The role of phone numbers in understanding cyber-crime

A. Costin * J. Isachenkova * M. Balduzzi +
A. Francillon * D. Balzarotti *

*Eurecom, Sophia Antipolis, France

+Trend Micro Research, EMEA

July 11, 2013

Introduction

Online/digital identifiers in cyber-crime

- Mail
- Domain name/Web site
- Social networks/Nicknames profile
  - Extensive studies: [LMK+10, TGM+11, KKL+08, Ede03, CHMS06]
- Phone numbers
  - Limited studies: [CYK10, STHB99, Pol05, Hyp]
  - Studied mainly in context of premium short number mobile frauds
  - *Our main focus*

Introduction

Phone number usages

- Mail signatures
- Extensively used in many businesses
- Offers less anonymization than other identifiers
- Links cyber domain to reality domain
- Commonly used in various online frauds, e.g.:
  - Premium numbers fraud
  - Scam fraud

Introduction

Importance of cyber-crime and phone numbers – example

- Banking *Trojan.Shylock* [symb]
- Injects code into banking websites
- Replaces telephone details into the contact pages of online banking websites

Injected Numbers		Original Numbers	
Calling from the UK	Calling from abroad	Calling from the UK	Calling from abroad
0800 310	+44 8705	08457	+44 8705
0800 310	+44 118 9	0845 3	+44 118 9
0800 310	+44 8705	08457	+44 8705

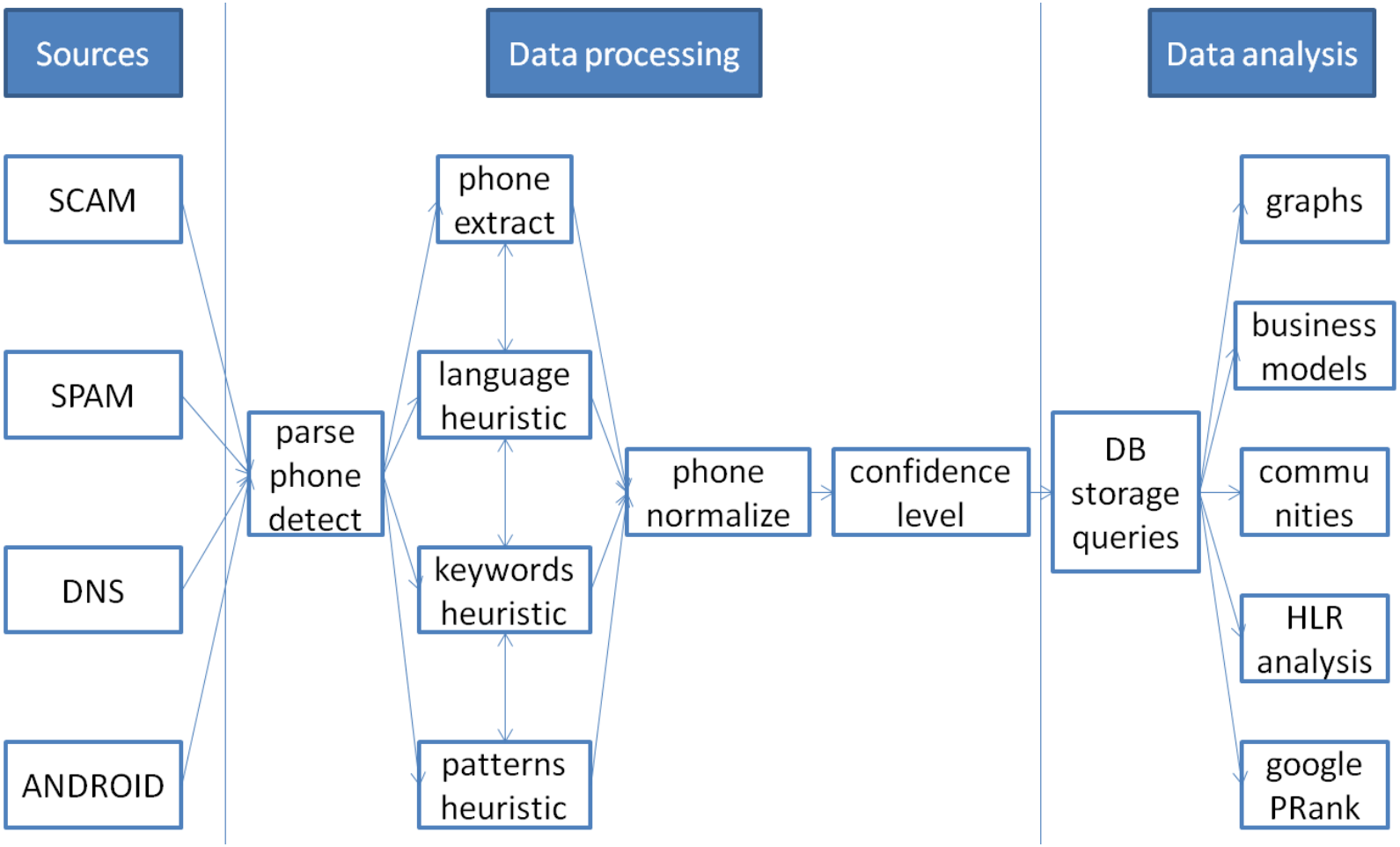
Hypothesis

- Phone numbers are used in cyber-crime activities
- Can we find telecom operators preference?
- Can we find geographical preference?
- Phone numbers can be a stronger identification metric vs. other identifiers

Goals

- Check those hypothesis against real data-sets
- Evaluate the reliability of automated phone numbers extraction and analysis
  - Identify challenges and limitations
- Automatically find patterns associated with recurrent criminal activities
- Automatically correlate the extracted information for
  - Telecom operator preference
  - Geographic area preference

Methodology



Datasets I

Data Sources initially considered

- SPAM
  - Large and extremely noisy dataset
  - Extremely challenging to extract and clean phone numbers
- WHOIS
  - Focused on malicious domains
  - High quality dataset (intl. format)
  - Phone numbers are dummy or replaced by CERTs' contact numbers

Datasets II

- ANDROID
  - Small and noisy dataset
  - Mainly contained *short premium numbers* – open problem
- SCAM
  - Large and high quality dataset
  - Phone numbers are an important part of business model
  - Focus on this dataset

Phone Number Extraction I

Success and Reliability of Extraction depend on

- How well formatted the number is
  - *Call: 0336 9505705 9 am - 5 pm*
  - Can be decoded as 2 valid numbers: *+44336950570*5 or *+33695057059*
  - We aim at obtaining:
    - Non ambiguous normalized number
    - Fully qualified international format number

Phone Number Extraction II

- How structured and easy to parse the information is
  - WHOIS records (easy) vs.
  - Malicious mobile binary (difficult)
- How noisy the data source is
  - Spam messages are very noisy (to defeat anti-spam filters)
  - Scam messages have almost no noise

Phone Number Extraction Challenges

Example Number obfuscation used [syma]

(Ч^95)1 ^2^ три ^40^ ОО
(495) один 2 3 – 4 0 – 00 } (495)123400

/495/ Ч;3 =2 I 00 0
(Ч~9~5) 43~2~ один~о~о~0 } (495)432100

(Москва) 1 ^2^ три ^40^ ОО
(Moscow) один 2 3 – 4 0 – 00 } (495)123400

(MOW) 4~3~2~1~0~ 00
(Мос) четерь 3 2;I=0О ноль } (495)432100

	English	Russian
1	I i l	N/A
2	Zz	N/A
3	N/A	ЗзЭэ
4	N/A	Чч
6	N/A	ЬьБб
0	Oo	Оо

	Russian	English
1	один	one
2	два	two
3	три	three
4	Четыре	four
5	пять	five
6	шесть	six
7	семь	seven
8	восемь	eight
9	девять	nine
0	ноль	ten

Scam Message Sample

BUSINESS PROPOSAL - Netscape Message

File Edit View Go Message Communicator Help

Get Msg | New Msg | Reply | Reply All | Forward | File | Next | Print

Subject: BUSINESS PROPOSAL
Date: Tue, 26 Jun 2001 22:13:10 -0700 (PDT)
From: emmanuel udo <emmanuel_z2@yahoo.com>
To: tim.richardson@senecac.on.ca

FROM: DR, EMMANUEL UDO.
TEL: 234 1 759 1549; FAX: 234 1 759 0379.
E-MAIL:emmanuel_z2@yahoo.com
BUSINESS PROPOSAL

ATTN: PRESIDENT / CEO,

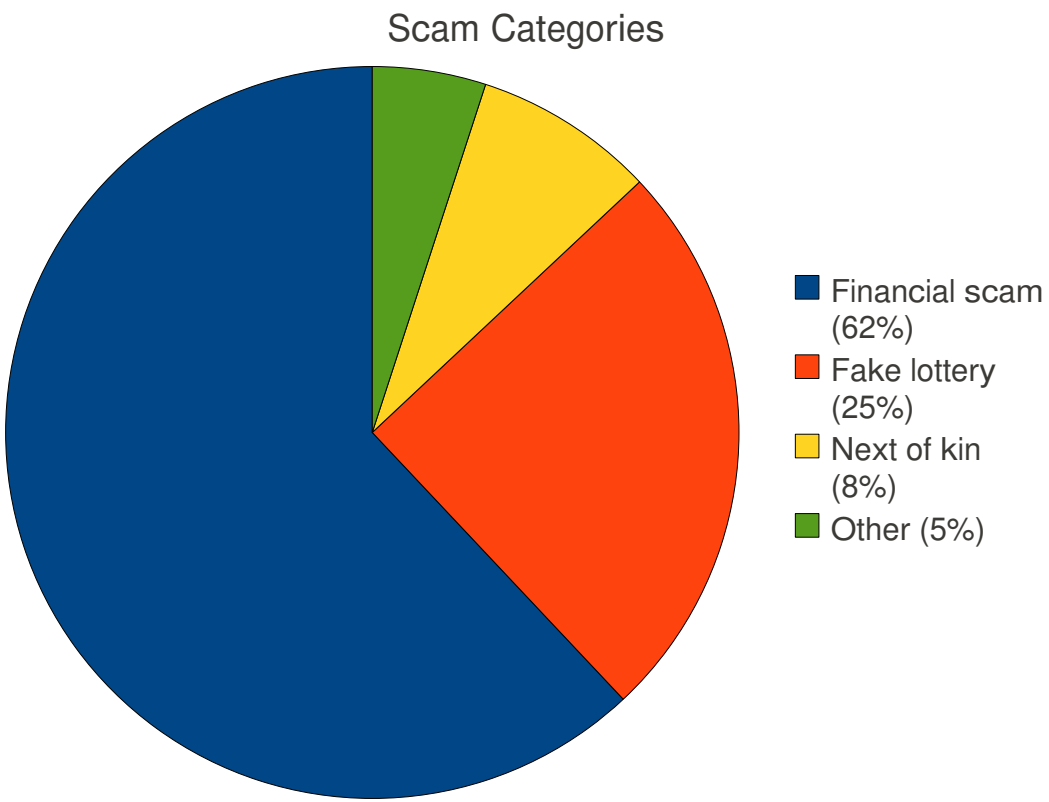
My name is EMMANUEL UDO, a member of the Presidential Task Force on Oil Spillage Clean-up. Early last year there was a major oil spillage in the Niger Delta Region of Nigeria which rendered over 70% of the communities homeless.The contract was handled by a foreign firm but because of the huge monetary profit we envisaged we decided to over-invoice the contract sum. Now the contract has been completed and the original contractor has since been paid,but the contract balance of US$38 million,which resulted from the over invoiced contract sum that has been left in a suspense account with the CENTRAL BANK of NIGERIA,is what me and my partners are planning to take out of

SCAM Dataset

- Used *user reports aggregator* 419scam
- Data timespan: January 2009 – August 2012
- Enriched and correlated with numbering plans (NNPC) databases
  - Free (*libphonenumber*)
  - Commercial (more detailed and updated)

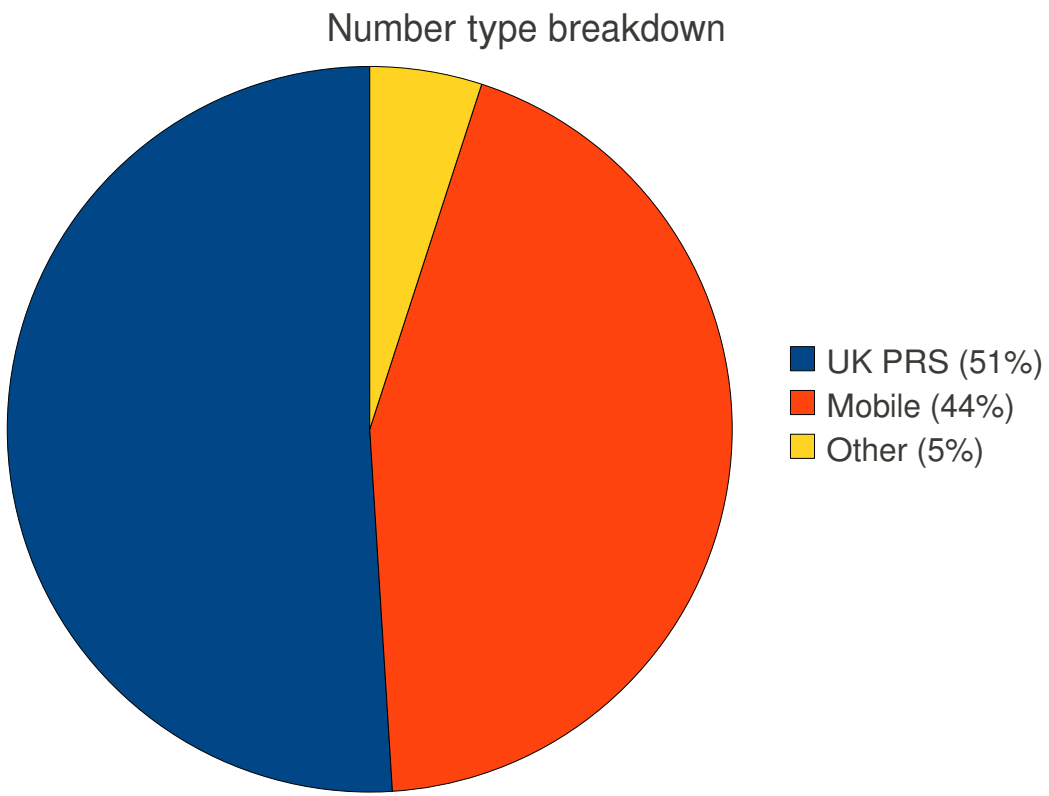
SCAM Email Categories

- Emails classified in 10 categories
- 3 categories cover over 90% of the data



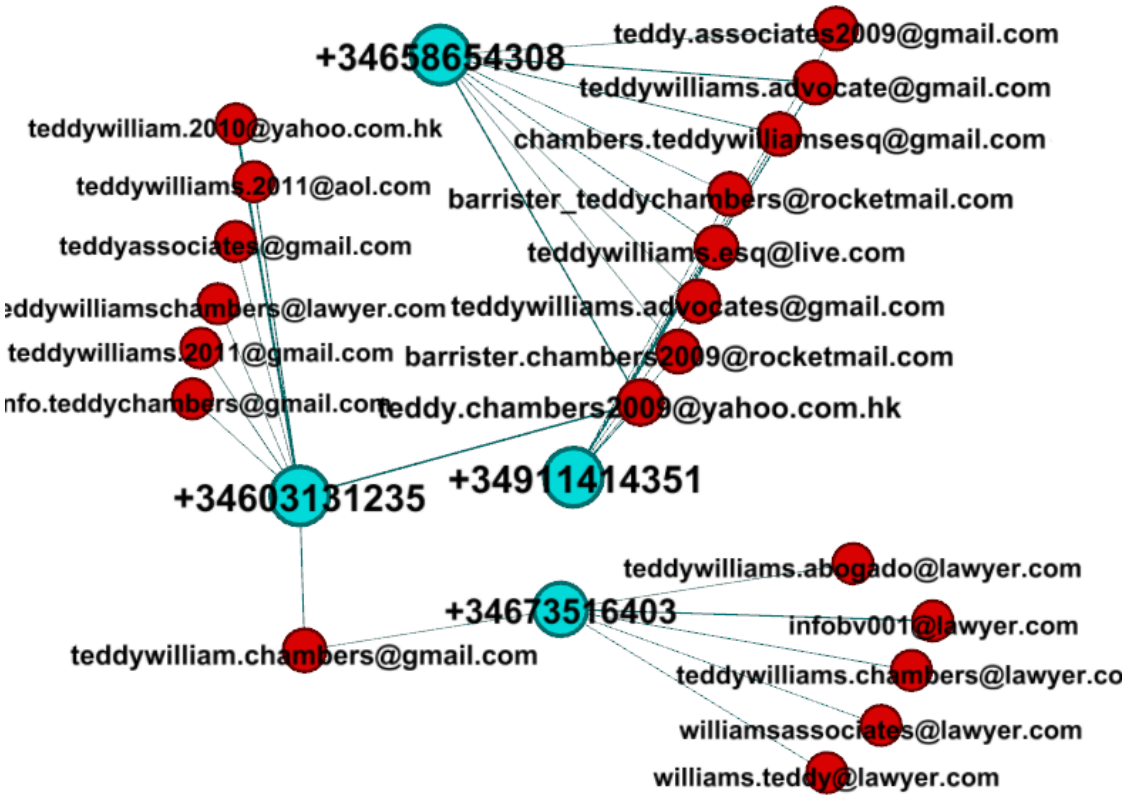
SCAM Phones Categories

- ~67k unique normalized phone numbers
- Classified using numbering plans (NNPC) databases



SCAM Communities/Identity Links

- Used clustering techniques, discovered *identity links*
- Identified 102 communities
- Supports the hypothesis that phone numbers are a good metric to study scammers



ANALYSIS OF MOBILE PHONE NUMBERS

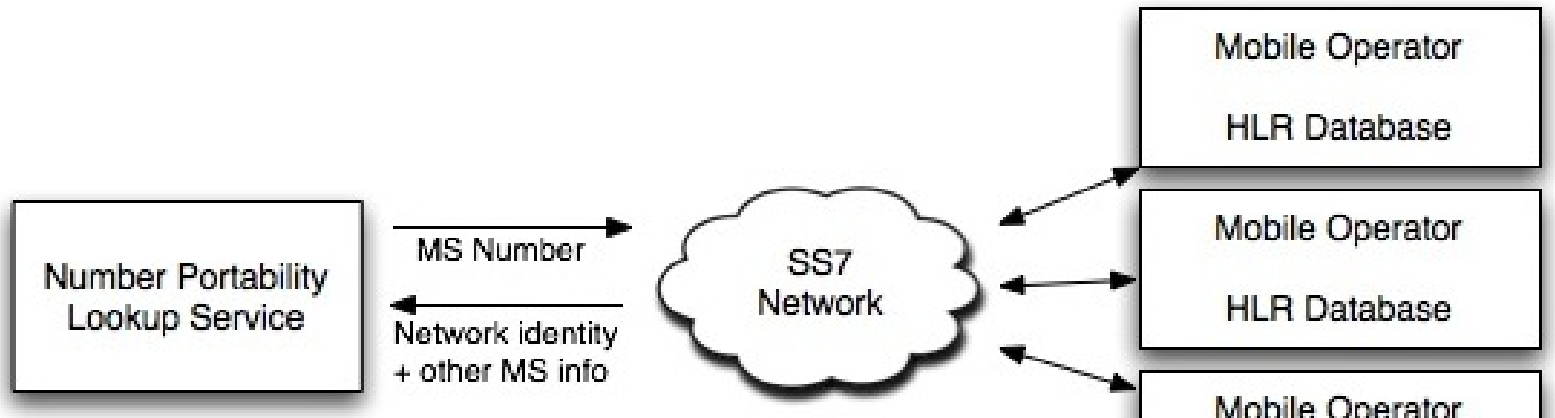
Questions and Hypothesis

- For how long are phone numbers used?
- Are phone numbers reused or discarded?
- If discarded, after how long?
- Are phone numbers used in roaming?
- If roaming, to which extent?

We try to answer these questions with *HLR queries*

HLR Querying

- HLR=Home Location Register
- Important component of Mobile Network Operators

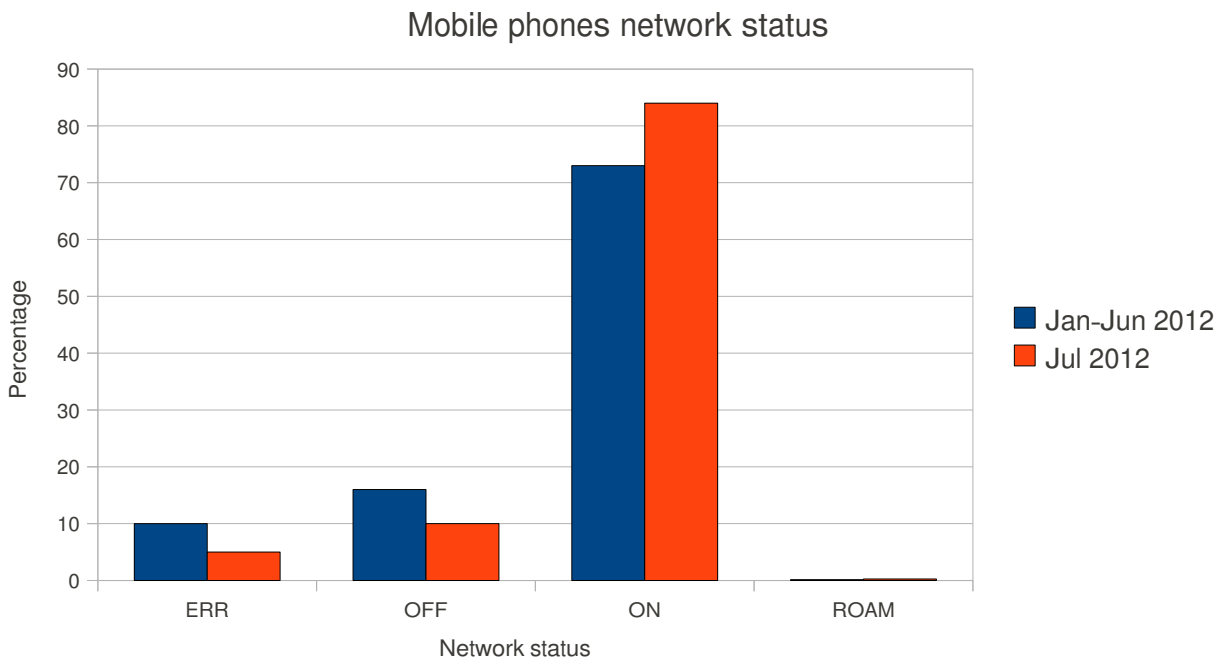


HLR Query - Diagram Copyright Comcetera Ltd, 2009
NumberPortabilityLookup.com

Single HLR Queries

In Aug 2012, querying once for all mobiles encountered in :

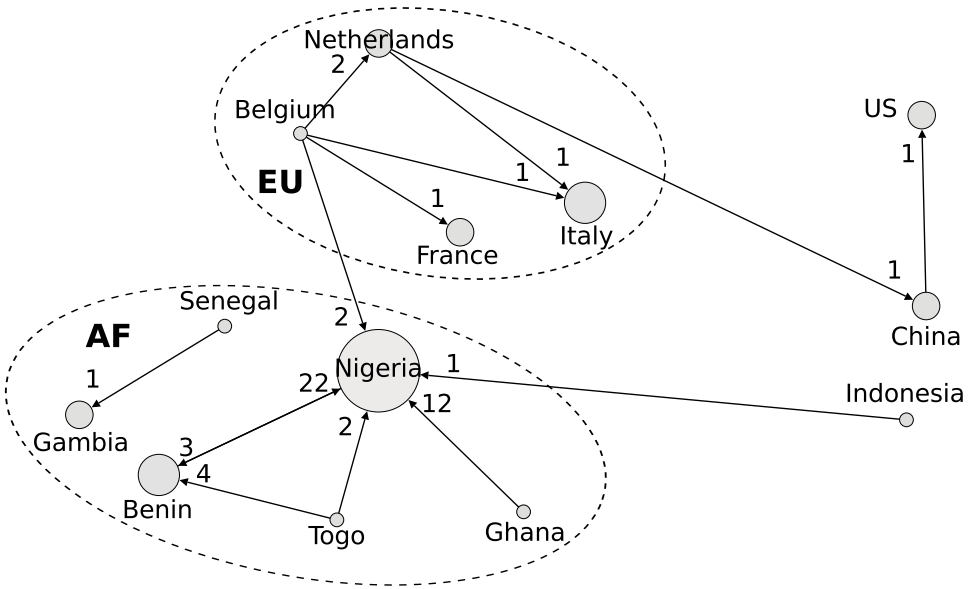
- Jan – Jun 2012
- Jul 2012



Repeated HLR Queries

Performed HLR queries

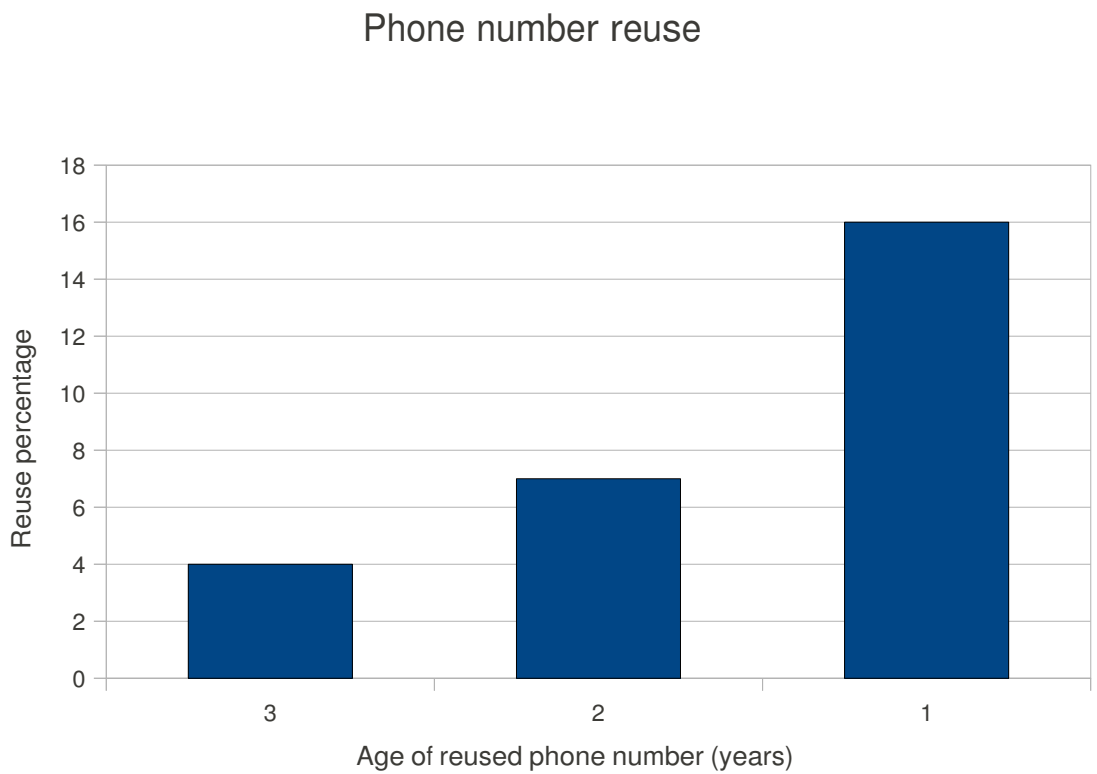
- For 1400 numbers
- Every 3 days
- During Jul – Aug 2012
- Hypothesis1: Possibility of a link with the Nigerian groups
- Hypothesis2: May be used to conceal location



Phone Numbers Reuse

Question:

- For how long a scam number is used?



ANALYSIS OF UK PRS PHONE NUMBERS

What are UK PRS numbers? I

Definition

- Premium rate services (PRS) are a form of micro-payment for paid content, data services and value added services that are subsequently charged to user phone bill
- UK PRS is a 800 Mil. GBP bussines (2009)

What are UK PRS numbers? II

Usages

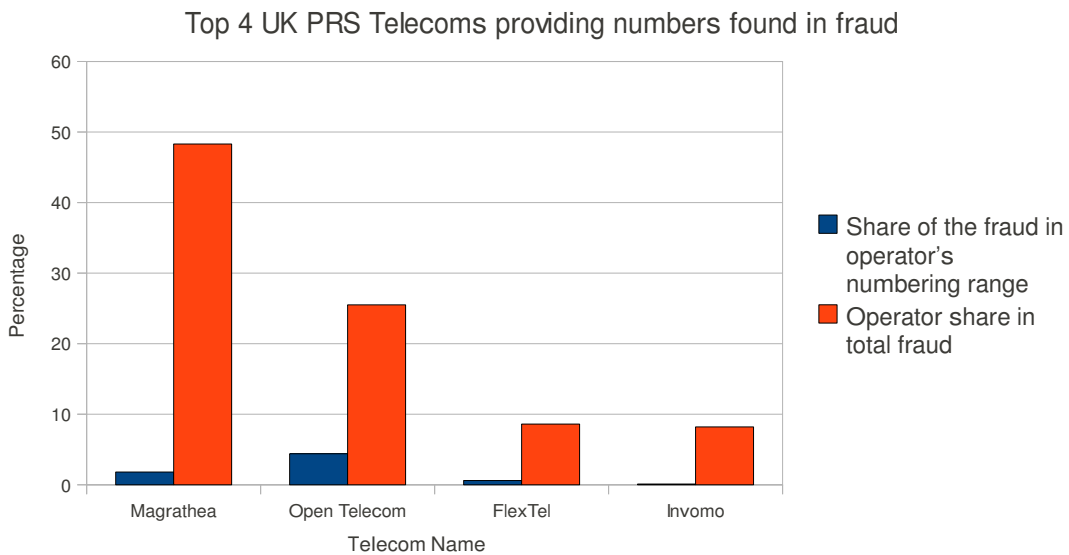
- Conceal geographic location of real phone, via *call forwarding*
- Earn revenue from calls to these numbers

Challenges

- Hard to trace the ”service provider”
- Hard to trace the real phone number behind forwarding
- Hard to detect or prove that fraud is involved

Range of UK PRS numbers

- ~34k unique phone numbers in UK range of 07x Premium Rate Services numbers
- 4 operators (out of 88) provide more than 90% of fraud-related UK PRS numbers
- ~5% of one operator allocated range is fraud-related



Conclusion – Results

- Phone numbers are a strong digital identifier in some cyber-crime activities
- Phone numbers help in automated scammer community detection
- HLR lookups help
  - in identifying identify recurrent cyber-criminal business models
  - to study phone numbers' geographical use and activity patterns

Conclusion – Future Work

- Phone number extraction is an open, non-trivial problem
  - Improve matching algorithms and their context-awareness
- PRS phone numbers are opaque
  - is a ”traceroute” of PRS phone numbers possible?
  - learn business models behind them
- Short number extraction and evaluation
  - Open and challenging, non-trivial problem
  - Becomes a growing concern with mobile malware

Questions?

Contacts:

- Software and System Security Group @ EURECOM
- S3.eurecom.fr

Thank you!

References I


Duncan Cook, Jacky Hartnett, Kevin Manderson, and Joel Scanlan, *Catching spam before it arrives: domain specific dynamic blacklists*, Proceedings of the 2006 Australasian workshops on Grid computing and e-research, ACSW Frontiers '06, vol. 54, 2006.


Nicolas Christin, Sally S. Yanagihara, and Keisuke Kamataki, *Dissecting one click frauds*, CCS '10, ACM, 2010.


Eve Edelson, *The 419 scam: information warfare on the spam front and a proposal for local filtering.*, Computers & Security **22** (2003), no. 5.

References II


Mikko Hypponen, *Malware Goes Mobile*, http://www.cs.virginia.edu/~robins/Malware_Goes_Mobile.pdf.


Christian Kreibich, Chris Kanich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage, *On the spam campaign trail*, LEET'08, 2008.


Olumide B. Longe, Victor Mbarika, M. Kourouma, F. Wada, and R. Isabalija, *Seeing beyond the surface, understanding and tracking fraudulent cyber activities*, CoRR **abs/1001.1993** (2010).

References III


Craig Pollard, *Telecom fraud: Telecom fraud: the cost of doing nothing just went up*, Network Security **2005** (2005), no. 2.


J. Shawe-Taylor, K. Howker, and P. Burge, *Detection of fraud in mobile telecommunications*, Information Security Technical Report **4** (1999), no. 1.


Evolution of Russian Phone Number Spam, http://www.symantec.com/connect/blogs/revolution-russian-phone-number-spam.

References IV

Trojan.Shylock Injects Phone Numbers into Online Banking Websites, http://www.symantec.com/connect/blogs/merchant-malice-trojanshylock-injects-phone-numbe

Kurt Thomas, Chris Grier, Justin Ma, Vern Paxson, and Dawn Song, *Design and evaluation of a real-time url spam filtering service*, Proceedings of the 2011 IEEE Symposium on Security and Privacy, 2011.