

Using Synthetic Decoys to Digitally Watermark Personally-Identifying Data and to Promote Data Security

Jonathan White and Dale Robert Thompson

Dept. of Computer Science/Engineering, University of Arkansas
{jlw09,drt}@uark.edu

Abstract

Identity theft continues to be an ever-present problem. Identity theft and other related crimes are becoming an unparalleled phenomenon that nearly everyone will have to deal with in some way in the coming years. As the number of people affected by identity theft and data spills has grown into the tens of millions, more needs to be done in the way of providing mechanisms to secure personally identifying data, including data consisting of social security numbers, names, addresses, and phone numbers.

One such mechanism that would enhance security is the use of realistic synthetic decoy records. These decoys would be inserted into the actual data in such a way that only the person or program that inserted the decoys can tell what is real and what is synthetic. Also, these decoys could also be created in such a way that they are probabilistically unique, making a kind of watermark for that particular dataset.

This paper examines a method by which identity theft can be combated by using decoys as described above. While decoys do not hide or encrypt the actual personally identifying data, it will be shown that they can be used to uniquely pin point particular data sources, making it possible to isolate what source was used in the theft. It will also be shown how realistic decoys make it much more difficult to use the actual data, because of the inability to distinguish between what is real and what is fake. Finally, we will show how we have implemented a system that is capable of producing these very realistic, personally identifying, decoy records.

1. Introduction

Identity theft is a wide reaching problem that is costing both individuals and businesses hundreds

of billions of dollars a year [1]. To illustrate, it has been estimated that various identity theft losses worldwide totaled greater than \$221 billion dollars in 2003 [2]. While this work is not focused on the various types of identity theft or its effects, it would serve well to illustrate a few examples of identity theft that were particularly devastating, because they affected so many people.

In May of 2005, Citigroup notified a group of over 3.9 million Americans of a data spill that resulted in the loss of several years worth of past loan data [3]. This data spill occurred in transit when a tape backup of this financial data was lost by a courier service. The data contained information of past and current customers, including personally identifying data such as social security numbers, names, and addresses. In a statement, Citigroup said that they had no reason to believe that this information had not been used inappropriately, nor had it received reports of unauthorized activity.

Citigroup has a lot of company when it comes to divulging large amounts of personally identifying data. In March of 2005, Time Warner announced that tapes containing information on 600,000 employees were lost by a firm that they subcontracted to for data archival [4]. Around the same time, Wachovia Corp. and Bank of America Corp. notified over 100,000 consumers by postal mail that their financial records may have been stolen by bank employees and sold to outside collection agencies. Police officials have said that over 700,000 bank customers may have been affected by this crime [5]. In another large data leak in April of 2005, Ameritrade said that they had informed over 200,000 customers about a missing data tape that contained personal information, including information on past financial transactions [6]. From these four

incidents alone, over 5.4 million Americans had their personal information potentially divulged. And these four incidents only spanned three months.

One thing that is of particular interest to this work is the above statement that many corporations who leak data make, that they have no reason to believe that the data they leaked had been used for criminal purposes. They can say this because both they and the affected consumers have no real way to prove that the leaked personal information was used as a result of this particular leak. Even if a consumer has their identity stolen and they have been previously notified that their identity had been potentially divulged, the consumer still can not say for sure that this data leak caused the crime.

This lack of knowledge about what source was used in an identity theft crime is a large problem. It would be a great boon for everyone involved if corporations had the ability to prove whether their leaked data was used or not. It would give police investigators another tool in prosecuting identity thefts.

In this paper, we will show a method by which corporations can watermark their data by inserting very realistic synthetic decoy records into their data. We will show how these decoy records can make each personally identifying database unique; giving corporations and consumers the ability to pinpoint what data spill was used in an identify theft crime. We will also show how these decoys make the original data more secure.

This paper is organized as follows. In section 2, we explain what decoy records are, how they are used, and some of the benefits of decoy records. In section 3, we will elaborate on the literature existing about decoy records. In section 4, we show how we designed a program that can make the type of realistic synthetic decoy records that have been mentioned. Section 5 covers the downsides of using decoy records, and how the bias that is caused by there presence can be removed. In section 6, we show the results of our program and point out areas of improvement. Section 7 contains our works cited.

2. Decoys

A decoy is something that serves to confuse or distract an opponent. A good decoy will look real, at least at a distance, so that the opponent will not be able to tell whether or not it is a decoy right away. The opponent must waste time and effort sorting out what is a decoy and what is not. In the database field, a decoy is a realistic record that has been purposely inserted into real data in order to distract someone who has access to that data. If the decoys cause enough distraction, an attacker may not be able to separate out what is real and what is only a decoy, protecting the data from misuse.

So, decoy records are used to both obfuscate and protect real data by making it more difficult to find and use the real thing. Also, decoys do not do any harm or otherwise change the actual real data. They will change the properties of the data as a whole, but the original data is still there, fully intact.

Decoys have several interesting and very useful attributes. Since decoys are purposefully made by someone who potentially knows what an opponent is looking for, the decoys can be designed to look like a gold mine of information. The designer of the decoys can control all of the decoy's properties, including how complex they are, how many of them exist in the data, where they are placed in the data, and whether or not they are unique.

If the decoys are created in such a way to be unique across every database in the world to a certain high percentage, they can serve to watermark the data that they are in. This is another way in which decoys can be used. However, in order to be a good watermark, the decoy must appear realistic, otherwise an attacker will be able to easily tell what is fake, and the will never attempt to use the information that the decoy contains.

This uniqueness property is something that is very important, because this is a characteristic that is rarely true for databases consisting of personally identifying data. It is very difficult to alter every

individual record in a database to ensure that each record is unique, not only in that particular database, but also every other database in the world. Altering the actual data can also make it more difficult for trusted users to use the data, because they need to know not only what records were altered, but also how each record was altered. This is a problem that data with inserted synthetic decoys face as well. However, we will show in section 5 how unique decoys are actually easier to remove than correcting altered actual data.

3. Decoys in the Literature

Decoys have been used to increase the confidence in distributed computations where trust in the client machines can not be assured [7]. This confidence is achieved by overwhelming an attacker with the combinatorial problem of differentiating live data with decoy data. The live data and some decoys are sent out from a trusted server to the client machines. If an attacker gets access to some client machines, in certain areas, it is very difficult to use any of the data, because it takes a lot of time and effort to separate out the fake data [7]. This combinatorial problem is specifically difficult to solve in applications such as matrix multiplications, where the problem breaks down to essentially factoring large numbers.

Decoys have also been used to maintain interaction with a calling thread or process that violates a software component that the decoys are tasked to defend [8]. Once a violation is noticed, instead of allowing the actual system to interact with the violator, a decoy is used. This frees the actual system to do necessary work. The decoy's goal is to learn as much as possible about the violator and report this information back to a trusted agent. The decoys can terminate the interaction once the true nature of the intruder has been determined. Since the decoys typically do have access to the information the intruder may be looking for, this helps to assure the safety of the actual data.

Decoys have been employed to protect digital intellectual property, such as music and movies

[9]. If the copyright owner fears that their property may be pirated, they can flood various channels (such as file sharing networks) with data that looks like their property, but in reality is just a decoy with the same file name and file size. If a pirate attempts to download a music or movie from one of these channels, they will be forced to waste time and effort sorting between the decoys that are out there. This reduces the likelihood that the actual intellectual property will be pirated.

One of the properties of good decoys was that they are realistic, at least at a distance. Making decoys that mimic personally identifying data such as names, addresses, phone and social security numbers would be very valuable to corporations who wished to use them in their data. We did not come across any instances where someone had tried this in the literature. We decided to make a program that could make decoys that looked like Americans.

4. The American Dataset Program

In response to this need, we have made a program that we call the American Dataset Program (ADP). ADP was designed to produce realistic synthetic decoy records that could be inserted into personally identifying datasets. Each record in the created dataset has the same attributes that a real American person would have – such as a first name, last name, social security number, phone number, full address, and an occupation. Every field in the synthetic dataset is plausibly and semantically meaningful; none are random fields. Each field is statistically accurate when compared to the real population distribution of America. The fields that ADP can produce are detailed in section 4.2.

4.1 How ADP works

The ADP program works by allowing the user to specify the properties of the decoy records. The user is allowed to specify how many decoy records they want, what fields should be present, and even if the records should be sorted by some key field. The user is also prompted for a key, which is used to seed the random number generator that is used in the field generation

process. The key gets appended to a code that represents the selections they made.

The individual fields are generated based on statistics that were gathered about their frequency in the American population. For example, the program produces the last name field by using a file that has over 89,000 last names and the percentage that they occurred. This percentage adds up to 100%. The dataset that is created attempts to model this distribution by multiplying the number of requested decoys times the percentage for each field, resulting in the total number that would be expected for each of the 89,000 last names. Then, the program selects at random from this pool of last names, removing a last name from the pool whenever it selects one. Since the last name *Smith* is the most popular American last name, it is highly likely that this will be the most common last name in the created decoy dataset, as long as the created dataset consists of a fairly large amount of decoy records. Most of the fields are produced in this manner. Details of the fields that ADP can currently produce follows.

4.2. Field information

The following paragraphs detail the current 16 fields that ADP can produce. They cover what statistics were used as data sources, how many values they contained, and some other general details about them.

4.2.1. Zip code, city, state, longitude, latitude

- Source: U.S. Census Bureau [10].
- Description: The file that was used was a compilation of statistics from the 2000 census. The Census Bureau found the population that lived in every zip code, making the population distribution accurate.
- Unique Values: Around 30,000

.0045 60623 IL CHICAGO 87.71 41.84 112047
.0043 11226 NY BROOKLYN 73.95 40.64 111396
.0041 NY NEW YORK 73.95 40.76 106564
.0039 NY NEW YORK 73.96 40.79 100027
.0038 CA BELL GARDENS 118.17 33.96 99568

Figure 1: The most common zip codes in America. Field 1 is the percentage of America that lives in that zip code, field 2 is the zip code, field 3 is the state, field 4 is the city, field 5 is the longitude, field 6 is the latitude, and field 7 is the population in that zip code.

4.2.2. First name, last name, and gender

- Source: U.S. Census Bureau [11 - 13].
- Description: The data source that was used was based on a sample taken from the 1990 census. The Census Bureau sampled 7.2 million of the names on the census roles and formed a table listing the most common American first and last names. This file already had the population distribution calculated. The Census Bureau did not include names that very rarely occurred, so only 90% of the 7.2 million names were represented in the file.
- Unique values: 89,000 last names and 5,500 first names with the associated gender.

<u>Last names:</u>
1.006 SMITH
0.810 JOHNSON
0.699 WILLIAMS

<u>First names:</u>
1.6590 JAMES M
1.6355 JOHN M
1.3145 MARY F

Figures 2 and 3: The most common first and last names in America. Field 1 is the percentage of occurrence, field 2 is the name, and, for the first names, field 3 is the gender.

4.2.3. Occupation, yearly salary, and hourly salary

- Source: Bureau of Labor and Statistics [14] and www.payrate.com [15].
- Description: Two data sources were used to generate occupations with the associated pay. The data from the Bureau of Labor and Statistics listed all the jobs that were available as options in the 2000 census with how many indicated that they had those jobs. This file also listed the national average hourly pay and yearly pay for each of those positions. Some data from payrate.com was used to find out

how much, on average, people from different states made in relation to the national average.

- Relation to other fields: What occupation you have is dependent on your age (people younger than 21 have a very limited selection of available occupations). However, how much you make in whatever occupation you have (both yearly and hourly) is dependent on what your particular occupation is and where you live. For example, a professor in California will make more than a professor in Arkansas, but a surgeon in Arkansas will still make more than a professor in California.
- Unique values: Over 700 different occupations. Each occupation has 50 different salaries available (each state has a different pay rate).

0.04044 Surgeon 91.15 189590
0.01943 Anesthesiologist 87.22 181420
0.01565 Obstetricians/Gynecologist 86.37 179640

Figure 4: The four highest paying jobs on average. Field 1 is the percentage of occurrence, field 2 is the occupation, field 3 is the national average hourly salary and field 4 is the national average yearly salary.

4.2.4. Birth month and birth year

- Source: Center for Disease Control [16] and the Census Bureau [17].
- Description: Two files from the Bureau of Labor and Statistics and the Center for Disease Control were used that had information on when people in America were born. The first file gave the percentage of Americans that are currently alive who were born in discrete five-year spans going back to 1900. The other file used information about births in 2000 and listed the births by month, allowing the calculation of the population distributions for birth months.
- Unique values: 105 birth years, 12 birth months.

Birth years:

1.6 1961
1.6 1962
1.6 1963

Birth months:

8.98 August
8.72 July
8.55 October

Figures 5 and 6: The most common birth years and birth months. Field 1 is the percentage and field two is the year or birth month.

4.2.5. Street names

- Source: U.S. Census Bureau by way of a public library in California [18].
- Description: This file contained the 75 most common street names in America. This file was made by the Census Bureau in 1990 as a piece of promotional work for the Bureau. The Bureau no longer had this data available through their website, but this data was able to be found in an online archive from a library in California.
- Unique values: 75.

3.46695 Second
3.23243 Third
3.15809 First

Figure 7: The most common streets in America. Field 1 is the percentage and field two is the street name.

4.2.6. Social Security Number

A social security number is typically given at birth in America. This nine-digit number serves to uniquely identify the individual. Social security numbers are used for bookkeeping purposes by the various governmental divisions within America, and they are one of the most misused pieces of information around. The social security number (SSN) attribute is different from the previously mentioned fields in that the SSN is guaranteed to be unique for each individual person. This adds a significant amount of complexity to the generation process.

An American social security number consists of three parts [19]. The first three digits are determined by what state the person applied for the social security number in (typically the state they were born in). The middle two digits range from 01 to 99, but they are not given out consecutively. The middle two digits depend on how many people have ever applied for a social security number in that state. For example, if 100,000 people have applied for social security numbers in Alaska, they will have much smaller middle two digits when compared to a state that has had several million people apply for social security numbers. The final four digits serve to make the entire social security number unique. According to the Social Security Bureau, the last four digits are given out at random.

Through some research at the Social Security Bureau's website, a data file that had all the first three digits that are available for each state in America was found. Since 1971, the first three digits that a person gets are randomly selected from all the possible first three digits that are available for the state. In all, there are around 750 possible first three digits. So, when ADP gives out an SSN, the first process that is performed is to select an appropriate first three digits at random from all the available first three digits for the state that the person was born in.

A data file from the Social Security Bureau's website that gave the order in which the middle two digits are given was used. The middle two digits are not given out consecutively. They are given out starting out at 01, then 03, 05, 07, 09 are given, then all the evens greater than 10 are given, then all the evens less than 10 are given, and finally all the odds from 11 to 99 are given.

When the Social Security Bureau has given out all the available unique numbers for a particular first three – middle two pair, they simply increase the middle second two digits to the next available digit. For example, if the middle two digits were 03, they would next give out 05. If the current middle two digits were 48, they would next give out 50. The Social Security Bureau gives out SSNs with particular first three – middle two combinations until there are no more remaining unique SSNs for that pair [19].

So in reality the middle two digits are a function of how many people have been born in each state. People who were born when the Social Security Bureau was first formed would tend to have the lowest available middle two digits, and people born fairly recently would tend to have the highest available middle two digits. A data file from the Social Security Bureau that had a list of the highest middle two digits that had currently been given out for each of the available first three digits was found. It was also known how the population of America was distributed age wise, so this information was able to be used to give very realistic middle two digits.

When ADP generates the middle two digits of an SSN, it takes in as parameters the first three digits of the SSN and the age of the person. Based on the highest middle two digits that have currently been given out for those particular first three digits, the program calculates what middle two digits a person with that age would likely have had given the age distribution of people in the United States. For example, if the highest middle two digits for SSN 123 was 99 (the highest possible given), a person born in 2005 would receive the middle two digits 99. A person born in 1905 would probably receive the middle two digits 01 and a person born in 1950 would receive middle two digits of 60. This happens because the population distribution of the U.S. is skewed. If however, the highest middle two digits for SSN 123 was 09 (only 5 are actually available, 01, 03, 05, 07, and 09), then a person born in 1905 would still receive 01, a person born in 1950 would receive 05 or 07, and a person born in 2005 would receive 09.

The final four digits of the SSN are what make the SSN unique. ADP generates the last four digits starting at a random number between 1 and 9999. Each of the possible first three – middle two – final four combinations starts at a different random number, which is generated dynamically when the program begins. These are stored an array of size 700 by 100. Therefore, *ADP will never generate the same social security numbers twice*. This correctly models how the Social Security Bureau gives out social security numbers.

4.2.7. General details of the SSN attribute

- Source: Social Security Bureau [19].
- Description: These files contain information on how social security numbers are given out in America including what first three digits are associated with each state, the order in which the middle two digits are given out, the highest middle two digits that have been given out for each possible first three digits. The final source file contains information on the age makeup of the US. It was the same file that was used to generate the year of birth.
- Relation to other fields: SSN depends on state of birth and birth year of the individual.
- Unique values: Every generated SSN is unique. The number of possible SSNs is around 700 million or so.

Idaho: 518 519 Vermont: 008 009 WestVirginia: 232 233 234 235 236

Figure 8: Examples of the first three digits that are available for some various states.

4.2.8. Telephone number

Various phone companies throughout the United States give out telephone numbers. Each phone number is unique to a telephone line, so by calling a number you can be assured that you are reaching a particular place or person. The North American Numbering Plan Administration (NANPA), which is a part of the U.S. government, is in charge of making sure that no phone company gives out the same number to two different people [20]. A telephone number consists of an area code, a three digits prefix, and a four digit house code. Data from NANPA that had every area code and prefix for most cities in America was used to generate realistic phone numbers. The area code is determined by the zip code that the person lives in. Each zip code has only one area code, which models real life. This makes it very easy to generate the area code for an American.

The prefix of a telephone number depends on the city that the particular person lives in. There are many available prefixes per city, sometimes even over one hundred in the case of a large city like Dallas or Chicago. Given the city, ADP selects an available prefix. ADP treats every prefix as being equally likely in the population.

The last four digits are almost exactly the same as the last four digits of the social security number. The last four digits for a given area code – prefix pair serve to make the telephone number unique, and they are given out at random. Like SSNs, ADP will not give out the same telephone number to more than one person.

4.2.9. General details of the telephone number attribute

- Source: NANPA [20].
- Description: This file contains information on what area codes and prefixes are assigned to various cities in America.
- Relation to other fields: The telephone number attribute depends on zip code of the person. A large city like Chicago can have a few area codes and several available prefixes for each area code. Each zip code - area code pair has several equally likely available prefixes.
- Unique values: Every generated telephone number is unique to a household. There are well over 3 billion possible telephone numbers that ADP could generate.

413 Belchertown Maine 213 252 323 460 413 Chester Maine 354 556 907 Wrangell Alaska 874

Figure 9: Examples of some cities with their actual area codes and available prefixes. Field 1 is the area code, field 2 is the city, field 2 is the state, and the following fields are the available prefixes.

5. Downsides

There are downsides to using decoy records in personally identifying datasets. It takes a lot of effort to produce realistic decoys, as evidenced by the American Dataset Program. It took several months of research to locate all of the data that was needed to produce each field. This information also had to be processed into a usable form in order to incorporate it into the program. This is a very labor intensive task.

Also, once the decoys are inserted, they must more than likely be removed at some point. The decoys can skew the properties of the actual data, and this bias can be very significant in certain applications [21].

However, we considered this problem when we designed the American Dataset Program. When the user specifies what they want their decoys to look like, they are prompted for a key. This key is concatenated with other information, and this information can be used to generate the exact same dataset again. Given the same key and other properties, ADP will generate the same dataset every time. A small change in the key results in a large change in the produced dataset, making it difficult to produce a similar decoy dataset without knowing the key.

So, when a user wishes to use the actual data, they can regenerate the decoy records just by knowing the key. Then, since the decoy records are guaranteed to be unique to a high percentage, they can be removed from the underlying data with a simple SELECT statement. What remains is the actual data, in its original form. Besides this waste of time, another downside to this solution is that the program that generated the decoys must be available to anyone who wants to use the data that has the decoys in it. The American Dataset Program takes up about 40 megabytes of hard drive space, and it can generate over 30,000 records a minute. While this is not a lot, it is still a penalty that must be considered.

6. Conclusions and Further Research

The results of our ADP experiment were very encouraging. We were able to produce synthetic decoy records that have very realistic properties in large quantities. The current ADP implementation can produce synthetic decoys with 16 meaningful fields. The fields that ADP can produce are some of the most common fields that occur in personally identifying data, and it is a good start to having the type of decoys that could be used to watermark this type of data.

There are still several areas where ADP could be improved. Due to the lack of information, we are only using a file that has statistics on 75 street names. This would be the first thing that we would improve. The program also uses information which is a little dated; this would be another good improvement. Also, the program does not take into account the implied ethnicity of a last name or first name, so currently names like *Abdul Smith* and *Enrique Harris* will be generated, because no ethnicity is associated with the name pairing. In the future, we would like the program to be able to produce realistic drivers' license numbers and even simulated account numbers.

As mentioned before, the decoys can make every personally identifying data source different. If a data spill occurs, the corporation that spills the data could reproduce the decoys and announce them missing to the various groups that might encounter them, such as the credit reporting agencies. If these unique decoys are ever encountered by these agencies, the corporation can be assured that it was their data that was being used maliciously. This is an ability that corporations currently do not have, and it is something that would add to the confidence that consumers have in their data being secure. This is the most promising aspect of using data decoys in personally identifying databases.

7. References

- [1] Abdullah, A.K. "Protecting Your Good Name: Identity Theft and its Prevention." *Proceedings of the 1st annual conference on Information security curriculum development*, October 8, 2004, pp. 102 – 106.

- [2] Aberdeen Group. "Identity Theft: A 2 Trillion Dollar Industry in 2005." May 2003.
- [3] McMillian, R. "Citigroup Loses Data on 3.9 Million Customers." *IDG News Service*, June 6, 2005. <http://www.pcworld.com/news/article/0,aid,121178,00.asp>.
- [4] Daurat, C. "TimeWarner Reports Loss of Personal Data on 600,000 Employees." *Bloomberg News*, May 3, 2005, pp. E02.
- [5] Associated Press. "More Than 100,000 Notified of Possible Record Theft." *USA Today*, May 23, 2005.
- [6] Sullivan, B. "Ameritrade Warns 200,000 Clients of Lost Data." *MSNBC Interactive*, April 19, 2005. <http://www.msnbc.com/id/7561268>.
- [7] Jackson, T., Hart, D. "Data Decoys for Confidentiality in a Distributed Computation: Matrix Multiplication." *Proceedings of ACMSE 2004*, April 2, 2004, pp. 307 – 308.
- [8] Michael, J. "On the Response Policy of Software Decoys: Conducting Software-Based Deception in the Cyber Battlespace." *Proceedings of the 26th International Computer Software and Applications Conference on Prolonging Software Life*, August 2002, pp. 957 – 962.
- [9] Kushner, D. "Copy Protection: Digital Decoys." *IEEE Spectrum*, Volume 40, Issue 5, May 2003, pg. 27.
- [10] U.S. Census Bureau; "Census 2000 Gazetteer Files," published September 25 1996; <http://www.census.gov/geo/www/gazetteer/places2k.html>.
- [11] U.S. Census Bureau, "Last Name Files," published December 20 1999, maintained by Laura Yax. <http://www.census.gov/genealogy/names/dist.all.last>
- [12] U.S. Census Bureau, "Female First Name Files," published December 20 1999, maintained by Laura Yax. <http://www.census.gov/genealogy/names/dist.female.first>
- [13] U.S. Census Bureau, "Male First Name Files," published December 20 1999, maintained by Laura Yax. <http://www.census.gov/genealogy/names/dist.male.first>
- [14] Bureau of Labor and Statistics. "2000 National Labor Statistics," published 2000; <ftp://ftp.bls.gov/pub/special.requests/oes/oes02nat.zip>.
- [15] Payrate.com. "Real-Time Salary Survey Information," published 2005; <http://www.payscale.com/salary-survey/aid-9257/rid-79/fid-6886/RANAME-SALARY>.
- [16] Center for Disease Control. "National Vital Statistics Reports," published 2002; http://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_02.pdf.
- [17] U.S. Census Bureau. "Population Distribution and Composition, 2000," published 2000; <http://www.census.gov/population/pop-profile/2000/chap02.pdf>.
- [18] Santa Cruz Public Libraries. "Most Common Street Names," published 1990; <http://www.santacruzpl.org/readymref/files/q-s/stnames.shtml>.
- [19] Social Security Online. "SSN- Order of Issuance," published 2005; <http://www.ssa.gov/employer/ssnweb.htm>.
- [20] North American Numbering Plan Administration. "United States Central Office Code Plan," published 2005; www.areacode-info.com/COC/codeddownload.htm
- [21] Adams, N., Worthmann, J. "Security-control methods for statistical databases: a comparative study." *ACM Computing Survey*, Volume 21, Issue 4, December, 1989, pp. 515 – 556.