Нигеријска превара

Рачунарство и друштво

Математички факултет

Универзитет у Београду

Павле Савић

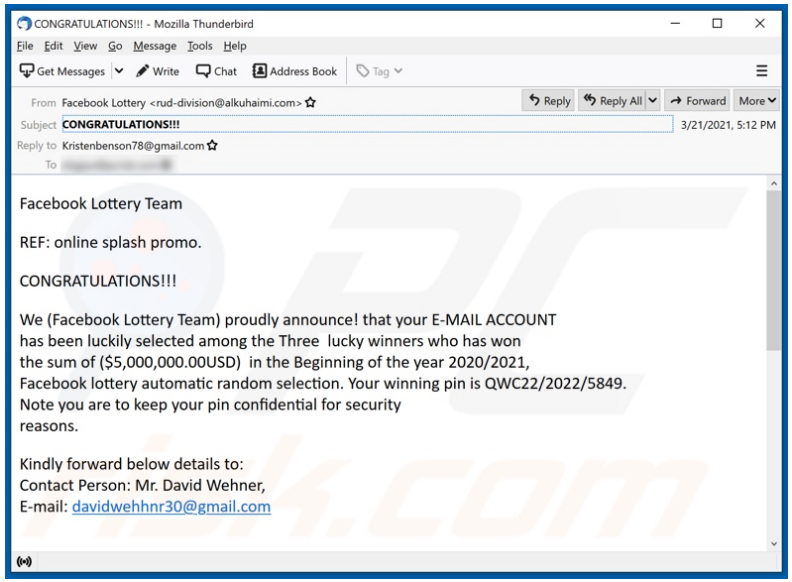
mi17169@alas.matf.bg.ac.rs

Београд, 2021.

Увод

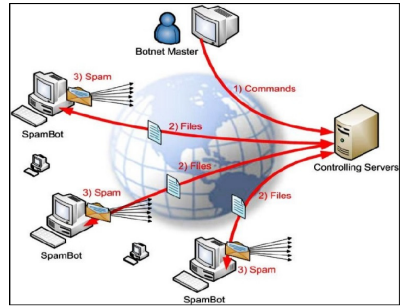
- Превара 419 и нигеријски кривични законик
- Авансне провизије, лажне лутрије, црни новац
- Пошта -> Факс -> Е-пошта
- 419scam.org
- Различита у односу на нежељену пошту (eng. *spam*)
- 59% нежељене поште садржи *URL*-адресе
- Избегавање традиционалних *spam* филтера е-поште
- 8% нежељене поште чиниле су Преваре 419 (2009-2014)

Слика 1: Црни новац



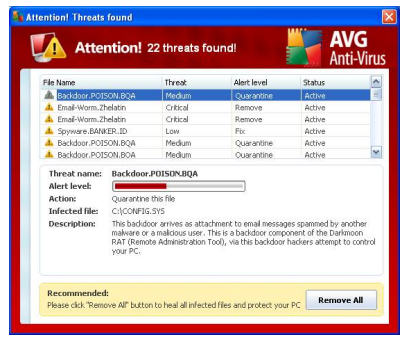
Слика 2: Лажна лутрија

- Употреба телефонских бројева током дужег периода потенцијално шанса за откривање *скамера*
- *Кластеровање* - омогућује повезивање мејлова који деле заједничке особине, потенцијално припадају истој кампањи (?)
- Телефонски бројеви - камен темељац приликом повезивања
- *Macro-cluster*-и потенцијално указују на кампање већег обима
- *Кластеровање* се експериментално показало као успешно и за истраге других шема сајбер криминала



Слика 3: Ботнет

Слика 4: Циљани напад



Слика 5: Лажни антивирус

Скуп података

- *419scam.org - a 419 scam aggregator* (јануар 2009-август 2012.) претпроцесирани подаци
- Телефонски бројеви могу означавати географску локацију
- Бројеви мобилних телефона коришћени за нигеријску превару прецизно указују на државу пребивалишта нападача, пронађено неколико случајева роминга

Табела 1: Опште статистике

Опис	**Број**
Скам поруке	36.761
Јединствене поруке	26.250
Укупно мејл адреса	112.961
Јединствене мејл адресе	34.723
Укупно телефонских бројева	41.320
Јединствени телефонски бројеви	11.768
Број држава	12

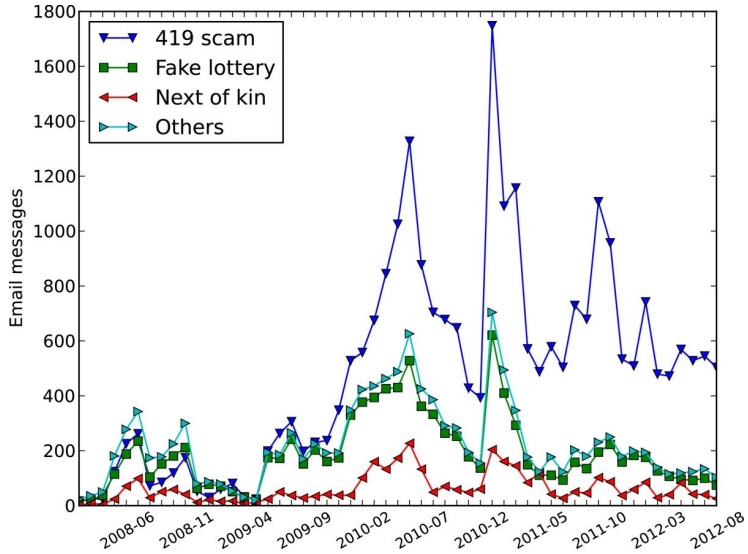
- Број адреса е-поште три пута већи од броја телефонских бројева
- Свака порука у скупу може садржати до 5 адреса е-адреса (*from*, *reply*, и адресе наведене у тексту поруке)
- Дистрибуција уједначена у посматраном трогодишњем периоду
- Скуп ограничен на европске и афричке регионе
- 71% адреса е-поште коришћено је само током једног дана, преостале у просеку 79 дана
- 51% телефонских бројева коришћено је само током једног дана, остатак у просеку 174 дана
- Телефонски број значајна особина приликом *кластеровања*

Табела 2: Телефони по државама

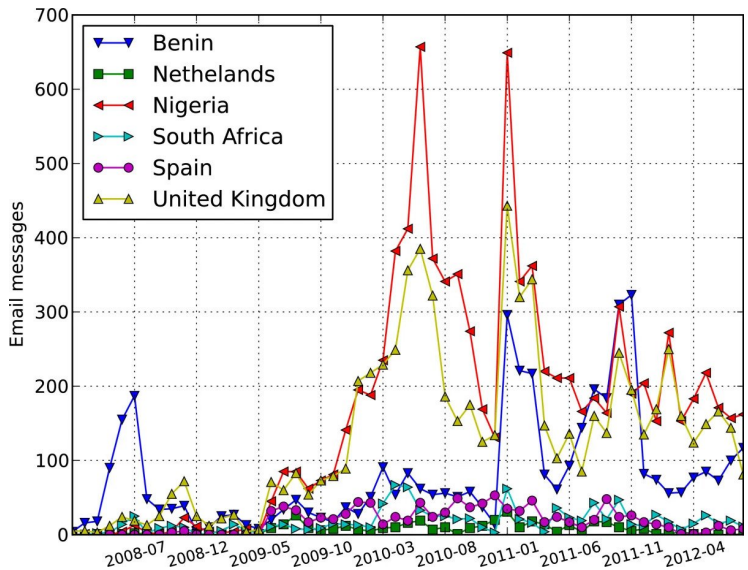
Држава	**Укупно телефона**	**Укупно (%)**
Уједињено Краљевство	4.499	43
Нигерија	3.121	30
Бенин	1.448	14
Јужна Африка	562	5
Шпанија	372	4
Холандија	263	3
Обала Слоноваче	89	1
Кина	68	1
Сенегал	47	0.5
Того	11	0.1
Индонезија	1	0.01

- Сви бројеви из УК у скупу података припадају *personal numbering services*
- Укупно у скупу 44% *personal numbering services* бројева, 44 % бројева мобилних телефона, 12% фиксних линија, мање од 1% су непостојећи бројеви

- Порукама које су препознате као 419 превара се додељује и поткатегорија
- Категоризација врсте преваре заснива се на фреквенцији речи у порукама
- 64% сврстано у категорију *419 scam*, 24% у категорију *Fake lottery*
- *Fake lottery* повезана са европским телефонским бројевима, код *419 scam* порука нападачи користе подједнако често британске и нигеријске бројеве



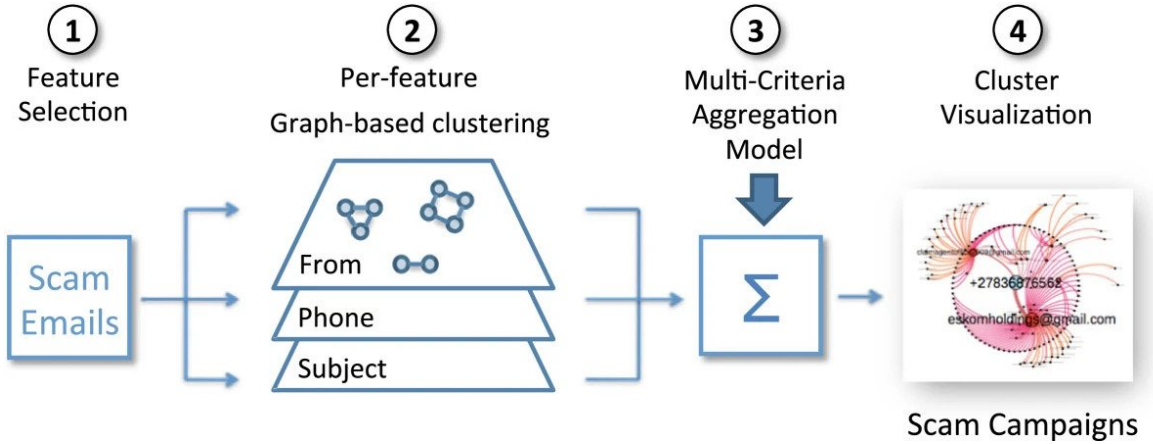
Слика 6: Scam email categories over time



Слика 7: *419 scam* category phone numbers over time by countries

Методе анализе података

- Да би се идентификовале групе порука које имају изгледа да буду део исте кампање врши се *кластеровање*
- Користи се *TRIAGE* радни оквир за безбедносно истраживање података
- Критеријуми за груписање засновани на подскуповима заједничних особина
- *TRIAGE* идентификује сложене обрасце у подацима што помаже при одређивању **тактика**, **техника** и **процедура** којима се нападачи служе
- *TRIAGE* показао употребљивост и у контексту других безбедносних истрага



Слика 8: *TRIAGE* workflow on scam dataset

- Први корак - одабирање карактеристика (адреса е-поште пошиљаоца, наслов поруке, датум слања, адреса одговора, број телефона, адресе е-поште у телу поруке...)
- Други корак - грађење односа међу узорцима у односу на изабране особине користећи **метрике сличности**

Табела 3: Weights of individual features (total = 1)

Feature	**Importance**
Phone	0.30
From	0.12
Reply	0.18
Subject	0.25
Email body	0.1
Date	0.05

- Трећи корак - агрегирање појединачних сличности карактеристика, омогућено је додељивање тежина особинама
- Четврти корак - агрегирана вредност се користи као улаз за класичне *graph clustering* алгоритме
- Дефинисање тежина према *Regular Increasing Monotone - RIM* квантификатору омогућава нам да моделирамо стратегије

Резултати кластеровања

- Идентификовано 1.040 кластера који се састоје од најмање 5 скам порука
- Величине кластера у просеку мале, ово би могло одражавати тежњу скамера да остану *испод радара*

Табела 4: Global statistics of the top 250 clusters

Statistic	**Average**	**Median**	**Maximum**
Number of emails	38	28	376
Number of from	13.9	9	181
Number of replies	6.2	5	56
Number of subjects	9.9	7	114
Number of phones	2.5	2	34
Duration (in days)	396	340	1.454
Number od dates(distinct)	27.9	22	259
Compactness	2.5	2.4	5.0

Процена резултата кластеровања

- Кластеровање је класификација без надзора, ваљаност резултата се процењује објективним критеријумима
- Екстерна и интерна валидација (*Adjusted rand index*)
- Испитана укупна компактност резултата, разврстана по појединачним карактеристикама
- Компактност је индекс ваљаности кластера који показује колико су кластери хомогени
- *TRIAGE* чува све појединачне везе између сличних особина порука унутар кластера, на тај начин пружа увид у *стабилне* карактеристике

Табела 5: Top coalitions of features across all clusters

Coalition	**Percentage (%)**
(phone, subject, from, reply, email body)	13
(phone, reply, email body)	12
(phone, subject, reply, email body)	11
(phone, from, reply, email body)	7
(phone, subject)	6
(phone, from)	5
(phone, reply)	4
(phone, reply, subject)	4
(phone, reply, subject, from)	4
others	33

- Особине одговорне за повезивање скам порука у кластерима укључује телефонске бројеве у 88% случајева, *reply* е-адресу у 66% случајева, *from* е-адресу у 46% случајева

Макро-кластери

- Настају спајањем међусобно слабо повезаних кластера
- У нашем контексту, указују на могуће скам кампање већег обима, сачињене од међусобно слабо повезаних поткампањи вођених од стране истих нападача
- Значајне особине су само адресе е-поште и телефонски бројеви, тражени су кластери који деле најмање једну адресу е-поште и/или телефонски број
- Идентификован скуп од 845 изолованих кластера и други скуп од 195 повезаних кластера (62 макро-кластера)
- Ове везе шема одлучивања сматрала је преслабим, па су стога ти скамови груписани у одвојене кластере
- На аналитичару је да процени колико су такве везе заправо значајне
- Пракса је да се почне од скупа смислених кластера, а затим да се прагови постепено смањују до тачке у којој више нема основа за приписавање кампањи истој групи

Закључак

- Телефонски бројеви и адресе е-поште имају кључну улогу у превари 419 за разлику од других шема сајбер криминала
- Нигерија и Велика Британија најзаступљеније државе порекла ових порука
- Велика разноликост начина увлачења жртве у превару и начина вођења кампање
- Скамери се труде да увек буду актуелни