ICT security analysis

Kristiina Köster, Tõnis Rahe, Mai Ristioja

Introduction

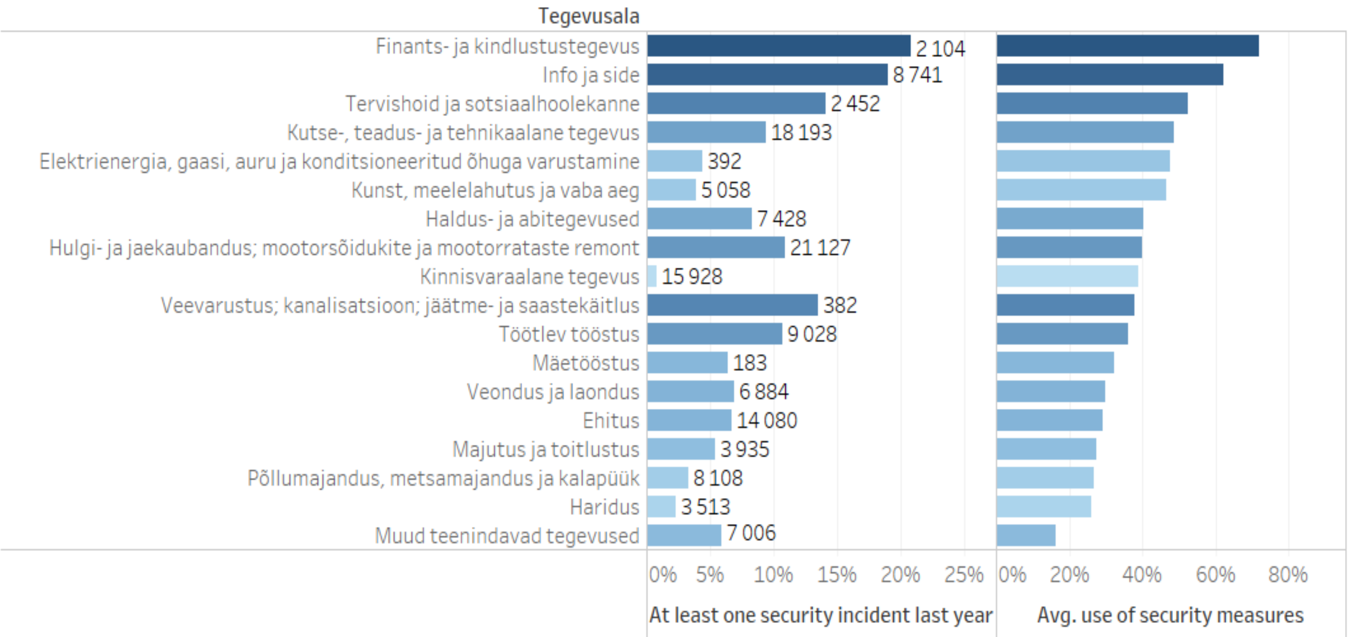
Security breaches may lead to great losses to companies in terms of money, trust, information etc. The goal of this project was to find and show correlations between different security measures and the volume of security breaches, thus gaining insight how effective different security measures might be. We worded three hypotheses:

1. Using positive security practices correlates with more security incidents being discovered.
2. Security incidents are more likely being discovered in companies with ICT specialists
3. More security measures are being implemented in companies of information, communications and finance sectors.

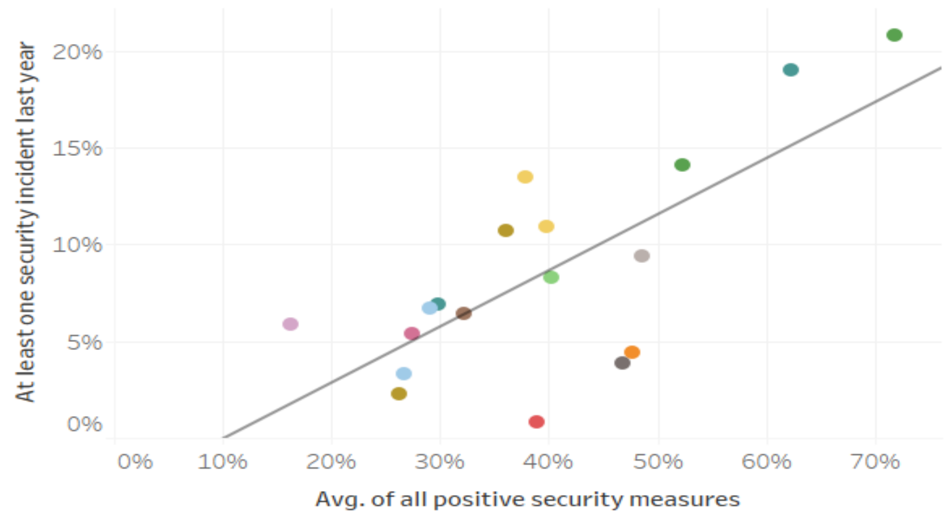
Data

Our analyses was based on two datasets found from the database of Statistics Estonia (Statistikaamet). Both consist of data about different ICT security aspects in companies (companies with more than 50 workers are all represented, smaller companies are chosen randomly) according to the fields of activity and the sizes of the companies. We analysed values from the year 2019. All input values are aggregated and shown as percentages of the total number of companies. Separately, we also extracted from the Estonian tax authority dataset the number of companies in each different field of activity.

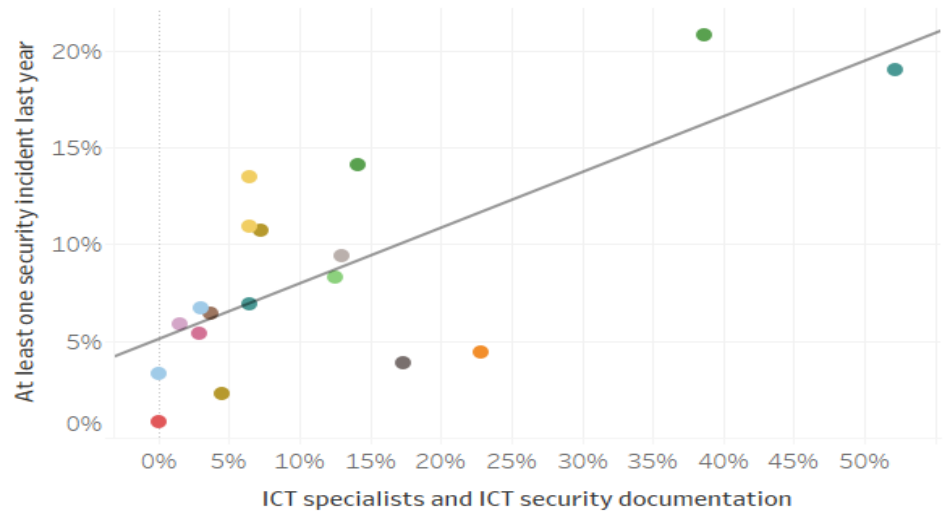
Incidents By Activity



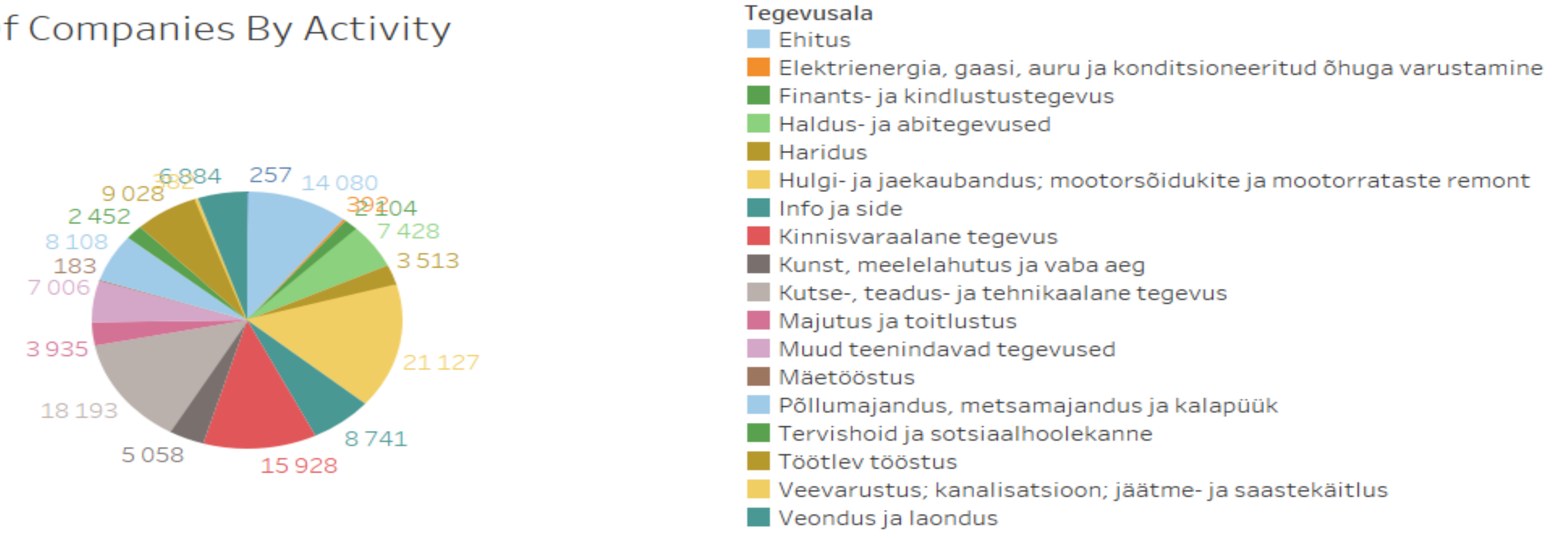
All Positive Practices By Activity



ICT Specialists By Activity



Numbers Of Companies By Activity



Link to our Github repository: https://github.com/mairistioja/ICTsecurityAnalysis

Methods

We found correlations between different attributes and used permutation tests to check their significance.

We performed visual analysis of existing data using Tableau Desktop.

Additionally, we also tried to find new correlations by brute force in attempts to find any other interesting patterns in input data. These were also filtered by using permutation tests.

Results & Conclusions

As of our first hypothesis, using positive security practices did correlate with more security incidents being discovered, but we could not prove its significance. The correlation coefficient was 0.718 and even though the correlation could be guessed from the visualisation, it did not pass the permutation test which gave a threshold of 0.797. The same happened with our second hypothesis about the discovery of security incidents being more likely in companies with ICT specialists. The correlation coefficient of 0.722 did not exceed a permutation test threshold of 0.832. Hence we could not prove these correlations to be statistically significant.

Our third hypothesis about more security measures being implemented in companies of information, communications and finance sectors was easily provable directly from the input data, even by visual inspection.

Because we were dismayed about our hypothesis, not finding anything interesting and but still wanting to discover something interesting from the dataset, we set out to find new interesting correlations by brute force (made possible by modern computing power). For each relevant pair of attributes in our input data we calculated their correlation and run a permutation test to filter out the significant correlations. However, this only produced strong correlations (|R| > 0.90) for attributes which are by common sense known to be related, related by definition (e.g. between the attributes "Enterprises with ICT specialists" and "Enterprises which provided ICT training for ICT specialists in the previous year") or were otherwise uninteresting correlations between security practices being used together.