Cluster Analysis of IT Security Risks in Chosen Sectors

Ante Buljan
European Central Bank, Germany
Mario Spremić
Faculty of Economics and Business, University of Zagreb, Croatia

Abstract

The problems of digitalisation and transition of companies into the digital markets has become one of the crucial issues in contemporary business. Digital transformation is changing markets and interactions. These trends impose a question on how secure is this environment and how companies are combating this issue. This new environment shows us how knowledge is dispersed across a global market and in individual, national, markets. The goal of the research is to investigate the differences between countries in Europe according to how their companies tackled the challenges of IT security. Clustering is conducted by the use of simple k-means method using the data on European countries available in Eurostat. The digital divide has been found among European countries according to their usage of investigated IT security practices.

Keywords: information technology security, digital markets, Europe, business transformation, clustering, k-means
JEL classification: O57

Introduction

Digital transformation has impacted various markets and changed the ways we think about and conduct business overall. These changes are effecting the security for all those who interact with the markets. Buttarelli (2017) shows how these changes have had an impact on society in general. According to him the social changes are so huge that all of the basic human rights would need to be redefined. This transformation is showing the need to protect the information in the digital environment. This is done through the use of IT security by implementation of protective, effective control measures and policies.

The information technology security (IT security) has become a competitive advantage in this condition, while some economic sectors have already shown a big step forward in the standardisation of good practices, such as banking and insurance industries. The banking and insurance industries are heavily regulated by the controlling agencies. They are required to use at least the acceptable secure practices in digital environment. The companies lacking relative IT security management will be exposed to wide range of IT-threats. This would lead to negative impact on their customers, business partners, employees and the entire ecosystem. For these reasons inappropriate IT security practices would have negative impact on reputation.

There is a high level of consensus regarding the fact that digital transformation has already begun and is proceeding to gain momentum as we can see from the works from Spremić (2017b), Shaughnessy (2018) and Mićić (2017). It can be stated that the digital environment is usually unknown to the companies that are trying to digitalise

their business and that they are unprepared for these new technologies especially for information security (Simpson, 2016; Spremić, 2017a; Zerzan, 2009). IT security also has an impact on the markets themselves (Ding, Yan and Deng, 2016), but also we can see there exists an impact of the market on IT security as well (Kolfal, Patterson and Yeo, 2013). Zerzan (2009) shows that at least some understanding of IT security is beneficial. Innovations like cloud computing are changing the landscape and introducing new dangers (Loske, 2015). Sørensen and Puigvert Gutiérrez (2006) discuss the harmonisation of financial markets in the European Union. Knotek (2014) and Christensen (2011) investigate trends in the harmonisation of the markets and future tendencies. As Warfield (2012) argues, private ownership causes a higher level of incompliance and lower standards in IT security in general. Therefore, Cain (2010) stresses the need for regulations and standardisation. Steffee (2010) and Semer (2012) investigate the awareness of IT security in the context of the human element, which is less considered in the research than the technological in companies. Unfortunately, that leaves a big gap for malicious activities. Military sector has also shown a significant interest in IT security, and security aspects and potential uses can be found detailed in NATO conference research papers (Kowalik, Gorski, and Sachenko, 2004) as well as reinforced confirmation of interest a few years later by Yim, Castiglione and You (2014).

The main objective of this research is to investigate the state of IT security and compare it in three sectors of the economy. We use k-means cluster analysis, with the goal to investigate the digital divide between European countries according to their utilization of IT security.

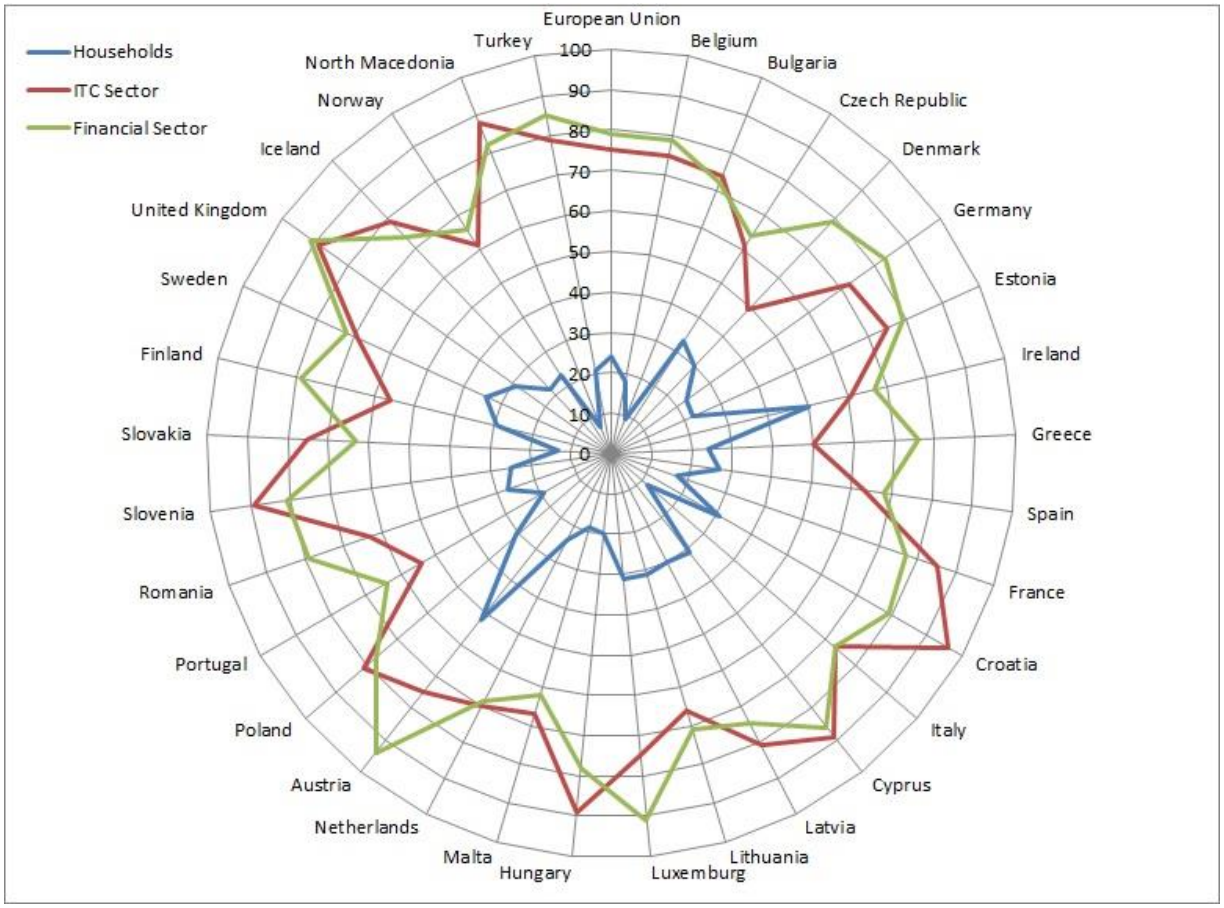
In our research, we use the data about IT security component in the digital environment, that is available from Eurostat. This research discusses the percentage of individual entities that have not reported a security incident in past 12 months. We compare three sectors (households, ICT sector, and financial sector) and by using the cluster analysis, we will show grouping of countries by showing their tendencies toward the development of these components. In this part of the text, we will review a choosing of data and the data itself. Further, we will explain the choice of the used methodological approach of the cluster analysis and explain the methodology of this approach itself. At the end, we show the results of this analysis and further discus them and produce the conclusions of our analysis.

Methodology

Data

We limit our analysis to selected European countries. Data from Eurostat has been used for the year 2010, about the persons and companies who did not experience a security indicates in the 2010 year in three sectors: households, ICT sector, and the financial sector. European countries are used for this analysis due to their historical common heritage and ongoing integrating processes. The dataset had the highest availability of data for countries in European Union and countries in immediate surroundings. Figure 1 indicates large differences in IT security between business sectors and households. The business sectors how very high percentage of unaffected companies. At the same time households are indicating very low percentage of unaffected households.

Figure 1
Percentage of Subjects Who Did Not Experience a Security Incident, 2010



Source: Authors' work based on Eurostat (2019a; 2019b)

Cluster analysis

K-means clustering is when using heuristic methods like Lloyd's algorithm, easy to implement and apply to large sets of data. Thus is successfully used in different areas, from market segmentation, computer geostatic, and astronomy to agriculture. It is commonly used as a pre-processing step for other algorithms, like finding starting configuration.

K-means clustering is used as a step for partially supervised learning. In this use, clustering is conducted in a large data set, which needs to be marked. After that supervised learning is conducted and for each marked pattern distance of each of k learnt central clusters is computerised as to become *k extra* characteristic for the pattern. Characteristics can be Boolean with value 1 for closed centres or some smooth transformation for far away transforming the pattern of clusters through Gauss RBF. It contains hidden layers of radial base network function.

In a given set of observations (x_1, x_2, …, x_n), where each observation is a d-dimensional realistic vector, *k-means clustering* aims to partition n observations in k sets ($k \leq n$) $S = \{S_1, S_2, \ldots, S_k\}$, in such a way that it minimises the *within-cluster sum of squares* – WCSS

$$\arg\min_{s} \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| x_j - \mu_i \right\|^2 \qquad (1)$$

μ_i is the main point in S_i.

Most commonly used algorithm uses repetitive clearing techniques. Although high representation is called *k-means* algorithm, and also Lloyd's algorithm, especially in the computer branch. After assigning an initial set of *k-means* m1(1),…, mk(1) to the algorithm, algorithm alternates between two steps:

- Assigning step: Each cluster is assigned observation whose significance is closest to it (observations are partitioned according to Voronoi's diagram generated by significance).

$$S_i^{(t)} = \{x_p: \left\| x_p - m_i^{(t)} \right\| \leq \left\| x_p - m_j^{(t)} \right\| \forall 1 \leq j \leq k\} \quad (2)$$

  Where each x_p is assigned exactly one $S^{(t)}$, and if possible, it is assigned to two or more.
- Updating step: Recalculates new significance which needs to become the centre of new observation in a cluster.

$$m_i^{(t+1)} = \frac{1}{S_i^{(t)}} \sum_{x_j \in S_i^{(t)}} x_j \quad (3)$$

- The algorithm stops in a step when there are no more changes.

In comparison to computing complexity, *k-means clustering* problem of observations in d dimensions is: (i) NP-weights in Euclidean space d even for two clusters; (ii) NP-weights for generalised number of clusters, and (iii) If k and d (dimensions) are corrected, problem can be precisely solved in time O (ndk+1 log n), where n represents the number of units that need to be grouped

While the possible variations of this algorithm are as follows: (i) Clustering by the method of phases of C-median values is softer version of K-means where every point of data has a Fuzzy degree of belonging for each cluster; (ii) Gauss model of mixture in combination with expected minimisation algorithm (EM algorithm) reflects the probability of assigning a cluster; (iii) Few methods have been suggested for choosing the best starting clusters. One of the newer proposed methods is k-means++; (iv) Purification algorithm uses a K-D tree for accelerating every k-means step; (v) Some methods intend to accelerate each k-means step by using corset or triangle inequality; (vi) Spherical k-means clustering algorithm is used for directional data, and (vii) Minkowski metric weighted k-means is facing noise problems.

Results

The cluster analysis has been conducted based on 32 European countries. The average % of subjects that encountered at least one IT security incident of all countries for each sector is as follows: (i) Households - 25.4375%; (ii) ICT Sector - 72.7813%; and (iii) Financial Sector - 76.0938%. The clustering has been conducted based on all three sectors in each country. The average results for each of the clusters are as follows in Table 1. We can notice that in all of our clusters, the households are showing drastically smaller results than business sectors. The cluster distribution of the countries is shown as described in Table 2.

Table 1
Average of % of Subjects Who Did Not Experience a Security Incident Across Clusters

	Total	Cluster A	Cluster B	Cluster C	Cluster D
Households	25.4375	34.8	27.3	27.5	19.7692
ICT Sector	72.7813	84.8	64.7	54.5	80
Financial Sector	76.0938	88.2	66.6	78.25	78.0769

Source: Authors' work based on Eurostat (2019a; 2019b)

Table 2
Distribution of Countries Across Clusters

Cluster	Country
Cluster A	Croatia, Cyprus, Luxemburg, Austria, United Kingdom
Cluster B	Belgium, Bulgaria, Germany, Estonia, France, Italy, Latvia, Hungary, Poland, Slovenia, Iceland, North Macedonia, Turkey
Cluster C	Denmark, Greece, Romania, Finland
Cluster D	Czech Republic, Ireland, Spain, Lithuania, Malta, Netherlands, Portugal, Slovakia, Sweden, Norway

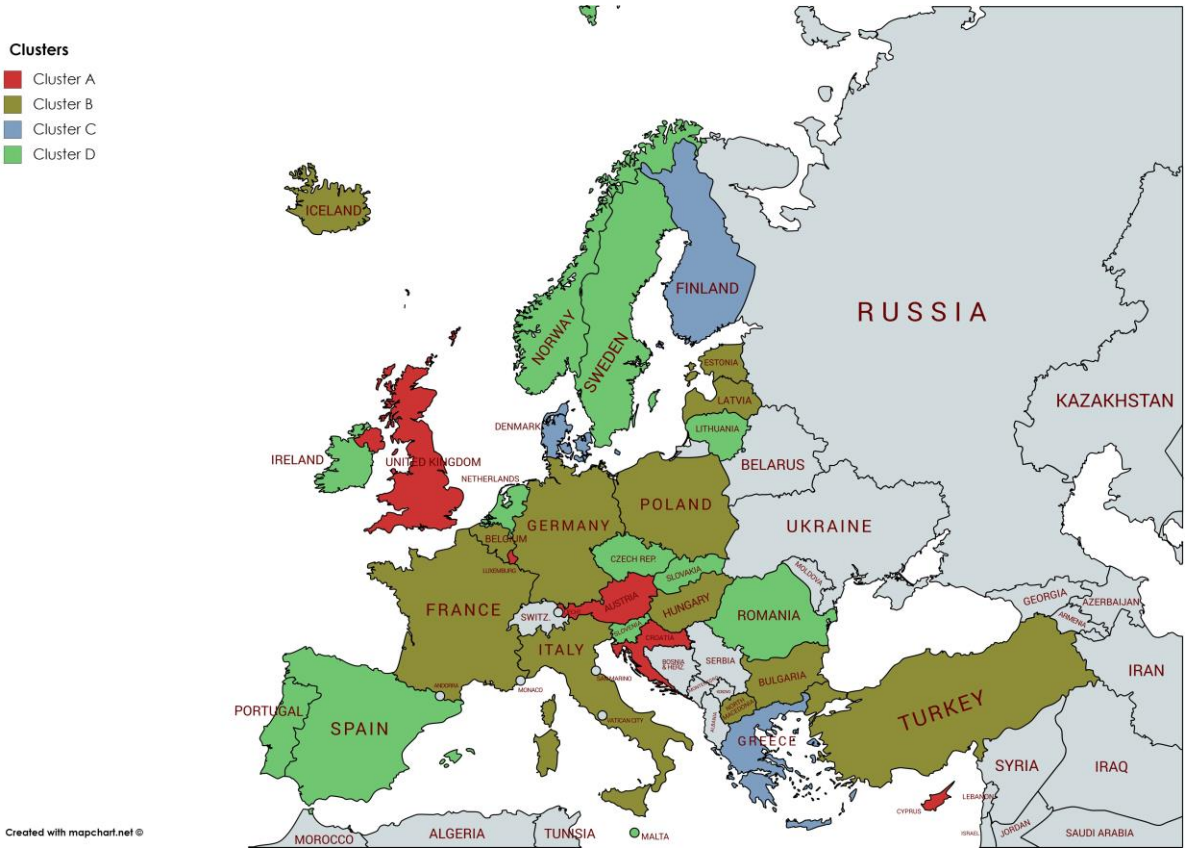
Source: Authors' work based on Eurostat (2019a; 2019b)

Figure 2 presents the geographical distribution of clusters. These clusters are better shown on a map of Europe as described by Figure 2. In this way, we can easily see the dispersion of the clusters across Europe.

Regarding the clusters themselves, we can consider the Cluster A as the best cluster as it shows the highest comparative values in all three reviewed sectors. Countries in this cluster show highest harmonisation and effectiveness of IT security policies. This can be explained through common factor that they all had a relatively developed infrastructure. This is due to the historical investments in infrastructure or relative size of the country. This could be considered a prerequisite for advancement and development of IT security as well as digital markets in general. Cluster B contains the countries that have fallen behind but have the same development regarding the households sector. Cluster C contains the countries that have a disproportion in the business sector thus. This cluster shows countries that are not developed but are still in line with policies and regulation (due to higher result in the financial sector). Obviously, cluster D has harmonised business sectors but shows the lowest results in the households sector comparatively to all other clusters.

All clusters and average values show a significant difference between households and business sectors. Households represent both a source of income and work for business sectors. As households represent employees they can also be considered directly connected to the business sectors. The employees have a tendency to disregard their good practices in their own private environment. Through this connection malicious intent can be reflected directly to the companies they work for. Risk from households can be thus transferred to the businesses. Awareness on IT security Is not only important in the work place but in private environment as well. Particular malicious methods (social engineering for instance) are more easily conducted when people feel safe and confident.

Figure 2
Clusters in Europe



Source: Authors'' work using according to the cluster analysis and the information in Table 2, using mapchart.net

Conclusion

IT security will certainly be considered a necessity in the future for both business and household sectors since innovations and development (in technological advancement but also malicious capabilities) as well as the transformation of markets into the digital environment is stronger than ever and does not show any sign of slowing down.

Our result from cluster analysis actually shows a very high harmonisation in the whole of Europe. The highest level of harmonisation is as expected in the financial industries. This is due to the regulations that exist in this industry regarding IT security. The ICT sector, which would be expected to show the leading role in this field as an innovator but also a source of best practices, does not necessarily follow this rule.

The results show a big difference in averages in households and business sectors. This indicates that harmonisation between them is not high. Businesses in digital environments should consider further investments in employees IT security. In such a way spill over of negative effects from households could be avoided.

The limitations of the dataset available have forced us to disregard some European countries due to lack of data, which is not available for all countries. Further limitation is usage of the data for 2010 due to the fact that harmonised data for households, financial sector and ICT sector was available only for the year 2010. Therefore, further research in regards to expanding to other possible economic sectors should also be considered.

References

1. Buttarelli, G. (2017), "Privacy matters: updating human rights for the digital society", Health and Technology, Vol. 7, No. 4, pp. 325-328.
2. Cain, A. (2010), "Impact of Regulation Is Top Concern", Internal Auditor Journal, Vol. 67, No. 5, p. 14.
3. Christensen, J. F. (2011), "Industrial evolution through complementary convergence: the case of IT security", Industrial and Corporate Change, Vol. 20, No. 1, pp. 57-89.
4. Ding, W., Yan, Z., Deng, R. H. (2016), "A Survey on Future Internet Security Architectures", IEEE Access, Vol. 4, pp. 4374-4393.
5. Eurostat (2019a), "Security incidents and consequences (isoc_cisce_ic) dataset", European Commission, available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_cisce_ic&lang=en (5 July 2019)
6. EUROSTAT (2019b), "Security related problems experienced through using the internet for private purposes (isoc_cisci_pb) dataset", European Commission, available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_cisci_pb&lang=en (5th of July 2019)
7. Knotek, P. (2014), "Banking sectors in EMU – Cluster analysis banking sectors in EMU", European Scientific Journal, Vol. 10, No. 34, pp. 60-71.
8. Kolfal, B., Patterson, R. A., Yeo, M. L. (2013), "Market Impact on IT Security Spending", Decision Sciences, Vol. 44, No. 3, pp. 517-556.
9. Kowalik, J. S., Gorski, J., Sachenko, A. (2004), "Cyberspace Security and Defense: Research Issues", Proceedings of the NATO Advanced Research Workshop, Gdansk, Poland, Springer.
10. Loske, A. (2015), IT Security Risk Management in the Context of Cloud Computing, Darmstat, Springer.
11. Mićić, L. (2017), "Digital Transformation and Its Influence on GDP", Economics, Vol. 5, No. 2, pp. 135-147.
12. Semer, L. J. (2012), "Evaluating the Employee Security Awareness Program", Internal Auditor Journal, Vol. 69, No. 6, pp. 53-56.
13. Shaughnessy, H. (2018), "Creating digital transformation: Strategies and steps", Strategy & Leadership, Vol. 46, No. 2, pp. 19-25.
14. Simpson, W. R. (2016), "Securing Information Systems in an Uncertain World Enterprise Level Security", Journal of Systemics, Cybernetics and Informatics, Vol. 14. No. 2, pp. 83-90.
15. Sørensen, C. K., Puigvert Gutiérrez, J. M. (2006), "Euro area banking sector integration using hierarchical cluster analysis techniques", Working paper No. 627, European Central Bank, Frankfurt am Main.
16. Spremić, M. (2017a), Sigurnost i revizija IS-a u okruženju digitalne ekonomije (Security and IS revision in digital economy environemnt), Faculty of Business and Economics, Zagreb.
17. Spremić, M. (2017b), Digitalna transformacija poslovanja (Digital transformation of business), Faculty of Business and Economics, Zagreb.
18. Steffee, S. (2010), "Employees Ignoring IT Security", Internal Auditor Journal, Vol. 67, No. 5, pp. 14-16.
19. Warfield, D. (2012), "Critical Infrastructures: IT Security and Threats from Private Sector Ownership", Information Security Journal: A Global Perspective, Vol. 21, No. 3, pp. 127-136.
20. Yim, K., Castiglione, A., You, I. (2014), "Prosperity of IT security technologies in homeland defense", Journal of Ambient Intelligence and Humanized Computing, Vol. 5, No. 2, pp. 169-171.
21. Zerzan, A. (2009), "New Technologies, New Risks? Innovation and Countering the Financing of Terrorism", World Bank Working Paper No. 174, World Bank, Washington, DC.

About the authors

Ante Buljan, MA, is an IT specialist working for European Institutions. He received his MA at Faculty of Business and Economics at the University of Zagreb. During his studies, he was also collaborating with the Department of Informatics and Department of Macroeconomics and Development in the capacity of a student assistant. His main research interests are Information security, digital communication channels, electronic and distance learning, the digital transformation of business, and macroeconomic aspects of digitalisation. Ante is currently employed as an IT Specialist at European Central Bank, Germany. The author can be contacted at **antebuljan1994@gmail.com**.

Mario Spremić is a full professor at the Department of Informatics, Faculty of Economics and Business (FEB), University of Zagreb, Croatia, and a guest lecturer at several international institutions (such as Imperial College London). He holds a B.Sc. in mathematics, and M.Sc. in IT management and Ph.D. in business, all from the University of Zagreb. He joined FEB (Zagreb), in 2000, with previous corporate experience as a computer programmer and project manager. Mario has participated in executive education programs at MIT Sloan School of Management and EFMD Executive Academy. He has broad experience in international accreditation of higher education institutions (EQUIS, AACSB, EPAS peer-review). His main research interest areas are digital computing, the digital economy, ICT governance cyber security, and IT auditing. The author can be contacted at **mspremic@efzg.hr.**