Stephen Coggeshall

Chief Technology Officer
ID Analytics

Economic Crime and the Online World
November 3, 2011

Overview

- ida:labs – what is it?
- Identity Resolution – what is it?
- Research into Child ID theft
- Research into Identity Manipulation
- Horizon research – Who’s who online?

ID Analytics and ida:labs

- Who is ID Analytics?
  - We are a private company that provides identity theft prevention services to institutions (banks, wireless, retail credit…)
- What is ida:labs?
  - id:a labs is a multidisciplinary group of mathematicians, computer scientists, economists, financial experts, cognitive scientists and advisors from ID Analytics and other respected institutions.
- Why is it?
  - id:a labs reveals important trends in consumer behavior by examining identity use, to help organizations better manage both the risk and opportunity of an individual consumer.

id:analytics.

Identity Resolution

Identity Resolution – Figures out “who is this?”

- Identity Resolution is a system of fuzzy algorithms with specially organized data
- Input is some fragment of “SNAPD”:
  - **S**SN, **N**ame, **A**ddress, **P**hone, **D**ate of birth
- Output is
  - A unique person label: “Identity Number”
  - A confidence factor that this is who it is
  - Validity flags for the input fields
  - “Best SNAPD”
- System is robust to
  - Missing fields
  - Dirty data
  - Typos

Child ID Theft

Child Identity Theft Study – Highlights

- ID Analytics monitors 2.5 million consumers, 172,000 children (under 18)
- We examined the alerts for all 172,000 enrolled children
- When an alert occurs for a child it is 7 times more likely to result in harm than for an adult
- Projecting to the entire U.S. children we find at least 140,000 children are victims each year

Note – this approach would not see parents stealing their children's identities

New research:

- We find about 6 million parents and children sharing identity information
- Likely 500,000 children victims (under 18) over the past few years
- Indications of 2 to 3 million elderly parent abuse – perhaps the bigger problem!

Examples of Parents/Children Sharing SSNs

Likely victims: Young children

Child Name	Age	Parent Name	Age
Sarah Schmidt	4	Larry Schmidt	23
Julie Broxton	3	Gary Broxton	28
Jonathan Viscosi	12	Carol Viscosi	40

Likely perps: Credit-seeking parents

Likely perps: Credit-seeking children

Child Name	Age	Parent Name	Age
Brooke Garrison	59	Laura Garrison	90
Mary Parks	59	Tyler Parks	92
Richard Caudill	59	William Caudill	91

Likely victims: Elderly parents

Victim or perp??

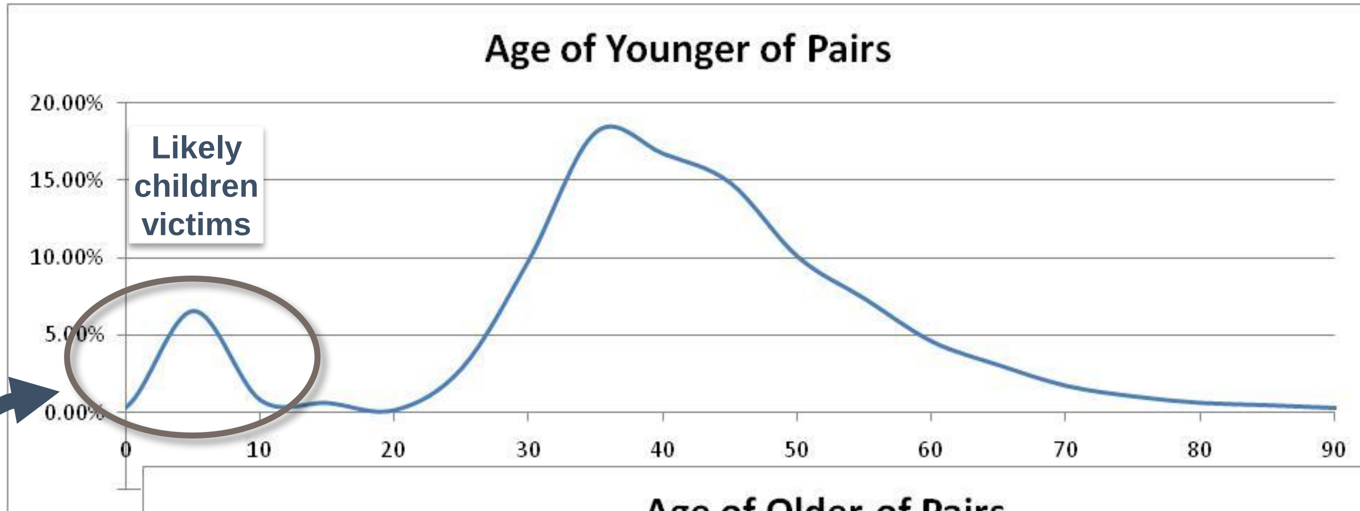
Child Name	Age	Parent Name	Age
Patricia Smith	29	Sebastian Smith	57
Valerie Bechtel	30	Fred Bechtel	55
Janice Latlock	32	Charles Latlock	51

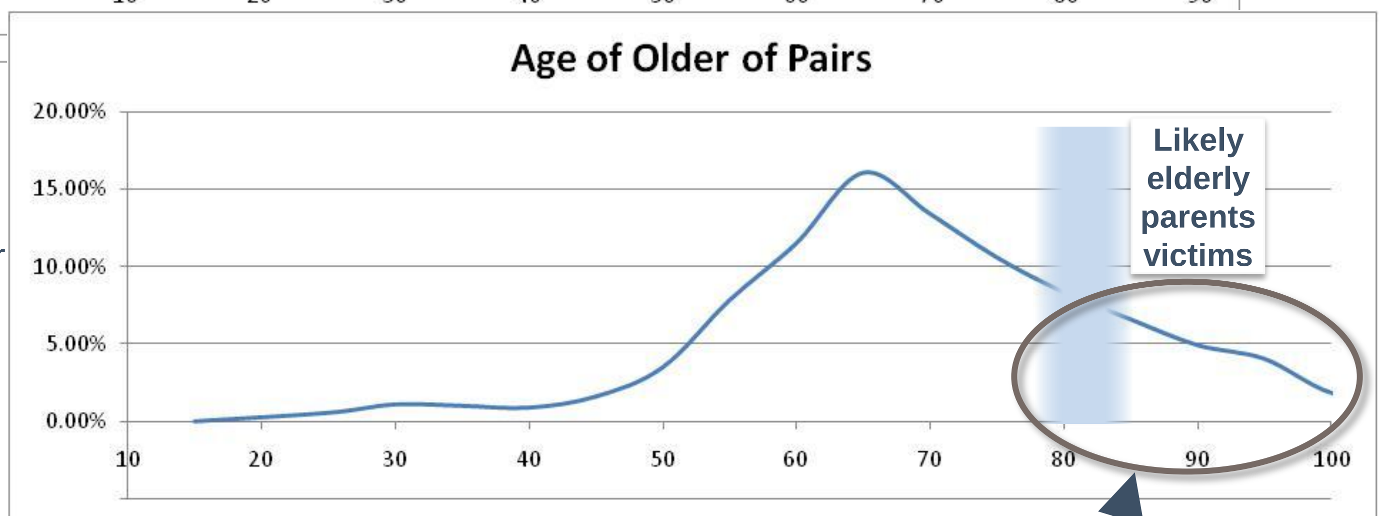
Victim or perp??

All identity information has been modified with the internal relationships preserved

id:analytics.

Examine Age Distributions of Pairs





More than 500,000 cases of parents using their children's identity information

2 – 3 million elderly parents victims

id:analytics.

Identity Manipulation

High-Level Results for Identity Manipulation

- About **17%** of the U.S. have deliberately manipulated their identities
  - Deliberately means ***repeatedly*** and ***systematically*** (not typos)
- **8 million** people deliberately used 2 or more SSNs
- **16 million** people use multiple DOBs
- About **10 million** spouses share their identity information
- About **6 million** parents/children share identity information
- About **45 million** people in the U.S. manipulate their identities. This ranges from relatively benign:
    - (“John Robert Townsend” also writing “JR Townsend”)

  to clearly criminal:

    - James in LA uses 10 SSNs, 14 different DOBs, 8 different last names

id:analytics.

Example 1 of Identity Variation

First Name	Last Name	SSNs	DOBs	Address	Zip	Phone
ANITRA	JOHNSON	815015642	1/11/1980	307 GRANADA DR	72994	8476644253
LATASHA	MCWILLAN	815115642	2/21/1980	1401 WALKER AVE #331	72091	8474511519
MCCLELAND	MCWILLIAMS			1095 SPRING MEADOW	71946	3950056410
	MCMILLAN			303 EAST SHR	71830	3920056710
	MCNULTY			1002 5TH ST	71323	
				317 WINONA ST	71323	

3 different First Names

5 different Last Names

2 different SSNs

2 different DOBs

All identity information has been modified with the internal relationships preserved

id:analytics.

Example 2 of Identity Variation

9 different First Names

5 different Last Names

23 different SSNs

11 different DOBs

Suspicious address variation

First Name	Last Name	SSNs	DOBs	Address	Zip	Phone
IRENE	ALMONE	580530044	1/7/1968	3310 ALGONQUIN AVE	23983	7186498265
LAQUINTA	CALHONE	586530044	12/21/1969	400 SCRUB OAK CT	23892	6861217347
LAQUITA	CALHOON	589489998	12/27/1969	4828 TERRACE TRL	23886	2828888889
LAQUITE	THOMPSON	589499935	1/7/1969	2600 PARK BLVD	23885	2827179083
LAQUTA	TOMSON	589539044	1/16/1969	4600 FAIR PARK BLVD	23885	2826132420
LEQUITA		590030040	1/17/1969	PO BOX 60011	23885	2825861415
QUITA		590490035	1/20/1969	2129 KINGSDALE DR	23881	2825802499
RENEE		590499937	1/27/1969	3211 MARYANN DR	23881	2825336941
RICHARD		590499938	1/27/1970	3229 KNOX ST	23881	2825330343
		590529641	1/27/1979	3424 FALCON DR	23881	2824637753
		590529941	1/27/1980	34321 FALCON DR	23881	2824637005
		590529994		3628 GLEN PARK CIR	23881	2824636015
		590530014		4709 LEONARD ST	23881	2824319994
		590530035		4719 LEONARD ST	23881	2824319964
		590530036		5161 DORMAN ST	23881	2824317894
		590530037		5163 DORMAN ST	23881	2824315964
		590530040		PO BOX 15840	23881	2822067559
		590530081		3008 GALEMEADOW DR	23877	2822028819
		590530244		1313 GLASGOW RD	23866	2821537027
		590538044		3052 BIRDSONG DR	23860	2821533359
		590929664		7063 MEADOWS DR	23860	2820798274
		590930043		7800 HILL DR TRLR 168	23860	2820612298
		590960044		4412 KEETER DR	23820	
		590980044		64 FOREST GLN	22660	

All identity information has been modified with the internal relationships preserved

id:analytics.

Summary Examples of Bad Identity Manipulators

City	First Name	#SSNs	#DOBs	# First Name	Male/ Female?	# Last Names
NY	Frank	146	7	7	N	5
Detroit	Wendy	23	5	4	N	10
Atlanta	Jody	21	2	17	Y	5
Phoenix	Michael	24	10	9	Y	8
Houston	Mary	32	10	5	Y	9
Atlanta	Clinton	21	13	3	N	2
San Fran.	Augustina	27	18	17	Y	10
Atlanta	Dawn	24	15	3	Y	3
NY	Robert	22	12	2	N	3
Detroit	Theresa	22	21	14	Y	14
DC	Thomas	21	3	4	N	5
Detroit	Linda	46	25	5	Y	10
DC	Smithton	33	5	5	N	7
Miami	Trisco	43	4	8	N	3
Atlanta	Corey	39	5	4	Y	3

City	First Name	#SSNs	#DOBs	# First Name	Male/ Female?	# Last Names
NY	Brent	21	12	7	N	9
Miami	Latavo	29	3	7	N	2
DC	Anthony	44	7	3	N	5
Miami	Latoya	32	2	4	Y	4
NY	William	100	4	3	N	6
Miami	William	100	2	3	N	6
Phoenix	Robert	40	3	3	N	4
Miami	William	34	4	3	N	4
Phoenix	Samuel	21	9	4	N	4
Cleveland	Jamal	106	12	6	N	5
NY	William	69	14	6	N	5
Houston	Susan	21	3	5	Y	5
Phoenix	Lisa	29	21	3	N	2
Caseyville	Paula	101	7	5	Y	9
Bakersfield	Patrick	74	5	4	Y	13

id:analytics.

List of the Worst 3-Digit Zip Code Regions

Population We See	3-digit zip	Region	Bad Ranking Out Of All 3-Digit Zips**
131,300	777	N of Houston, TX	0.1%
592,308	799	El Paso, TX	0.2%
1,103,528	482	Detroit, MI	0.3%
202,614	485	Flint and Detroit, MI	0.5%
226,857	392	Jackson, MS	0.6%
143,952	464	Gary, IN	0.7%
150,907	489	Lansing and Flint, MI	0.8%
720,179	774	Houston, TX	0.9%
514,775	700	New Orleans, LA	1.0%
2,482,956	770	Houston, TX	1.1%
812,869	785	Mcallen, TX	1.2%
269,773	784	Corpus Christi, TX	1.4%
610,405	751	Dallas, TX	1.5%
120,276	387	Greenville, MS	1.6%
832,039	483	Detroit, MI	1.7%
923,012	775	Houston, TX	1.8%
318,117	797	Midland, TX	1.9%
34,706	048	Rural Maine	2.0%
252,280	776	Houston, TX	2.1%

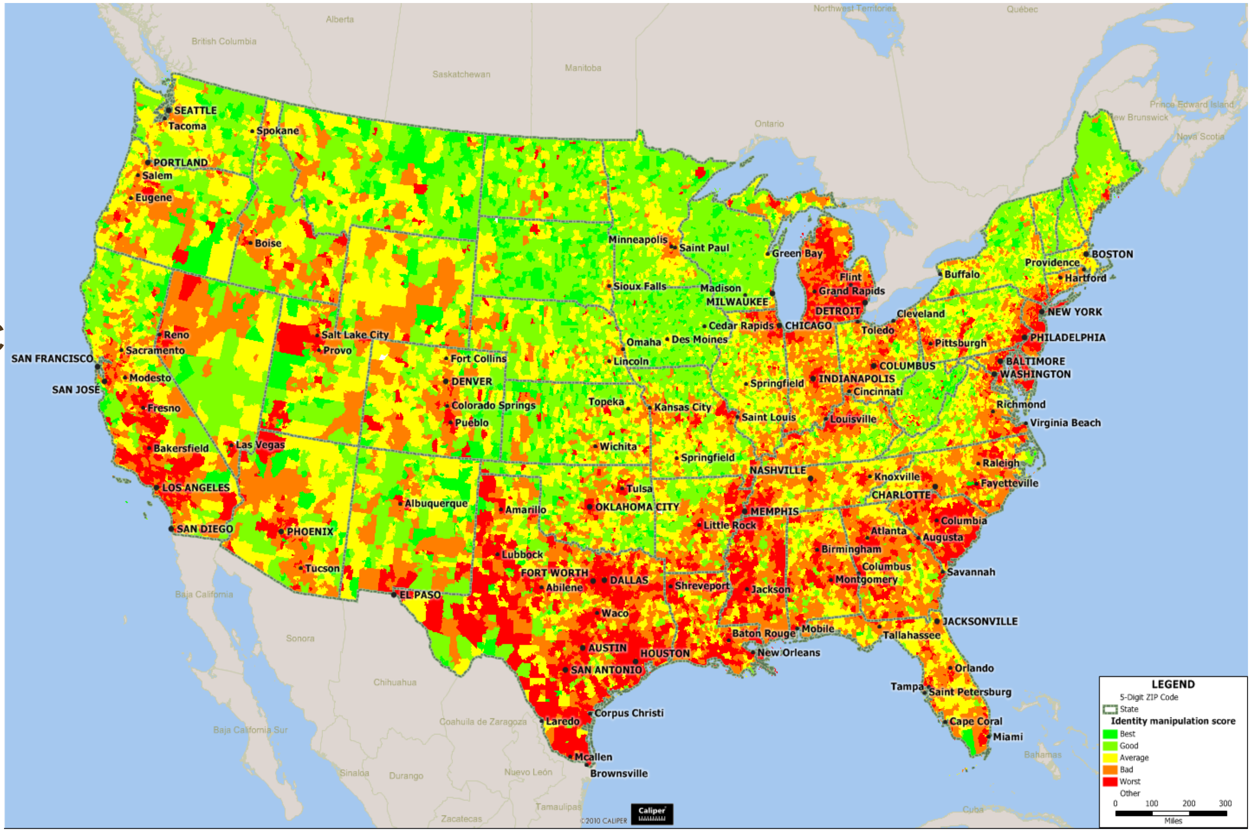
Population We See	3-digit zip	Region	Bad Ranking Out Of All 3-Digit Zips**
1,589,534	750	Dallas, TX	2.4%
259,271	197	Wilmington, DE	2.5%
637,634	114	Queens, NY	2.6%
393,785	780	San Antonio, TX	2.7%
382,069	701	New Orleans, LA	2.8%
528,087	077	NE of Trenton, NJ	2.9%
1,290,589	782	San Antonio, TX	3.0%
1,564,829	191	Philadelphia, PA	3.2%
216,236	110	Queens, NY	3.3%
386,247	103	Staten Island, NY	3.4%
207,735	783	Corpus Christi, TX	3.5%
297,120	199	Wilmington, DE	3.6%
1,031,048	928	Anaheim, CA	3.7%
795,041	773	N of Houston, TX	3.8%
268,073	711	Shreveport, LA	4.0%
55,089	651	Jefferson City, MO	4.1%
1,113,769	926	Santa Ana, CA	4.2%
83,487	219	SW of Wilmington, DE	4.3%
708,306	381	Memphis, TN	4.4%

****A bad ranking percentage of e.g. 1% means only 1% are worse, 99% are better**

id:analytics.

Map Shows Where ID Manipulation Occurs

- Bad areas include
  - Michigan
  - Texas
  - Southern CA
  - Southeast: LA, MS, SC
  - East Coast: NJ, MD, NY City, Wash DC
  - …



3-digit zip code regions
Color compares local IM rate to US average

id:analytics.

Examples of Professionals Who Manipulate Their IDs

Identity Manipulation Score Can Help Find Rings

- Examined a small list of 43,000 identity manipulators across the U.S.
- Did any of these people share addresses? Yes, many.
- Example of one cluster in Fort Worth, TX:

	# SSNs	# DOBs	# First Names	# Last Names	Male/ Female?	# Addresses	# Phones
Latesha Chapman	7	6	5	3	Y	27	25
Thomas Milton	12	8	13	8	N	29	36
Dawn Thompson	11	5	4	2	N	17	22
Latoya Thompson	24	12	9	5	Y	25	22
Ryan Milton	7	8	3	2	N	16	13

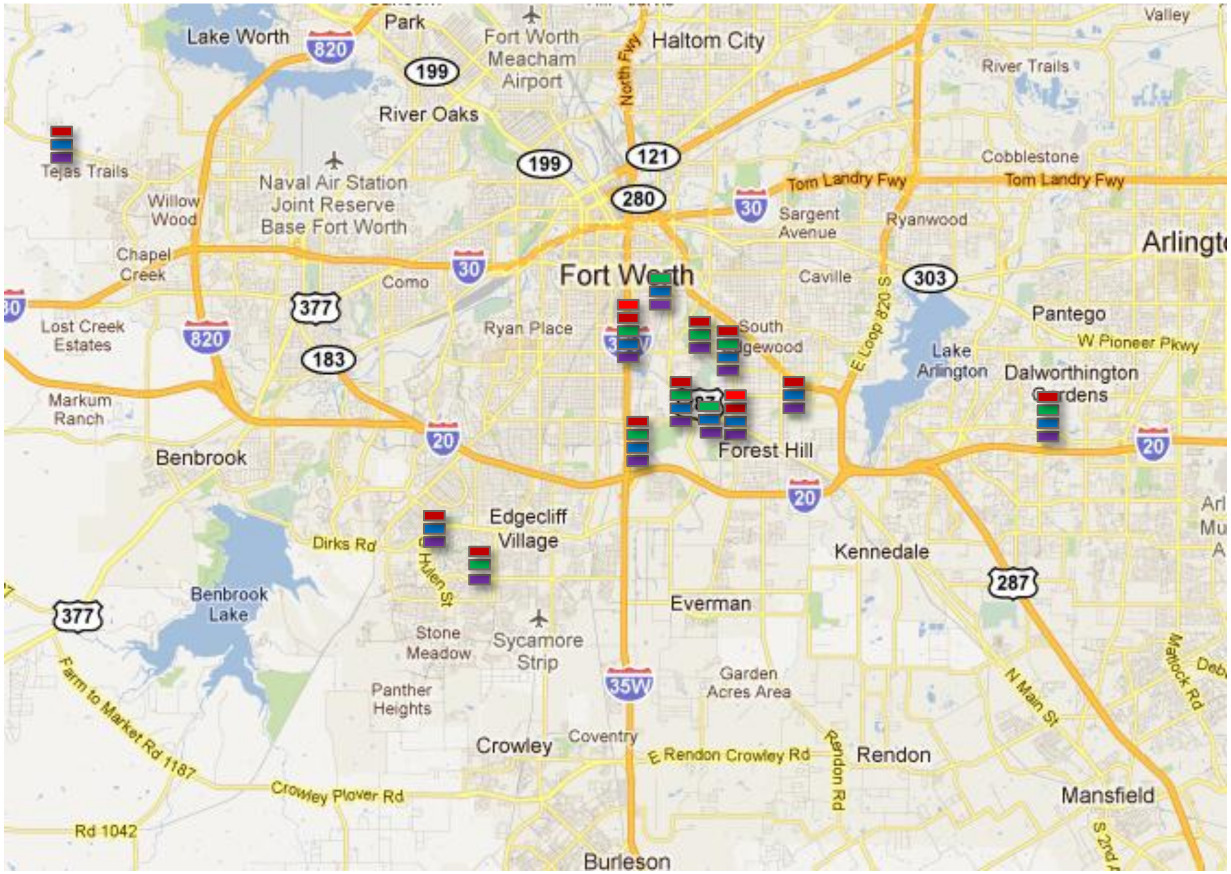
All these people shared

- several addresses
- fragments of names, SSNs and DOBs

All identity information has been modified with the internal relationships preserved

id:analytics.

Active Fraud Ring in Fort Worth



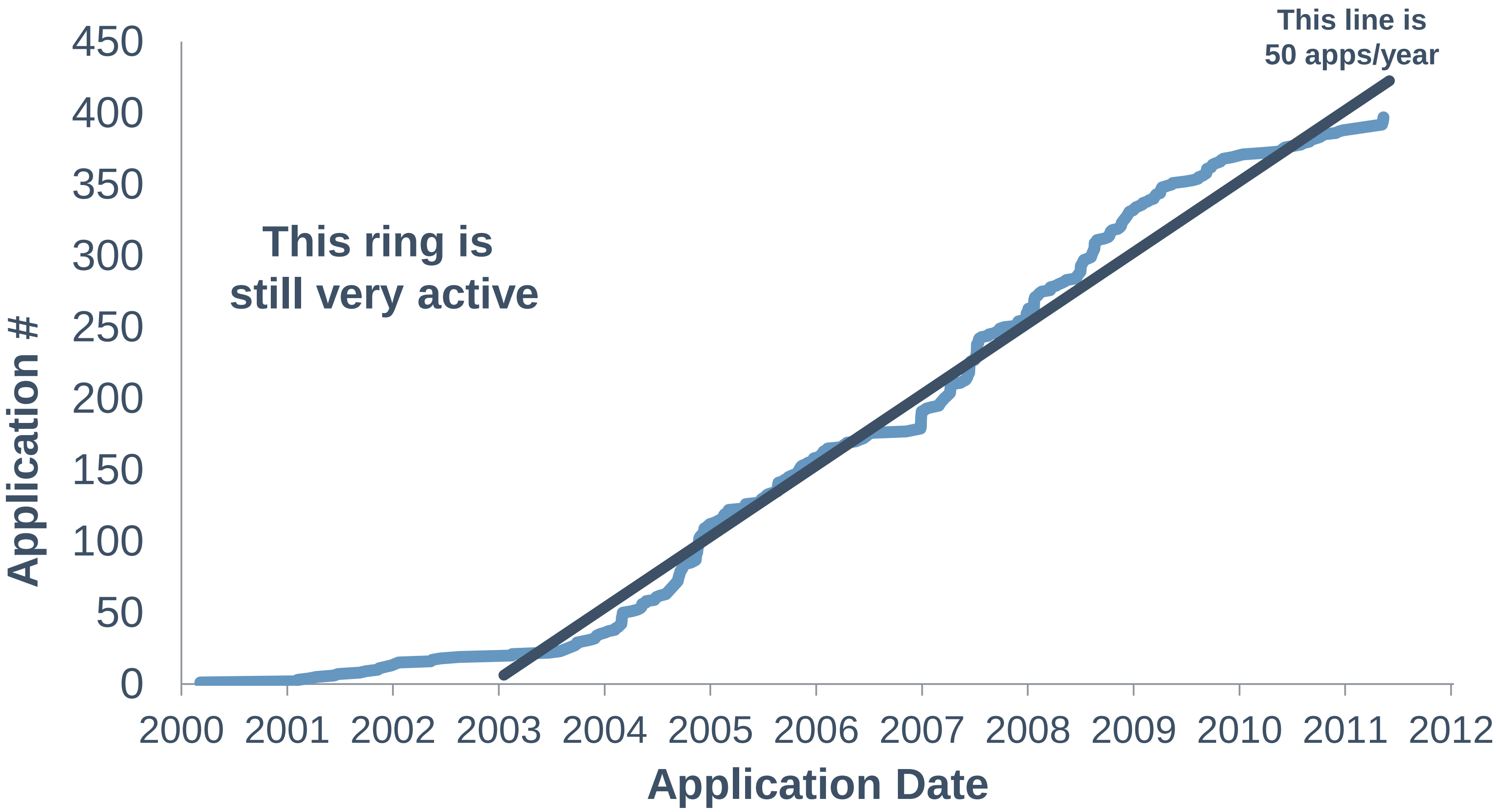
Legend

- Latesha Chapman
- Thomas Milton
- Dawn Thompson
- Latoya Thompson
- Ryan Milton

- Color indicates who is using which address
- These 5 identity manipulators share several addresses

id:analytics.

Application Activity Linked to Ring Addresses



Thin File/No Hit* is Really an Identity Manipulator (1)

	Name	SSNs	DOBs	Address	Zip
Application	Latoya McCarthy	461510032	Blank	2616 E 1ST ST	22687
A person we see uses these variations	Laticia McCarthy	461690032	7/06/1960	PO BOX 1272	23675
	Laticia Rivers	470507665	6/30/1971	1 BRIAN PL	23633
	Latoya Brooks	470507755		2616 E 1ST ST	22687
	Latoya Waters	470510032		5784 REMINGTON	22666
	Robert Hill	470510725		5784 REMINGTON # 706	22666
	Robert Jones	470700032		2507 BAMBERRY DR 37 S	22665
	Robert Morrison			3700 SOUTH CT	22665
				5451 LUBBOCK AVE	22665
				6853 S CREEK DR	22665
				5757 TRUELSON DR	22665
				9005 TOPPERWIND CT	22664
				620 CREEK POINT DR	22664
					22619

*A Thin File/No Hit refers to a request for a credit pull at the credit bureau, and the result is a record with little or no information

All identity information has been modified with the internal relationships preserved

id:analytics.

Thin File/No Hit Really an Identity Manipulator (2)

	Name	SSNs	DOBs	Address	Zip
Application	Katreena Harnan	371235546	8/19/1980	703 DESOTA PL	50456
A person we see uses these variations	Cateena Jackson	370552901	8/30/1975	128 LEONARD LN	50456
	Cateena Harnon	370583501	8/30/1976	232 CARRIAGE CIRCLE DR	50456
	Katrina Hornan	370653599	8/01/1980	28 MCKINLEY ST	50456
	Monica Hornon	371201594	8/29/1980	703 DESOTA PL	50456
		371232901	8/30/1980	719 DESOTA PL	50456
		371232923	8/291981	720 DESOTA PL	50456
		371232971		721 DESOTA PL	50456
		371233501		1131 DOVER RD	50457
		371233594		227 PROSPECT ST	50457
		371233595		227 PROSPECT	50572
		371233597		584 MARTIN LUTHER KING JR BLVD S	50457
		371233598			50457
		371233600		486 S SAGINAW ST	50458
		371233601		109 STANDLEY ST	50458
		371233602		1233 COLONY LN # 3	50458
		371233623		1233 COLONY LN # 328	50458
		371235050		710 KENILWORTH AVE	50763
		371237472		805 KETTERING AVE	50763

All identity information has been modified with the internal relationships preserved

id:analytics.

Identity Manipulation Results

- We see millions of identity manipulators in great detail – who, where, what
- Even professionals manipulate their identities
- Identity manipulator score can help find rings
- Many thin file/no hits are really identity manipulators

Who’s Who Online

Who's Who Online

- Can we observe a person's partial and/or obfuscated identity information online and clarify who it is?
  - Social networks (Facebook, Myspace, Linkedin, Twitter…)
  - Loyalty programs (airlines, hotels, grocery stores...)
  - Online gaming (World of Warcraft, PlanetSide, Gaia Online…)
  - Blogs (Huffington Post, Gizmodo, BusinessInsider…)

- Why?
  - Age verification (online sales of restricted items – guns, alcohol…)
  - Identify criminals (does Facebook or MySpace want to know who's a sex offender?)
  - Identify people with improper multiple accounts/benefits or otherwise masquerading
  - Better understand who people are

Online Identities: a Many-to-Many Match Problem

- Build an algorithm to compute the confidence that people match
- Now, each cell contains a confidence number
- Want to find best choice of which IDA # goes with each Online Entity
- Mostly, each Online Entity maps to one and only one, but not quite
- More complicated than a constrained optimization/linear programming problem since its many-to-many

Hundreds of millions

Hundreds of millions

	IDA #1	IDA #2	IDA #3	IDA #4
Online Entity 1	C11	C12	C13	C14
Online Entity 2	C21	C22	C23	C24
Online Entity 3	C31	C32	C33	C34
Online Entity 4	C41	C42	C43	C44
...	...	...	...	...

Goal – Find the Most Likely Identity Match

- Online person # 1
- Online person # 2
- Online person # 3
- Online person # 4
- Online person # 5

…

is most likely

ID Analytics person # 937465746

ID Analytics person # 746287463

ID Analytics person # 364982647

ID Analytics person # 136957153

ID Analytics person # 376249382

…

Each of these assignments has a confidence factor

Summary

- **Ida:labs** – Perform applied research to further understand identity-related issues and to improve ID Analytics' products and services
- **Identity Resolution** – System to figure out who's who in the presence of sparse or messy data
- **Child identity theft study** – At least 500,000 children are victims; about 2 to 3 million elderly victims
- **Identity manipulation** – 10's of millions of people are criminally manipulating their identities. We know who they are, where they live and what they're doing.
- **Horizon research** – Who's Who online: Clarify who is the real person behind an online presence. Can identify inappropriate or criminal behavior.