Sulmi me frekuencë në shifrimin e Cezarit dhe me zëvendësim

Erjon Asllani, Fatbardh Gashi, Vullnet Gërvalla

Programi Shkencat Kompjuterike, Departmenti i Matematikës, Fakulteti i Shkencave Matematike-Natyrore, Universiteti i Prishtinës, Prishtinë, 10000, Kosovë.

Contributing authors: erjon.asllani@student.uni-pr.edu; fatbardh.gashi6@student.uni-pr.edu; vullnet.gervalla@student.uni-pr.edu;

Abstract

Nevoja për kriptografi dhe shifrim apo kriptim të mesazheve është ndjerë me mijëra vjet më parë dhe nga nevoja ka pasur mjaft disa përpjekje dhe arritje në këtë fushë gjithmonë duke ndërtuar mbi përpjekjet e mëparshme. Disa përpjekje në veçanti si shifrimi i Cezarit dhe shifrimin me zëvendësim kanë parë më shumë përdorim. Për të testuar sigurinë ne kemi bërë analizën dhe thyerjen e këtyre shifrimeve duke u bazuar në të dhëna statistikore të mbledhura dhe përfundimisht të nxjerrura nga punime literature të ndryshme në gjuhën Shqipe.

Keywords: Kriptografia, Shifrimi i Cezarit, Shifrimi me zëvendësim, Sulmi me frekuencë

1 Hyrje

Kriptografia, arti i shkrimit apo zgjidhjes së kodeve, ka qenë një aspekt i rëndësishëm i komunikimit dhe fshehtësisë për mijëra vjet. Nga egjiptianët e lashtë deri te kriptimi dixhital i ditëve moderne, metodat e përdorura për të enkriptuar dhe dekriptuar mesazhet kanë evoluar shumë me kalimin e kohës.

Në këtë punim kërkimor, ne do të diskutojmë dy nga format më të hershme të enkriptimit, shifrimin e Cezarit dhe shifrimin me zëvendësim, dhe se si ato janë të ndjeshme apo të dobëta ndaj sulmeve me frekuencë, por gjithashtu udhës duke vizualizuar procesin e enkriptimit dhe dektriptimit dhe përfundimisht duke përfshirë një program grafik që kryen të gjitha veprimet e lartpërmendura.

2 Historia, përdorimi dhe dobësitë e shifrimeve

2.1 Shifrimi i Cezarit

Shifrimi i Cezar-it, i njohur edhe si shifrimi me zhvendosje, është një metodë e thjeshtë kriptimi që përfshin zhvendosjen e secilës shkronjë të alfabetit nga një numër i caktuar pozicionesh. Shifrimi është emëruar sipas Jul Cezarit, i cili thuhet se ka përdorur këtë metodë për të komunikuar me gjeneralët e tij gjatë fushatave ushtarake. Ky shifrim nuk ishte një metodë veçanërisht e fortë kriptimi, por ishte mjaft efektive për të parandaluar armikun që të kuptonte lehtësisht mesazhet.

Për të enkriptuar një mesazh duke përdorur shifrën Cezar, ne thjesht zëvendësojmë çdo letër në mesazh me letrën që zhvendoset nga çelësi. Për shembull, duke përdorur një çelës prej 3, mesazhi ”HELLO” do të enkriptohej si ”KHOOR”. Për të dekriptuar mesazhin, thjesht zhvendosim çdo letër nga çelësi për të marrë mesazhin origjinal. Pavarësisht thjeshtësisë së tij, shifra Cezari ishte një metodë efektive kriptimi në kohën e saj. Megjithatë, ajo u thye përfundimisht nga kriptoanalistët që ishin në gjendje të shfrytëzonin dobësitë në shifrim. Një nga shembujt më të famshëm të kësaj është historia e historianit romak Suetonius, i cili shkroi në biografinë e tij të Jul Cezarit se gjenerali përdori një

shifër për të komunikuar me trupat e tij. Shifra ishte një zëvendësim i thjeshtë i shkronjave, me A të zëvendësuar nga D, B nga E, e kështu me radhë nëpërmjet alfabetit. Suetoni, duke qenë historian, e njihte mirë gjuhën greke, e cila e lejonte të njihte modelet në periferi.

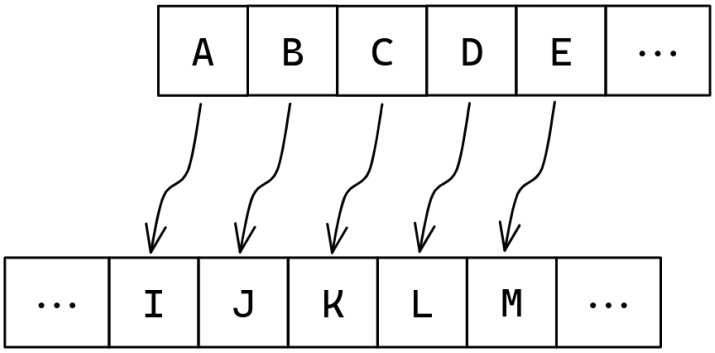


Fig. 1: Shifrimi i Cezarit (me zhvendosje)

Pavarësisht nga dobësitë e tij, shifra Cezari luajti një rol të rëndësishëm në zhvillimin e kriptografisë. Ai hapi rrugën për metoda më komplekse të enkriptimit, duke përfshirë edhe shifrën e zëvendësimit. Kodi i zëvendësimit është një lloj kriptimi që përfshin zëvendësimin e çdo letre në plaintext me një shkronjë ose simbol të ndryshëm në mesazhin e kriptuar. Kjo mund të bëhet duke përdorur një grup të para-rregulluar të zëvendësimeve, të tilla si A=B, B=C, C=D, dhe kështu me radhë. Në mënyrë alternative, zëvendësimet mund të jenë të rastësishme, me çdo shkronjë në mesazhë duke u zëvendësuar nga një shkronjë ose simbol i ndryshëm në mesazhin e koduar.

2.2 Shifrimi me zëvendësim

Shifra e zëvendësimit, ashtu si shifra e Cezarit, funksionon në parimin e zëvendësimit të çdo shkronje të tekstit të thjeshtë me një shkronjë ose simbol të ndryshëm në tekst. Megjithatë, në vend që t'i zhvendosë shkronjat me një numër fiks pozicionesh, shifra e zëvendësimit përdor një ndryshim të rastësishëm të shkronjave për të krijuar enkriptimin.

Shifra e zëvendësimit u përshkrua për herë të parë në shekullin e 9-të nga matematikani arab dhe polimatist Al-Kindi, i cili e quajti atë si "metoda e ngatërrimit". Ideja bazë e shifrës së zëvendësimit është krijimi i një harte një-për-një midis shkronjave të alfabetit dhe një grupi simbolesh ose shkronjash të tjera. Ky hartë mund të përfaqësohet si një tabelë ose matricë, e njohur si një tabelë zëvendësuese ose një çelës shifror.

Për të kriptuar një mesazh duke përdorur shifrën e zëvendësimit, teksti i thjeshtë ndahet fillimisht në shkronja individuale ose grupe shkronjash, në varësi të gjatësisë së tabelës së zëvendësimit. Më pas, çdo shkronjë zëvendësohet me simbolin ose shkronjën përkatëse në tabelën e zëvendësimit. Për shembull, nëse shkronja A vihet në hartë me simbolin *, atëherë çdo paraqitje e shkronjës A në tekstin e thjeshtë do të zëvendësohet me simbolin * në tekstin e shifruar.

Një nga shembujt më të hershëm të njohur të shifrës së zëvendësimit u përdor nga ushtria spartane në Greqinë e lashtë. Ata përdorën një pajisje të quajtur scytale, e cila ishte një shufër me një diametër të caktuar rreth së cilës mbështillej një rrip pergamenë ose lëkure. Mesazhi shkruhej për së gjati në shirit dhe kur shiriti hapej, mesazhi do të gërvishtej, duke e enkriptuar në mënyrë efektive.

Në mesjetë, shifra e zëvendësimit përdorej në forma të ndryshme nga shoqëritë sekrete dhe organizatat fetare për të komunikuar fshehurazi. Një organizatë e tillë ishte Kalorësit Templarë, të cilët dihej se kishin përdorur një shifër të njohur si shifra Pigpen, e cila zëvendësoi shkronjat me simbole të renditura në një rrjet.

Gjatë periudhës së Rilindjes, shifra e zëvendësimit u përdor nga gjykatat evropiane për korrespondencën diplomatike. Këto shifra ishin shpesh më komplekse se shembujt e mëparshëm, duke përdorur alfabete dhe simbole të shumta, dhe ndonjëherë kombinoheshin me teknika të tjera të enkriptimit, siç është transpozimi.

Një metodë tjetër e përdorur për të thyer shifrimin e zëvendësimit është i njohur si sulmi i frekuencës . Kjo përfshin analizimin e shpeshtësisë së shkronjave, çifteve të shkronjave ose grupeve të

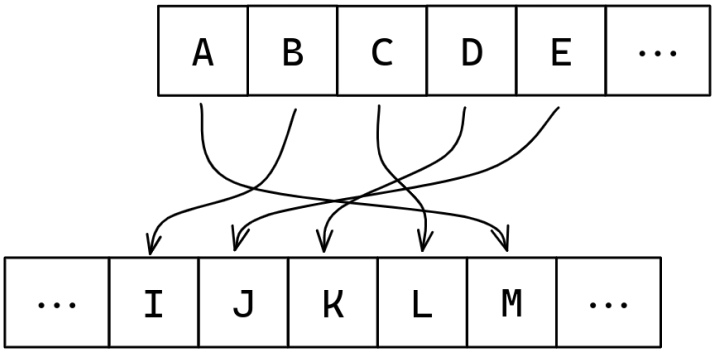


Fig. 2: Shifrimi me zëvendësim.

shkronjave në ciphertext dhe krahasimin e tyre me frekuencat e pritshme në gjuhën që përdoret. Për shembull, në gjuhën angleze, çiftet më të zakonshme të shkronjave janë "TH", "HE", "IN", "ER", dhe "AN". Duke analizuar frekuencën e këtyre çifteve në ciphertext, kriptonalizuesit mund të bëjnë hametime të arsimuara se cilat çifte të letrave kanë më shumë gjasa të përfaqësojnë cilat shkronja në plaintext. Metoda të tjera të përdorura për të thyer kodet e zëvendësimit përfshijnë ekzaminimin e Kasiskit, i cili përfshin gjetjen e sekuencave të përsëritura të shkronjave në ciphertext, dhe indeksin e koincidencës, i cili mat ngjashmërinë midis shpërndarjes së frekuencave të shkronjave në ciphertext dhe shpërndarjen e pritshme në gjuhën që përdoret.

3 Analiza dhe sulmi i frekuencës

3.1 Analiza e literaturës

Për të kryer një sulm të frekuencës duhet të bëjmë një analizë statistikore të një mostre të literaturës shqipe dhe të nxjerrim përqindjen e paraqitjes së karakterëve, në bazë të së cilëve krijojmë një tabelë frekuence ku çdo karakteri iu shoqërohet një përqindje të frekuencës së tij si në vijim: `'e' -> 7.12%, 'ë' -> 6.32%, ....` Pasi kemi kryer një analizë të tillë duke analizuar 5 punime letërsie në gjuhën shqipe, kemi mbërritur tek të dhënat të shfaqura në tabelën 1. Kjo tabelë na mundëson të kryejmë sulme frekuence në tekste të kriptuara me shifrimin e Cezarit apo me shifrimin me zëvendësim.

3.2 Sulmi i frekuencës

Pasi të kemi analizuar një sasi të mjaftueshmë të literaturës së një gjuhë atëherë mund të përdorim produktin e asaj analize për të sulmuar direkt tekstin e shiftuar. Mënyra se si mund ta realizojmë një sulm të tillë është duke kryer analizën e njëjtë në tekstin e shifruar si kemi kryer tek literatura në gjuhën shqipe, pra duke numëruar paraqitjen e karakterëve në tekstin e shifruar.

Pasi të kemi analizuar tekstin e shifruar dhe kemi gjetur paraqitjen e secilit karakter atëherë do të kemi mundësi t'i krahasojmë këto statistika me statistikat nga analiza e mostrës së literaturës Shqiptare dhe duke përdorur ato të krijojmë një tekst të ri ku karakteri më së shumti që paraqitet tek teksti i shifruar të zëvendësohet me karakterin më të paraqitur në mostrën tonë. P.sh nëse paraqitet në një tekst të shifruar më së shumti karakteri **'u'** atëherë në bazë të njohurive që kemi grumbulluar në tabelën 1 do të jemi mjaft të sigurt se në tekstin origjinal ai karakter ishte **'e'** ose **'ë'**. Iterojmë në këtë proces për disa nga karakteret më të shpeshta dhe do të kemi një rezultat i cili do të jetë i ngjashëm me tekstin origjinal para se të shifrohej.

3.3 Shifrimi, deshifrimi dhe analiza e tekstit në praktikë

Që të kryejmë këto analiza dhe gjithashtu të shifrojmë dhe deshifrojmë mesazhe ne kemi krijuar një vegël duke përdorur gjuhën **Java** dhe **Swing** për pjesën grafike të programit. Programi hapet duke shfaqur ndërfaqen që përmban një kuti hyrëse të të dhënave(kutia e majtë), një kuti dalëse e të dhënave(kutia e majtë), disa butona dhe një menu për zgjedhje të veprimit. Së pari ne duhet të zgjedhim nga menuja në veri të ndërfaqes veprimin që do të kryejmë, kemi dy opsione, **Encrypt** dhe **Decrypt**. Opsioni i parazgjedhur është për të enkriptuar dhe për të enkriptuar/shifruar një mesazh

atëherë vendosim tekstin e thjeshtë në kutinë e majtë dhe shtypim butonin **Encrypt** i cili do të na vendosë në kutinë e djathtë tekstin e enkriptuar me shifrim me zëvendësim të rastësishëm, si tek figura 3.

Table 1: Frekuenca e 25 shkronjave më të shpeshta

				Libri							
1		2		3		4		5		**Totali**	
char	perc	char	perc	char	perc	char	perc	char	perc	char	perc
e	8.4871	ë	9.7481	ë	9.2589	e	8.6442	e	8.6292	e	8.6602
ë	8.3750	e	8.2075	e	9.0506	i	7.9386	ë	8.3239	ë	8.2824
i	8.0504	a	7.7876	t	8.2423	t	7.5285	i	7.8193	t	7.8652
t	7.9528	t	7.2931	i	7.2135	ë	7.2270	t	7.7387	i	7.7404
a	7.7304	r	6.1209	n	6.5519	a	6.8654	r	6.5189	a	7.1214
r	6.4114	n	6.0905	a	6.4381	r	6.8086	n	6.0185	r	6.4997
n	6.3714	i	5.8842	r	6.3099	n	5.6382	a	5.6397	n	6.1751
h	5.4039	s	4.3547	s	5.1965	s	5.6104	s	5.1549	s	5.2385
s	5.0668	h	4.1594	h	5.1217	o	4.0607	o	3.8785	h	4.5852
u	4.0847	u	3.8123	u	4.3536	h	3.2134	m	3.4615	u	3.7900
d	3.7436	m	3.8086	d	4.2867	k	3.0902	k	3.3569	d	3.5072
k	3.3966	o	3.6134	j	3.7644	m	3.0623	h	3.2170	o	3.4474
m	3.3648	k	3.3666	k	3.6030	u	3.0315	d	2.9513	k	3.3507
o	3.1681	j	3.3390	o	3.0777	d	2.7141	u	2.7124	m	3.1678
j	3.1432	p	2.7588	m	2.8392	p	2.7141	.	2.6445	j	3.0828
l	2.4242	d	2.7128	p	2.5875	l	2.5377	j	2.6332	p	2.5370
p	2.3589	l	2.0645	l	2.0118	j	2.4750	p	2.5187	l	2.3479
g	1.4574	,	2.0369	g	1.5259	v	1.8676	l	2.3647	g	1.4474
b	1.4270	b	1.2551	,	1.2710	g	1.3867	-	1.7823	v	1.3399
.	1.3422	g	1.1998	v	1.1366	.	1.2961	g	1.6424	b	1.2656
,	1.2162	.	1.1768	q	1.1042	b	1.2318	v	1.3399	,	1.2475
v	1.0827	v	1.1464	b	1.0538	,	1.1789	,	1.2749	.	1.2472
q	0.9682	q	1.1215	.	0.8957	c	0.8491	b	1.0247	q	0.9324
f	0.7975	f	0.8278	f	0.7317	f	0.8461	c	0.9371	f	0.7991
y	0.7402	z	0.7468	z	0.6763	q	0.7414	f	0.8551	y	0.6399

Burimi: Këto rezultate i kemi nxjerrur duke e analizuar frekuencën e shkronjave në 5 libra të literaturës shqipe.

Shënim: Fjala **char** iu referohet karakterit ose shkronjës ndërsa **perc** iu referohet përqindjes së frekuencës të atij karakteri në literaturë.

[1]Ben Blushi - Të jetosh në ishull.

[2]Pëllumb Kulla - Rrëfenja nga Amerika.

[3]Ben Blushi - Shqipëria.

[4]Bardhyl Mahmuti - MASHTRIMI I MADH.

[5]Bardhyl Mahmuti - KURTHET TERMINOLOGJIKE NË FUNKSION TË MOHIMIT TË GJENOCIDIT NË KOSOVË.

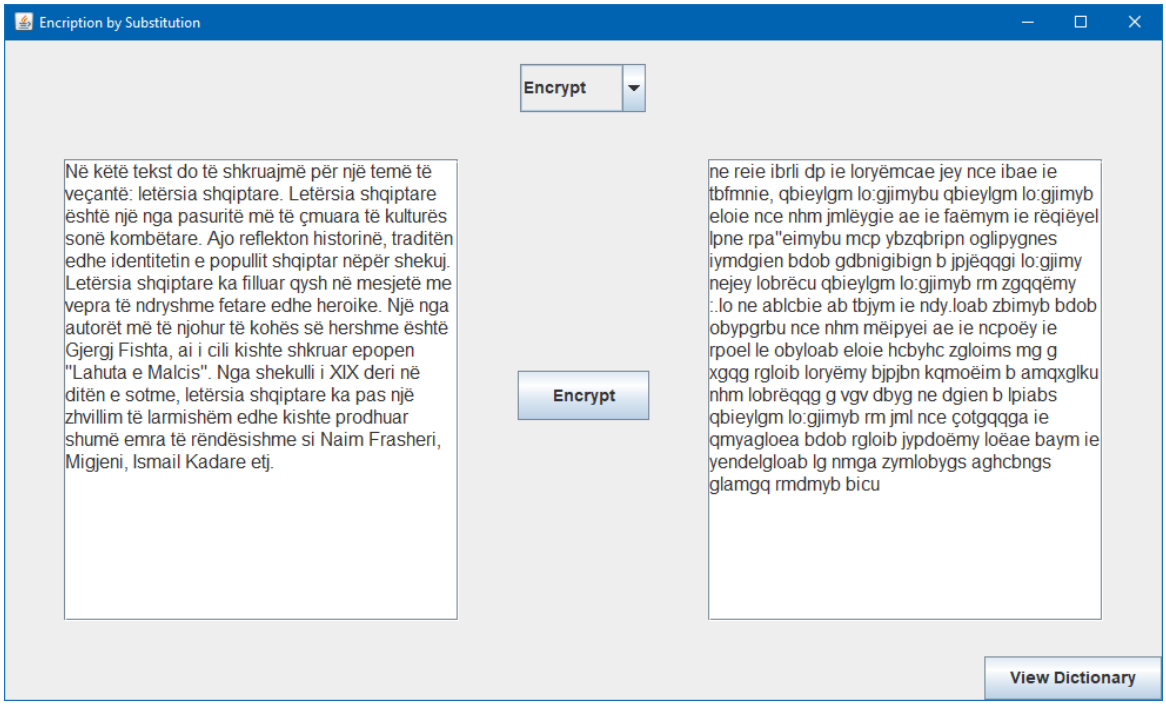


Fig. 3: Enkriptimi i një mesazhi me shfrimin me zëvendësim të rastësishëm.

Table

Original Letter	Encrypted Letter
a	m
b	"
"	k
c	x
d	d
e	b
f	z
ç	f
g	h
h	o
i	g
j	c
ë	e
k	r
l	q
,	s
m	a

Fig. 4: Tabela e shkronjave të enkriptuara.

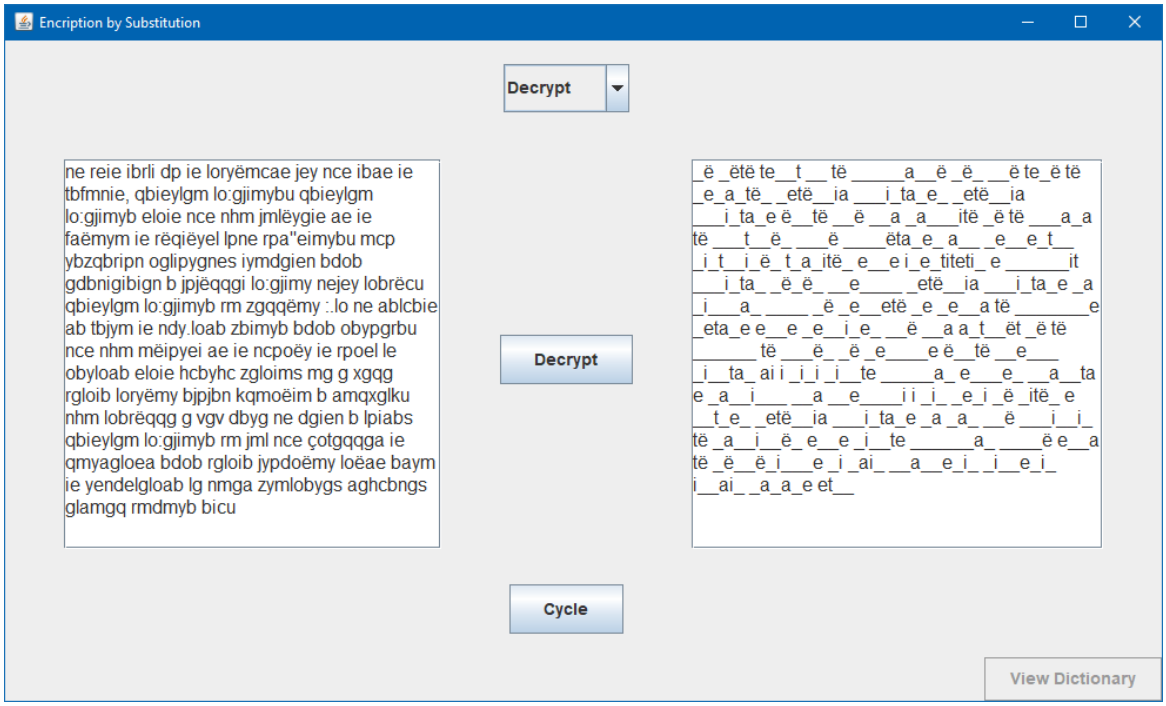


Fig. 5: Dekriptimi i mesazhit me metodën e analizës së frekuencave, ku janë të paraqitura vetëm 5 shkronjat më të përdorura në gjuhën shqipe.

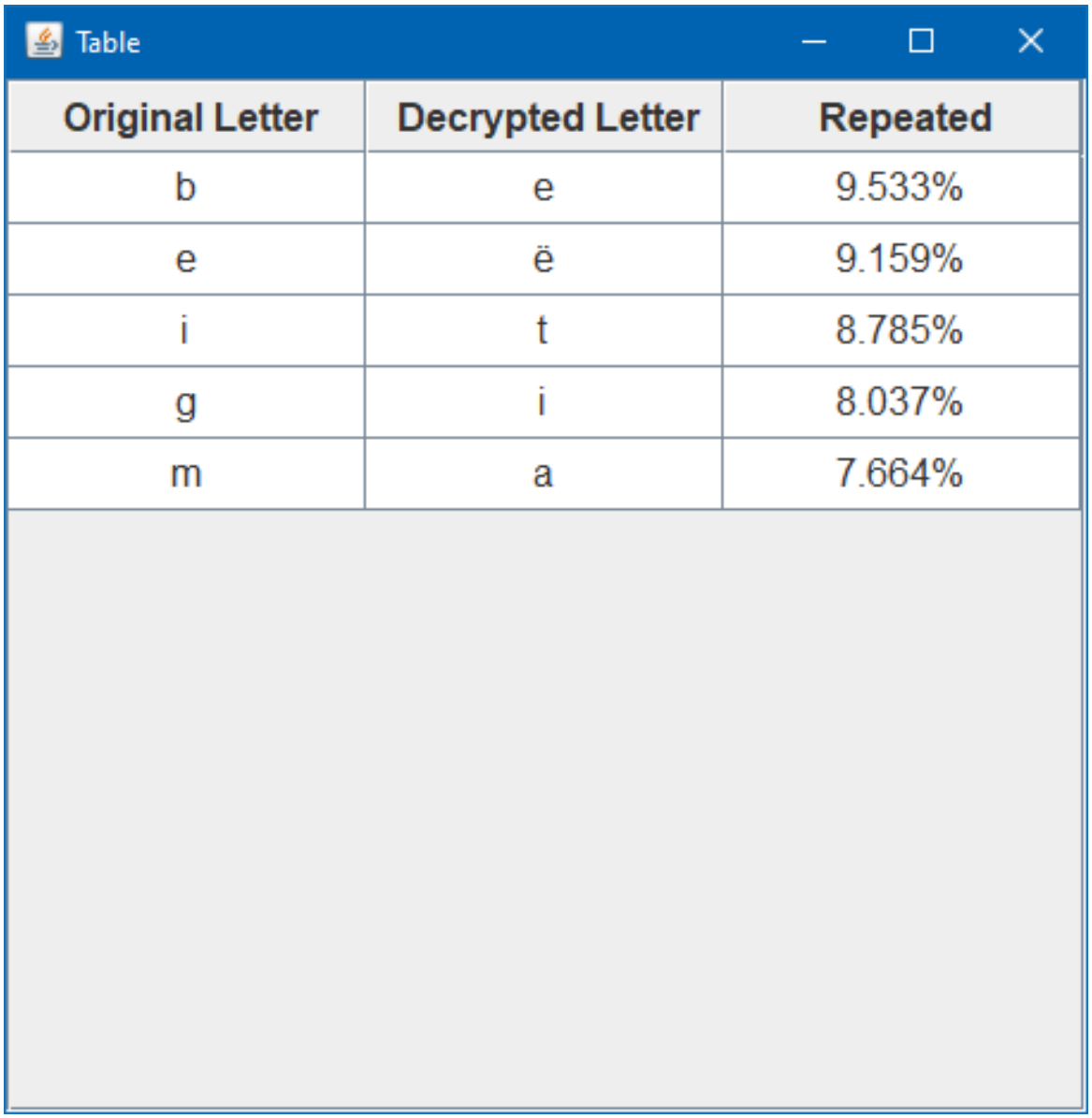


Fig. 6: Tabela e shkronjave të dekriptuara.

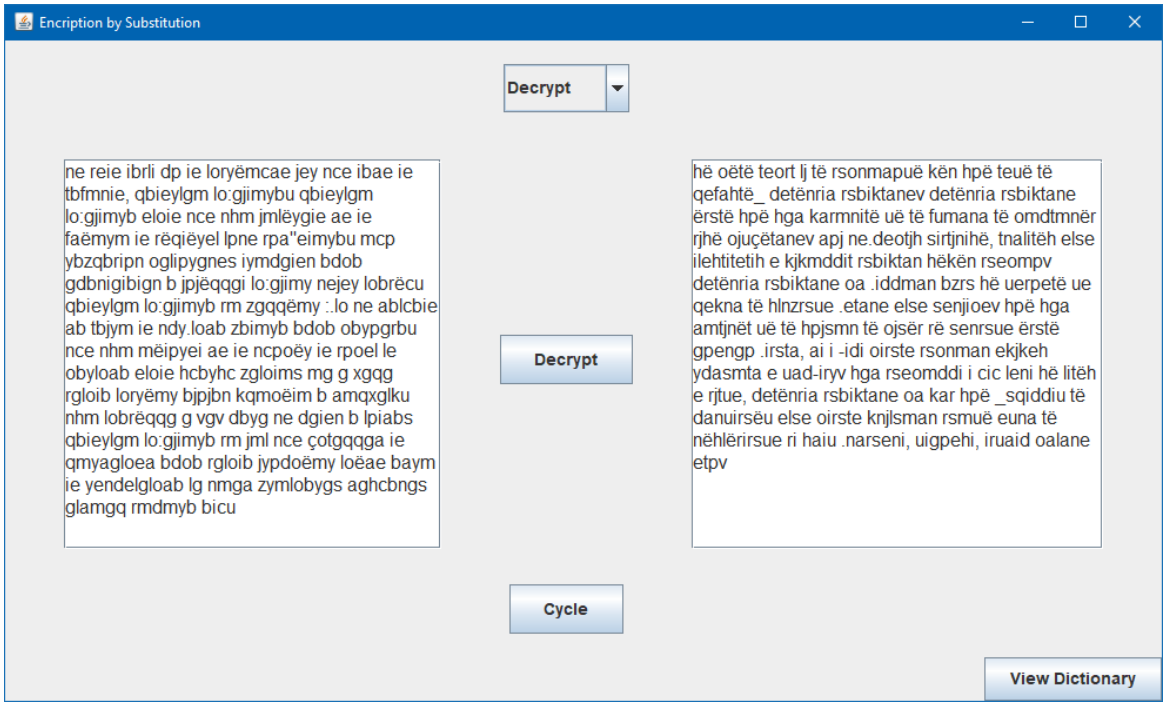
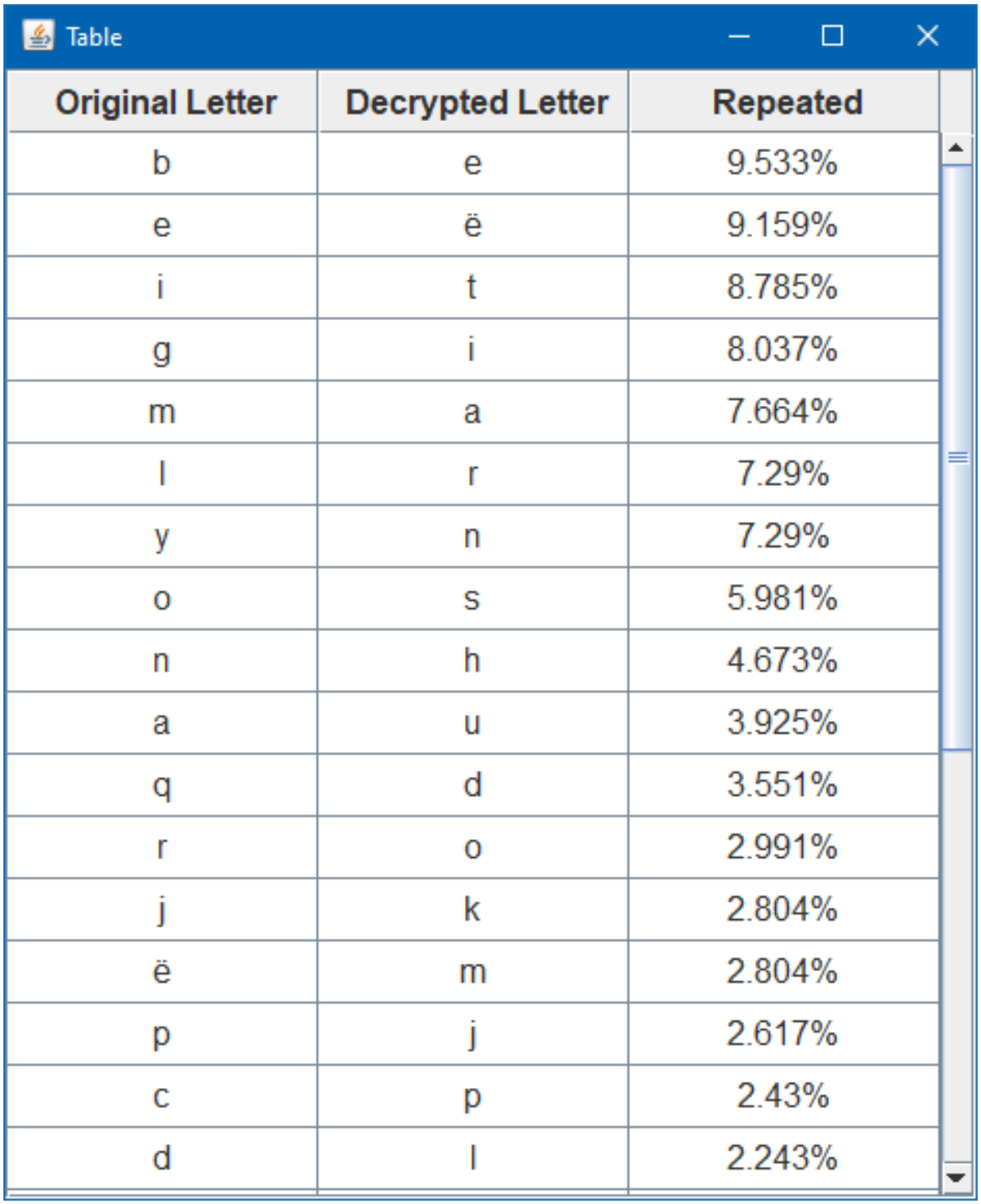


Fig. 7: Dekriptimi i mesazhit me metodën e analizës së frekuencave, ku janë të paraqitura të gjitha shkronjat më të përdorura në gjuhën shqipe.



Original Letter	Decrypted Letter	Repeated
b	e	9.533%
e	ë	9.159%
i	t	8.785%
g	i	8.037%
m	a	7.664%
l	r	7.29%
y	n	7.29%
o	s	5.981%
n	h	4.673%
a	u	3.925%
q	d	3.551%
r	o	2.991%
j	k	2.804%
ë	m	2.804%
p	j	2.617%
c	p	2.43%
d	l	2.243%

Fig. 8: Tabela e shkronjave të dekriptuara.