Exploratory study for the detection and analysis of links between prospective advance fee fraud emails in an intelligence perspective

Stéphane **Birrer**, *Crime Analyst* and Research Assistant***
Olivier **Ribaux**, PhD, *Associate Professor*** and *Crime Analyst**
Julien **Cartier**, *Crime Analyst**
Quentin **Rossy**, *Research Assistant***
Sébastien **Capt**, *Crime Analyst* and Research Assistant***
Mélanie **Zufferey**, Student**
* *Police cantonale vaudoise, Switzerland*
** *Forensic Science Institute, University of Lausanne, Switzerland*

I. Introduction

The so called advance fee fraud, or 419 scam, is an international criminal phenomenon that has dramatically evolved over the past years with the development of the Internet. The gist of the fraud resides in persuading prospective victims to pay a modest amount of money, through the false promise of receiving substantial benefits in return. Massive sending of emails is generally the technique used to sound the internet for finding promising gullible candidates who are tempted by quick and easy gain. Once the contact has been initiated, the particularly credible scammer will argue the potential victim into feeding an account with advance fee. There are a broad variety of scenarios, but most typically, the emails describe the need, under different pretexts, to move funds from West Africa. The person who will provide assistance by making his own account available for this transaction will be offered a generous commission. Scenarios

evolve with actual international events that are good pretexts for asking for help or suggesting a potential substantial gain.

These messages reach a large quantity of potential victims in any country. It has been demonstrated that the emails provenance is mostly West Africa and particularly Nigeria. However, scammers tend to install an activity also in the other continents. For instance, it has been detected that, in Europe, a substantial quantity of spam originate from the Netherlands (Edelson, 2003) (Schiffer, Birrer, Cartier, Capt & Ribaux, 2004). Emails sent from one country often propose contacts (phone number, fax, etc) in another country. This presumes some degree of organization: tasks are distributed over different people who have different roles and who are spread over a broad geographical area. More generally, the roots of the phenomenon are certainly to be found in the worrying worldwide development of organized crime coming from West Africa (ONU, 2005) which is also a destabilizing factor for this region (Viosca, Bergiel & Balsmeier, 2004). The amounts stolen are difficult to estimate but seem to be very significant by only considering specific cases reported which are only a small part of the reality (Edelson, 2003) (Smith, Holmes & Kaufmann, 1999) (National White Collar Crime Center and the Federal Bureau of Investigation, 2002). The impact of this pervasive phenomenon that monopolizes internet resources, in particular the nuisances caused to the web users is also not to be neglected. However, even if punctual strategic analysis provides indication about the extent and the development of this criminality, there are difficulties to

deeply understand its underlying mechanisms (Office Fédéral de la police, 2003).

Preventative approaches reside in the dissemination through different media of warning messages to the community. A successful situational method consists in spam filtering that avoid mailboxes invasion (Edelson, 2003). Repressive style of policing provides limited results. The efforts deployed by the criminal justice system lead only occasionally to condemnations of suspects. The international nature of the phenomenon and inadequate legal frameworks are mentioned as probable causes of this weakness (Edelson, 2003) (Oriola, 2005), but when a case is investigated in depth, the importance of the amount stolen and the money laundering mechanism appear clearly (Bécherraz, 2004a) (Bécherraz, 2004b). Despite those edifying examples, perpetrators are still very rarely prosecuted and when the case crosses jurisdictional boundaries, the chance to succeed dramatically decreases.

There is no global strategy for fighting against this phenomenon that rapidly evolves and extends to other types of frauds that use the spam technique. In particular, another form is illustrated by the “phishing” method that consists of inciting connections on websites that imitate an existing e-banking service with the intention to capture passwords and access codes.

The objective of the project is to promote an intelligence-led approach by systematically collecting, collating and analyzing the form, the content and the provenance of spam in an operational and

strategic perspective. The utility and feasibility of this approach has been tested through this preliminary study.

II. The project

Individual who feel that a fraudulent mechanism is behind a typical advance fee fraud message frequently forward the received email to the police. It is assumed that the form and content of these emails have an intelligence potential that is not systematically used, neither at a strategic, nor at an operational level. Punctual studies have already demonstrated some interest in reasoning on this information, mostly in order to evaluate the provenance of the messages (Edelson, 2003) (Schiffer et al., 2004), but there was still a need to find out original and useful forms of inferences that can be drawn through a deeper analysis of these emails. Specifically, the potential of this set of information for sorting out the activity of different groups of scammers has been tested through this study. Moreover, there was also an interest in detecting if an activity was physically installed within a country of interest (in this study: Switzerland). The evaluation of the obtained results has been considered to provide indications about the worth of undertaking other steps toward the design and implementation of a more ambitious intelligence process.

III. Material and method

The sample

A sample of 400 emails collected during two separated periods of three months has been manually collated in an iBase® database. Visualization facilities have also been implemented through the use of Analyst's Notebook® (I2 group inc.) in order to explore and represent detected links.

Links

Moreover, links will be assumed through similarities detected between emails, such as correspondence between scenarios, similitude in the contact details or other clues that could indicate some "proximity" between the senders of the emails. At this stage, knowledge is insufficient for a more formal classification. Thus, it has been provisionally decided to classify those hypothetical links into three categories or levels.

- Level 1, represented by a plain line: two emails are linked when there is a high degree of certainty that the same group of people are at the origin of the emails. It applies for instance when the same contact details (phone number, email address or fax number) appear in both texts.
- Level 2, represented by a dotted line: the validity of link is subject to doubt. It is used for instance to represent information related to the localization of the sender (see below).

- Level 3, represented as a dashed line: is an attempt to connect information. It applies for instance when two emails show some degree of proximity through similar scenarios or names appearing in the text.

Variables

The structure of the database first designed in a previous preliminary study (Schiffer et al., 2004) has been simplified. It has been chosen to concentrate on a limited set of attributes (figure 1) that have been assumed to have a good potential for linking:

- phone/fax numbers, address and name of people appearing in the scenario
- IP (Internet Protocol) numbers
- email addresses included in the header and the contents of the message
- dates and time

IP numbers have a technical flavor and their interpretation can cause some specific problems. They are to be further described.

IP Numbers

A unique number is attributed by Internet Service Providers (ISP) to each computer during each connection to the internet. This process assigns an identity to the machine that is technically used for reliably guiding transfers of information. The numbers distributed by each ISP themselves belong to a delineated list received from accredited organisms that organize the management of

IP numbers in the world in order to avoid collisions (different machines connected with the same IP number at the same time on the internet). These numbers are of particular interest from a forensic perspective, because it is theoretically possible to know which ISP a specific machine was connected to, from an IP number of interest collected at a certain time. Then, under certain conditions that depend on legal frameworks in the country where the ISP is located, the ISP can be asked by the justice to retrieve which computer was online during the connection of interest. As the IP number can appear in the header of each email, it is thus theoretically possible to identify the machine that sent the message.

Reliability and availability of IP numbers

Of course, this judicial procedure cannot be part of a systematic process as it is constrained by legal rules not directly applicable to the spam under examination and very complicated to apply when crossing jurisdictions. Thus, IP numbers as a trace will not deploy its full potential within the intelligence process under development. Only information about the ISP and its localization will be used to indicate the probable localization of the sender. But ISP can also sometimes cover wide geographical areas that can divert from the real localization of scammers. Another problem arises when considering IP numbers for linking emails when they are dynamically attributed by the ISP at each connection of a computer to the internet. In these situations,

more often than not, they are not the same for different connections for the same machine.

The reliability of IP number can also be questioned because it is technically very easy to hide or spoof the IP address when sending emails. In these circumstances, the localization and identification of the machine becomes very hard, if not impossible. Examples have shown scammers using this technique extensively.

Finally, in the process of collecting information, the headers of emails are frequently lost through a sequence of forwarding the messages. Thus IP numbers are available only in a limited set of situations. In this study, some IP addresses have been recovered through a specific process out of the scope of this paper (Schiffer et al., 2004). Moreover, in very specific case, combination of different techniques has been used to confirm the reliability of particular IP numbers.

Even under those restrictions, IP numbers stay an useful information within our intelligence process as it will be shown.

IV. Results

Links based on IP numbers

On the 400 messages, 155 have been attributed an IP number. Figure 2 shows examples of the links that have been inferred from this information.

Level 1 links

On the 400 messages, two groups of 7 emails, one group of 4 emails (see figure 3), six groups of 3 emails and 35 groups of 2 emails are found. Mostly, similarities have been also found through email addresses used.

Combined links

Links of different levels have highlighted series of connection between two very different types of scenarios: typical advance fee fraud and lottery games (figure 4).

Activity in Switzerland

A computer has been confirmed being used within Switzerland for sending emails (figure 5). This intelligence could be found with IP number, call number (+41 is the international calling codes for Switzerland) and Swiss domain names (.ch).

V. Discussion

The sample

The chosen sample contains a very small proportion of emails in circulation and concerns only those received in a specific geographical area covered by one police force. It has to be considered that, by its temporal and geographical shape, the sample was likely to contain links. Thus, the potential of a systematic approach cannot be straightforwardly evaluated by the generalization of the obtained results in particular because no link of level 1 could be established between the two periods

separated by one year. However, strategic and operational potential of the method have been clearly illustrated through this study.

Strategic and operational utility of detected links

The linking process has led to the detection of a high quantity of links. They provide a picture of groups of particularly active scammers and have shown that the same groups of perpetrators can use very different scenarios and a certain capability of adaptation. Intelligence about the localization of scammers has also been inferred from the available information, such as indication about the activity of perpetrators within Switzerland.

Exploitation of intelligence

When an activity is detected within a country, the time delay available to react is usually very short for identifying and localizing possible suspects. For the case detected during this study, a preliminary investigation has been carried out by the police in order to verify and consolidate the information obtained. However, due to the fragmentation of police forces and the distribution of competences in Switzerland, the case was not under the competency of this specific police. Moreover, there were hesitations about the possible qualification of the fraud (organized crime or not?) and thus about which structure will prosecute the case. Unfortunately, the time window necessary to launch such a judicial procedure at another level would have been too long to keep a chance of arresting the scammer. This example

emphasizes the need to carefully examine the relationship between the intelligence provided and possible associated measures that can be taken. Intelligence should be fully integrated into the decision process at each level. In this specific punctual example, the criminal justice system was not prepared to take into account such forms of indications.

In a further case, another activity has been detected within the same region, but it has been proven that the scammer was sending emails through the computer network of a school, by exploiting a security hole. This technique was probably used to by-pass anti-spam filters or simply to hide the real origin of the spam. At least, in this case, the intervention resulted rapidly in an update of the network security of this school.

Level of links

Despite the number of solid links detected, knowledge about how to interpret signs of proximity between emails is still lacking. Through the systematization of the process and experience gained by the analysis of specific cases, it can be assumed that the comprehension of links and of how to carry out an efficient comparison processes will iteratively improve. For instance, similarities have been observed between emails addresses of scammers that use mailboxes of very popular internet providers. These resemblances should be further studied in order to better understand some mechanisms related to the automatic generation of new emails accounts that are probably used for massive sending.

Obviously, the formalization of the comparison process will promote the combined use of all available variables. For instance, similarities in emails addresses such as “ejones” and “ericjones” have been confirmed through the comparison of phone numbers, scenarios and the proximity of the dates that indicate when the emails have been sent (see figure 6). Other analysts should have already inferred hypotheses from this information, but very few materials have been published on this subject. There is a need to stimulate exchange of knowledge about how to reason on links. The development of a whole intelligence process will give a chance for creating a community that will share and discuss these experiences.

Perspectives

Results obtained during this study with a very limited sample clearly plead in favor of the systematization of the process. It has been shown that modus operandi used by the scammers can rapidly evolve, but it can be assumed that the basic principle of massively sending unsolicited emails in order to attract people in a trap will stay relatively stable. Thus, the ambition of the project is to build an architecture that covers such situations and integrate for example the “phishing” method or other similar techniques. Within a stable framework, the architecture must be very flexible in order to take into account the evolution of the phenomenon.

The process will be designed around the phases of a very typical intelligence process: collection, collation, analysis and dissemination. Computerization of some part of the process is obviously a major objective in order to be able to treat substantial quantity of emails.

In this perspective, different projects are under development:

- the development of a computerized process that accepts, pre-treats, and collates emails in a database
- the organization of the collection of data with partners
- the classification and indexation through the automatic recognition of scenarios and variables
- the reproduction of similar studies with bigger samples at a wider geographical and temporal scale. Identification of new relevant variables
- the formalization of the comparison process, a better conceptualization of links and the evaluation of the utility of the intelligence provided
- the evaluation of text mining technologies for automatic analysis of the content of the emails
- the automatic recognition of patterns
- the development of graphical interfaces and visualization tools in order to favor data exploration

However, such a process will deploy its whole efficiency only if it is be well integrated into the criminal justice system, in particular:

- intelligence should be part of decisional processes

- analysis at a strategic and operational level will beneficiate from the use of very different types of knowledge, ranging from understanding about organized crime to the capacity of reasoning on digital traces.

VI. Conclusion

This exploratory study has shown the interest of systematically analyzing advance fee fraud's emails in an intelligence perspective. The analysis of a limited sample has shown great promises by the detection of links, the localization of scammers and the provision of indications about the evolution of the phenomenon. For such a system to be useful, it has also been demonstrated the necessity for a global approach that integrates intelligence within strategic and operative decisional process.

There is still a lack of formalization. Links should be further described and understood, as well as the comparison process that proceeds through the combination of the use of the different variables.

A flexible architecture for the whole process is under development through a series of preliminary projects for evaluating the degree of computerization that can be reasonably reached and the potential of new data mining techniques.

The development of such an intelligence process could promote knowledge sharing and improve interpretation models in an iterative way.

VII. References

Bécherraz, G.-M. (2004a, 08.09.2004). "Nigerian Connection" sur le gril. 24 Heures.

Bécherraz, G.-M. (2004b, 15.09.2004). "Nigerian Connection": escroc blanchisseur condamné. 24 Heures.

Edelson, E. (2003). The 419 scam: information warfare on the spam front and a proposal for local filtering. Computers & Security, 22(5), 392-401.

National White Collar Crime Center and the Federal Bureau of Investigation. (2002). IFCC 2002 Internet Fraud Report.

Office Fédéral de la police. (2003). Rapport sur la sécurité intérieure de la Suisse.

ONU. (2005). La criminalité transnationale organisée dans la région de l'afrique de l'ouest.Unpublished manuscript, New York.

Oriola, T. A. (2005). Advance fee fraud on the Internet: Nigeria's regulatory response. Computer Law & Security Report, 21, 237-248.

Schiffer, B., Birrer, S., Cartier, J., Capt, S., & Ribaux, O. (2004). Analyse de la forme, du contenu et de la provenance des courriers électroniques. Revue internationale de criminologie et de police technique et scientifique(2), 148-158.

Smith, R. G., Holmes, M. N., & Kaufmann, P. (1999). Nigerian Advance Fee Fraud. Trends & Issues in Crime and Criminal Justice(121).

Viosca, C., Bergiel, B. J., & Balsmeier, P. (2004). Effects of the Electronic Nigerian Money Fraud on the Brand Equity of Nigeria and Africa. Management Research News, 27(6), 11-20.

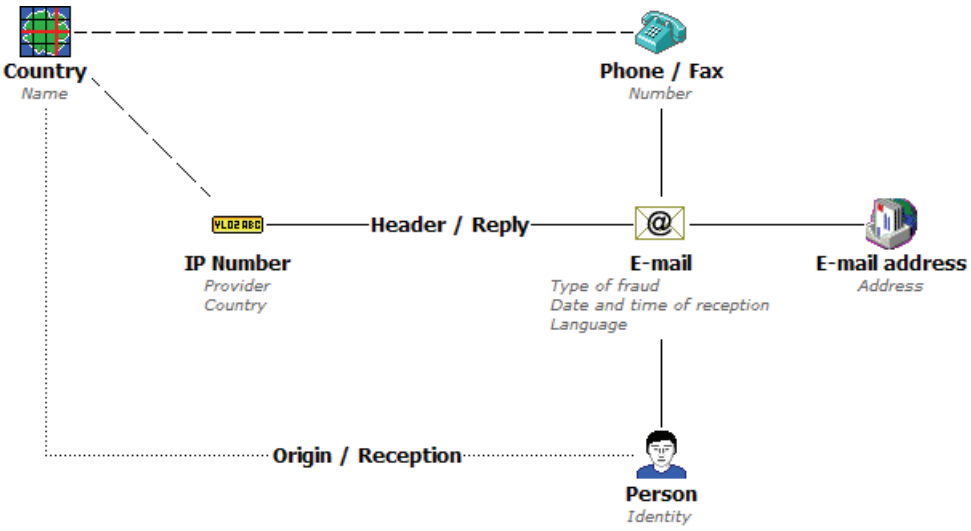


Figure 1 : Architecture of database

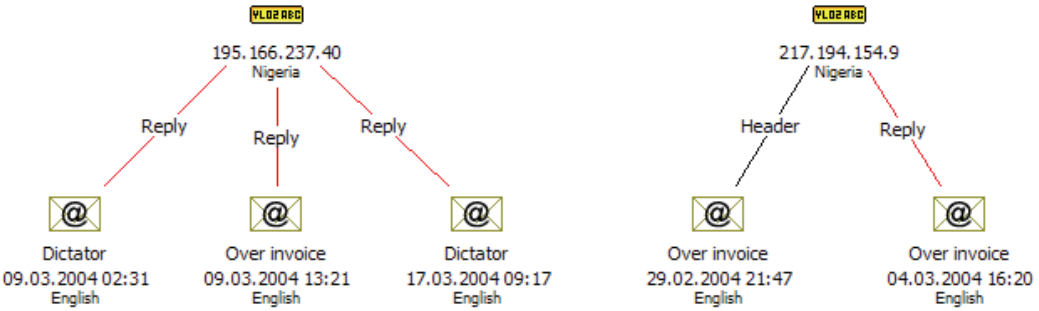


Figure 2 : Same IP number with several linked E-mails

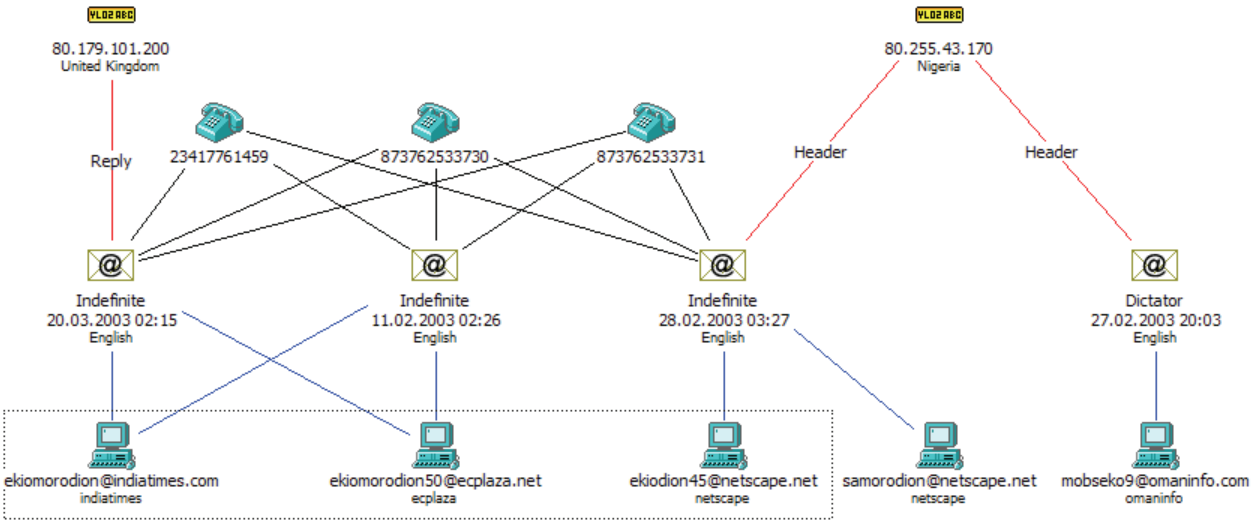


Figure 3 : Links of level 1

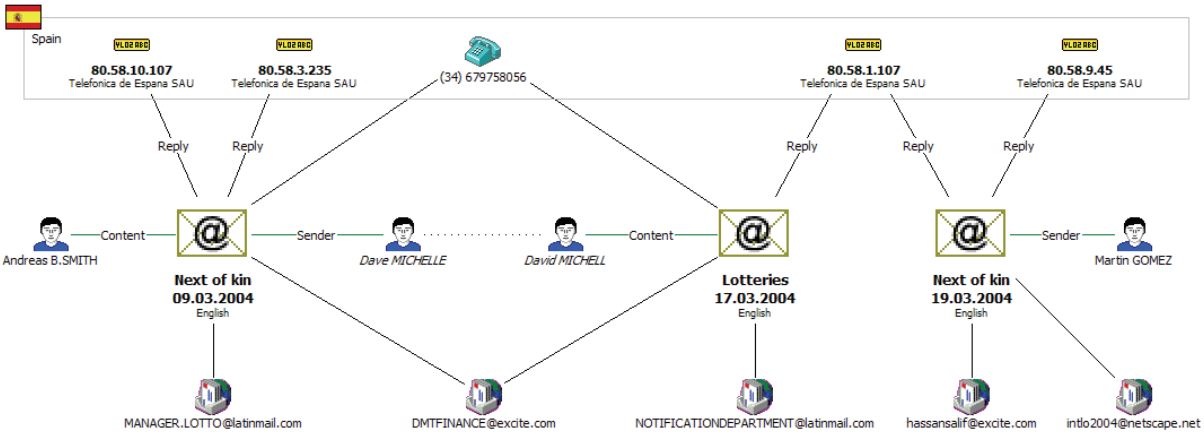


Figure 4 : Links between typical advance fee fraud and lottery game

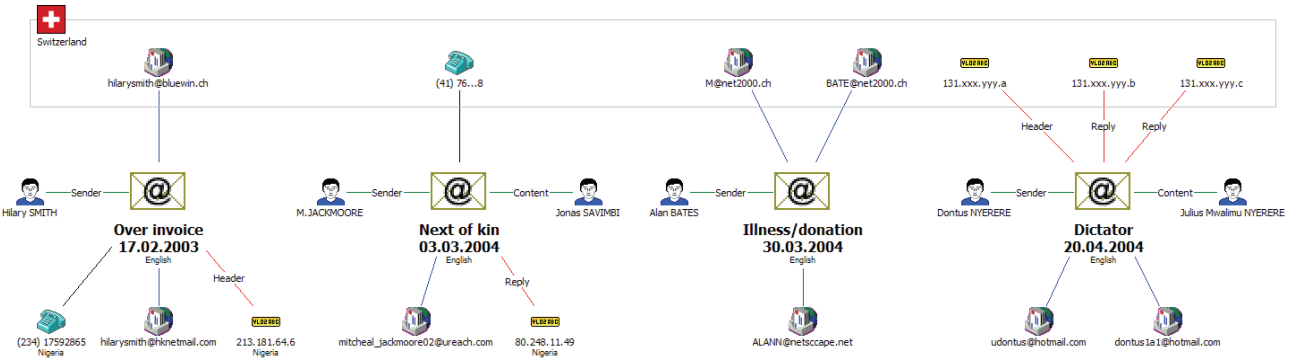


Figure 5 : E-mails linked with Switzerland

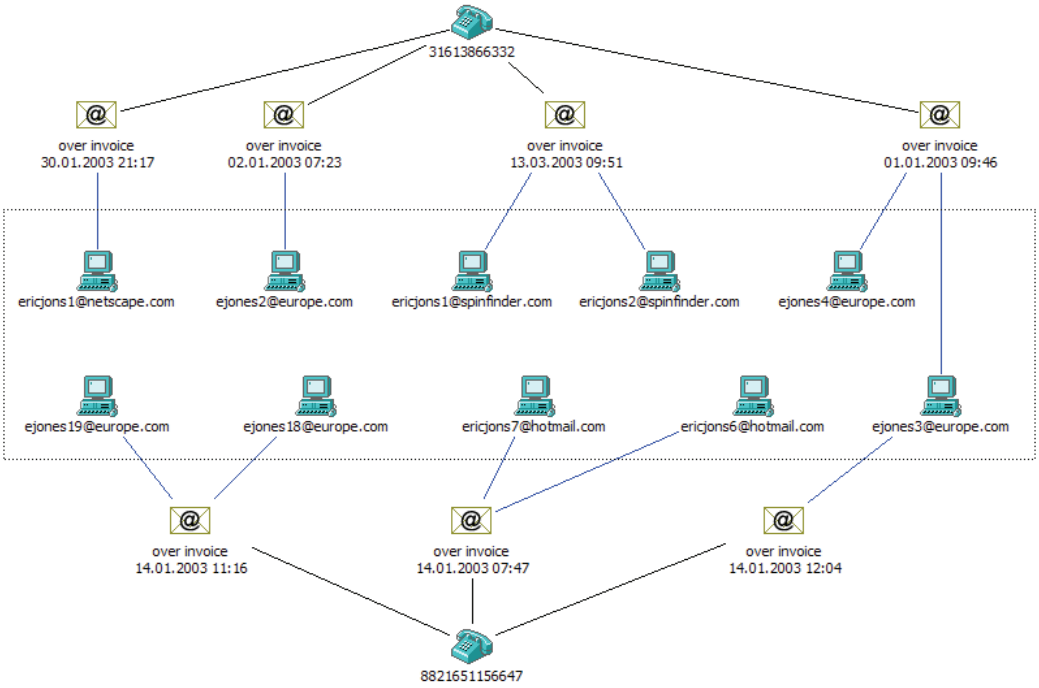
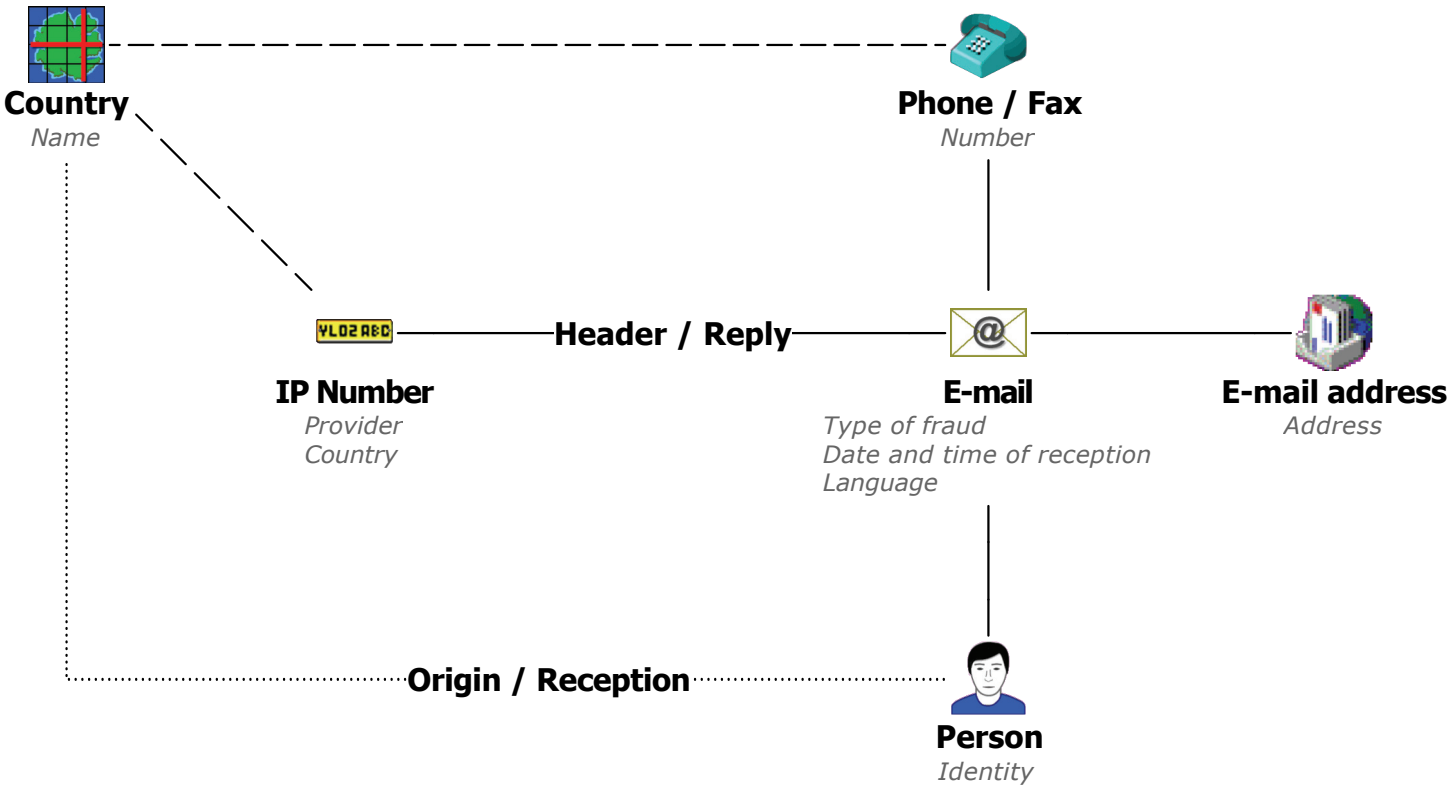
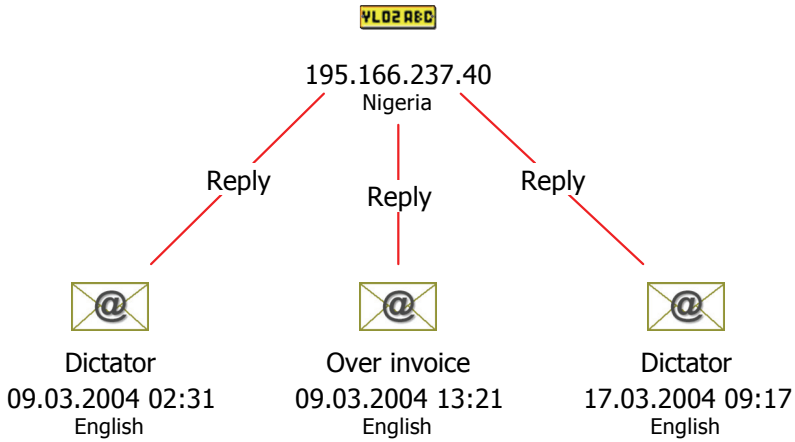
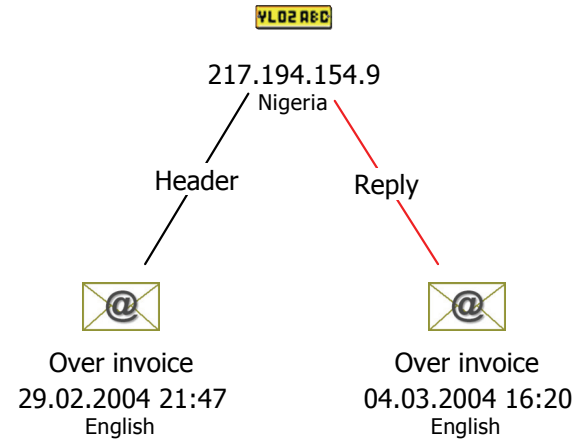
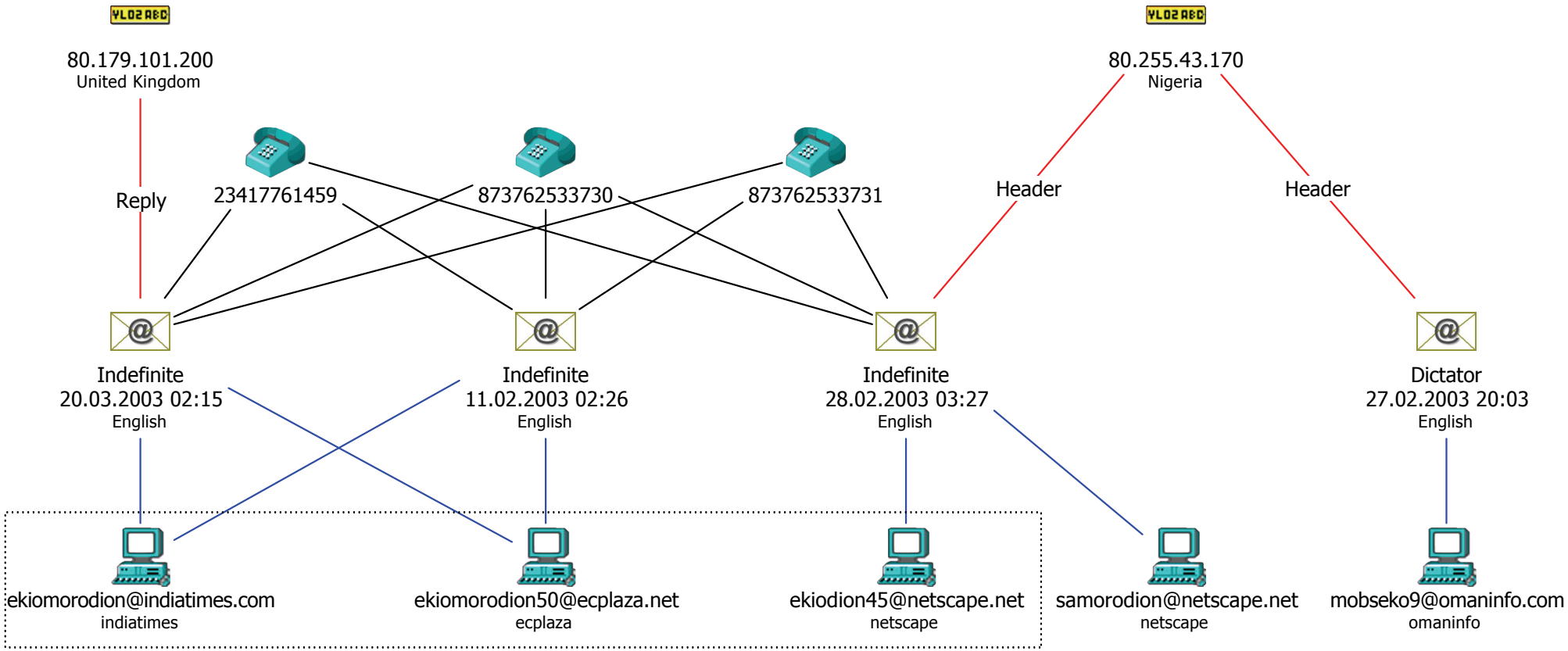


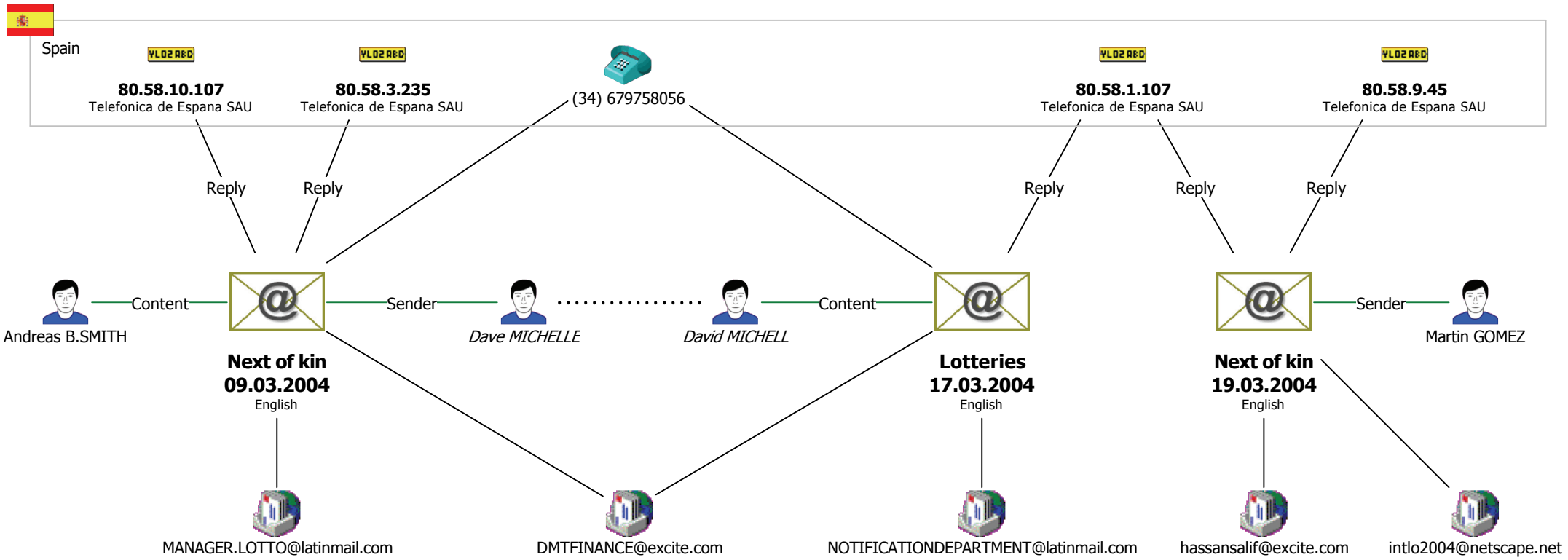
Figure 6 : E-mails addresses as link information

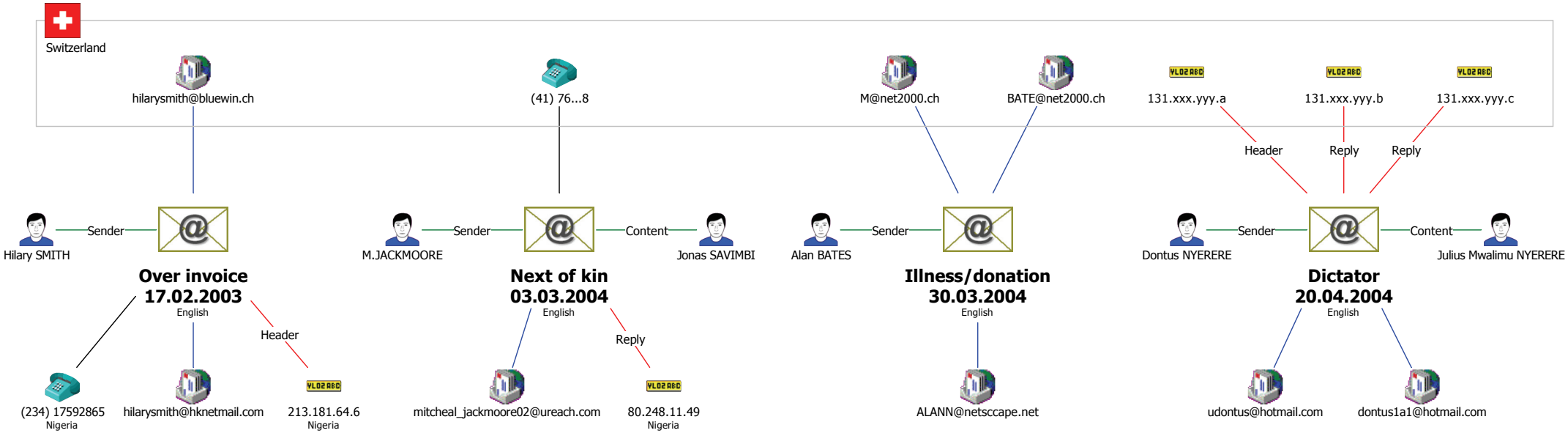

Country
Name
Phone / Fax
Number
YL02 ABC
Header / Reply
IP Number
Provider
Country
E-mail
Type of fraud
Date and time of reception
Language
E-mail address
Address
Origin / Reception
Person
Identity

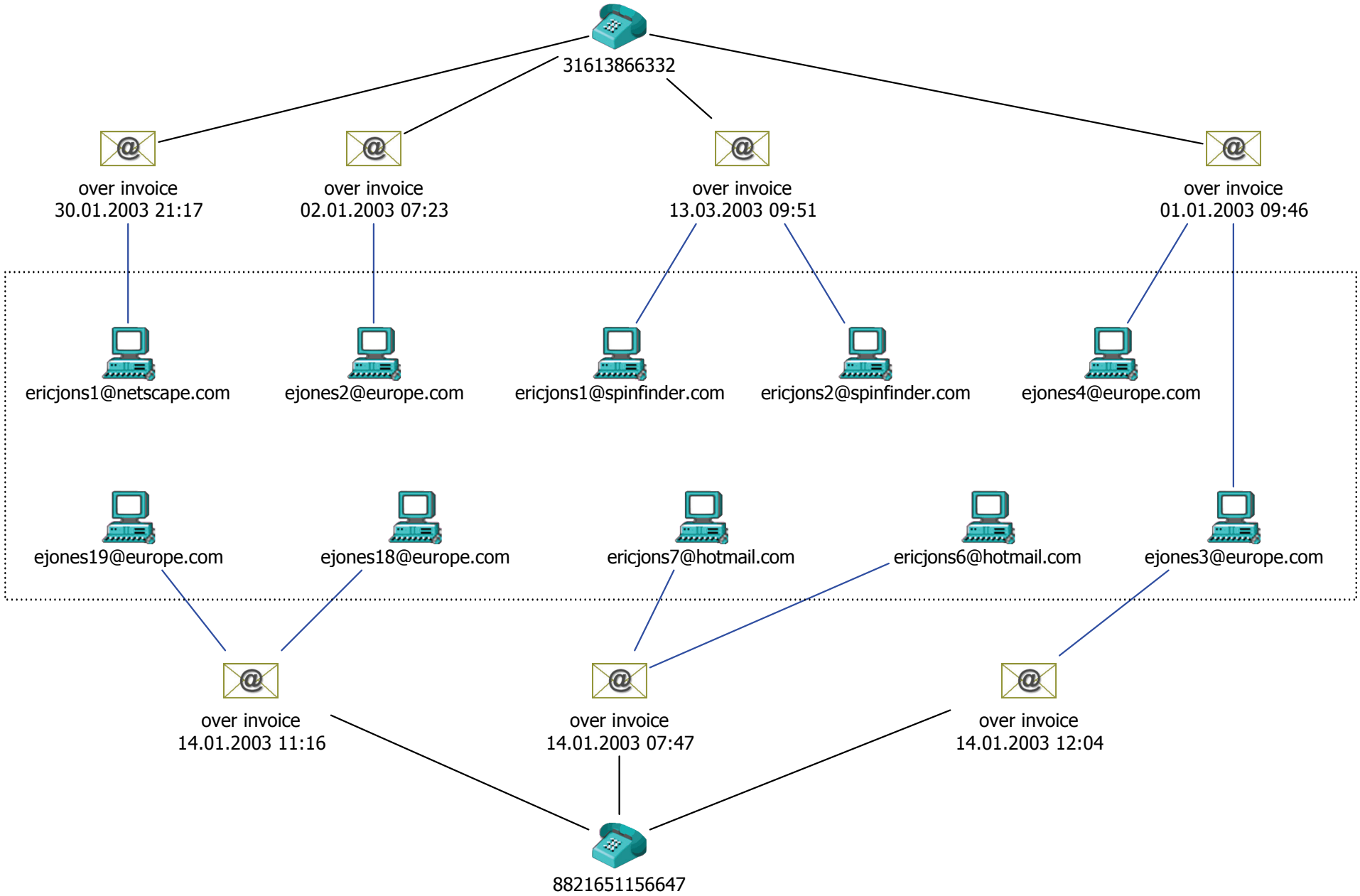

195.166.237.40
Nigeria
Reply
Reply
Reply
Dictator
09.03.2004 02:31
English
Over invoice
09.03.2004 13:21
English
Dictator
17.03.2004 09:17
English


217.194.154.9
Nigeria
Header
Reply
Over invoice
29.02.2004 21:47
English
Over invoice
04.03.2004 16:20
English


80.179.101.200
United Kingdom
80.255.43.170
Nigeria
Reply
23417761459
873762533730
873762533731
Header
Header
Indefinite
20.03.2003 02:15
English
Indefinite
11.02.2003 02:26
English
Indefinite
28.02.2003 03:27
English
Dictator
27.02.2003 20:03
English
ekiomorodion@indiatimes.com
indiatimes
ekiomorodion50@ecplaza.net
ecplaza
ekiodion45@netscape.net
netscape
samorodion@netscape.net
netscape
mobseko9@omaninfo.com
omaninfo


Spain
80.58.10.107
Telefonica de Espana SAU
80.58.3.235
Telefonica de Espana SAU
(34) 679758056
80.58.1.107
Telefonica de Espana SAU
80.58.9.45
Telefonica de Espana SAU
Reply
Reply
Reply
Reply
Reply
Content
Sender
Content
Sender
Andreas B.SMITH
Dave MICHELLE
David MICHELL
Martin GOMEZ
Next of kin
09.03.2004
English
Lotteries
17.03.2004
English
Next of kin
19.03.2004
English
MANAGER.LOTTO@latinmail.com
DMTFINANCE@excite.com
NOTIFICATIONDEPARTMENT@latinmail.com
hassansalif@excite.com
intlo2004@netscape.net


Switzerland
hilarysmith@bluewin.ch
(41) 76...8
M@net2000.ch
BATE@net2000.ch
131.xxx.yyy.a
131.xxx.yyy.b
131.xxx.yyy.c
Header
Reply
Reply
Hilary SMITH
Sender
M.JACKMOORE
Sender
Content
Jonas SAVIMBI
Alan BATES
Sender
Dontus NYERERE
Sender
Content
Julius Mwalimu NYERERE
Over invoice
17.02.2003
English
Next of kin
03.03.2004
English
Illness/donation
30.03.2004
English
Dictator
20.04.2004
English
Header
Reply
(234) 17592865
Nigeria
hilarysmith@hknetmail.com
213.181.64.6
Nigeria
mitcheal_jackmoore02@ureach.com
80.248.11.49
Nigeria
ALANN@netsccape.net
udontus@hotmail.com
dontus1a1@hotmail.com


31613866332
over invoice
30.01.2003 21:17
over invoice
02.01.2003 07:23
over invoice
13.03.2003 09:51
over invoice
01.01.2003 09:46
ericjons1@netscape.com
ejones2@europe.com
ericjons1@spinfinder.com
ericjons2@spinfinder.com
ejones4@europe.com
ejones19@europe.com
ejones18@europe.com
ericjons7@hotmail.com
ericjons6@hotmail.com
ejones3@europe.com
over invoice
14.01.2003 11:16
over invoice
14.01.2003 07:47
over invoice
14.01.2003 12:04
8821651156647