AUTOMATED NORMALIZATION AND CORRELATION OF MOBILE PHONE EXTRACTIONS USING THE STANDARD CASE

Martina Reif, Eoghan Casey, Quentin Rossy

PHONE DATA

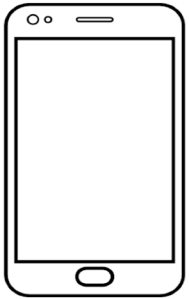
Data **sources**

- Call Detail Records (CDR), extractions, wiretaps, …
- Complementarity

Interest **in criminal investigations**

- Reconstruction of criminal events and activities
- Identify and locate targeted individuals
- Can contain traces of illegal activity

→ Analysis aims to provide support for the investigation and to help with decision making.



Volume growing

Variety of data and technology

Frequent changes of apps & devices

Connections with other devices

Rapid evolution of technology

Extraction tools

Different software with different data models

Updates to data models

Proprietary systems

Extracted data

Variations in content due to different capabilities of tools

Variations in formats

Variations in data Representation

Analysis systems

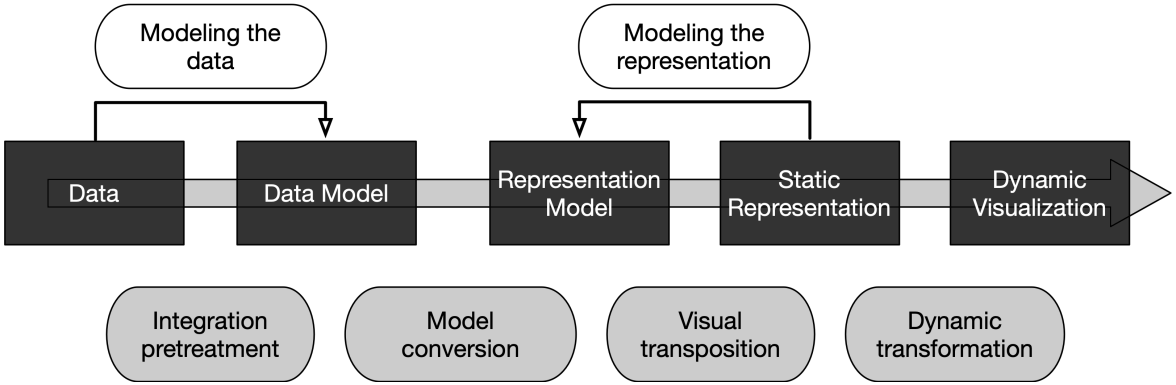
Different software with different data models

Variations in formats

Updates to data models

PROPOSED SOLUTION

→ Implement a structured method to the analysis of communication data extracted from mobile phones, using the standard CASE.



CASE

- Ontology-based **standard** for the **representation** of digital evidence
- Allows for **correlation** and **combination** of different data sources
- Destined for **criminal investigations**

CASE “ENTITIES”


```
2761        {
2762            "@id": "Account-522b7628-7fe7-11e9-b2b4-0c4de9c21b53",
2763            "@type": "Trace",
2764            "propertyBundle": [
2765                {
2766                    "@type": "Account",
2767                    "accounIdentifier": "identifier example"
2768                },
2769                {
2770                    "@type": "ApplicationAccount",
2771                    "application": "Application-528e17e2-7fe7-11e9-b2b4-0c4de9c21b53"
2772                },
2773                {
2774                    "@type": "DigitalAccount",
2775                    "displayName": "username example"
2776                }
2777            ]
2778        },
```

IDENTIFYING AND LOCATING TARGET DATA

- Communication data
- Identifiers (phone numbers, email addresses, etc.)
- Temporal indicators (timestamps, etc.)
- Spatial indicators (GPS coordinates, addresses, etc.)

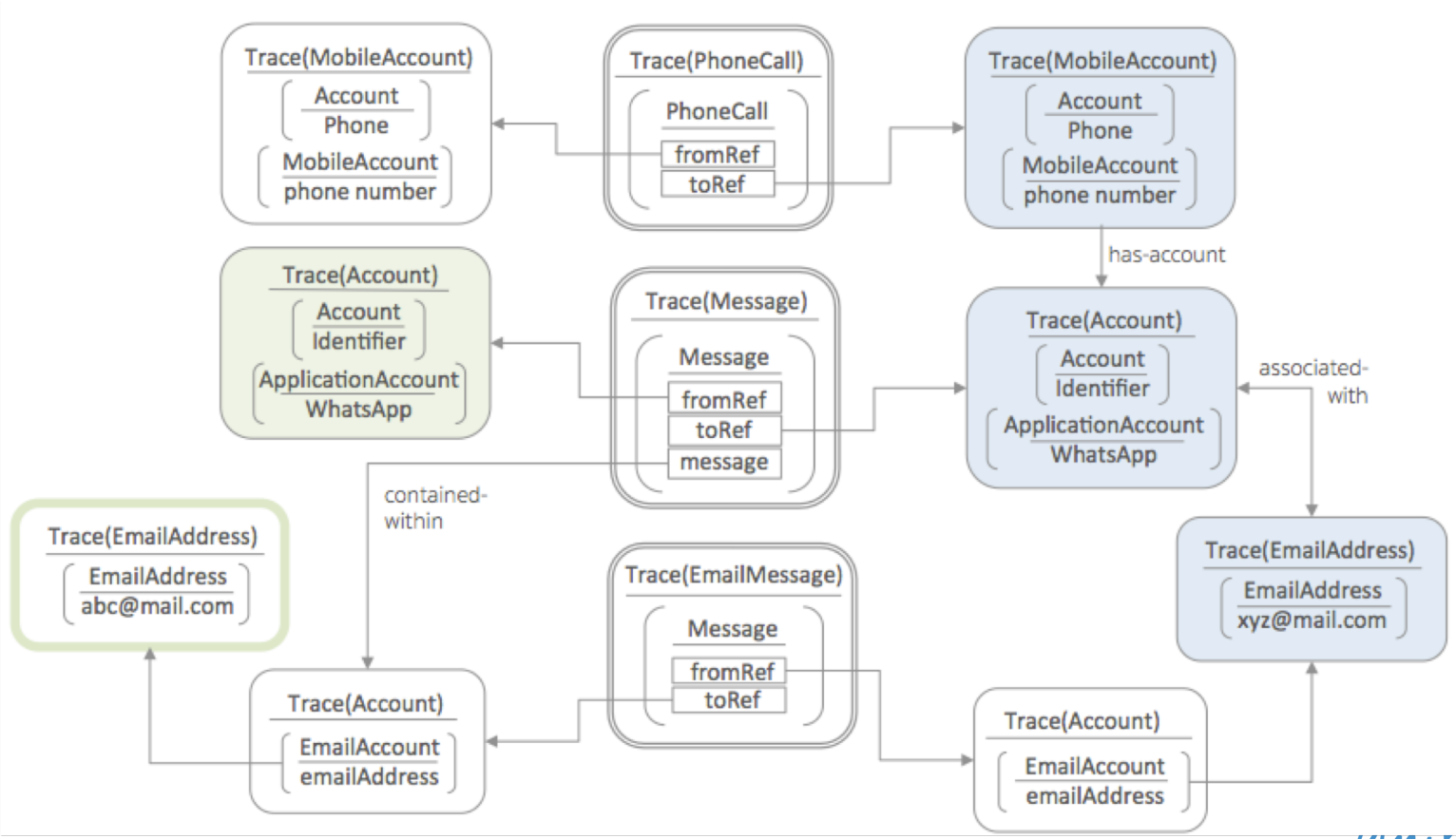
DATA

- Representation of communications
  - Indication of only one party
  - Different attributes
- Differences in format of identifiers (e.g. phone numbers, email addresses) and other information (timestamps)
- Duplicates
- Incomplete or encrypted data
- Missing links between entities (e.g. WhatsApp vs phone number)

MAPPING / MODELLING THE DATA

- Represent information with CASE vocabulary
  - Information present in extracted data
  - Link information inferred from the extracted data

- Create CASE objects and relationships between them


Trace(MobileAccount)
Account
Phone
MobileAccount
phone number
Trace(PhoneCall)
PhoneCall
fromRef
toRef
Trace(MobileAccount)
Account
Phone
MobileAccount
phone number
has-account
Trace(Account)
Account
Identifier
ApplicationAccount
WhatsApp
Trace(Message)
Message
fromRef
toRef
message
Trace(Account)
Account
Identifier
ApplicationAccount
WhatsApp
associated-
with
contained-
within
Trace(EmailAddress)
EmailAddress
abc@mail.com
Trace(EmailMessage)
Message
fromRef
toRef
Trace(EmailAddress)
EmailAddress
xyz@mail.com
Trace(Account)
EmailAccount
emailAddress
Trace(Account)
EmailAccount
emailAddress

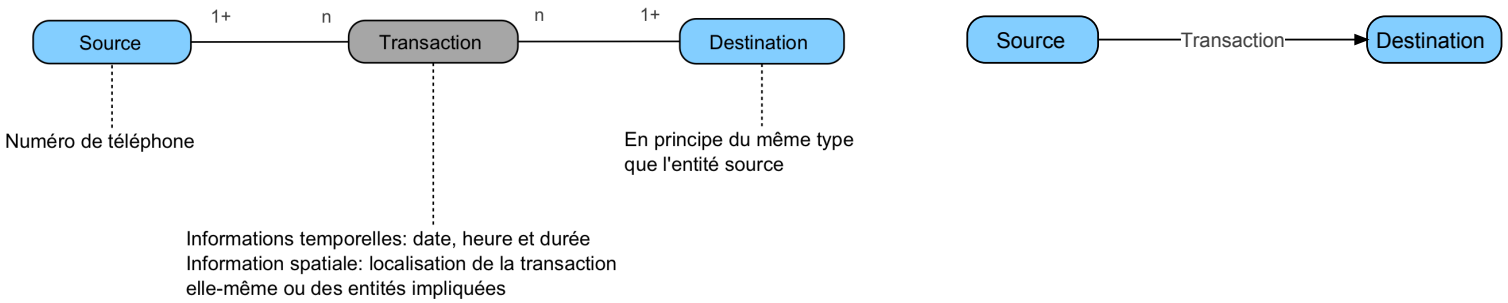
TRANSLATION OF DATA

Elaboration of python scripts

- Normalization and correlation of data
- Establishment of inferred relationships
- JSON output (translated data)

REPRESENTATION MODEL

- Communication data is usually represented as “transaction”



- Concept can be applied to all types of communication data

TRANSACTION OUTPUT

- Combined communications
- Chosen and normalized identifiers
- Common and specific attributes
- CSV

type | uuid | type(pb) | Timestamp | fromRef | toRef | duration | message | application| latitude | longitude

Série de cambriolages OP [redacted]

Schéma relationnel quantitatif produit sur la base de données téléphoniques

Auteur: DAF
Enquêteur: [redacted]
V1.0 11.11.2019

DOCUMENT DE TRAVAIL
N'EST PAS DESTINE AUX ACTES JUDICIAIRES
NE DOIT PAS FIGURER AU DOSSIER DU JUGE

[redacted]
[redacted] CTR
[redacted] CTR
[redacted]

25

39

[redacted]

11

[redacted]

2

[redacted]

8 32 1 298 405 6 114 4 8 4 132 11 29 3 4 20

[redacted]

21

9 55 1 313 7 1 1 1 29 7 2 141 1 368 6

SMS technique indiquant un appel manqué

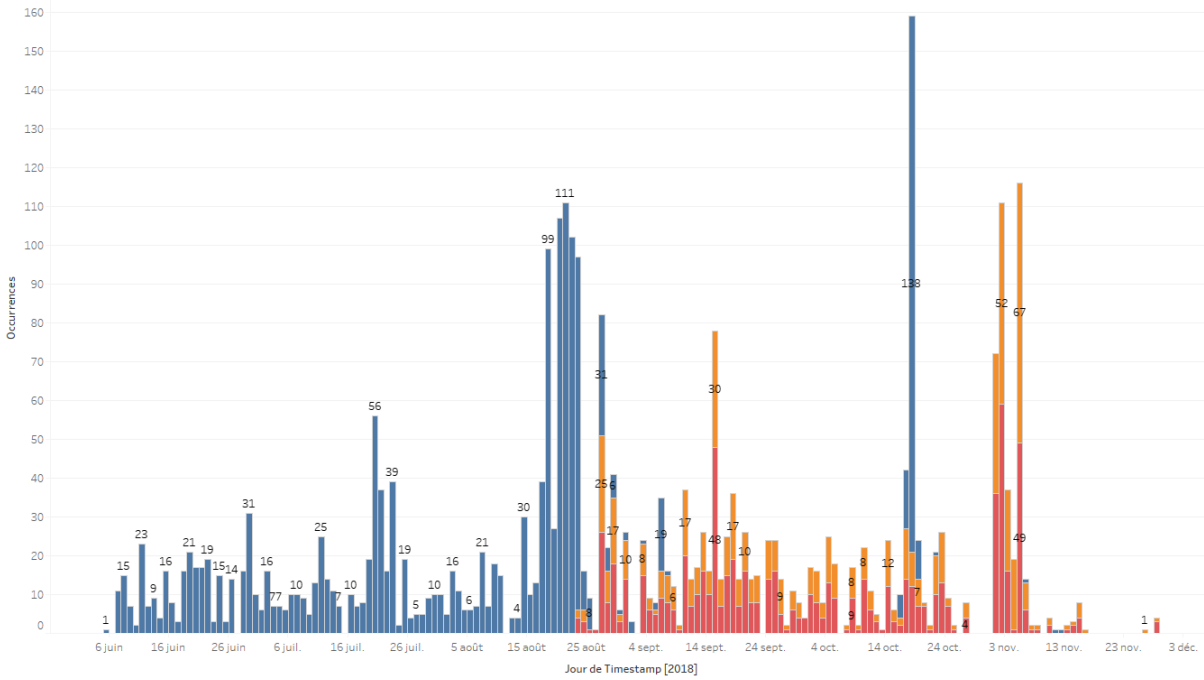
[redacted]

[redacted]

Type de communications

Type	Source	Direction	
Call	Téléphonie traditionnelle	Inconnu	9
		Entrant	202
		Sortant	235
	Facebook Messenger	Inconnu	17
	FaceTime	Entrant	39
		Sortant	74
	Skype: h*********	Inconnu	21
	Viber	Inconnu	761
	WhatsApp	Entrant	318
		Sortant	242
	WhatsApp Audio	Inconnu	11
		Entrant	30
		Sortant	9
Chat	iMessage	Entrant	13
		Sortant	7
	iMessage: *****@gmail.com	Inconnu	2
		Entrant	9
		Sortant	13
	iMessage: +41******	Inconnu	28
		Entrant	579
		Sortant	399
	iMessage: +41*******	Entrant	4
		Sortant	4
	Viber	Inconnu	18
		Sortant	2'440
	WhatsApp	Inconnu	129
		Entrant	79'377
		Sortant	48'340
MMS	Téléphonie traditionnelle	Entrant	108
		Sortant	59
SMS	Téléphonie traditionnelle	Inconnu	5
		Entrant	8'493
		Sortant	8'006

Juin - Novembre 2018



DISCUSSION

- Proposed method for the analysis of communication data

- Possibility to combine different datasets
- Possibility to integrate other types of evidence

Thank you for your attention.