, *wav* și *zip*. În Tabelul 4.2 avem generate componentele pentru $k = 2$ și $n = 5$. Dacă un adversar descoperă valoarea primului nod iar prima componentă este 14 atunci acesta știe cu certitudine că aceasta corespunde unui fișier *gif*. Dacă obține accesul nodului 4 și citește valoarea 205 atunci știe că fișierul este de tipul *rar*. Dacă citește valoarea 27 de pe primul nod atunci știe că poate fi un *wav* sau *zip*; dar poate să distingă cele 2 fișiere cu probabilitate 1 dacă dezvăluie o singură valoare de pe celelalte noduri (2, 3, 4 sau 5) pentru că valorile sunt distincte.

Detectarea de conținut

Table 4.1: Semnături de fișiere

Tip de fișier	Primii 4 octeți			
doc	D0	CF	11	E0
gif	47	49	46	38
pdf	25	50	44	46
png	89	50	4E	47
rar	52	61	72	21
wav	52	49	46	46
zip	50	4B	03	04

Table 4.2: Componentele primului bloc ($k = 2$)

Tip fisier	Nod 1 ($i = 1$)	Nod 2 ($i = 2$)	Nod 3 ($i = 3$)	Nod 4 ($i = 4$)	Nod 5 ($i = 5$)
doc	31	85	154	193	14
gif	14	213	156	120	49
pdf	117	133	213	126	46
png	217	41	121	210	130
rar	51	144	241	205	172
wav	27	192	137	109	36
zip	27	198	141	103	44

Multe documente urmează un anumit tipar precum contracte, chitanțe, bonuri fiscale sau curriculum vitae. Deoarece majoritatea conținutului rămâne neschimbat, există o probabilitate destul de mare ca multe componente să aibă aceeași valoare. O dată ce un adversar reușește să determine componentele unui nod, poate determina prin analogie tipul de conținut al fișierului original.

Fișierele vulnerabile sunt cele care conțin o secvență de octeți periodică (imagini cu un pattern repetitiv) sau cele care au preponderent octeți nuli (valoarea componentelor asociată majorității blocurilor fiind 0). Spre deosebire de prima metodă, aceasta nu necesită compararea cu un al 2-lea fișier, tratând componentele de sine stătător. Multiple componente identice indică existența unor secvențe repetitive în conținutul fișierului.

4.5.2 Implementare și rezultate practice

Pentru a arăta aplicabilitatea rezultatelor în practică, am implementat propunerea descrisă în Secțiunea 4.4 și am testat pe câteva cazuri.

În cadrul implementării am folosit limbajul Python 3.0 sub sistemul de operare ArchLinux. Python este un limbaj high-level, permițând programatorilor să exprime concepte în mai puține linii de cod față de C++ sau Java și este disponibil sub licență open-source [35]. Pentru a realiza comunicarea între procese am folosit Cerealizer iar distribuția componentelor a fost generată grafic cu ajutorul pachetului Matplotlib [36, 37]. De asemenea am folosit sistemul de versionare Git iar în prezent codul folosit este găzduit de GitHub [38, 39].

Având în vedere că articolul original nu menționează o metodă de padding, am considerat o metodă standard pentru a completa octeții ultimului bloc: alipim la sfârșitul lui \mathcal{S} octeții 80 00 ... 00 00 până când lungimea ultimului bloc ajunge la k octeți.

Menționăm această metodă doar pentru completitudine, aceasta neafectând rezultatele, care consideră doar secvențele de octeți din antet sau de la începutul

fișierului.

Detectarea tipului de fișier

Extindem analiza făcută în Secțiunea 4.5.1 asupra tipurilor de fișiere din Tabelul 4.1 pentru $k = 2$ și creștem valoarea indicelui i până când 2 componente devin egale. Fie f_l polinomul de gradul 1 asociat primului bloc al fișierului aflat pe linia l . Analog f_c polinomul de gradul 1 asociat primului bloc al fișierului aflat pe coloana c .

Tabelul 4.3 prezintă valoarea maximă a nodul i pentru care $f_l(i) \neq f_c(i)$. Valoarea -1 indică lipsă de coliziuni ale lui $f_c(x)$, $f_l(x)$ pentru i , $i = 1, 2, \dots, 255$ ($\nexists 1 \leq i \leq 255$ a.î. $f_l(i) = f_c(i)$)).

Deoarece pe diagonala principală toate componentele sunt identice pentru $k \leq 4$, poate fi ignorată. în Tabelul 4.3 observăm valoarea 0 pentru perechea (wav, zip) pentru că $f_{wav}(1) = f_{zip}(1) = 27$ (Tabelul 4.1).

Tabelele 4.4 și 4.5 prezintă rezultatele pentru $k = 3$ și $k = 4$. Pentru $k \geq 5$ este nevoie de un antet cu mai mult de 4 octeți.

Considerăm metoda de distribuire descrisă exact ca în Tabelul 4.6, și anume nodul de stocare i primește componenta $f(i)$. Dacă un atacator preia controlul nodului cu valoarea i acesta poate face distincția între tipul a 2 fișiere partajate cu probabilitate 1 dacă i e mai mic decât valoarea afișată în tabel.

Spre exemplu, un adversar care obține accesul unui singur nod de stocare, iar partajarea a fost făcută pentru $k = 2$ poate distinge cu probabilitate 1 între un fișier *doc* sau *pdf* dacă $i \leq 194$. Deoarece $n > 194$ nu se întâmplă în practică, acest atac funcționează. Faptul că multe valori sunt ridicate și majoritatea sunt -1 (Tabelul 4.4), indică un nivel scăzut de securitate al schemei. în cazul celor

Table 4.3: Indicele maxim i a.î. componentele primului bloc sa fie distincte($k = 2$)

Tip Fișier	doc	gif	pdf	png	rar	wav	zip
doc	-	169	194	209	170	206	110
gif	169	-	133	137	75	-1	133
pdf	194	133	-	-1	115	151	133
png	209	137	-1	-	229	147	195
rar	170	75	115	229	-	-1	42
wav	206	-1	151	147	-1	-	0
zip	110	133	133	195	42	0	-

Table 4.4: Indicele maxim i a.î. componentele primului bloc sa fie distincte($k = 3$)

Tip Fișier	doc	gif	pdf	png	rar	wav	zip
doc	-	63	-1	-1	-1	-1	-1
gif	63	-	-1	-1	-1	-1	-1
pdf	-1	-1	-	164	-1	119	-1
png	-1	-1	164	-	143	122	129
rar	-1	-1	-1	143	-	143	-1
wav	-1	-1	119	122	143	-	172
zip	-1	-1	-1	129	-1	172	-

Table 4.5: Indicele maxim i a.î. componentele primului bloc sa fie distincte($k = 4$)

Tip Fișier	doc	gif	pdf	png	rar	wav	zip
doc	-	-1	38	95	1	95	98
gif	-1	-	-1	-1	167	-1	-1
pdf	38	-1	-	12	11	119	70
png	95	-1	12	-	243	95	148
rar	1	167	11	243	-	-1	94
wav	95	-1	119	95	-1	-	-1
zip	98	-1	70	148	94	-1	-

cu -1 un adversar va câștiga întotdeauna cu probabilitatea 1, indiferent de indicele nodului de stocare pe care îl obține.

Precizăm că acest atac funcționează doar dacă nodurile de stocare își păstrează indexul în cazul partajării repetate a.î. aplicația disponibilă pe partea de client calculează cele n valori $f(i_1), \dots, f(i_2)$ pentru i_1, \dots, i_n distincte și păstrează nodul j asociat lui i_j .

Detectarea de conținut

Considerăm scenariul pentru detectarea conținutului unui fișier, demonstrând practic cum un adversar poate să facă diferența între 2 componente stocate pe același nod aparțin unor documente similare. Atacul presupune accesul la un singur nod, indiferent de mărimea pragului k .

Pentru experiment am ales 3 fișiere *pdf* disponibile online la [40]:

- Europass Curriculum Vitae - BG, Bulgaria;
- Europass Curriculum Vitae - DK, Dannemark;
- Europass Mobility - RO, Romania.

Primele două fișiere reprezintă șablonul pentru un CV european în limba bulgară, respectiv daneză. Observăm că nu doar conținutul șabloanelor diferă, dar și limba în care au fost traduse. Cel de-al treilea fișier este complet diferit de primele două, fiind un document personal în limba română, folosit pentru înregistrarea cunoștințelor dobândite într-o țară europeană.

Pentru a arăta vulnerabilitatea sistemului Alouneh et al. [33], partajăm cele 3 fișiere folosind schema (2, 4). Experimentul reprezintă o implementare practică a metodei prezentate în Fig. 4.6.

Fig. 4.8, 4.9, 4.10, 4.11 conțin 2000 de componente generate pentru cele 3 fișiere, stocate în baza de date corespunzătoare nodurilor 1, 2, 3 și 4. Presupunem că indicele i este fixat, polinoamele fiind evaluate în același nod i . Un adversar care obține componentele unui nod poate deduce cu ușurința graficul lor.

Considerăm că un adversar reușește să obțină componentele din Fig. 4.8. Poate deduce ușor că primele două fișiere au un număr mare de componente identice iar cel de-al 3-lea este diferit de primele.

Din cauza șablonului pe care îl respectă cele două CV-uri, un adversar are nevoie în cazul de față doar de primele 500 de componente pentru a constata

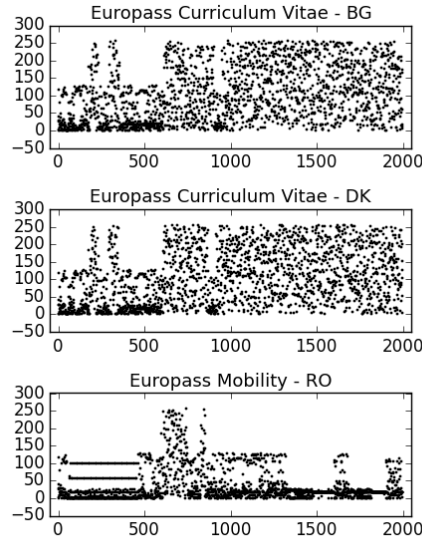


Figure 4.8: Nod 1: Graficul componentelor pentru primele 2000 de blocuri

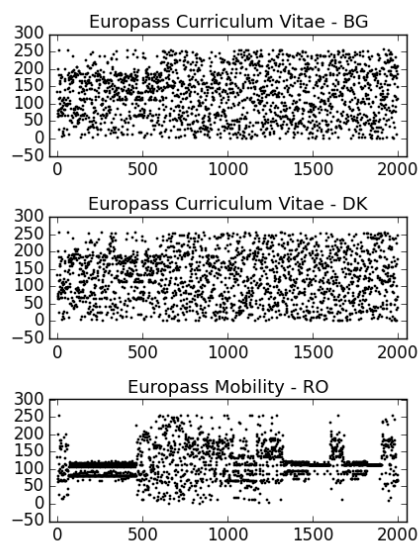


Figure 4.9: Nod 2: Graficul componentelor pentru primele 2000 de blocuri

similiaritatea celor 2 cu o probabilitate ridicată. Diferența vizuala este evidentă

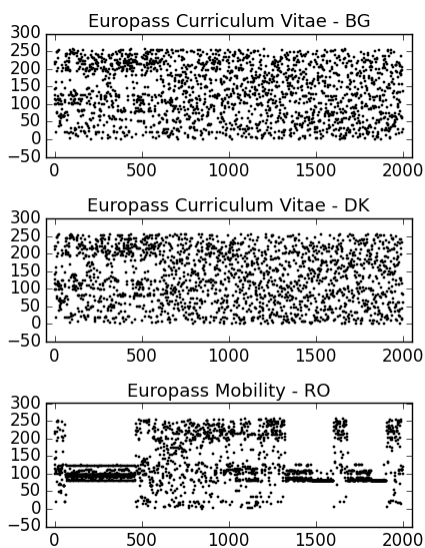


Figure 4.10: Nod 3: Graficul componentelor pentru primele 2000 de blocuri

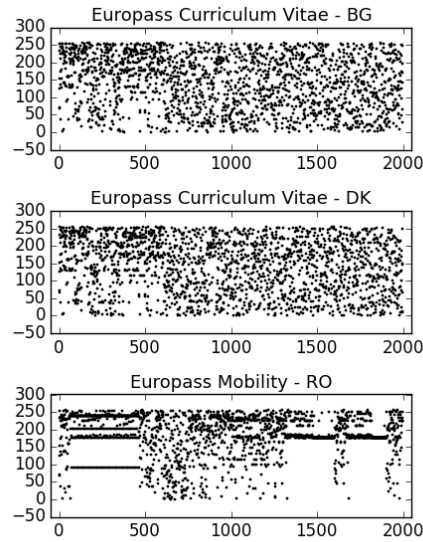


Figure 4.11: Nod 4: Graficul componentelor pentru primele 2000 de blocuri

între oricare din cele 4 grafice, deci un adversar poate reuși să afle informații despre fișierele partajate indiferent de nodul de stocare vulnerabil.

Considerăm în continuare un alt scenariu, în care partajăm o imagine pentru care informația se repetă. Fie imaginea din Fig. 4.12 partajată folosind schema din Secțiunea 4.4 având 4 noduri iar informația stocată pe cel puțin 2 din ele pot reconstitui imaginea partajată inițial. 4.6. La fel ca în exemplul precedent, Fig. 4.13, 4.14, 4.15, 4.16 reprezintă grafic componente imaginii din Fig. 4.12. Un adversar constată cu ușurință prin obținerea informațiilor stocate aflate pe un singur nod (indiferent de mărimea pragului k) că fișierul original conține un pattern care se repetă.

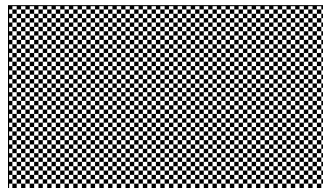


Figure 4.12: Imagine conținând pattern-uri

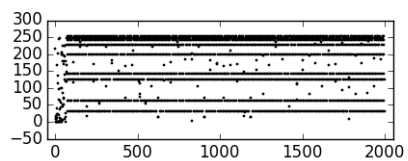


Figure 4.13: Nod 1: Graficul componentelor pentru primele 2000 de blocuri

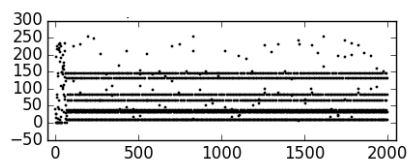


Figure 4.14: Nod 2: Graficul componentelor pentru primele 2000 de blocuri

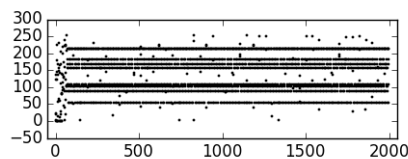


Figure 4.15: Nod 3: Graficul componentelor pentru primele 2000 de blocuri

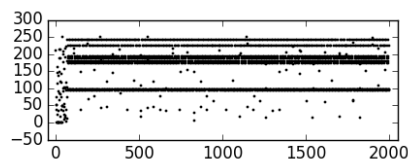


Figure 4.16: Nod 4: Graficul componentelor pentru primele 2000 de blocuri

Capitolul 5

Implementare alternativă a sistemelor de stocare

5.1 Arhitectură bazată pe scheme de partajare

5.2 Instanțe Amazon EC2

5.3 Reconstituirea fișierelor

Referințe

- [1] Katz, J., Lindell, Y.: Introduction to Modern Cryptography (Chapman & Hall/Crc Cryptography and Network Security Series). Chapman & Hall/CRC (2007)
- [2] Storer, M.W., Greenan, K.M., Miller, E.L., Voruganti, K.: Potshards - a secure, recoverable, long-term archival storage system. TOS **5**(2) (2009)
- [3] R.F.Olimid, D.A.Rotaru: An immediate multi-party generalization of ID-NIKE from constrained PRF. Theory Day in Computer Science (2014)
- [4] R.F.Olimid, D.A.Rotaru: On the security of a backup technique for database systems based on threshold sharing. Journal of Control Engineering and Applied Informatics (in recenzie) (2015)
- [5] WebSite: The Caesar cipher (2015) Ultima accesare: Februarie, 2015, <http://www.cs.trincoll.edu/~crypto/historical/caesar.html>.
- [6] WebSite: Hacking the Vigenere cipher (2015) Ultima accesare: Februarie, 2015, <http://inventwithpython.com/hacking/chapter21.html>.
- [7] WebSite: Enigma cipher (2015) Ultima accesare: Februarie, 2015, <http://practicalcryptography.com/ciphers/enigma-cipher/>.
- [8] Martin, K.M.: Challenging the adversary model in secret sharing schemes. Coding and Cryptography II, Proceedings of the Royal Flemish Academy of Belgium for Science and the Arts (2008) 45–63
- [9] Goldreich, O., Goldwasser, S., Micali, S.: How to construct random functions. Journal of the ACM (JACM) **33**(4) (1986) 792–807
- [10] Dierks, T.: The transport layer security (tls) protocol version 1.2. (2008)
- [11] Boneh, D., Waters, B.: Constrained pseudorandom functions and their applications. In: Advances in Cryptology-ASIACRYPT 2013. Springer (2013) 280–300
- [12] Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. Journal of the ACM (JACM) **56**(6) (2009) 34

- [13] Diffie, W., Hellman, M.E.: New directions in cryptography. *Information Theory, IEEE Transactions on* **22**(6) (1976) 644–654
- [14] Joux, A.: A one round protocol for tripartite diffie–hellman. In: *Algorithmic number theory*. Springer (2000) 385–393
- [15] Boneh, D., Silverberg, A.: Applications of multilinear forms to cryptography. *Contemporary Mathematics* **324**(1) (2003) 71–90
- [16] Shamir, A.: Identity-based cryptosystems and signature schemes. In: *Advances in cryptology*, Springer (1985) 47–53
- [17] Paterson, K.G., Srinivasan, S.: On the relations between non-interactive key distribution, identity-based encryption and trapdoor discrete log groups. *Designs, Codes and Cryptography* **52**(2) (2009) 219–241
- [18] Boneh, D.: The decision Diffie-Hellman problem. In: *Algorithmic number theory*. Springer (1998) 48–63
- [19] Blakley, G.: Safeguarding cryptographic keys. *Proceedings of the 1979 AFIPS National Computer Conference* (1979) 313–317
- [20] Shamir, A.: How to share a secret. *Commun. ACM* **22**(11) (1979) 612–613
- [21] Ito, M., Saito, A., Nishizeki, T.: Secret sharing scheme realizing general access structure. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* **72**(9) (1989) 56–64
- [22] Benaloh, J., Leichter, J.: Generalized secret sharing and monotone functions. In: *Proceedings on Advances in Cryptology. CRYPTO '88*, New York, NY, USA, Springer-Verlag New York, Inc. (1990) 27–35
- [23] Karnin, E.D., Member, S., Greene, J.W., Member, S., Hellman, M.E.: On secret sharing systems. *IEEE Transactions on Information Theory* **29** (1983) 35–41
- [24] Chor, B., Goldwasser, S., Micali, S., Awerbuch, B.: Verifiable secret sharing and achieving simultaneity in the presence of faults. In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, IEEE (1985) 383–395
- [25] Feldman, P.: A practical scheme for non-interactive verifiable secret sharing. In: *Proceedings of the 28th Annual Symposium on Foundations of Computer Science. SFCS '87*, Washington, DC, USA, IEEE Computer Society (1987) 427–438
- [26] Shor, P.W.: Algorithms for quantum computation: discrete logarithms and factoring. In: *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*, IEEE (1994) 124–134
- [27] Beimel, A.: Secret-sharing schemes: a survey. In: *Coding and cryptology*. Springer (2011) 11–46

- [28] Wylie, J.J., Bigrigg, M.W., Strunk, J.D., Ganger, G.R., Kiliççöte, H., Khosla, P.K.: Survivable information storage systems. *Computer* **33**(8) (2000) 61–68
- [29] Subbiah, A., Blough, D.M.: An approach for fault tolerant and secure data storage in collaborative work environments. In: *StorageSS*. (2005) 84–93
- [30] Patterson, D.A., Gibson, G., Katz, R.H.: A case for redundant arrays of inexpensive disks (raid). *SIGMOD Rec.* **17**(3) (June 1988) 109–116
- [31] Chen, P.M., Lee, E.K., Gibson, G.A., Katz, R.H., Patterson, D.A.: Raid: High-performance, reliable secondary storage. *ACM Comput. Surv.* **26**(2) (June 1994) 145–185
- [32] Schwarz, T.J.E., Miller, E.L.: Store, forget, and check: Using algebraic signatures to check remotely administered storage. In: *26th IEEE International Conference on Distributed Computing Systems (ICDCS 2006)*, 4-7 July 2006, Lisboa, Portugal. (2006) 12
- [33] Alouneh, S., Abed, S., Mohd, B.J., Kharbutli, M.: An efficient backup technique for database systems based on threshold sharing. *JCP* **8**(11) (2013) 2980–2989
- [34] Vaudenay, S.: Security flaws induced by cbc padding - applications to ssl, ipsec, wtls. In: *Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques: Advances in Cryptology. EUROCRYPT '02*, London, UK, UK, Springer-Verlag (2002) 534–546
- [35] WebSite: Python Programming Language - official website (2015) Ultima accesare: Februarie, 2015, <https://www.python.org/>.
- [36] Hunter, J.D.: Matplotlib: A 2D graphics environment. *Computing In Science & Engineering* **9**(3) (2007) 90–95
- [37] WebSite: Cerealizer package (2015) Ultima accesare: Februarie, 2015, <https://pypi.python.org/pypi/Cerealizer>.
- [38] WebSite: Github - Official Website (2015) Ultima accesare: Februarie, 2015, <https://www.github.com>.
- [39] WebSite: Cod Github - Official Website (2015) Ultima accesare: Februarie, 2015, <https://www.github.com/rdragos/splitting-scheme>.
- [40] WebSite: Europass - download examples (2014) Ultima accesare: November, 2014, <http://www.europass.cedefop.europa.eu/ro/home>.