Towards a Multidimensional Model for Terrorist Attacks Analysis and Mining

Firas Saidi
RIADI Laboratory
National School of Computer Sciences, Manouba,Tunisia
e-mail: firas.saidi@ensi.rnu.tn

Zouheir Trabelsi
College of Information Technology, UAE University, Abu Dhabi, UAE
e-mail: Trabelsi@uaeu.ac.ae

Henda Ben Ghezala
RIADI Laboratory
National School of Computer Sciences, Manouba,Tunisia
e-mail:henda.benghezala@riadi.rnu.tn

Abstract— **The number of terrorist attacks have grown steadily in recent times, causing dire human and material damages. Analyzing potential terrorism data is without doubt an indispensable task, not only to understand these terrorist events, the terrorist implicated actors, their communities and their operation methods and tactics, but also to predict future attacks. However, terrorist events are commonly conducted differently and rarely fit the models of conventional attacks. Thus, terrorist events are usually hard to analyze, mine and investigate. In this paper, we propose a multidimensional model, named Terrorist Data Warehouse (TerDW) for allowing terrorist investigators to perform interesting and specific analysis for decision-making objectives. Practically, while consulting gathered information relative to specific terrorist attacks, such as attacks' locations, dates, used weapons, and implicated attackers, investigators can query the proposed data warehouse (TerDW) using multidimensional queries. The proposed model is based on consistent dimensions and measures. Terrorist attacks data are extracted from the Global Terrorism Database (GTD), a well-known database that includes information about terrorist events that took place from 1970 to 2017. Queries examples have been conducted to evaluate the efficiency of the proposed model. The queries' results demonstrate clearly that the proposed model allows investigators to carry out more appropriate multidimensional analysis, investigations and decisions.**

Keywords- Terrorist Attack; GTD; TerDW; MDX; OLAP analysis.

I. INTRODUCTION

Terrorist organizations try to multiply their malicious activities and to change their tactics, especially with the emergence of Web accessible devices. In fact, investigators are trying to gather useful information about these groups to devise a prevention strategy. Many researchers on terrorism constructed very interesting data sets about terrorist incidents around the word such as: Global Terrorism Data Base, John Jay and ARTIS Transnational Terrorism Database (JJATT), and Chicago Project on Security and Terrorism [1].

Thus, analyzing potential data related to terrorist events is without doubt an indispensable task, not only to understand the terrorism phenomenon, terrorist implicated actors, their operation methods and tactics, but also to predict and prevent future attacks.

However, terrorist events are commonly conducted differently and rarely fit the models of conventional attacks. Thus, terrorist events are hard to analyze, mine and investigate. To overcome this issue, we propose a multidimensional model, named Terrorist Data Warehouse (TerDW), for allowing terrorist investigators to perform interesting and specific analysis to help their decision-making. Practically, data gathered about terrorist attacks and incidents such as attacks' locations, dates, used weapons, and implicated actors are interesting to be visualized as multidimensional queries' results for decision- making goals. TerDW model includes consistent dimensions and indicators to analyze terrorist data extracted from the GTD database [2]. Interesting queries examples performed using TerDW are discussed in this work.

This paper is organized as follows: Section 2 describes related works on terrorist groups and activities identification and analysis. Section 3 presents and discusses the proposed framework architecture. In Section 4, the multidimensional model TerDW is discussed in details. Section 5 discusses how TerDW can be used for OnLine Analytical Process (OLAP) analysis and presents results of queries examples. Section 6 discusses the advantages and limitation of the proposed model. Section 7 concludes the paper and outlines perspectives for future research works.

II. RELATED WORKS

Various research works focused on the detection, analysis and mining of terrorist organizations and their malicious activities. These works are categorized into two approaches, namely Social Networks Analysis (SNA)-based approaches and hybrid approaches [1].

In [3], the SNA methodology was applied on the Jemaah Islamiyah cell that was responsible for the Bali bombings in 2002. Furthermore, an intelligent analysis framework was proposed to understand the structure of such a cell and to

 ISBN: 978-1-61208-729-0

predict with real time analysis the outcomes of these terrorist subgroups.

In [4], the authors studied Al Qaeda network characteristics and activities. In fact, the authors argued that Ben-Laden occupied a central position within the Al Qaeda network based on the computation of some measures.

The work in [5] presented a list of techniques that may analyze new summaries of terrorist attacks from the GTD database. The authors emphasized the use of text mining to extract critical concepts from free text. In a second phase, they used different learning algorithms including Naive Bayes, decision tree and support vector machine, to learn the terrorism incident, and especially, to detect its types from the news.

In [6], the authors proposed a dimensional model used in analyzing cybercrime data. Furthermore, they presented some interesting queries and the types of cybercrime analysis that can be performed based on the proposed model.

In [7], the authors introduced a theoretical framework which has two aims: first, to discover criminal organizations in networks issued from phone calls records and, second, to help investigators to analyze their structure and to discover hidden roles and relationships. Thus, this work contributes to terrorist network visualization from mobile phone calls and to community detection and analysis.

In [8], the authors introduced a technique for terrorist network destabilization based on two main steps, namely, first, communities detection and, second, the analysis of these groups by the identification of key actors and potential links.

III. TERRORIST DATA WAREHOUSE (TERDW) MODEL

In order to analyze terrorist attacks and incidents, we use a multidimensional model of data warehouse. Practically, we describe the methodology used to design the TerDW model. In fact, we refer to the well-known methodology of Kimball [9] to design the proposed Data Warehouse. Indeed, this process follows four main steps, as shown in Table I.

A fact table operates with dimensions and holds the data to be analyzed. A dimension table stores data about the ways in which the data in the fact table can be analyzed. Furthermore, we note that dimensions and measures of the multidimensional model are inspired from the GTD data base structure [2].

Figure 1 describes the proposed TerDW model, which is a galaxy model with two fact tables, namely Terrorist incident and Terrorist attack tables. Practically, the Terrorist attack table is used to analyze attacks, while the Terrorist incident table is used to analyze crimes at each incidence level.

We note that the **Galaxy schema** (Figure 1) is a Data Warehouse model, which contains two or more fact tables sharing some dimension tables. Sharing dimension tables can reduce database size, especially where shared dimensions have many possible values.

The proposed model can be exploited for decision making by investigators and terrorism specialists.

We identify two facts, namely, **Terrorist Incident fact** and **Terrorist Attack fact,** used to analyze terrorist events following different axis, i.e., dimensions and using different computed values, i.e., measures.

TABLE I. KIMBALL METHODOLOGY FOR TERDW MODEL IMPLEMENTATION

#	Steps	Description
1	Select the business process	Our business process is terrorist attack investigation.
2	Declare the grain, which establishes exactly what a single fact table row represents	In the Global Terrorism Database (GTD) [2], there are two types of terrorist attacks, namely: incident and attack. The incidence is a suspicious activity that may or may not result in a terrorist attack. **Facts:** Terrorist Incident (F1) and Terrorist Attack (F2).
3	Identify the dimensions used to analyze the fact table	**Dimensions:** Date, Attacker, Perpetrator organization, Location, Incident information, Attack information, Weapon, Target (victim), Demographic and Social information.
4	Identify measures of the fact table	Measures: Total number of fatalities, Number of Perpetrator Fatalities, Total Number of Injured, Number of Perpetrators Injured, Value of Property Damage (in USD), Total Number of Hostages/ Kidnapping Victims, Hours of Kidnapping / Hostage Incident, Days of Kidnapping / Hostage Incident, Total Ransom Amount Demanded, Total Ransom Amount Paid, Number Released/Escaped/Rescued

 ISBN: 978-1-61208-729-0

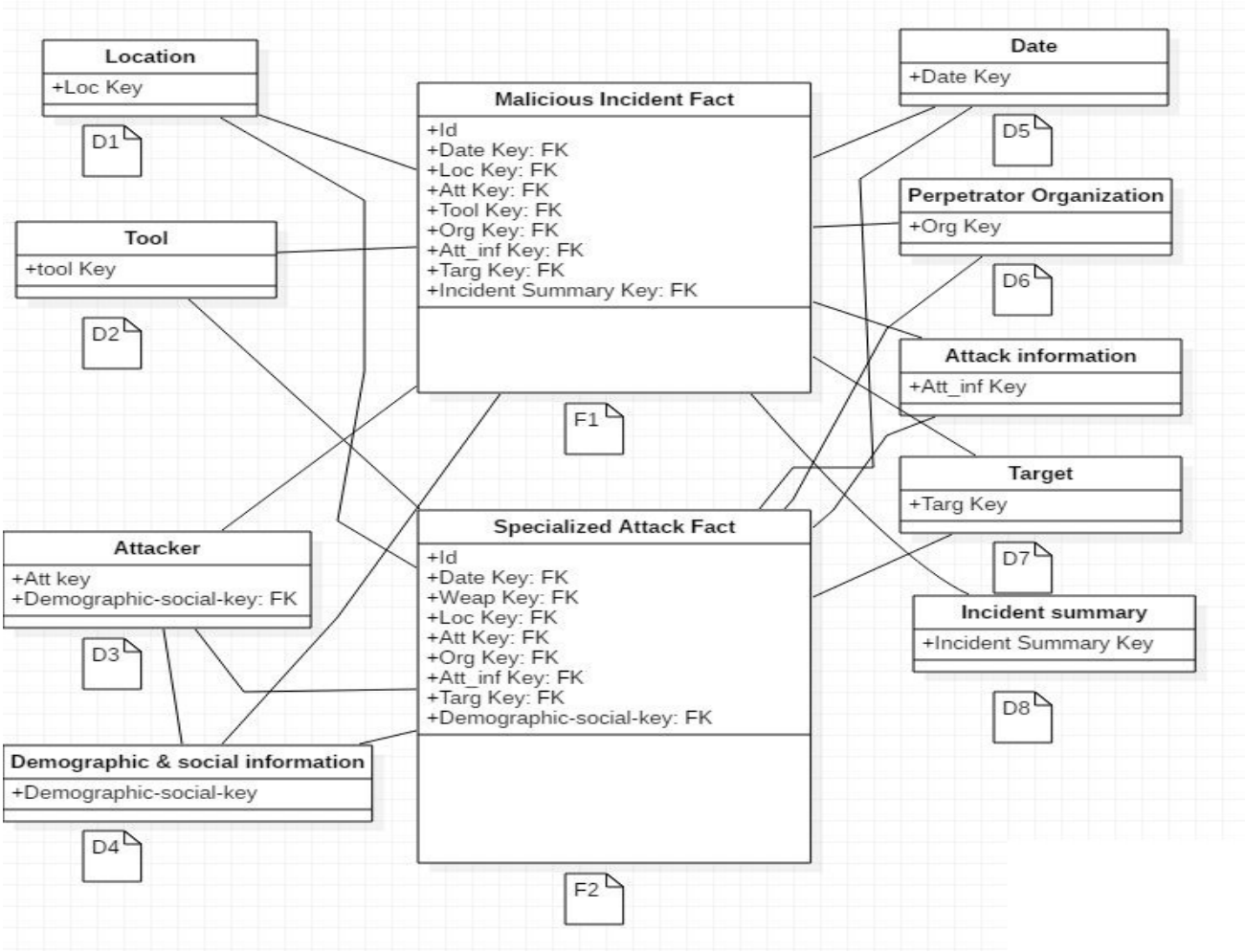

Figure 1. Galaxy Terrorist Data Warehouse (TerDW) schema

IV. TERRORIST ATTACKS ANALYSIS PROCESS

In this section, we present the process of terrorist attacks analysis in Figure 2. We apply the Extract Transform Load process on GTD database. Data are loaded on a TerDW, which is used to analyze and mine terrorist attacks based on various dimensions.

Investigators can ask the OnLine Analytical Processing (OLAP) cubes [10] constructed from the TerDW and can perform multidimensional analysis or apply data mining techniques in order to extract valuable knowledge about terrorist attacks and their related dimensions. Practically, the data warehouse is powered by data from GTD data base [2]. Indeed, data are extracted, transformed and loaded using an open source as Extract Transform Load (ETL) tool.

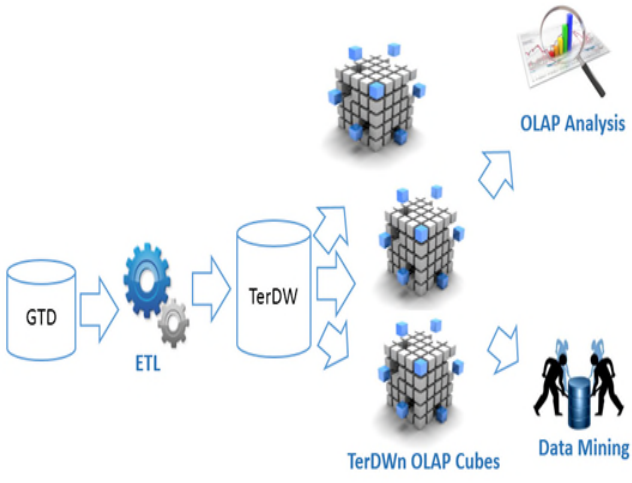

Figure 2. Terrorist attacks analysis process

V. OLAP ANALYSIS OF TERRORIST ATACKS

In this section, we present some multidimensional analysis based on TerDW to help investigators in predicting terrorist attacks and destabilizing terrorist cells (see Figure 3).

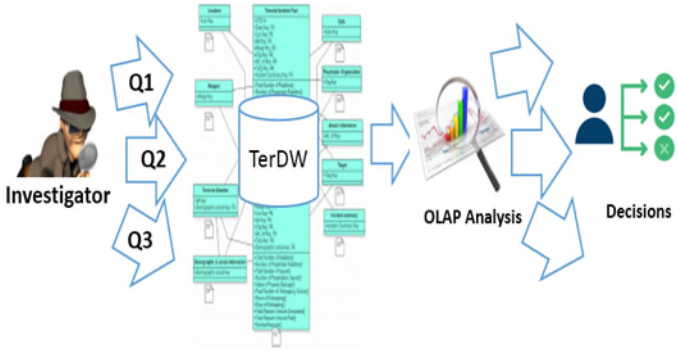

Figure 3. Investigator's decision making process

Based on the dimensions and measures defined in Figure 1, many interesting Multidimensional Expressions (MDX) (queries) can be executed to analyze data of terrorist attacks data. Figure 4 presents an example of MDX query which aims to extract the total number of injured and rescued (Total Number and Rescued Number) of 2011 and 2016 attacks (Date dimension) in Africa (Location dimension) from TerDW-1 OLAP cube (includes Terrorist Attack fact).

```
SELECT
   { [Measures].[Total Number of
Injured], [Measures].[Number Rescued] },
{[Date].[Attack
information].[Year].&[2011],[Date].[Att
ack information].[Year].&[2016],}     ON
1
   FROM [TerDW1]
   WHERE ([Location].[Africa])
```

Figure 4. Example of MDX query

The TerDW model allows investigators to submit interesting queries on specific terrorist events and actors (Figure 5).

Investigators can have an idea about establishments and targeted persons, most frequent attacks, used weapons, periods of frequent attacks, demographics and social information of attackers, information about terrorist organizations and can quantify different human and material damages.

Figures 6, 7 and 8 present the results of three queries applied on the proposed TerDW data model. Figure 6 shows the most frequent terrorist attack's type that took place between 1995 and 2012. The most frequent attacks are bombing and explosion, armed assault and infrastructure attacks respectively.

Figure 7 shows statistics about the most targeted victims during the terrorist attacks in the GTD database. Military, police and business actors are the most targeted in terrorist attacks.

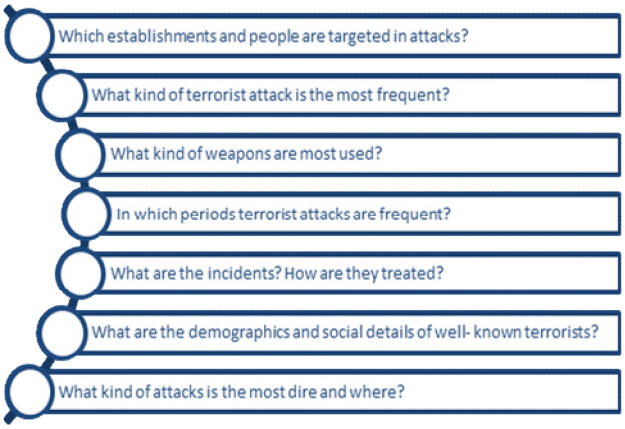

Figure 5. Queries samples for OLAP analysis

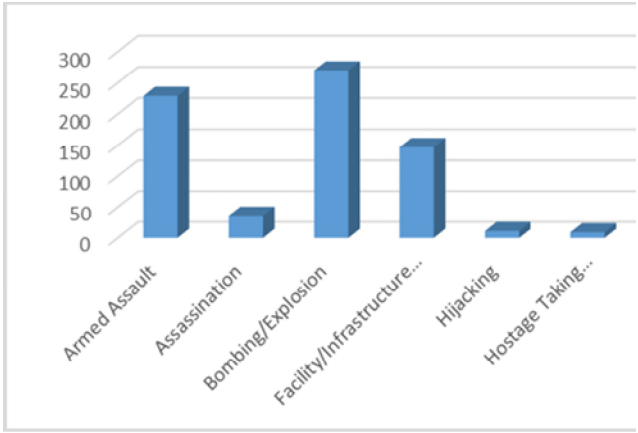

Figure 6. Terrorist attack type frequency between 1995-2012

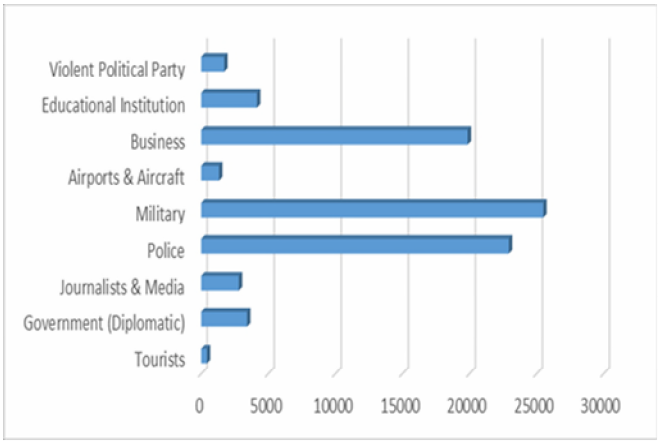

Figure 7. Most targeted victims in terrorist attacks

Figure 8 shows the percentages of the most used weapons during terrorist events in the GTD database. Firearms, explosive/bombs and dynamite and incendiary devices are the most used weapons in terrorist attacks, respectively.

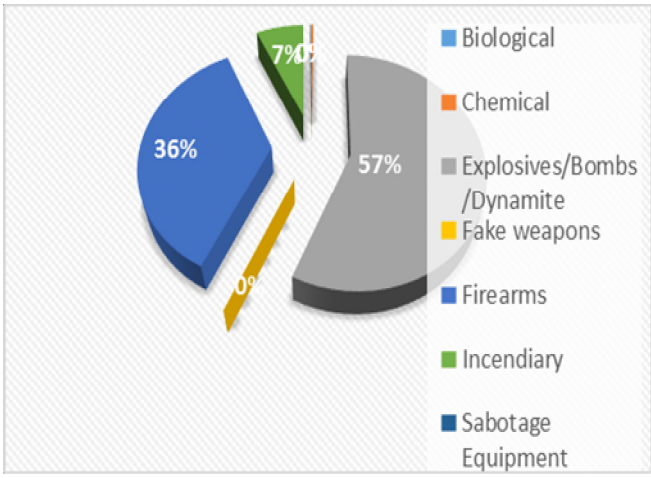

Figure 8. Most used terrorist weapons

Thus, investigators can extract data and statistics about terrorist attacks, used attack tools and techniques and the attack actors. In fact, terrorists might be key actors of cells or criminal organizations and might not belong to any terrorist community. This information is important to identify the well-known ‘’lone wolves” actors [11]. A “lone wolf” actor is a terrorist who prepares and executes terrorist acts alone, i.e., outside of any command organization and without the assistance of other terrorist actors. In fact, this type of terrorists is very dangerous since it is difficult to identify and to predict.

Implementation and Evaluation

In order to experiment the proposed multidimensional model, we used SQL Server BI framework. We created TerDW following the conceptual model described in the paper. After that, we used the SQL Server Integration Services (SSIS) as Extract Transform Load (ETL) tool. From the GTD database, data were extracted, transformed and loaded to the TerDW. For OLAP analysis, we used SQL Server Analysis Services (SSAS) tool.

Investigators can create various interesting reports in an easy manner using SQL Server Reporting Services (SSRS).

VI. DISCUSSION

The proposed Data Warehouse TerDW is a multidimensional model used for terrorist attacks and incidents analysis. Practically, investigators can query the TerDW to answer various interesting questions about attacks, attackers and used techniques. Indeed, this pack of information can be indispensable to understand terrorist organizations following different axes, i.e., dimensions, their topology, tactics and operation methods.

However, this model is limited and cannot provide deep information and correlations using simple MDX queries. Hidden relationships between data can provide interesting knowledge. Thus, TerDW must be valorized by data mining techniques in order to extract valuable knowledge and to discover new information not only used to understand terrorist events, but also to predict future attacks.

VII. CONCLUSION

In this paper, we proposed a data warehouse (TerDW) which is a multidimensional model for decision-making. We explained how TerDW can be used for OLAP analysis to help investigators in their work not only to understand the terrorism phenomenon, but also to predict others related criminal events. Furthermore, we detailed how this model can answer some interesting questions and we presented different results of the aforementioned OLAP queries. As an interesting perspective of this work, data mining techniques can be used to extract valuable knowledge about terrorist events and actors from TerDW. Moreover, a machine learning process can be applied on GTD to predict future attacks.

REFERENCES

[1] F. Saidi, Z. Trabelsi, K. Salah and H. B. Ghezala. ”Approaches to analyze cyber terrorist communities: Survey and challenges”. Computers Security, vol. 66, pp. 66-80, 2017.

[2] G. LaFree and L. Dugan, “Introducing the global terrorism database”, Terrorism and Political Violence, vol. 19, no. 2, pp. 181-204, 2007.

[3] S. Koschade, “A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence”, Studies in Conflict & Terrorism, vol. 29, no. 6, pp. 559-575, 2006

[4] E. Wu, R. Carleton, and G. Davies, "Discovering bin- Laden’s replacement in al-Qaeda, using social network analysis: a methodological investigation”, Perspectives on Terrorism, vol. 8, no. 1, pp. 57-73, 2014.

[5] S. Nizamani and N. Memon, “Analyzing News Summaries for Identification of Terrorism Incident Type”. Educational Research International, vol. 3, no. 4, 2014

[6] I. Y. Song, J. D. Maguire, K. J. Lee, N. Choi, X. Hu and P. Chen, “Designing a data warehouse for cyber crimes”, Journal of Digital Forensics, Security and Law, vol. 1, no. 3, pp. 1, 2006.

[7] J. J. Xu and H. Chen. ”Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks”. Decision Support Systems, vol. 38, no. 3, pp.473-487, 2004.

[8] D. Anggraini, S. Madenda, E. P. Wibowo and L. Boumedjout, “Network Disintegration in Criminal Network”, In 11th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), pp. 192-199, 2015.

[9] R. Kimball and M. Ross, “The data warehouse toolkit: the complete guide to dimensional modeling”, John Wiley & Sons. 2011.

[10] T. Niemi, M. Niinimäki, J. Nummenmaa and P. Thanisch, “Constructing an OLAP cube from distributed XML data”, In Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP , pp. 22-27, 2002.

[11] P. J . Phillips, “Lone wolf terrorism”, Peace Economics, Peace Science and Public Policy, vol. 17, no. 1, 2011.

 ISBN: 978-1-61208-729-0