Subjektive Risikobewertung – über Datenerhebung und Opinion Pooling

Stefan Rass[1] · Jasmin Wachter[1] · Stefan Schauer[2] · Sandra König[2]

[1] Alpen-Adria Universität Klagenfurt
Forschungsgruppe Systemsicherheit
{stefan.rass | jasmin.wachter}@aau.at

[2]AIT Austrian Institute of Technology GmbH
Center for Digital Safety & Security
{stefan.schauer | sandra.koenig}@ait.ac.at

Zusammenfassung

Die subjektive Bewertung von Risiko stellt Experten oftmals vor die Herausforderung einer Quantifizierung unscharfer Größen (wie etwa „Security“, „Gefahr“ oder Ähnliches) auf Grundlage unvollständiger Informationen. Die dafür empfohlene kategoriale (qualitative) Risikobewertung erleichtert die Arbeit der ExpertInnen, erschwert jedoch gleichermaßen die Weiterverarbeitung der Risikodaten auf Grundlage mathematischer oder statistischer Modelle. In diesem Beitrag stellen wir eine Form der Datenerhebung vor, welche eine Spezifikation qualitativer Risikowerte unter (gleichzeitiger) Angabe von Unsicherheit über die geäußerte Meinung (im Sinne einer Quantifizierung) ermöglicht. Aus den so erhobenen Daten lassen sich sehr einfach und direkt statistische Modelle für die Risikobewertung ableiten, welche in vielfältiger Form weiterverarbeitet werden können. Insbesondere ermöglichen geeignet erhobene Daten ein sogenanntes „Opinion Pooling“, welches langfristig zu einer Konsensfindung über eine Risikosituation führen kann. Dies ermöglicht eine Verbesserung der Risikobewertungen im Rahmen des Plan-Do-Check-Act-Zyklus auf Grundlage statistischer Modelle.

1 Einführung

Die Bewertung von Risiken und das Risikomanagement im Allgemeinen stellen heute einen zentralen Aspekt in der strategischen Führung von Unternehmen dar. Vor allem im Bereich der Informations- und Kommunikationstechnologie (IKT) gibt es in den letzten Jahren eine steigende Anzahl an Bedrohungen und Zwischenfällen. Diese haben gezeigt, dass sowohl Klein- und Mittelbetriebe (KMB bzw. KMU) als auch große Konzerne und kritische Infrastrukturen im Fokus von Angreifern stehen (siehe Vorfälle wie die WannaCry Ransomware [Bren17] oder der Ausfall des Stromnetzes in der Ukraine [Zet16], etc.). Dabei ist es jedoch nicht genug, technische und organisatorische Maßnahmen zu ergreifen, um diese Angriffe abzuwehren. Vielmehr ist es wichtig, vorab die bestehenden Bedrohungen und die damit verbundenen Risiken zu bewerten und zu priorisieren, um präventive Maßnahmen zu identifizieren und in den Unternehmen umzusetzen.

Betrachtet man in diesem Kontext die gängigen (teilweise standardisierten) Vorgehensmodelle für Risikomanagement (wie etwa die ISO 31000 [ISO09], Cobit 5 for Risk [ISAC13] oder

P. Schartner · A. Baumann (Hrsg.) · D•A•CH Security 2017 · syssec (2017) 225-237.

Octave [CSYW07]) so geht der Trend weg von einer quantitativen Bewertung der Eintrittswahrscheinlichkeiten sowie der Auswirkungen eines Ereignisses und hin zu einer Einschätzung basierend auf kategorialen Skalen (wie bereits durch das BSI vorgeschlagen [Mün12]). Dabei werden sowohl die Auswirkungen als auch die Wahrscheinlichkeiten anhand einer fixen Anzahl von vorgegebenen Kategorien (z.B. „niedrig“ „mittel“ und „hoch“) eingestuft. Diese Kategorien können abhängig vom Sicherheitsziel (finanzielle Mittel, Reputation, etc.) unterschiedlich definiert werden und unterschiedliche Semantik besitzen. Insbesondere im IKT-Bereich ist ein qualitativer Ansatz für die Risikobewertung vorzuziehen, da hier meist die notwendigen Informationen über Zwischenfälle aus der Vergangenheit fehlen, um exakte quantitative Aussagen über Wahrscheinlichkeiten und Auswirkungen treffen zu können. Zudem gibt es vor allem im Kontext von intentionalen Angriffen unterschiedlichste Faktoren (wie etwa die finanziellen Mittel oder die Motivation eines Angreifers), welche diese Größen beeinflussen können.

Ähnlich wie bei einem quantitativen Ansatz (also bei der Arbeit mit numerischen Werten) können auch bei einem qualitativen Ansatz mathematische Methoden und Modelle zur Risikobewertung eingesetzt werden. Zu beachten ist, dass hierbei der Umfang und die Qualität der kategorialen Daten ausschlaggebend für die Qualität der Ergebnisse ist. Ein aktuelles Beispiel bietet das EU FP7 Projekt HyRiM[1], in welchem qualitative Daten (z.B. aus Expertenmeinungen oder aus Simulationen) für eine spieltheoretische Risikooptimierung verwendet werden. Ein zentraler Punkt dabei ist die *verlustfreie* Aggregation der verfügbaren Daten in Form von Verteilungsfunktionen. Dies birgt den Vorteil, dass alle verfügbaren Daten mit in die Bewertung einfließen und nicht nur Extremwerte betrachtet werden (wie es etwa im Maximum-Ansatz der Fall ist, welcher bei vielen Standard-Frameworks zum Risikomanagement Anwendung findet). Dieser Ansatz der verlustfreien Aggregation wird auch im derzeit laufenden nationalen Förderprojekt CERBERUS[2] (FFG/KIRAS Programmlinie) angewendet und weiterentwickelt.

Ein entscheidendes Manko von qualitativen Bewertungen ist, dass diese von ExpertInnen getroffen werden, welche ihre subjektive Wahrnehmung in die Bewertung einbringen. So können etwa bestimmte Personen ein Szenario eher vorsichtig oder zu pessimistisch einschätzen, was die Wahrnehmung in Bezug auf das Risiko des Szenarios beeinflusst. Im Gegensatz dazu kann natürlich eine eher unbedarfte Bewertung ebenfalls in einer zu geringen Einschätzung resultieren. Hier ist es wichtig, einen Mittelweg zu finden, um die unterschiedlichen Bewertungen zusammenzuführen und zu einem konsolidierten Ergebnis zu kommen. Einen Ansatz, der diese Harmonisierung umsetzt, wollen wir in diesem Artikel vorstellen.

Zusätzlich zu den subjektiven Einschätzungen ist ebenfalls noch zu beachten, dass so erhobene Daten auch mit einer gewissen Unsicherheit behaftet sind. Einzelne Personen haben zum Beispiel nicht den gesamten Überblick über einen Bereich im Unternehmen und können daher den vollen Umfang der Auswirkungen eines Ereignisses nur bedingt einschätzen. Klassische Ansätze nötigen ExpertInnen jedoch häufig dazu, sich für einen festen Wert auf einer qualitativen Skala zu entscheiden. Die Unsicherheit im Bewertungsprozess kann hierin bisweilen nicht erfasst und damit auch nicht berücksichtigt werden. Wir stellen nachfolgend einen Ansatz vor, der den Grad der Unsicherheit bei der Einschätzung intuitiv leicht spezifizieren lässt, sodass die Angaben über subjektive Unsicherheit in weiterer Folge Berücksichtigung finden. Insbesondere wird die Angabe über die Unsicherheit beim Opinion Pooling berücksichtigt. Konkret

[1] HyRiM – Hybrid Risk Management for Utility Networks, Online: www.hyrim.net

[2] CERBERUS - Cross Sectoral Risk Management for Object Protection of Critical Infrastructures

führen wir die erhobenen Daten in der Form einer Risikobewertung mit Unsicherheit („Toleranz“) einer zweifachen „Harmonisierung“ zu: im ersten Schritt werden subjektive Effekte (Bias) einer Risiko-Über- bzw. -Unterschätzung ausgeglichen, und im zweiten Schritt ein Risiko-Konsens (im Sinne einer repräsentativen Gesamteinschätzung der Lage) gesucht.

2 Datenerhebung

Um ExpertInnen eine möglichst intuitive Möglichkeit zu bieten, Unschärfe bei Risikodaten sowohl bei der Eingabe als auch bei der Ausgabe zu berücksichtigen, ziehen wir eine ansonsten für die Ausgabe verwendete Darstellungsform als graphisches Eingabesystem in Betracht. Im Rahmen der Risiko-Priorisierung werden Bedrohungen zumeist in zweidimensionalen Diagrammen (auch *Risiko-Matrix* genannt) dargestellt, worin die Achsen jeweils qualitative Skalen für die Eintrittswahrscheinlichkeit bzw. die Auswirkung der Bedrohung angeben (das Produkt der beiden Faktoren liefert dann die entsprechende Risikobewertung). In gleicher Form kann das Risiko auch für die Eingabe quantifiziert werden, etwa indem der/die ExpertIn einen rechteckigen Bereich (Box) spezifiziert, welcher die folgende Semantik besitzt:

- Der Mittelpunkt der Box markiert per x- und y-Koordinate die jeweilige Bewertung der Parameter „Impact“ und „Likelihood“.
- Die Höhe bzw. Breite der Box stellt die Unsicherheit über die entsprechende Aussage dar.

Zur Erleichterung der Eingabe bei kategorialen Bewertungsskalen bietet sich zusätzlich die Anzeige einer textuellen Beschreibung der Kategorie an, sodass bei einer unsicheren Bewertung, d.h. (subjektiv) unklaren Zuordenbarkeit einer Bedrohung zu einer einzelnen Kategorie, die Box sich über mehrere Risikokategorien erstrecken kann. Abbildung 1 zeigt die vorgeschlagene Eingabeform für die Risikospezifikation.

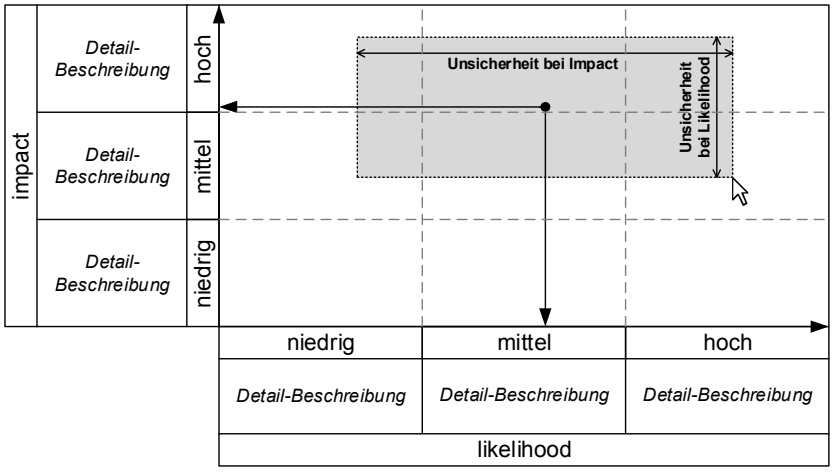

Abb. 1*:* Graphische Risikoeingabe mit Unsicherheit

Die hier beschriebene Datenerhebung kann etwa in Form von *(Online-)Umfragen* erfolgen, sodass insbesondere eine Reihe qualitativer Vorteile entstehen:

- *Asynchrone Bewertung*: Persönliche Treffen und Diskussionen (insbesondere die damit verbundene Terminfindung und persönliche Anwesenheit bei Besprechungen) können im Vorfeld vermieden werden. Eine Diskussion wird erst beim Treffen einer Entscheidung

erforderlich. Die Abgabe (persönlicher) Risikoeinschätzungen kann zeitlich beliebig und damit asynchron von anderen befragten Personen erfolgen.

- *Anonyme Datenerhebung*: Im Gegensatz zu persönlichen Meetings können bei lokal und zeitlich unabhängigen (Online-)Umfragen Rückschlüsse auf individuelle Meinungen verhindert werden. Dies vermeidet etwa sozial oder kulturell bedingte Effekte einer gehemmten oder veränderten Meinungsäußerung in Anwesenheit anderer Personen (wie etwa Vorgesetzten).
- *Ausnützen einer Matrix-Organisation*: Die Befragung unterschiedlicher Personen zu vielfältigen Aspekten eines Risikos (etwa Kosten, Reputation, etc.) erlaubt es, die eigenen Angaben auf Aspekte zu beschränken, die qualifiziert bewertet werden können. Ebenso kann der Personenkreis auch über die Organisationsgrenzen hinaus ausgeweitet werden (z.B. bei Befragung von KundInnen in Bezug auf Reputation).
- *Multiple Sicherheitsziele*: Bei Erhebung von mehreren Risikodimensionen (finanzieller Schaden, Reputation, etc.) besteht die Möglichkeit einer multikriteriellen Entscheidungsfindung [RaRa14, RSPG13]. Wir gehen auf diese Variante nachfolgend nicht näher ein und beschränken uns auf die Abfrage eines einzelnen Risikoaspekts (Betrachtung eines einzelnen Sicherheitsziels).

Um subjektiv unterschiedliche Sichtweisen im Rahmen der Dateneingabe „auszugleichen" besteht die Möglichkeit, die Skala nach den eigenen Vorstellungen „anzupassen", etwa indem die Bereiche für niedriges, mittleres oder hohes Risiko gemäß den eigenen Vorstellungen verschieden breit eingestellt werden können. Wird etwa – am Beispiel monetärer Verluste – der Bereich zwischen 10.000 und 1.000.000€ als mittlerer Verlust (für das Unternehmen der gegebenen Größe) empfunden, so kann der Bereich entsprechend groß gestaltet werden, wenn die vorgegebene Skala bis 1.5Mio € reicht. Eine andere Person mag Verluste > 500.000€ bereits als hoch empfinden, und kann somit den mittleren Bereich auf der angebotenen Skala entsprechend kleiner wählen. Hierdurch kann die angebotene Skala der eigenen Vorstellung angepasst werden, was genauere Risikoangaben unterstützt.

3 Opinion Pooling

Wir betrachten nachfolgend ein klassisches Verfahren zur Konsensfindung und erweitern dieses um die Berücksichtigung von Unsicherheiten (hybrides Pooling).

3.1 Klassisches Konsens Opinion Pooling

Das in Abschnitt 2 beschriebene (technische) Datenerhebungsverfahren liefert pro Bedrohung vier Parameter: Impact $\mu_I \pm$ Toleranz t_I und Likelihood $\mu_L \pm$ Toleranz t_L. Diese Angaben über Unsicherheit und Unschärfe lassen sich etwa als Parameter von Wahrscheinlichkeitsverteilungen interpretieren. Im konkreten Fall verwenden wir Gauß-Dichten mit entsprechenden Mittelwerten μ_I, μ_L und Standardabweichungen $\sigma_I := \frac{t_I}{3}$ und $\sigma_L := \frac{t_L}{3}$. Durch diese spezielle Wahl der Varianzen sind mehr als 99.7% der Wahrscheinlichkeitsmasse innerhalb der Box konzentriert, was der Angabe des Mittelwertes μ innerhalb der Toleranzgrenzen $\mu \pm 3\sigma$ entspricht.

Man beachte, dass dieses Vorgehen bewusst von einer streng kategorialen Modellierung abweicht, da den AnwenderInnen eine stetige Skala für ihre Bewertungen geboten wird, bei der lediglich die Bereiche zwischen Teilstrichen zu qualitativen Kategorien korrespondieren (vgl.

Abbildung 1). Insbesondere entsteht hierdurch keine konzeptuelle Inkonsistenz bei der Verwendung einer stetigen Verteilung zur Modellierung von kategorial bemessenem Risiko, jedoch wird eine einfache und semantisch konsistente Weiterverarbeitung der Daten auf statistischen Grundlagen ermöglicht. Diese Weiterverarbeitung dient der Konsensfindung im Sinne einer repräsentativen Risikoangabe auf Grundlage vieler Expertenmeinungen. Diese aggregierte Einschätzung (etwa ein gewichteter Mittelwert) lässt sich dann wieder Kategorien zuordnen, sodass eine kategoriale Gesamtbewertung wiedergewonnen werden kann (gemeinsam mit einer Aussage über die Unschärfe der Bewertung, welche sich aus der Varianz der aus den Daten konstruierten Verteilung errechnet).

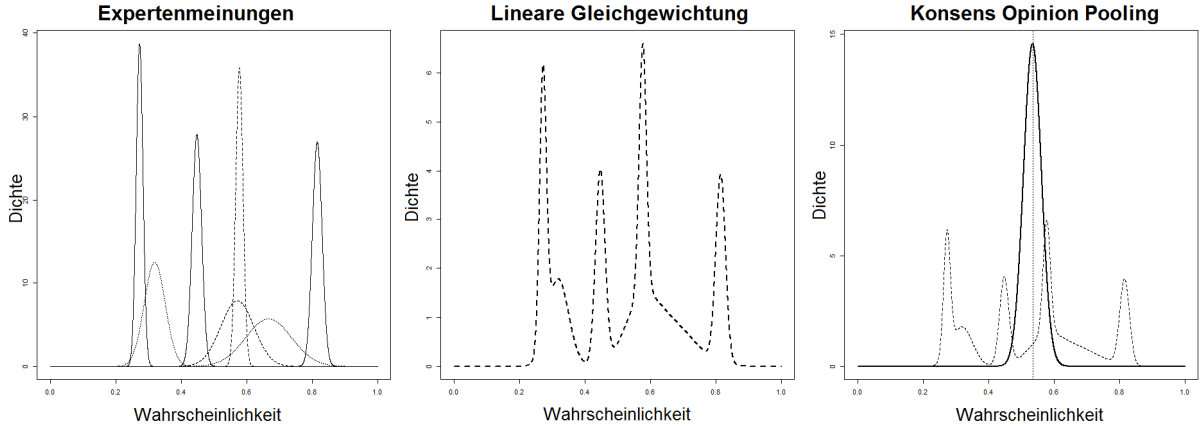

Abb. 2*:* Opinion Pooling

Abbildung 2 zeigt im linken Teil eine Menge von Expertenmeinungen, welche als Gauß-Dichten mit entsprechenden Parametern (auf der hier willkürlich gewählten Skala [0,1]) festgelegt wurden. Das mittlere Bild zeigt die Verteilung, welche sich durch die (äquivalentgewichtete) Linearkombination der entsprechenden Dichten ergibt. Dabei sei angemerkt, dass die Verwendung einer einheitlichen Varianz für alle Verteilungen hier zu einem klassischen Kerndichteschätzer führen würde. Das rechte Bild zeigt das Ergebnis eines bestimmten Opinion-Pooling-Verfahrens (*Konsens Opinion Pooling*), bei dem ExpertInnen ihre Meinung in Abhängigkeit von der „Distanz" zu den Meinungen anderer Experten wiederholt anpassen (die ursprünglichen Meinungen sind hier durch die gepunktete Linie zusammengefasst dargestellt). Dem Pooling liegt die Intuition zu Grunde, dass Personen umso eher ihre Meinung derer anderer Personen angleichen, je näher die fremde Meinung an der eigenen liegt. Umgekehrt ist die Neigung sich einer stark abweichenden Meinung anzuschließen erwartungsgemäß geringer. Es kann für dieses Verfahren sehr einfach eine Konvergenz gegen eine gemeinsame Meinung nachgewiesen werden [CaLa13]. Anzumerken ist, dass bei diesem Verfahren nur *ein* repräsentativer Wert berechnet wird (durchgezogene senkrechte Linie im rechtesten Bild) und keine Verteilung, wie Abbildung 2 suggeriert. Die dort eingezeichnete Glockenkurve besitzt eine Breite, die der Varianz des ermittelten Konsenswertes entspricht, und soll die Unsicherheit über den ermittelten Wert graphisch darstellen.

3.2 Hybrides Opinion Pooling

Lineares Opinion Pooling bzw. Opinion Pooling im Allgemeinen hat jedoch einen erheblichen praktischen Nachteil: die Wahl der Gewichte beeinflusst die Qualität der Vorhersage massiv und es gibt kaum praktische Ansätze zur Bestimmung dieser, weshalb üblicherweise eine Gleichgewichtung der Expertenmeinung, d.h. eine einfache Mittelwertbildung, vorgenommen

wird. Bei normalverteilten und unabhängigen Schätzwerten mit gleicher Unsicherheit ergibt der gewöhnliche Mittelwert den gepoolten Wert mit der geringsten Varianz, liefert also optimale Ergebnisse. Sind die Experten jedoch nicht unabhängig, so führt eine Gleichgewichtung nicht zwingend zum optimalen Ergebnis. Zusätzlich berücksichtigt der so gewichtete Mittelwert die unterschiedlichen (Un-)Sicherheiten der Experten bezüglich ihrer Schätzung nicht. Das oben beschriebene Konsens Opinion Pooling Verfahren ist ein praktischer Ansatz, der das Problem der Wahl der Gewichte löst, jedoch wird in diesem Verfahren die Unsicherheit der Experten nicht mit einbezogen.

Wir schlagen daher eine *adaptierte iterative Methode* vor, die auf obigen Verfahren basiert, jedoch zusätzlich die Unsicherheit der Experten über ihre Schätzangabe berücksichtigt. Demnach adaptieren sehr unsichere Experten eher die Meinung anderer Experten, die sich ihrer Einschätzungen sehr sicher sind. Zusätzlich werden, wie im oben beschriebenen Verfahren, Meinungen von Experten, die sich nahe der eigenen befinden, eher angenommen, als jene, die sich sehr von der eigenen unterscheiden. Im Detail nehmen wir an, dass N Expertenprofile $(\mu_1, \sigma_1), \dots, (\mu_N, \sigma_N)$ gegeben sind, die z.B. über die graphische Risikoeingabe mit Unsicherheit ermittelt wurden. In jeder Iteration adaptiert jeder Experte j seine Einschätzung als gewichtete Summe aller Experteneinschätzungen, d.h., $\mu_j^{(i+1)} = c_{j1}^{(i)} \cdot \mu_1^{(i)} + \cdots c_{jN}^{(i)} \cdot \mu_N^{(i)}$. Die Wahl der Gewichte $c_{jk}^{(i)}$ wird über die Sicherheit der Experten über ihre Einschätzung und die Nähe der Meinungen der Experten j, k zueinander bestimmt. Im Folgenden erläutern wir, wie man die Gewichte ermittelt, wenn man nur die Nähe der Schätzungen zueinander betrachtet. Danach widmen wir uns dem Fall, dass nur die Unsicherheit der Experten die Gewichtung beeinflusst. Schließlich stellen wir unseren *hybriden Ansatz* vor, demnach sowohl Nähe, als auch Unsicherheiten in der Gewichtung berücksichtigt werden.

Wir beginnen mit der Annahme, dass Experten eher Meinungen adaptieren, die „nahe" ihrer eigenen Einschätzungen liegen. Nähe wird in diesem Fall über die absolute Abweichung der Einschätzungen gemessen, d.h. $d(\mu_j, \mu_k) = |\mu_k - \mu_j|$, wobei auch andere Abstandsmaße in Frage kommen würden. Die Gewichtung, die Experte j den einzelnen Expertenmeinungen zuordnet, erfolgt nun proportional zur inversen Distanz, d.h.

$$w_{jk}^{(i)} = \frac{\alpha_j^{(i-1)}}{\epsilon + d\left(\mu_j^{(i-1)}, \mu_k^{(i-1)}\right)}$$

mit einem zuvor gewählten kleinen Tuning-Parameter $\epsilon > 0$. Der Wert $\alpha_j^{(i-1)}$ normiert dabei die Gewichte so, dass $\sum_{k=1}^{N} w_{jk}^{(i-1)} = 1$ gilt. Über $\mu_j^{(i+1)} = w_{j1}^{(i)} \cdot \mu_1^{(i)} + \cdots w_{jN}^{(i)} \cdot \mu_N^{(i)}$ erhält man das in [CaLa13] beschriebene iterative konsensuelle lineare Opinion Pooling.

Als nächstes betrachten wir die Annahme, dass Experten eher Schätzungen von Experten adaptieren, die sich sehr sicher über ihre Werte sind, als von jenen, deren Einschätzungen mit großen Unsicherheiten behaftet sind. Dies entspricht einem bayesschen Update, in dem der Experte j seine eigene Einschätzung (μ_j, σ_j) als $\mathcal{N}(\mu_j, \sigma_j^2)$ normalverteilte a-priori Verteilung betrachtet und die übrigen Expertenmeinungen $(\mu_k, \sigma_k), i \in \{1, \dots, n\}, k \neq j,$ als normalverteilte Beobachtungen mit gegebener Varianz $\sigma_i{}^2$. Die Maximum a-posteriori Schätzer μ_{MAP} und σ_{MAP}^2 ergeben sich dann als

$$\sigma_{MAP}^2 = \left(\frac{1}{\sigma_j^2} + \sum_{k \neq j} \frac{1}{\sigma_k^2}\right)^{-1} \text{ und } \mu_{MAP} = \left(\frac{\mu_j}{\sigma_j^2} + \sum_{k \neq j} \frac{\mu_k}{\sigma_k^2}\right) \cdot \sigma_{MAP}^2 .$$

D.h. es handelt sich hierbei um einen Linearen Opinion Pool, wobei die Gewichtungen der Expertenschätzungen proportional zu ihrer inversen Varianz gewählt werden.

In unserem hybriden Verfahren werden beide Ansätze kombiniert. Das Update jedes Expertenprofils (μ_j, σ_j) erfolgt in jedem Schritt über

$$\sigma_j^{2\,(i+1)} = \left(\frac{N \cdot w_{jj}^{(i)}}{\sigma_j^{2(i)}} + \sum_{k \neq j} \frac{N \cdot w_{jk}^{(i)}}{\sigma_k^2}\right)^{-1} \text{ und}$$

$$\mu_j^{(i+1)} = \left(\frac{N \cdot w_{jj}^{(i)} \mu_j^{(i)}}{\sigma_j^{2(i)}} + \sum_{k \neq j} \frac{N \cdot w_{jk}^{(i)} \mu_k^{(i)}}{\sigma_k^2}\right) \cdot \sigma_j^{2\,(i+1)}.$$

Man beachte hierbei, dass das Superskript „2“ in σ_j^2 für die Varianz steht, und „$(i+1)$“ die Iterationen zählt.

Für $\epsilon \to \infty$ konvergieren die Gewichte $w_{jk}^{(i)}$ gegen eine Gleichgewichtung, d.h. $\lim_{\epsilon \to \infty} w_{jk}^{(i)} = \frac{1}{N}$. In diesem Fall entspricht die Gewichtung dem bayesschen Update. Für $\sigma_j^2 \to \frac{1}{N}$ entspricht die Gewichtung dem Konsens Opinion Pooling. Demnach interpoliert unser hybrides Verfahren zwischen den beiden Ansätzen über die Wahl des Parameters $\epsilon > 0$ sowie die Unsicherheiten der Experten σ_j^2 für $j = 1, \ldots, N$.

Algorithmus 1 Ablauf des hybriden Opinion Pooling zur Konsensfindung

Input: N Experteneinschätzungen $\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_N^{(0)}$; N Standardabweichungen $\sigma_1^{(0)}, \sigma_2^{(0)}, \ldots, \sigma_N^{(0)}$; Tuning Parameter $\epsilon > 0$.

for $i = 1$ **to** ∞ **do** // Konvergenz garantiert für $i \to \infty$; praktisch: Abbruch nach einer „ausreichenden“ // Anzahl von Iterationen bis zur gewünschten Genauigkeit der Approximation

  for $j = 1$ **to** N **do**

    for $k = 1$ **to** N **do**

      $w_{jk}^{(i)} \leftarrow \alpha_j^{(i-1)} \,/\, \left(\epsilon + d\left(\mu_j^{(i-1)}, \mu_k^{(i-1)}\right)\right)$

    end for

    $\sigma_j^{2\,(i+1)} \leftarrow \left(N \cdot \frac{w_{jj}^{(i)}}{\sigma_j^{2(i)}} + \sum_{k \neq j} N \cdot \frac{w_{jk}^{(i)}}{\sigma_k^{2(i)}}\right)^{-1}$

    $\mu_j^{(i+1)} \leftarrow \left(N \cdot w_{jj}^{(i)} \cdot \mu_j^{(i)} / \sigma_j^{2(i)} + \sum_{k \neq j} N \cdot w_{jk}^{(i)} \cdot \mu_k^{(i)} / \sigma_k^{2(i)}\right) \cdot \sigma_j^{2\,(i+1)}$

  end for

end for

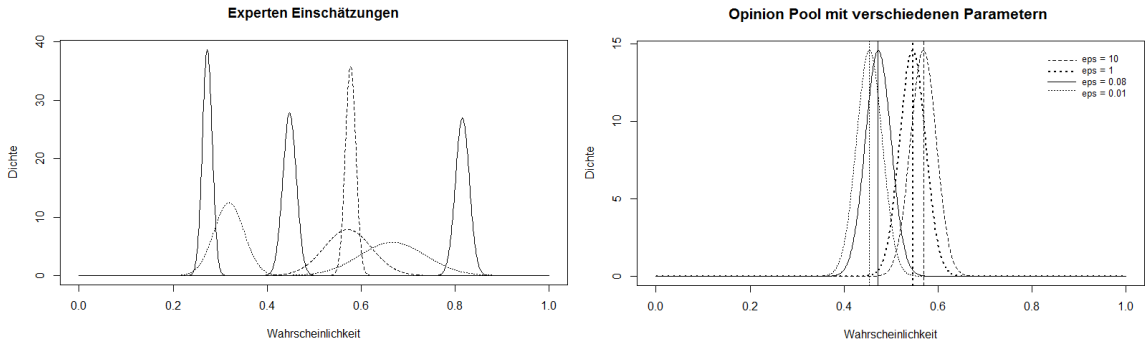

Abb. 3: Opinion Pooling (mit Unsicherheiten)

Abbildung 3 zeigt im linken Teil eine Menge von Expertenmeinungen, das rechte Bild zeigt die repräsentativen Werte (horizontale Linien). Die Unsicherheit (Varianz im Ergebnis) wird durch die gepoolte Varianz ausgedrückt.

Es kann hilfreich sein, Verfahren des Opinion-Pooling bzw. der Konsensfindung danach zu unterscheiden, ob diese informationserhaltend oder verlustbehaftet sind. Eine (in der Praxis häufig eingesetzte) Mittelwert- oder Medianbildung wäre hierbei ein Beispiel einer *verlustbehafteten Aggregation*, da die Risikoangaben vieler ExpertInnen in einem einzigen gemeinsamen Wert „verschmelzen". Eine solche einzelne Zahl verbirgt naturgemäß mögliche Schwankungen bei den (zufälligen) Auswirkungen einer Maßnahme, erleichtert jedoch gleichzeitig die Entscheidungsfindung, indem etwa jene Maßnahme empfohlen werden kann, welche das geringste Risiko (gemessen als „impact × likelihood") besitzt. Dies zeigt den „Informationsverlust", bei dem etwa die möglichen Schäden trotz geringeren Mittelwerts (Risikos) durchaus nach oben und unten stark abweichen können (und dies mit einer potentiell hohen Wahrscheinlichkeit auch tun). Es erscheint demnach im Beispiel in Abbildung 4 plausibler, die Maßnahme mit dem höheren erwarteten Schaden, jedoch dem „stabileren" Potential für Schaden zu wählen.

Alternativ hierzu bieten sich daher *verlustfreie* (informations-erhaltende) Verfahren der Datenaggregation an, wie etwa die Verwendung der gesamten Verteilung (siehe mittleres Bild in Abbildung 2) in Kombination mit einer stochastischen Ordnung [HyR17], oder die zusätzliche Verwendung der Varianz als Maß für die Unsicherheit (siehe rechtestes Bild in Abbildung 2). Diesen Ansatz verfolgen wir im vorliegenden Beitrag, sowie im CERBERUS Projekt (für den Einsatz einer stochastischen Ordnung und verlustfreier Risiko-Datenaggregation sei auf die Literatur [RKS16, ShSh06] verwiesen).

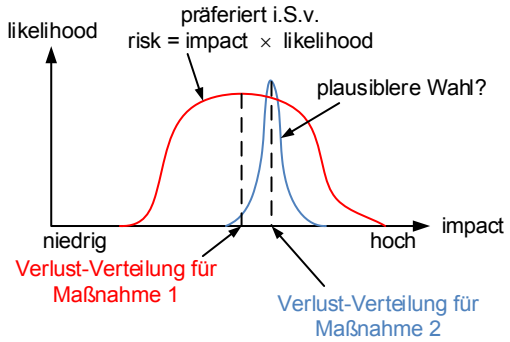

Abb. 4: Beispiel für Informationsverlust bei Daten-Aggregation
(Verteilungen sind hier zur besseren Illustration nicht maßstabsgetreu dargestellt)

4 Datenbereinigung und Ausgleich von Risikotypen

Für die manuelle Bewertung von Risiko spielen neben der eigenen fachlichen Expertise auch Aspekte der Persönlichkeit der jeweiligen ExpertInnen eine Rolle. So werden identische Risiken gemäß der persönlichen Risikoaversion bzw. Risikofreude, aber auch abhängig von der Art der Problemstellung bzw. Szenarioformulierung von verschiedenen Experten unterschiedlich bewertet (sinngemäß unter- bzw. überschätzt). Um die sich hieraus ergebenden Verzerrungen zu minimieren, d.h. die Daten geeignet zu bereinigen, beschreiben wir in diesem Abschnitt sowohl ein manuelles als auch ein automatisiertes Vorgehen zum Ausgleich der subjektiv bedingten Unschärfe bei der Risikobewertung. Beide Verfahren können als Datenbereinigung und ähnlich zu der (auch in anderen Bereichen erforderlichen) Elimination von Ausreißern angesehen werden. Anders als bei letztgenannter, werden offenbar falsche Schätzungen im vorliegenden Kontext jedoch nicht entfernt, sondern korrigiert.

4.1 Manueller Ausgleich

Man könnte meinen, dass Personen sich im Hinblick auf ihre Fähigkeiten zur Risikoschätzung in unterschiedliche Kategorien einteilen lassen (eventuell auch in Mischform per anteilsmäßiger Zugehörigkeit zu mehreren Kategorien). Eine empirische Studie zeigt jedoch [BBM98], dass der tatsächlich begangene Schätzfehler nur zu einem relativ geringen Teil von der Person selbst abhängt, als vielmehr vom eigenen Erfahrungsschatz bestimmt wird. Hierbei kommt kürzlich Erlebtem stärkerer Einfluss bei der Risikoeinschätzung zu, was sowohl zu einer Unter- als auch einer Überschätzung des Risikos führen kann. Hingegen werden die Unterschiede in der Risikoeinschätzung nur zu einem sehr geringen Anteil durch kulturelle oder soziale Faktoren erklärt (etwa 6% der Varianz ist hierauf zurückzuführen [BBM98]).

Wesentliche Unterschiede in der Bewertung sind auch durch die Domäne induziert, so werden etwa Risiken im Finanzbereich (von derselben Person) anders bewertet als Risiken der Gesundheit. In [WBB02] werden hier fünf Bereiche unterschieden: Finanz, Gesundheit/Sicherheit, Freiheit, Ethik und Soziales.

Bei einer manuellen Korrektur von Risikobewertungen ist insbesondere zwischen der Risikowahrnehmung und der Risikoeinstellung (i.S.v. Risikoaversion vs. Risikoappetit) zu unterscheiden. Dies ist bei Befragungen zu berücksichtigen, ebenso wie die Vermeidung von Anreizen (Incentives), welche die Befragungsergebnisse verzerren können [CaHo99]. Ähnliche Effekte sind im Kontext von Risikobewertung denkbar, beispielsweise, wenn die Wichtigkeit des eigenen Unternehmens oder Vorhabens durch bewusst hoch angesetztes Risiko unterstrichen werden soll.

Ein wesentlicher Faktor, welcher die Risikobewertung verzerren kann, ist das Bewusstsein über die möglichen Auswirkungen, welches abhängig von kürzlich erlebtem geschärft oder abgestumpft sein kann. So können kürzlich gesehene Berichte in den Medien das Risikobewusstsein für bestimmte Bedrohungen (etwa Terroranschläge) erhöhen, was zu einer tendenziellen Überschätzung der tatsächlichen Terror-Wahrscheinlichkeit führt. Umgekehrt werden Risiken mit denen in jüngster Vergangenheit keine Erfahrungen gemacht wurden potentiell eher unterschätzt, wie etwa die Gefahr von Identitätsdiebstahl nach Preisgabe persönlicher Daten in sozialen Medien.

Tab. 1: Einflussfaktoren auf Risikoabschätzung (vgl. [SELS13])

Faktor	Auswirkung
Medieneffekt	Wird die Thematik häufig oder aktuell in einem sensationellen Kontext in den Medien diskutiert erhöht dies tendenziell die Einschätzung.
Erlebnisse	Personen, die das zu untersuchende Risiko selbst erlebt haben oder jemanden, kennen der das Risiko erlebt hat, tendieren zur Überschätzung.
Organisatorische Aufteilung	Die Beschäftigung mit negativen Szenarien führt zu tendenziellen Überschätzungen des Risikos, während dort, wo positiven Erfahrungen überwiegen unterschätzt wird.
Emotionen	Die komplexe Frage, welchen Schaden ein Ereignis verursacht, kann bisweilen durch die einfache Frage, welches Gefühl man bei dem Ereignis hat, ersetzt werden. Emotionen wirken sich i.d.R. stark auf die Risikoquantifizierung aus
Nutzen	Menschen gehen intuitiv von einem negativen Zusammenhang zwischen Nutzen und Risiko aus. Die Betonung eines hohen Nutzens beeinflusst die Bewertung negativ.
Zeitdruck	Wenn schnell entschieden werden muss und rationales Denken eingeschränkt wird, werden oben genannte Effekte verstärkt.

Zum Ausgleich der so induzierten Tendenzen erscheint es ratsam, der Datenerhebung eine Reihe von Eingangsfragen voranzustellen, welche sich auf die oben genannten Faktoren beziehen. So kann etwa gefragt werden, ob kürzlich Medienberichte über das aktuell abgefragte Risiko gesehen oder gelesen wurden. Wird dies positiv beantwortet, so kann dies ein Hinweis sein, die nachfolgende Risikoangabe ggf. nach unten zu korrigieren (bei negativer Antwort kann die entsprechende inkrementelle Korrektur erfolgen). Eine Auswahl weiterer Faktoren, die Einfluss nehmen können, sind in Tabelle 1 angeführt. Die hier angeführten Faktoren dienen jedoch nur als Indikatoren, da sie zu keiner statistischen Klassifikation von Personen, die das Risiko tendenziell über- bzw. unterschätzen, geeignet sind. Insofern muss eine Korrektur manuell erfolgen nach Bewertung aller Daten über die Person und den gegebenen Kontext und die aktuelle Informationslage.

Die in Abschnitt 2 beschriebene graphische Eingabeform kann auch als Darstellungsvariante zur Datenbereinigung dienen, etwa indem alle abgegebenen Meinungen gleichzeitig dargestellt werden, sodass „Ausreißer“ oder nicht plausible Angaben manuell korrigiert werden können.

4.2 Automatischer Ausgleich

Hierfür teilen wir die Personenmenge in drei Kategorien ein, nämlich solche, die das Risiko unterschätzen, solche die das Risiko überschätzen, und jene, die das Risiko in etwa korrekt einschätzen. Bei einer Gesamtheit von N Risikoeinschätzungen $r_1, \dots, r_N$ seien $p \cdot N$ Werte zu hoch angesetzt, $q \cdot N$ Werte zu niedrig, und $(1-p-q) \cdot N$ Werte seien in etwa korrekt (bedürfen also keiner Korrektur). Ein zu hoher Wert r_i wird hierbei (additiv) um einen (konstanten) Wert Δ^* zu $r_i \leftarrow r_i - \Delta^*$ nach unten korrigiert; analog wird ein zu niedriger Wert r_j zu $r_j \leftarrow r_j + \Delta_*$ korrigiert. Ziel ist, die Varianz der Gesamtschätzung zu minimieren. Dabei gelte für den tatsächlichen (unbekannte) Risikowert r: Personen, die das Risiko überschätzen liefern Werte die wir als Zufallsvariable R^* auffassen können mit dem Erwartungswert $E(R^*) = r + \Delta^*$. Analog dazu liefern Personen, die das Risiko unterschätzen Realisierungen der Zufallsvariable R_* mit Erwartungswert $E(R_*) = r - \Delta_*$. Korrekte Einschätzungen sind Realisierungen der Zufallsvariable R mit dem Erwartungswert $E(R) = r$.

Die Gesamtheit der N Risikobewertungen stellt also eine Menge von Samples der drei Variablen R^*, R_* und R dar, wobei wir lediglich wissen, dass der Anteil von Samples von R^* gleich p, und der Anteil der Samples von R_* gleich q ist (die übrigen Werte werden als Samples von R aufgefasst). Die Bewertungen werden als (stochastisch) unabhängig angenommen, d.h. die Varianzen von R, R^* und R_* addieren sich, und stellen die Gesamt-Unsicherheit (Schwankung) der Risikoangaben dar, die es für eine möglichst „stabile" Schätzung zu minimieren gilt. Diese Minimierung erfolgt durch eine geeignete Verteilung der Korrekturwerte Δ_* und Δ^*, sowie der 0 (für „keine Korrektur"), in den jeweiligen Anteilen der Grundgesamtheit N.

Bezeichnen wir die korrigierten Risikoangaben als $s_i = r_i + b_i$ und $b_i \in \{\Delta_*, \Delta^*, 0\}$ für $i = 1, \dots, N$, so ist die zu minimierende Zielfunktion die empirische Varianz. Sei $\overline{s} = \frac{1}{N}(s_1 + s_2 + \cdots + s_N)$ arithmetische Mittel, so ist das zu lösende Optimierungsproblem

$$\frac{1}{N-1}\sum_{j=1}^{N}\left(s_j - \overline{s}\right)^2 \to \min$$

unter den Nebenbedingungen

$s_i = r_i + b_i$ für x_{y^2};	(die Korrektur ist additiv)
$b_i \in \{0, \Delta^*, -\Delta_*\}$ für $i = 1, 2, \dots, N$;	(die Korrektur ist entweder $+\Delta^*$, $-\Delta^*$ oder 0)
$\lfloor p \cdot N \rfloor \leq \lvert\{i: b_i = \Delta^*\}\rvert \leq \lceil p \cdot N \rceil$;	(nur ein relativer Anteil von p Werten wird nach oben korrigiert)
$\lfloor q \cdot N \rfloor \leq \lvert\{i: b_i = -\Delta_*\}\rvert \leq \lceil q \cdot N \rceil$.	(nur ein relativer Anteil von q Werten wird nach unten korrigiert)

Dieses Problem ist quadratisch und konvex (wie unmittelbar aus einer Eigenwertabschätzung der Hesse-Matrix der Zielfunktion unter Anwendung des Kreisesatzes von Gershgorin folgt; siehe [GovL96]), jedoch nicht streng konvex. Somit ist die ermittelte optimale Risikoschätzung eindeutig bestimmt, jedoch nicht notwendigerweise die Zuordnung der Korrekturwerte.

Das obige Problem lässt sich mit etwas mehr Aufwand auch in Matrix-Schreibweise und mit binären Nebenbedingungen formulieren, sodass es herkömmlichen Lösungsalgorithmen in Software zugänglich wird. Eine weitere Vereinfachung kann erreicht werden, indem die oben angegebenen Schranken durch geeignetes Runden in Gleichheitsbedingungen umgeformt werden, etwa indem $\lvert\{i: b_i = \Delta^*\}\rvert = m^*$ und $\lvert\{i: b_i = -\Delta_*\}\rvert = m_*$ verlangt wird, wobei $m^* \approx p \cdot N$ und $m_* \approx q \cdot N$ geeignet gerundet werden müssen. In dieser Form lassen sich die Nebenbedingungen via Lagrange-Multiplikator in die Zielfunktion einbetten, sodass ein quadratisches ganzzahliges Optimierungsproblem ohne Nebenbedingungen entsteht. Damit vergrößert sich die Palette der hierfür zur Verfügung stehender Lösungsalgorithmen. Obgleich ganzzahlige Optimierung i.A. NP-hart ist, waren experimentelle Instanzen des Problems ausnahmslos effizient lösbar. Unabhängig davon stehen im Kontext subjektiver Risikobewertungen i.d.R. keine sehr großen Datenmengen zur Verfügung (die Zahl N wäre die Anzahl der ExpertInnen, die tatsächlich eine Einschätzung abgeben), was den Aufwand zur Lösung des Optimierungsproblems in der Praxis in einem handhabbaren Rahmen halten wird.

Das Optimierungsproblem ist durch die Werte für p, q, Δ_* und Δ^* parametriert, welche empirisch ermittelt werden können. Hierbei würde ein Fragebogen ein Szenario beschreiben, und um eine Risikoeinschätzung bitten (etwa auf einer Skala von 1 bis 5). Nach Auswertung der

Fragebögen möge hiernach ein Datensatz $x_1, \dots, x_k \in \{1, 2, \dots, 5\}$ zur Verfügung stehen. Diese Daten werden im Anschluss mit einem *Referenzwert* y verglichen, welcher das „objektive" Risiko des beschriebenen Szenarios darstellt (insoweit dieses ermittelt werden kann, bzw. falls dieses idealerweise aus der Realität des Szenarios sogar bekannt ist, etwa der tatsächlich aufgetretene Schaden des – für die Befragung anonymisierten – Fallbeispiels). Aus den so erhobenen Daten ergeben sich die Parameter für die Optimierung durch Anzahl und Mittelwertbildung, wie folgt:

$$\hat{p} = \frac{|\{k: x_k \geq r\}|}{k}, \hat{q} = \frac{|\{k: x_k < r\}|}{k}, \widehat{\Delta}_* = \frac{1}{k} \sum_{i: x_i < r} (y - x_i) \text{ und } \widehat{\Delta}^* = \frac{1}{k} \sum_{i: x_i \geq r} (x_i - y).$$

Der ^-Akzent deutet hierbei an, dass es sich um statistische Schätzungen handelt, deren Güte naturgemäß mit der Anzahl k der Daten steigt. Für die oben beschriebene Optimierung wären die so ermittelten Schätzwerte zu verwenden.

Falls die Schwankung in den Über- bzw. Unterschätzungen explizit berücksichtigt werden soll, kann die additive Korrektur $s_i = r_i + b_i$ in eine multiplikative Korrektur $s_i = r_i \cdot b_i$ mit $b_i \in \{1 + \rho_*, 1 - \rho^*, 1\}$ geändert werden, worin ρ_*, ρ^* die jeweiligen relativen Fehler beim Über- bzw. Unterschätzen sind (die obigen arithmetischen Mittelwertschätzer sind dann durch geometrische Mittelwerte zu ersetzen).

5 Ausblick

Die wesentlichen Schwierigkeiten bei der Bewertung von Risiko entstehen aus den bisweilen stark unterschiedlichen Meinungen und dem Vorwissen über das zu bewertende Risikoobjekt. Wie in diesem Beitrag gezeigt, stehen statistische Verfahren zur Verfügung, um unterschiedliche Meinungen zu konsolidieren und zu harmonisieren, und können im Rahmen der Risikobewertung eingesetzt werden. Zudem bietet es sich zur Reduktion subjektiver Tendenzen bei der Abgabe von Risikobewertungen an, durch geeignete „Eingangsfragen" eine für alle teilnehmenden Personen ähnliche Ausgangshaltung zu erzeugen, indem etwa nach der Schilderung des Szenarios Fragen darüber gestellt werden (um den Verständnisgrad der Materie zu bewerten) und Indikatoren abgefragt werden, welche Hinweise auf eine Über- bzw. Unterschätzung des Risikos liefern (etwa kürzlich erlebtes, gehörtes oder gelesenes, das mit dem Szenario in Zusammenhang steht). Ebenso gibt es datengetriebene Ansätze, die eventuelle Über- und Unterschätzungen detektieren und korrigieren können. Die Praxistauglichkeit des Ansatzes wird in den kommenden Monaten im Zuge des Projekts CERBERUS weiter analysiert.

Danksagung

Dieser Beitrag wurde durch das FFG/KIRAS Projekt „CERBERUS - Cross Sectoral Risk Management for Object Protection of Critical Infrastructures" (Projekt-Nr. 854766) finanziert.

Literatur

[BBM98] J. Brenot, S. Bonnefous, C. Marris (1998): "Testing the Cultural Theory of Risk in France". Risk Analysis, Vol. 18, No. 6, 1998, p. 729 – 739.

[Bren17] B. Brenner: "WannaCry: the ransomware worm that didn't arrive on a phishing hook", https://nakedsecurity.sophos.com/2017/05/17/wannacry-the-ransomware-worm-that-didnt-arrive-on-a-phishing-hook/ (2017)

[CaHo99] C. F. Camerer, R.M. Hogarth: "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework", Journal of Risk and Uncertainty (1999) 19: 7. doi:10.1023/A:1007850605129

[CaLa13] A. Carvalho, K. Larson: „A Consensual Linear Opinion Pool", https://arxiv.org/pdf/1204.5399.pdf (2013).

[CSYW07] R. A. Caralli, J. F. Stevens, L. R. Young, W. R. Wilson: „Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process", CMU/SEI-2007-TR-012, Software Engineering Institute, Carnegie Mellon University (2007).

[GovL96] G. H. Golub, C.F. van Loan: "Matrix Computations", Baltimore: Johns Hopkins University Press, p. 320, ISBN 0-8018-5413-X (1996).

[HyR17] The HyRiM Consortium: „Hybrid Risk Management for Utility Networks", https://hyrim.net, EU Project, FP7 Grant No. 608090 (2017)

[ISAC13] ISACA: „Cobit 5 for Risk", Rolling Meadows, USA (2013)

[ISO09] International Standardization Organization: „ISO 31000: Risk Management – Principles and Guidelines", Genf, Schweiz (2009)

[Mün12] I. Münch (2012): Wege zur Risikobewertung. In: P. Schartner und J. Taeger (Hg.): DACH Security 2012: syssec, S. 326–337.

[RaRa14] S. Rass, B. Rainer: „Numerical Computation of Multi-Goal Security Strategies". In: Radha Poovendran und Walid Saad (Hg.): Decision and Game Theory for Security: Springer (LNCS 8840), S. 118–133 (2014).

[RKS16] S. Rass, S. König, S. Schauer: „Decisions with Uncertain Consequences-A Total Ordering on Loss-Distributions". In: PLoS ONE 11 (12), e0168583 (2016). DOI: 10.1371/journal.pone.0168583.

[RSPG13] S. Rass, S. Schauer, A. Peer, J. Göllner: Sicherheit auf Basis Multikriterieller Spieltheorie. in: P. Schartner, P. Trommler (eds.): DACH Security 2013, pp. 289-301, ISBN 978-3-00-042097-9

[SELS13] R. Sachs, E. Eller, E. Lermer, B. Streicher: „Psychologische Einflüsse auf die individuelle Einschätzung von Risiken", Emerging Risk Discussion Paper, Munich Re (2013).

[ShSh06] M. Shaked, J. G. Shanthikumar: „Stochastic Orders", Springer (2006).

[WBB02] E. U. Weber, A.-R. Blais, N.E. Betz: "A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviours", Journal of Behavioral Decision Making, Vol. 15, p. 263-290 (2002).

[Zet16] K. Zetter: Inside the Cunning, Unprecedented Hack of Ukraine′s Power Grid https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/ (2016).