"PKI based Data Coloring for securing and sharing open-data in Cloud Computing Environments

Mary-Jane Sule

University of Jos

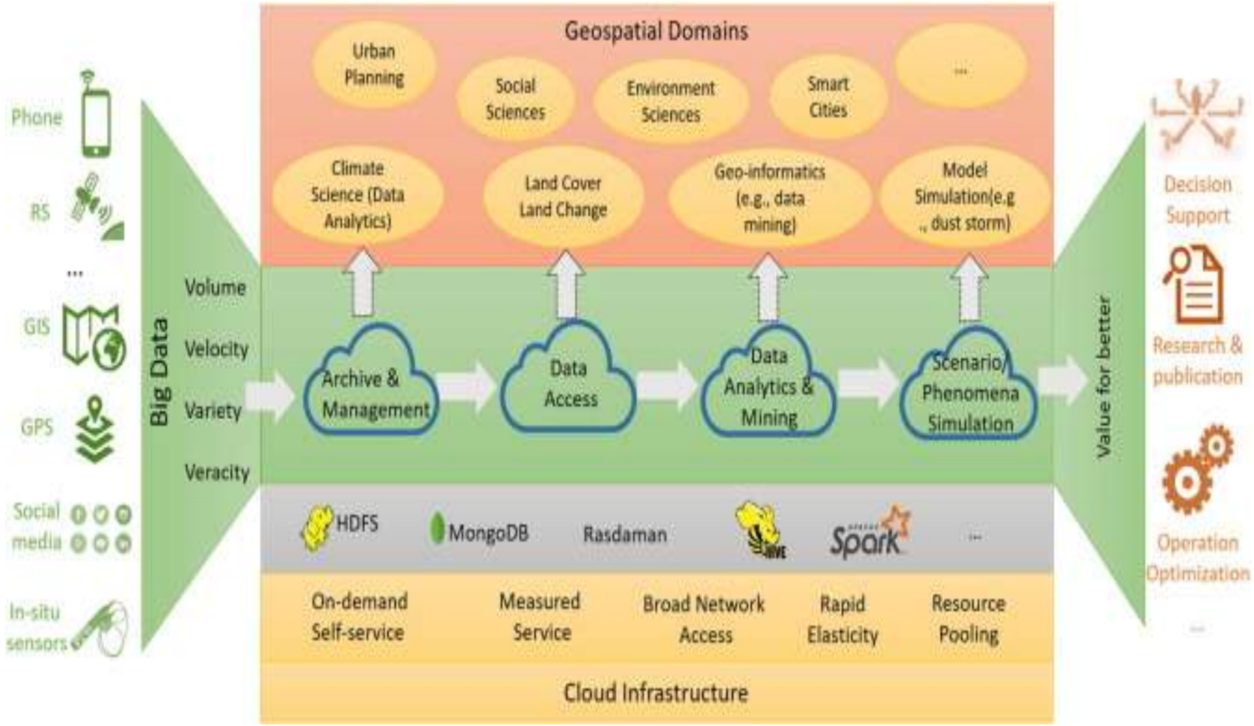
Nigeria

Content

Cloud Computing
Security
Public Key Infrastruture
Data Coloring
Data Coloring Implementation

Introduction

- Climate Research is receiving a lot of interest as it has direct as well as indirect impact on human life
- It is requiring a vase amount of computational resources
- Cloud Computing resources would accelerate the climate related researches and other mission critical services – faster result generations for large Data and availability of resources
- Expertise for researchers who are not computing experts

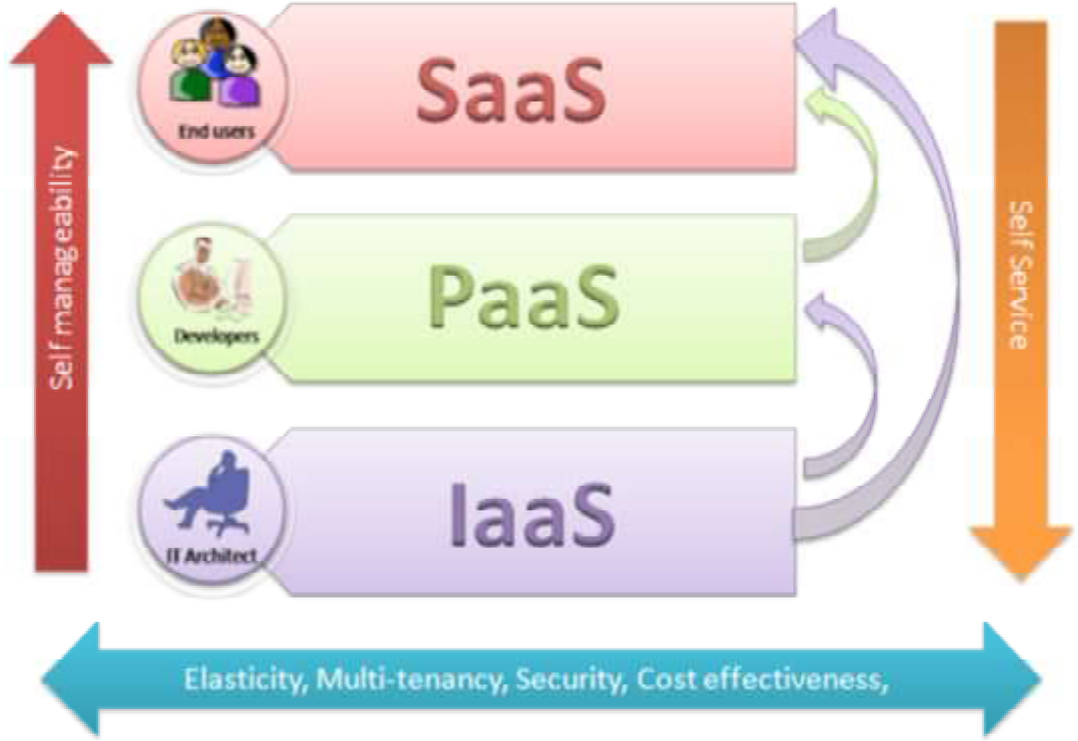


Cloud Computing

- Provides scalable computing resources with economic scale

- Made up of 3 fundamental layers that provide services and may be deployed in a number of ways – Public, Private, Community and Hybrid

- With Cloud Computing, there is
  - Multi-tenacy and shared resources
  - Improved Expertise
  - Only pay for what you use
  - There is some set standards which would improve interoperability



Challenges of Cloud Computing

- There are some challenges,
  - ➢Finance
  - ➢Data Integrity and Security
  - ➢Availability
  - ➢Lack of expertise
  - ➢Interoperability
- These are not entitely new:

Everyday Data Security

Every automated system needs to be protected to preserve its integrity, availability and confidentiality

PCs

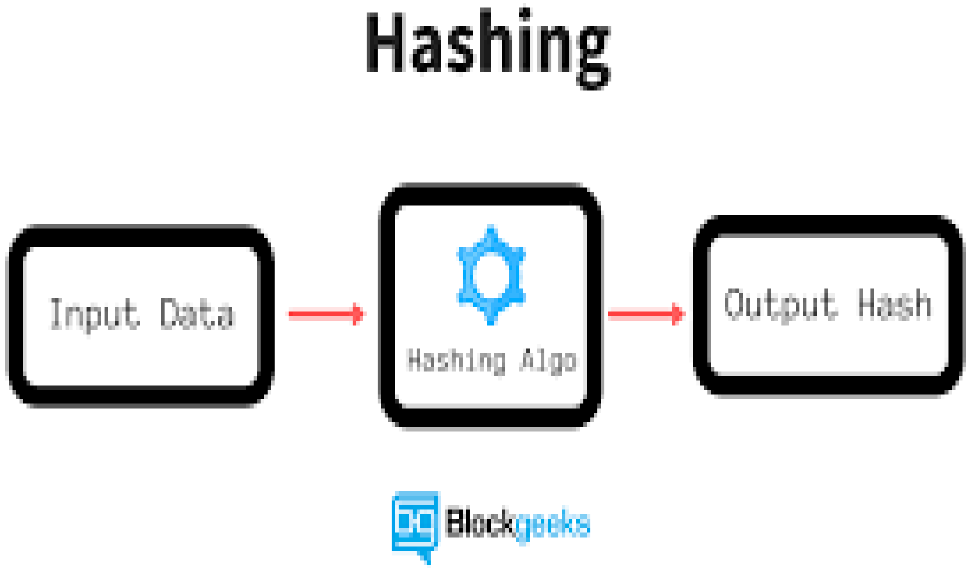
- Back-Up to External
- Restricted: Password
- Hashing
- Obfuscation / Steganography
- Encryption / Cryptography

Networked

- Doesn’t mean backed up
- Restricted : Password, Firewalls
- Hashing
- Obfuscation / Steganography
- Encryption / Cryptography

Hashing

- A unique size of string of numbers unique to the file
- Provides integrity of the Data like finger prints, same Data same hash
- Change in message, change in hash value
- storage of hash value external to data
- Creating a hash has low creation overhead
- impossible to generate a message from the hash value.
- Techniques used include SHA1 and MD5
- https://www.md5hashgenerator.com/
- https://passwordsgenerator.net/sha1-hash-generator/



https://blockgeeks.com/what-is-hashing-digital-signature-in-the-blockchain/

Obfuscation

- Making data to be illegible or meaningless / unclear
- Using known mechanisms to change color of images or order of letters.
- It does have medium computational overhead
- If mechanism is known it can be relatively easy to undo

http://ithare.com/wp-content/uploads/BB_part180_BookChapter29f_v10.png

Watermarking and Copyright (Wikipedia)

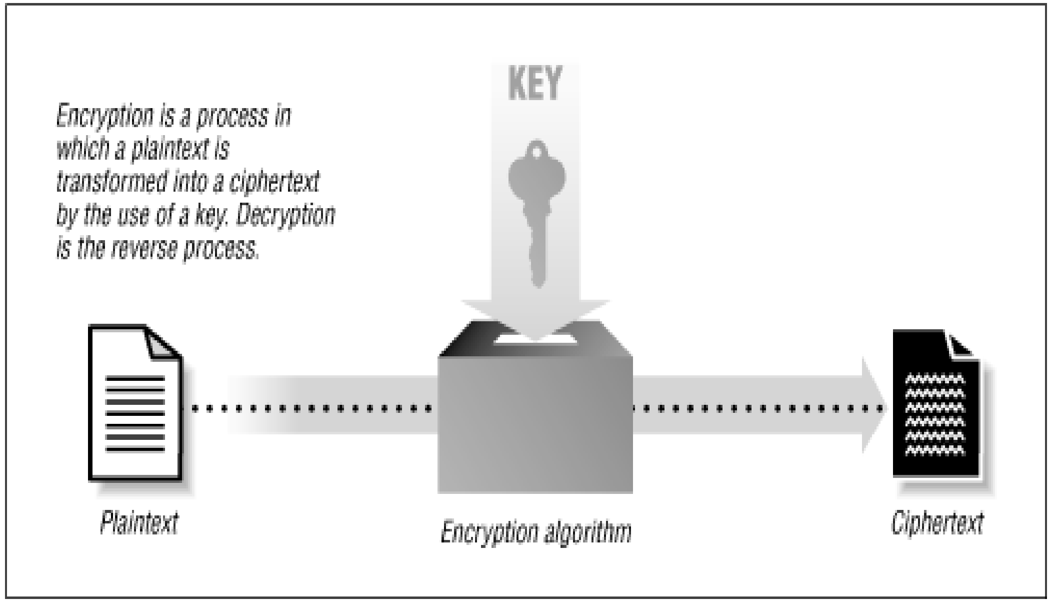
- Watermark is an indentifying image on paper when viewed by light to discourage counterfeiting.
- Digital watermark is a marker embedded in a noise-tolerant signal (eg audio, video or image) data to identify ownership of the copyright.
- Watermarking is a process of hiding the digital information
- Digital watermarks is used for tracing copyright infringement thereby verifying Data’s authenticity or integrity does nor change the size of the file

Steganography

- Act of hidden writing
- Embeds and hides existence of another message within another message from third party
- Requires the use of empty spaces
- Different from Cryptography as it does not make the message unreadable
- Doesn’t attract undue attention as the hidden message is invisible
- Though sometimes hidden message maybe encrypted or compressed before embedding

Encryption

- The process of converting data to allow only authorized persons access.
- It involves using keys to process it into one form cipertext (encrypt) and to plaintext (decrypt).
- Only secure as long as key is secured


KEY
Encryption is a process in which a plaintext is transformed into a ciphertext by the use of a key. Decryption is the reverse process.
Plaintext
Encryption algorithm
Ciphertext

https://www.cs.ait.ac.th/~on/O/oreilly/tcpip/puis/ch06_02.htm

Cryptography

- It’s a means / technique used to secure communication in the presence of adversaries (third party) – preventing the public from reading a private message
- Cryptographic systems provide the following
  - ➢ Confidentiality : concealment of data from unauthorized users (Privacy)
  - ➢ User Authentication : the user is who they say they are
  - ➢ Data Origin Authentication : proof of origin
  - ➢ Data Integrity : assurance data has not been tampered with
  - ➢ Non-repudiation : can proof the sender did indeed send the message

Public Key Infrastructure

- PKI encomprises set of roles, policies and procedures needed to create, manage, store and revoke Digital Certificate and mange pubic key encryption (Wikipedia)
- PKI using cryptography allows for secure communication on an insecure network
- PKI provides both public key encryption and use of digital signatures
- PKI consists of some important componets

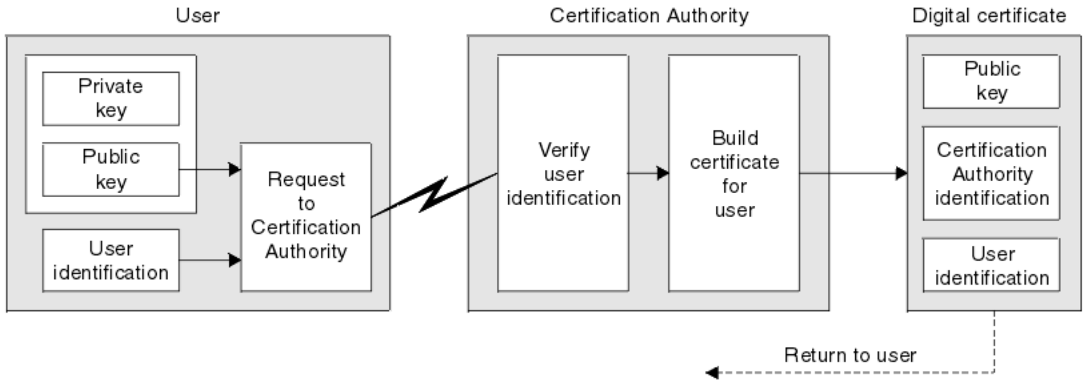
✓ Crytographic components : encryption and hash functions which provides data confidentiality, integrity and authenticates the sender

✓ Keys and random number generation

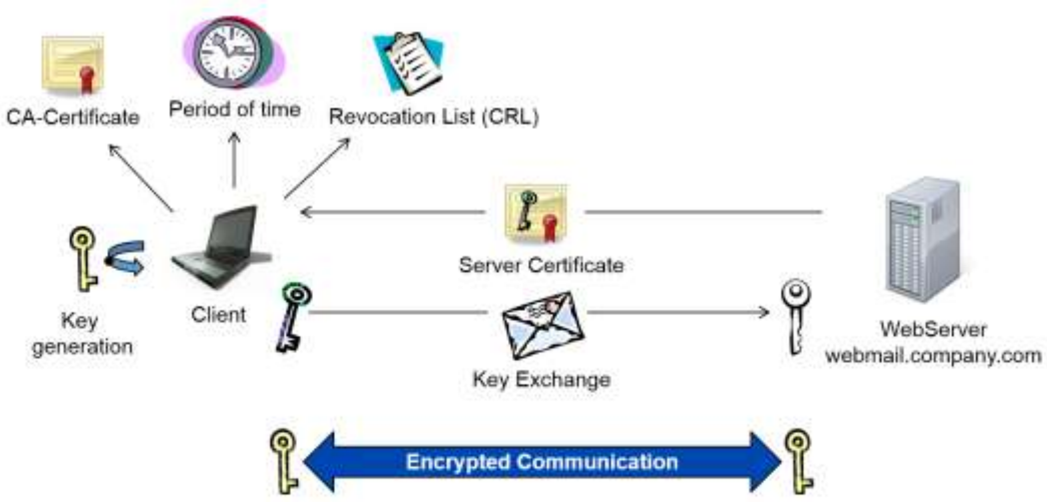
✓ Digital Certificate

Certificate Authority and Process

- A CA is third party trusted by all users that creates, distributes, revokes and manages digital certificates

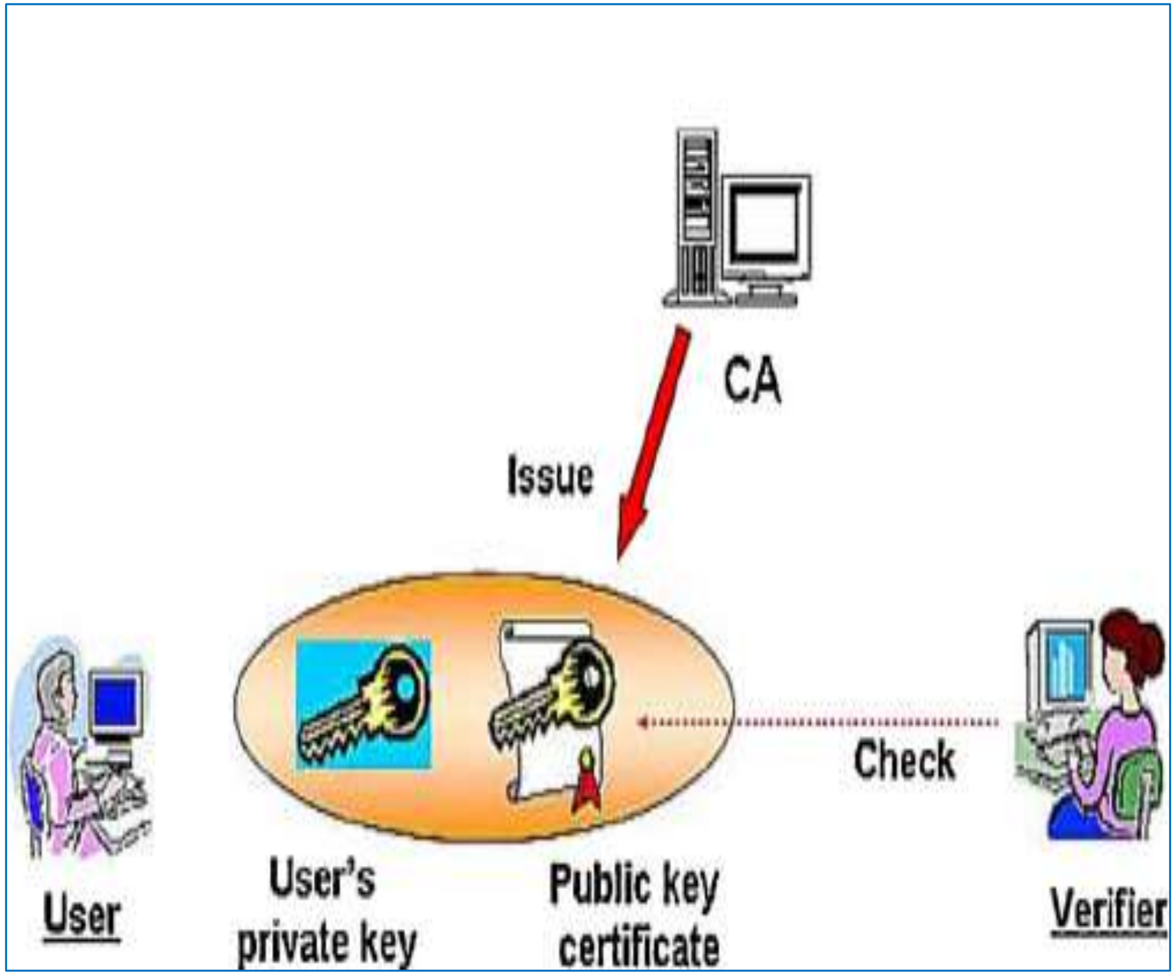


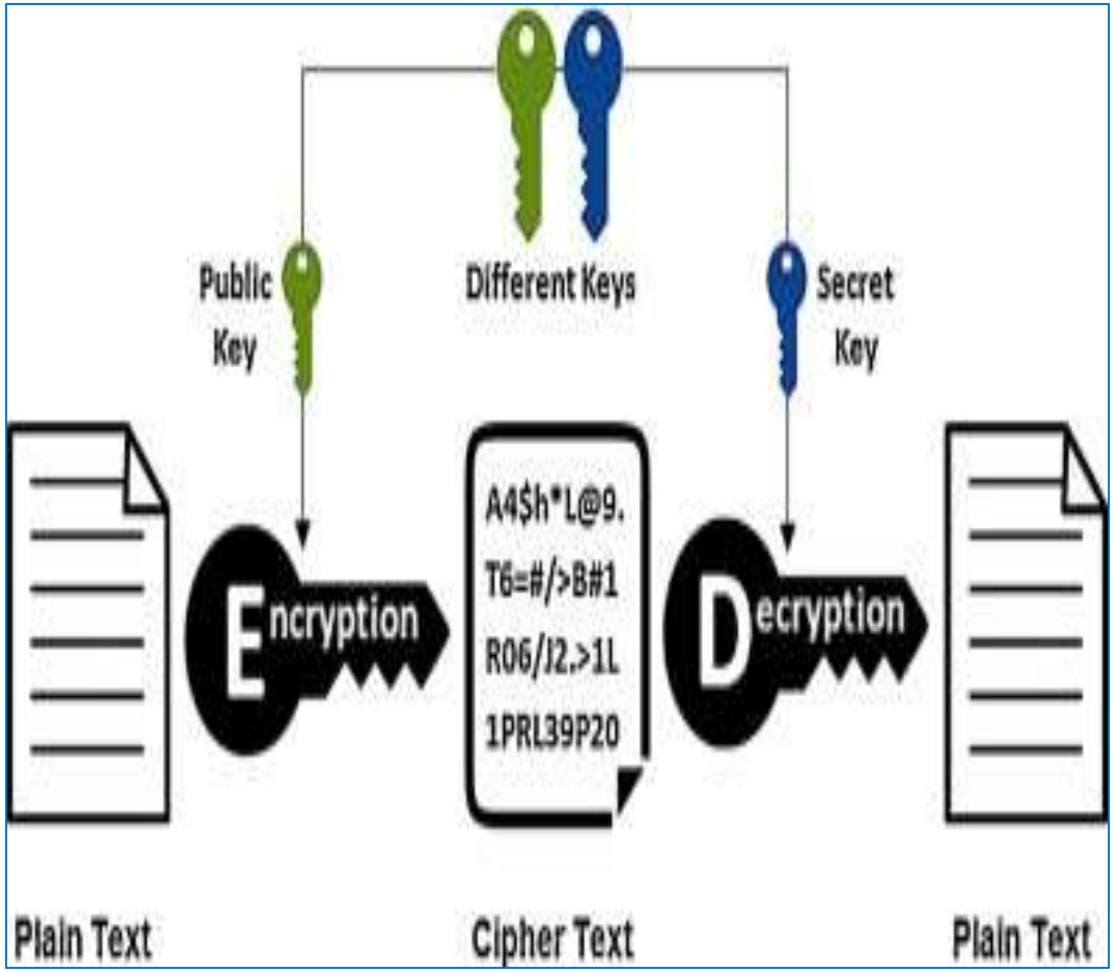
https://www.ibm.com/support/knowledgecenter/en/SSFKSJ_7.1.0/com.ibm.mq.doc/sy10590_.htm



CA issues digital certificates to users and certifies ownership of a public key. Certificates contain the public key of the owner(user)

Certificate Authority and Process (2)



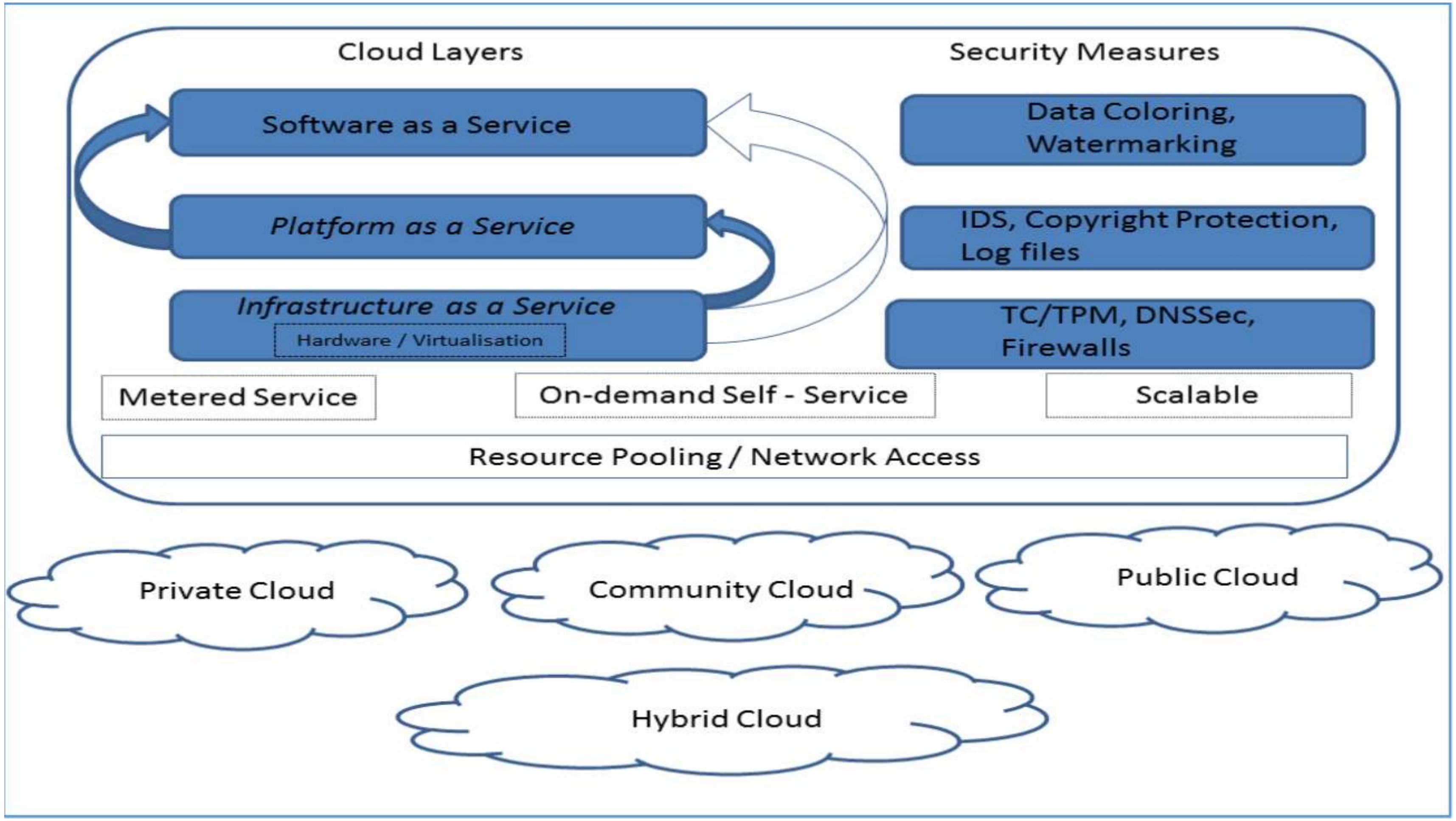


DS / Cloud User

A cloud end-user wants

- To be confident and trust that the data is secured on the CI
- Resources are available all the time
- to attest the Data integrity
  - tamper proof

Cloud Layers and Possible Security Measures


Cloud Layers
Security Measures
Software as a Service
Platform as a Service
Infrastructure as a Service
Hardware / Virtualisation
Data Coloring, Watermarking
IDS, Copyright Protection, Log files
TC/TPM, DNSSec, Firewalls
Metered Service
On-demand Self - Service
Scalable
Resource Pooling / Network Access
Private Cloud
Community Cloud
Public Cloud
Hybrid Cloud

Multilayer Security for CI

Layer	Security Challenges	Security Mechanisms
IaaS	DoS, DDoS	TC
PaaS	Alterations of Binary	IDS
SaaS	Data Loss / Data integrity	Data Coloring / Encryption

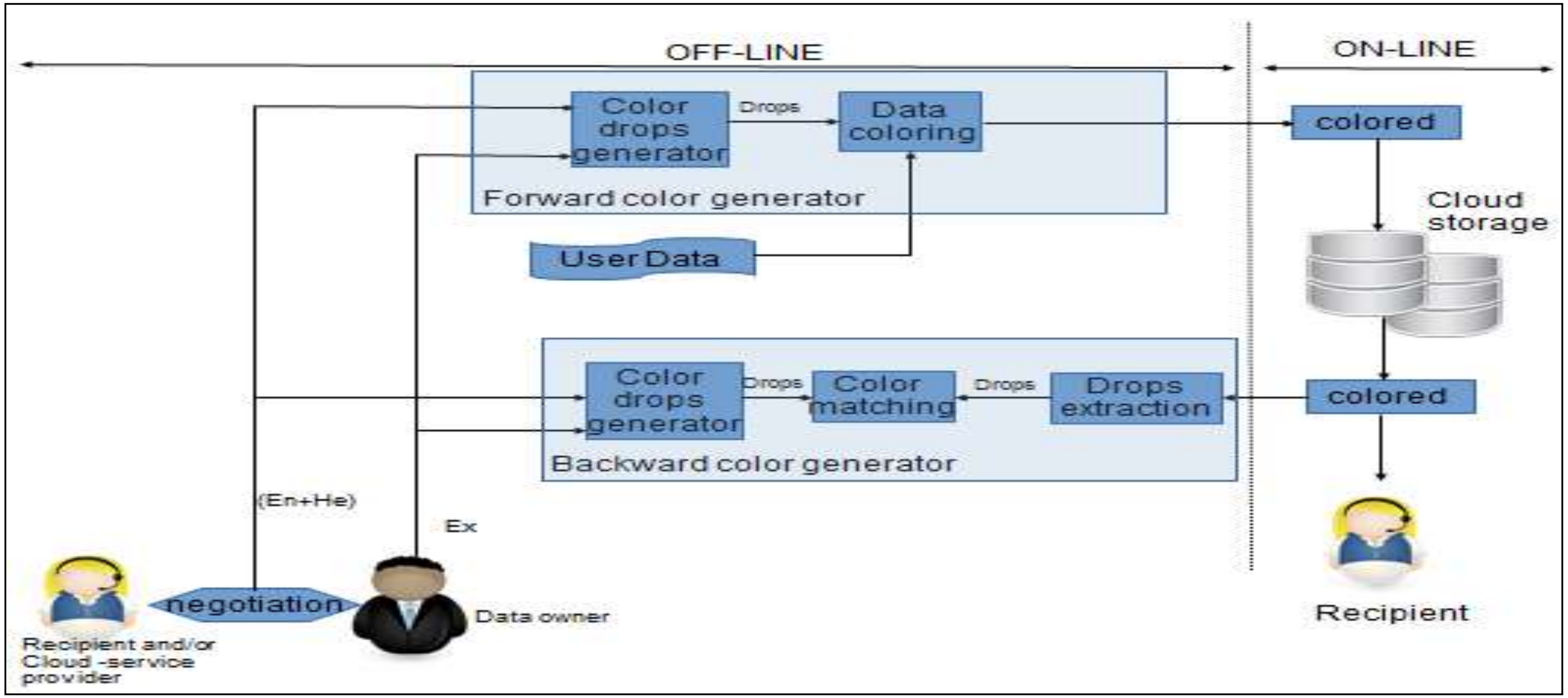
- 3 main objectives of Security is to provide Confidentiality, Integrity and Availability
- May not stop attacks but would make them less likely. No single measure can ensure complete security
- Among several solutions for cloud security are PKI and the use of multiple cloud solutions. Singh's (2017)
- Protecting Data at rest and in transit use also encrypted connections like HTTPS, SSL, TLS, FTPS

Data Colouring

- It is a technique for protecting data stored on a cloud system against un-wanted tampering and unauthorized access (especially by CSPs).
- Conceived as a watermarking software
- It is very similar to digital watermarking
- Claims to work across various types of files
- It can be achieved through cloud watermarking and encryption

- Establishes and uses concatenated fingerprints for watermarking through steganography.
- This highlights possible path of data loss or theft

Data Colouring Process



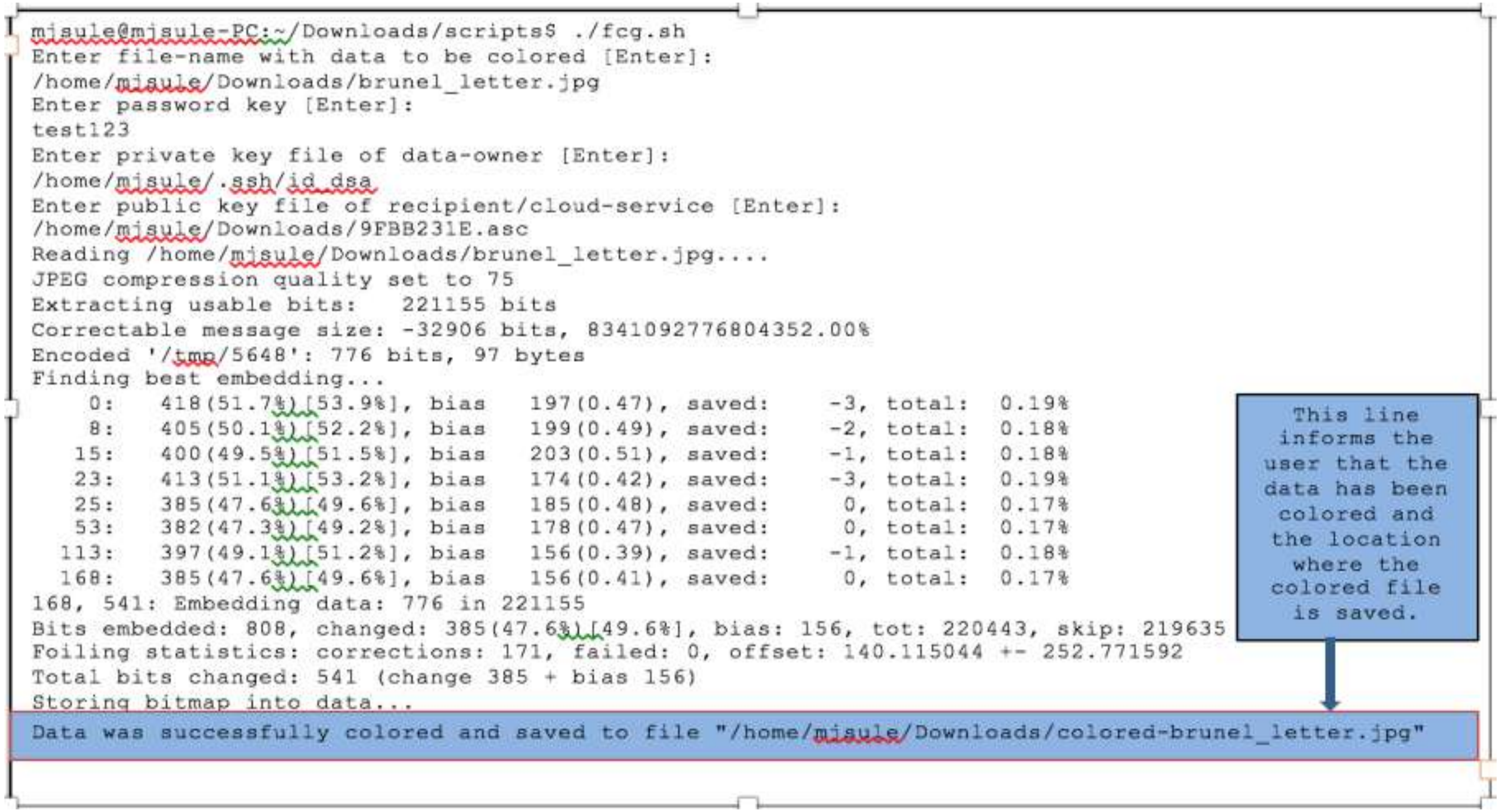
Comparism of Features

	Hashing	Encryption	Obfuscation	Data-colouring
Example	MD5, SHA	Code- table/cipher	Minimization, compression	Watermarks, fingerprinting
Association/ Technique	mathematical	Cryptographic transformation	Entropy reduction/ transformation	Mathematical/ embedding
Creation overhead	Low	High	Medium	Low
Resulting data can be directly processed on clouds	Yes	Yes (if decrypted and encrypted after processing)	Yes (if process is reversed and recreated after processing)	Yes
Security /features	Data integrity	Making data/content inaccessible to unauthorised access.	Making data/content illegible	Data ownership
Notes	Storage of hash is external to data.	Difficult to undo without original code-table	Relatively easy to undo and redo	Embedding/ distribution of watermark/ fingerprint inside data. may be difficult to detect and remove if steganography is used

Outguess - Data Coloring Implementation

- Free and Open tool for Steganography
- Relies on Data handlers to identify and modify redundant bits
- Handles different data formats once appropriate handlers are available

DC Implementation

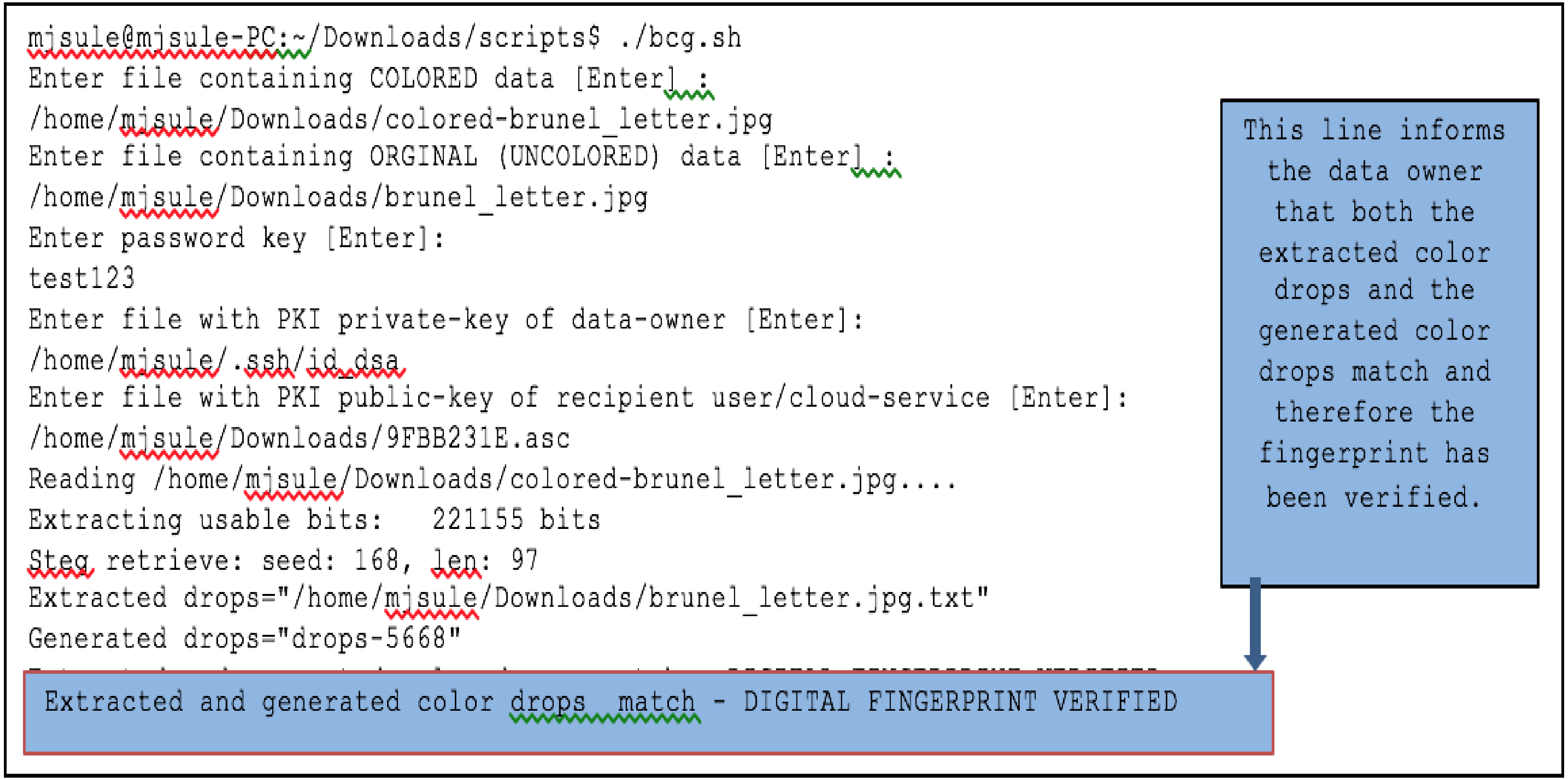


```
mjsule@mjsule-PC:~/Downloads/scripts$ ./fcg.sh
Enter file-name with data to be colored [Enter]:
/home/mjsule/Downloads/brunel_letter.jpg
Enter password key [Enter]:
test123
Enter private key file of data-owner [Enter]:
/home/mjsule/.ssh/id_dsa
Enter public key file of recipient/cloud-service [Enter]:
/home/mjsule/Downloads/9FBB231E.asc
Reading /home/mjsule/Downloads/brunel_letter.jpg....
JPEG compression quality set to 75
Extracting usable bits:   221155 bits
Correctable message size: -32906 bits, 8341092776804352.00%
Encoded '/tmp/5648': 776 bits, 97 bytes
Finding best embedding...
    0:   418(51.7%)[53.9%], bias   197(0.47), saved:    -3, total:  0.19%
    8:   405(50.1%)[52.2%], bias   199(0.49), saved:    -2, total:  0.18%
   15:   400(49.5%)[51.5%], bias   203(0.51), saved:    -1, total:  0.18%
   23:   413(51.1%)[53.2%], bias   174(0.42), saved:    -3, total:  0.19%
   25:   385(47.6%)[49.6%], bias   185(0.48), saved:     0, total:  0.17%
   53:   382(47.3%)[49.2%], bias   178(0.47), saved:     0, total:  0.17%
  113:   397(49.1%)[51.2%], bias   156(0.39), saved:    -1, total:  0.18%
  168:   385(47.6%)[49.6%], bias   156(0.41), saved:     0, total:  0.17%
168, 541: Embedding data: 776 in 221155
Bits embedded: 808, changed: 385(47.6%)[49.6%], bias: 156, tot: 220443, skip: 219635
Foiling statistics: corrections: 171, failed: 0, offset: 140.115044 +- 252.771592
Total bits changed: 541 (change 385 + bias 156)
Storing bitmap into data...
Data was successfully colored and saved to file "/home/mjsule/Downloads/colored-brunel_letter.jpg"
```

This line informs the user that the data has been colored and the location where the colored file is saved.

Figure: Generating colour drops (fcg.sh running)

DC Implementation 2

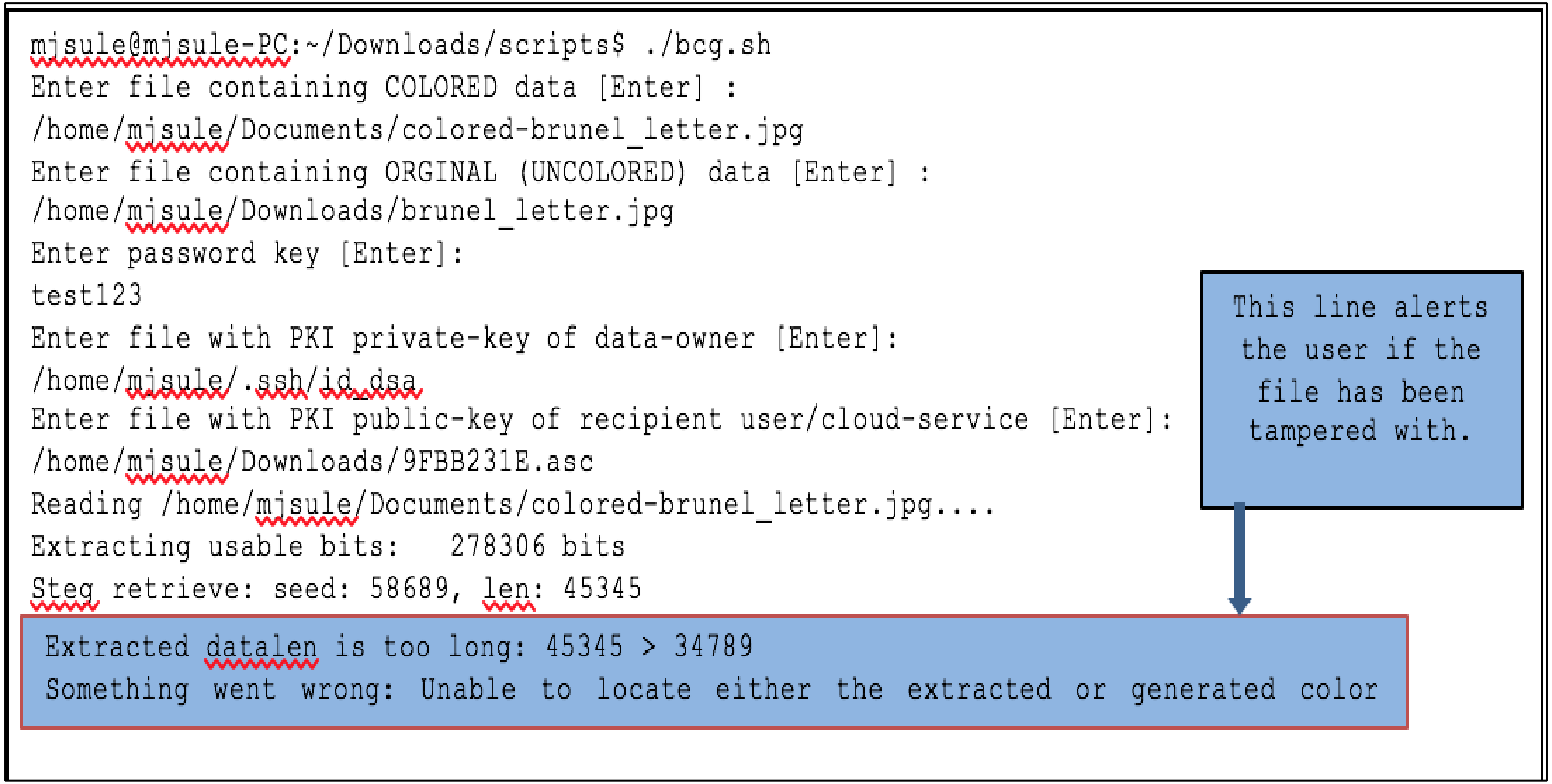


```
mjsule@mjsule-PC:~/Downloads/scripts$ ./bcg.sh
Enter file containing COLORED data [Enter] :
/home/mjsule/Downloads/colored-brunel_letter.jpg
Enter file containing ORGINAL (UNCOLORED) data [Enter] :
/home/mjsule/Downloads/brunel_letter.jpg
Enter password key [Enter]:
test123
Enter file with PKI private-key of data-owner [Enter]:
/home/mjsule/.ssh/id_dsa
Enter file with PKI public-key of recipient user/cloud-service [Enter]:
/home/mjsule/Downloads/9FBB231E.asc
Reading /home/mjsule/Downloads/colored-brunel_letter.jpg....
Extracting usable bits:   221155 bits
Steg retrieve: seed: 168, len: 97
Extracted drops="/home/mjsule/Downloads/brunel_letter.jpg.txt"
Generated drops="drops-5668"
Extracted and generated color drops  match - DIGITAL FINGERPRINT VERIFIED
```

This line informs the data owner that both the extracted color drops and the generated color drops match and therefore the fingerprint has been verified.

Figure: Extracting colour drops with bcg.sh

DC Verification



```
mjsule@mjsule-PC:~/Downloads/scripts$ ./bcg.sh
Enter file containing COLORED data [Enter] :
/home/mjsule/Documents/colored-brunel_letter.jpg
Enter file containing ORGINAL (UNCOLORED) data [Enter] :
/home/mjsule/Downloads/brunel_letter.jpg
Enter password key [Enter]:
test123
Enter file with PKI private-key of data-owner [Enter]:
/home/mjsule/.ssh/id_dsa
Enter file with PKI public-key of recipient user/cloud-service [Enter]:
/home/mjsule/Downloads/9FBB231E.asc
Reading /home/mjsule/Documents/colored-brunel_letter.jpg....
Extracting usable bits:   278306 bits
Steg retrieve: seed: 58689, len: 45345
Extracted datalen is too long: 45345 > 34789
Something went wrong: Unable to locate either the extracted or generated color
```

This line alerts the user if the file has been tampered with.

Figure: Verifying colour drops with bcg.sh

Sources for Colour Drops Generation

ITEM	CONTRIBUTION
Data-file to be coloured	Fingerprint to detect unauthorized modifications to content
Private Key of Data Owner	Fingerprint to indentify data owner
Public Key of Recipient or Cloud-Service	Fringerprint to trace path of data-loss/theft

Theft / Loss Responsibilities

	Private Key of data-owner	**Public key of cloud- service**	**Public key of data recipient**	**INFORMATION OBTAINED FROM DROPS**
1	YES	NO	NO	Identity of data-owner
2	NO	YES	NO	Identity of CSI
3	NO	NO	YES	Identity of recipient (CSP)
4	YES	YES	NO	Identity of both owner and CSI
5	NO	YES	YES	Identity of both CSI and recipient (CSP)
6	YES	NO	YES	Identity of both owner and recipient (CSP)
7	YES	YES	YES	Identity of owner, CSI and recipient (CSP)

Hands-On

ref

- Vincent Lozupone, Analyze encryption and public key infrastructure (PKI),International Journal of Information Management, Volume 38, Issue 1,2018,Pages 42-44,ISSN 0268-4012,https://doi.org/10.1016/j.ijinfomgt.2017.08.004.(http://www.sciencedirect.com/science/article/pii/S0268401217303195)
- Chaowei Yang, Manzhu Yu, Fei Hu, Yongyao Jiang, Yun Li, Utilizing Cloud Computing to address big geospatial data challenges, Computers, Environment and Urban Systems,Volume 61, Part B,2017,Pages 120-128,ISSN 0198-9715,https://doi.org/10.1016/j.compenvurbsys.2016.10.010.(http://www.sciencedirect.com/science/article/pii/S0198971516303106)
- https://rath.asia/2017/02/https-and-related-technology-explained/https-blog-06/
- A Deployment Model for Cloud Computing using the Analytic Hierarchy Process and BCOR Analysis - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/Visual-model-of-NIST-working-definition-of-cloud-computing_fig1_259972968 [accessed 11 Aug, 2018]
- https://en.wikipedia.org/wiki/Digital_watermarking
- https://en.wikipedia.org/wiki/Watermark
- Charlie Obimbo and Behzad Salami (May 16th 2012). Using Digital Watermarking for Copyright Protection, Watermarking Mithun Das Gupta, IntechOpen, DOI: 10.5772/38184. Available from: https://www.intechopen.com/books/watermarking-volume-2/using-digital-watermarking-for-copyright-protection
- https://blockgeeks.com/what-is-hashing-digital-signature-in-the-blockchain/