Zentrum für sichere Informationstechnologie – Austria
Secure Information Technology Center – Austria

A-SIT

Seidlgasse 22 / 9, 1030 Wien
Tel.: (+43 1) 503 19 63–0
Fax: (+43 1) 503 19 63–66

Inffeldgasse 16a, 8010 Graz
Tel.: (+43 316) 873-5514
Fax: (+43 316) 873-5520

http://www.a-sit.at
E-Mail: office@a-sit.at

DVR: 1035461 ZVR: 948166612 UID: ATU60778947

SICHERE EDGE-COMPUTING TECHNOLOGIEN

Lukas Alber – lukas.alber@iaik.tugraz.at

Abstrakt/Kurzbeschreibung:
Komplexe Aufgaben werden gern in die Cloud ausgelagert, weil viele kleine und mobile Geräte nicht über die nötige Rechenleistung für solche Aufgaben verfügen. Da die Cloud aber meist netzwerktechnisch weit entfernt liegt und somit die Verbindung eine erhöhte Latenz aufweist, wurde in den letzten Jahren viel am sogenannte Edge-Computing geforscht. Zumal sich bei diesem Paradigma die Recheneinheit nur einen Verbindungssprung (z.B. 5G-Basisstationen) entfernt befindet, ist die Latenz im Vergleich relativ gering. In diesem Bericht wollen wir uns eine Übersicht über die Funktionsweise von Edge-Computing verschaffen. Anschließend wird eine Sicherheitsanalyse durchgeführt und Bedrohungen für das Edge-Computing Paradigma identifiziert. Darauf folgend schauen wir uns Sicherheitsmechanismen und -maßnahmen an, die für sicheres Edge-Computing ergriffen werden müssen.

Inhaltsverzeichnis

1. Einleitung

Kleine und mobile Geräte jedweder Art besitzen nur begrenzt Rechenleistung, da sie meist Batterie betrieben und/oder handlich sein müssen. Da Nutzer aber zunehmend mehr Rechenleistung benötigen, andererseits Applikation eine gute Verbindungslatenz benötigen um interaktiv und nutzerfreundlich zu bleiben, wurde in den letzten Jahren geforscht wie diese Gegensätze überwunden werden können. Edge-Computing (EC) ist ein Lösungsansatz der sich diesem Problem annimmt. Wie der Name bereits verrät geht es um die Kante des Internets: also exakt der Ort, wo die angeforderten Daten auf den Bereich des einzelnen Nutzers übergehen. Das sind zum Beispiel die Basisstationen der 5G-Masten.

Ein erstes Mal wurde der Name Edge-Computing im 2013 von Nokia und IBM genutzt. Sie beschrieben damit eine Plattform, welche Berechnungen an der Basisstation durchführen konnte. Diese Plattform beschränkte sich aber auf eine lokal begrenzte Arbeitsweise [1].
Als ETSI ein Jahr später die Industry Specification Group (ISG) für „Mobile Edge Computing“ (MEC) ins Leben rief, wurde die Definition erweitert: Die Zusammenarbeit von mehreren Basisstationen bzw. mobilen Geräten und der fliegende Wechsel zwischen Basisstationen sollten ermöglicht werden[2].

Aber Edge-Computing ist nicht nur für mobile Geräte von Nutzen: Dank einer niederen Latenz, guter Verlässlichkeit und Skalierbarkeit ist der Einsatz der Technologie auch abseits von mobilen Geräten bzw. Mobilfunknetzwerken geeignet. Darum entschied sich die ISG, das von ihnen geprägte Kürzel MEC (das ursprünglich für „Mobile Edge Computing“ stand), in „Multi-Access Edge Computing“ umzumünzen[1].

1.1. Abgrenzung zu Cloud-Computing

Vom allgemeinen Cloud-Computing (CC) mit verteilten Rechenzentren unterscheidet sich Edge-Computing hauptsächlich in folgenden Eigenschaften: Während EC sich hauptsächlich auf Latenz-sensible Nutzer konzentriert, versucht CC sich aller Nutzer anzunehmen. CC arbeitet mit globalen Informationen, während EC mit den lokalen Daten der genutzten Basisstation (inklusive Kontextinformationen wie den Standort) auskommen muss. Auch sind Rechenressourcen und Speicher an der Edge nicht in der Skalierbarkeit wie in Cloud-Rechenzentren zur Verfügung. Des Weiteren spielt die netzwerktechnische Distanz eine wesentliche Rolle: Während die Edge nur einen Hop vom Nutzer entfernt ist, erhöht sich die Latenz zu der Cloud um jeden Hop im IP-Netzwerk. EC-Rechnerknoten sind regional verteilt auf Basisstationen und können von lokalen Betreibern gewartet werden. CC-Center andererseits sind zwar über die Welt verteilt, sind aber im Vergleich stark zentralisiert und werden direkt von großen Firmen betrieben[3].

1.2. Abgrenzung zu Fog-Computing

Das Fog-Computing Paradigma ist sehr ähnlich dem Edge-Computing, wurde aber schon vor Letzterem im Jahre 2012 von Cisco Systems beschrieben. Es ist definiert als eine Erweiterung des Cloud-Computing Paradigma um Rechenleistung, Speicher und Netzwerkdiensten irgendwo zwischen Endnutzer und den Cloud-Servern[4]. Und hier liegt der Hauptunterschied: Edge-Computing ist Netzbetreiber gebunden und das Computing findet z.B. an deren 5G-Basisstationen statt. Bei Fog-Computing kann das Computing zwar auch in der Basisstation vonstattengehen, aber auch in entfernteren Gefilden zwischen Edge und Cloud-Servern, z.B. an Multiprotocol Label Switching Backbones[5]. Dadurch kann die Fog-Computing Infrastruktur von beliebigen Betreibern, nicht nur den Netzbetreibern, betrieben werden[3].

2. Herausforderungen

In diesem Kapitel wird eine Sicherheitsanalyse durchgeführt, um Bedrohungen zu ermitteln. Darauffolgend werden Fehlerquellen erörtert und Gegenmaßnahmen diskutiert. Das genutzte Schema ist in Abbildung 1 veranschaulicht.

[1] https://www.etsi.org/images/files/ETSInewsletter/etsinewsletter-issue2-2017.pdf, zugegriffen am 30.12.2021

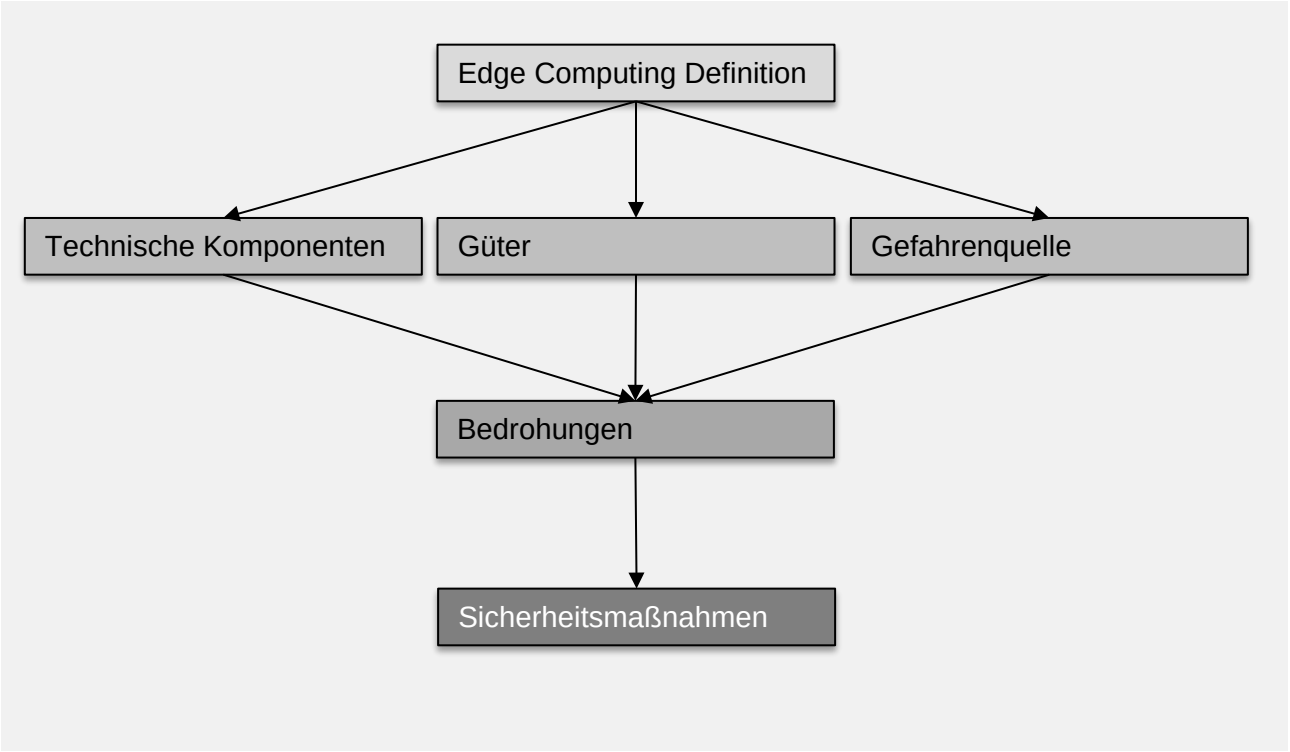

Abbildung 1: Vorgehensweise für eine systematische Erörterung der Herausforderungen.

2.1. Sicherheitsanalyse

Um eine Sicherheitsanalyse für Edge-Computing zu erarbeiten, haben wir folgende Annahme getroffen: Verbindung zwischen Endgerät und EC-Infrastruktur muss genügend gesichert sein (adäquat authentifiziert und verschlüsselt). Des Weiteren müssen sogenannte Schutzziele von Vornherein definiert sein. Es wird dafür die sogenannte CIA-Triade in erweiterter Form heranziehen [6]:

Vertraulichkeit: Das Ziel beschreibt die Absicht, dass vertrauliche Daten für Unbefugte zu keiner Zeit einsehbar sind.
Integrität: Das Ziel ist Daten und Funktionsweisen der Systeme vor Einflussnahme von Unbefugten zu schützen. Man unterscheidet zwischen Daten- und Systemintegrität.
Verfügbarkeit: Bestimmte Daten und Dienste müssen möglichst allzeit zur Verfügung stehen, so das Ziel.
Authentizität: Die Quelle der Daten soll überprüfbar echt und vertrauenswürdig sein, so das Ziel.
Verbindlichkeit: Das Ziel ist, dass gewisse ausgeführte Handlungen einer Entität unabstreitbar bleiben.

2.1.1. Technische Komponenten

Um strukturiert die Gefahren zu modellieren, wird Infrastruktur rund um Edge-Computing betrachtet und einzelne angreifbare Komponenten identifiziert. In Abbildung 2 werden die einzelnen Komponenten der Edge-Computing Infrastruktur dargestellt.

Netzwerkinfrastruktur: Edge-Computing Paradigma nutzt sowohl das mobile Kernnetz als auch die Internetinfrastruktur um Komponenten und Entitäten untereinander zu verbinden.

Edge-Rechenknoten: Hier laufen die Virtualisierungsserver und zusätzliche Managementservices. Durch die zur Verfügung stehenden APIs, ist es ein beliebtes Ziel für externe Angreifer.

Kerninfrastruktur: Im Edge-Computing Paradigma wird ein Edge-Knoten auch mal von der Cloud gestützt. Involviert werden dann vom mobile Kernnetz bis hin zu zentralisierten Cloudservices

verschiedenste Internetinfrastrukturen. Es soll hier aber nicht das Ziel sein, Bedrohungen für Cloudlösungen zu identifizieren (eine Taxonomie für Cloudbedrohungen findet man folgend [6]).

Virtualisierungsinfrastruktur: Um beliebige Services an der Edge zu nutzen, wird eine Virtualisierungsinfrastruktur auf den Edge-Rechenknoten genutzt. Aber nicht nur die Infrastruktur ist potenzielles Ziel von Angreifern, sondern auch die darauf laufenden, virtuellen Maschinen selbst.

Endgerät: Vom Endnutzer kontrolliert, stellt das Endgerät und die darauf laufende Software den Profiteur der Edge-Services dar. Aber das Endgerät ist nicht nur Verbraucher, sondern trägt aktiv Daten zur Ermöglichung verschiedener, crowd-sourced Edge-Services bei.

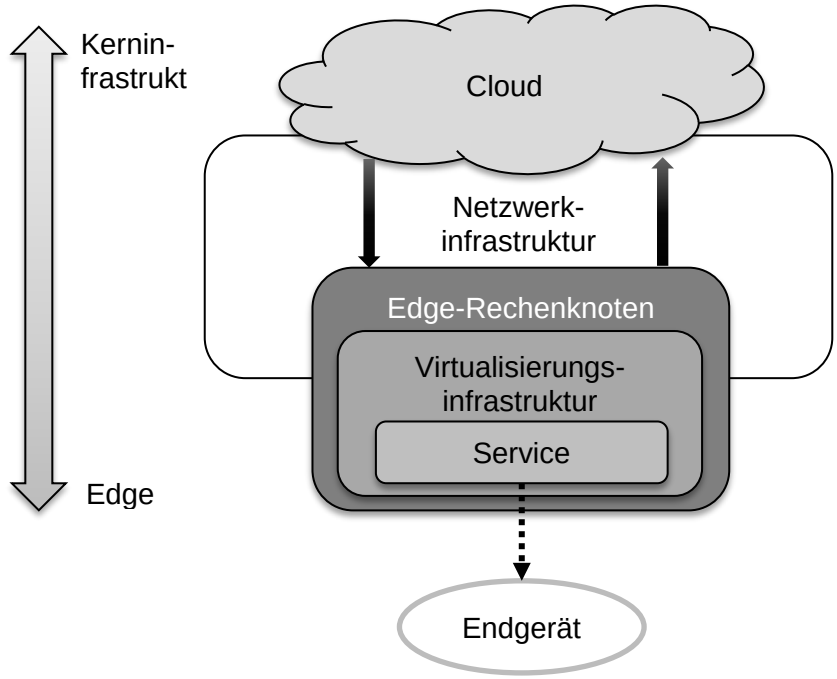

Abbildung 2: Komponenten der Edge-Computing Infrastruktur.

2.1.2. Güter

In diesem Schritt werden gefährdete Güter identifiziert, auf welche ein potenzieller Angreifer es abgesehen haben könnte und die darum unserer Schutzziele bedürfen:

Daten: Beim Edge-Computing sind verschiedenste Daten schützenswert: Manche Daten da sie vertraulich gehalten werden müssen (z.B. meist Nutzerdaten, aber auch Log-Daten), andere da von ihrer Verfügbarkeit die Funktionsfähigkeit des Systems abhängt. Bei wieder anderen Daten ist die Integrität wichtig oder die Authentizität muss gegeben sein. Welche Daten wie geschützt werden müssen hängt hauptsächlich vom Anwendungsfall ab.

Funktionsweise: Auch die ausgeführten Berechnungen müssen vor Einflussnahme geschützt werden. Ist die Integrität dieser kompromittiert, kann das schwerwiegende Auswirkungen auf eine ganze Reihe von resultierenden Daten haben.

Algorithmen: Nicht nur die Berechnungen selbst sind schützenswerte Güter, oft ist auch der Algorithmus hinter den Berechnungen schützenswert (z.B. bei geschlossenem oder proprietären Quellcode).

2.1.3. Gefahrenquellen

Obwohl Edge-Computing im Vergleich zu Fog-Computing ein relativ genau und geschlossen definiertes System ist, wird es nicht von einer einzelnen Partei kontrolliert. Einzelne Edge-Datenzentren können unabhängigen Parteien gehören, müssen aber nahtlos zusammenarbeiten.
Es gilt das „anywhere, anytime“ Prinzip, bekannt aus dem Internet-of-Things-Bereich. Das heißt jede Komponente der Infrastruktur kann jederzeit angegriffen werden.
Anderseits werden die Services der Edge-Datenzentren nur lokalen Endgeräten angeboten. Das schwächt das „anywhere“ ab; wobei zu beachten ist, dass Virtuelle Maschinen mit einem mobilen Endnutzer zum Edge-Rechenknoten in nächster Nähe mitspringen[1].

Der oft zitierte externe Angreifer hat durch das „anywhere, anytime“-Prinzip eine große Angriffsfläche zur Verfügung: Der Angreifer kann eine oder mehr Komponenten der bestehenden Infrastruktur unter seine Gewalt bringen (der externe Angreifer kontrolliert anfänglich per Definition keinen Teil der Infrastruktur), oder mit eigenen Infrastrukturkomponenten sich in Ecosystem einschleusen.

Interne Angreifer andererseits befinden sich bereits im Ecosystem und haben Zugang zu einzelnen Infrastruktur Komponenten: So kann zum Beispiel der Wartungsbeauftragte einer Basisstation leicht Einfluss nehmen, da er/sie direkten Zugang hat. Schwieriger hat es ein böswilliger Endnutzer, welcher die Funktionsweise eines Service beeinflussen will oder gar über die Virtualisierungsinfrastruktur auf Daten anderer Endnutzer zugreifen will.

2.2. Bedrohungen

Folgende Bedrohungen konnten dank der Sicherheitsanalyse (Unterkapitel 2.1) und zusätzlicher Literaturrecherche identifiziert werden[8], [9]:

Eine Bedrohung, die von einem betrügerischen Betreiber eines EC-Rechenknotenaus geht, ist die Offenlegung sensitiver Daten, Berechnungen oder Algorithmen. Aber auch die Verfälschung von Daten und Berechnungen. Typische Angriffe hier sind zum Beispiel Authentifizierungs- oder Autorisierungsangriffe oder Privilegien Eskalation.

Auch ein externer Angreifer kann einen EC-Knoten angreifen. Wie zuvor können Daten, Berechnung oder Algorithmen offengelegt oder verfälscht werden. Typische Angriffe von externen Angreifern sind hier z.B. Malware-Injektion oder ein Authentifizierungs- oder Autorisierungsangriff. Hier ist zu erwähnen, dass auch Log-Daten zu den gefährdeten Gütern gehören, da diese bei einer Offenlegung sensible Information enthalten kann.

Weiteres kann ein böswilliger Nutzer eines Endgeräts versuchen den Service so zu beeinflussen, damit er/sie Daten oder Berechnungen anderer Nutzer offenlegt oder verfälscht. Auch schützenswerte Algorithmen können von einem solchen Angriff offengelegt werden. Angreifer können hier z.B. direkten Zugriff durch Sicherheitslücken erlangen, einen Sidechannel-Angriff oder eine Privilegien Eskalation ausführen. Weiters kann das Senden falscher Sensorinformationen kollaborative Services die Berechnung anderer Endgeräte beeinflussen (z.B. selbstfahrender Verkehr). Auch kann ein Angreifer, als Nutzer eines Endgerätes, Schwachstellen in der API von Services und der Virtualisierungsinfrastruktur ausnützen.

Kann der Angreifer die Virtuelle Maschine bestimmen und damit auch den auszuführenden Code (z.B. als Nutzer eines Endgeräts), dann ist auch die Manipulierung durch Hardware-Exploits z.B. Bitflips durch Rawhammer [10], im Bereich des Möglichen. Auch könnte der Angreifer sich nach verletzbaren Geräten in der Reichweite des Edge-Knoten umsehen. Zusätzlich könnte der Angreifer auch Code für DDos-Angriffe ausführen(z.B. getarnt als Cloudzugriffe) vorausgesetzt der Angreifer kann die Virtuelle Maschine auf vielen EC-Knoten zugleich starten.

Der Angreifer kann aber auch die Verfügbarkeit angreifen. So kann er/sie als Nutzer eines Endgerätes absichtlich die Ressourcen des EC-Knoten überbeanspruchen (z.B. DDos-Attacke). Weiters kann ein gewollt oder ungewolltes Fehlverhalten des Wartungspersonals die Ausfälle provozieren. Aber auch eine physische Manipulation des EC-Knotens oder dessen Verbindung ist denkbar. Auch mit Verbindungsabbrüche zwischen der EC-Infrastruktur und dem Kernetz bzw. der

restlichen Internetinfrastruktur kann ein Angreifer die Verfügbakeit beeinflussen, da viele EC-Services Daten aus Cloud-Datenzentren nachladen.

3. Sicherheitsmaßnahmen und -mechanismen

Um die Identifizierten Bedrohungen aus der Sicherheitsanalyse weit möglichst abzuwenden, sind geeignete Sicherheitsmechanismen und -maßnahmen nötig. In den nachfolgenden Absätzen dieses Kapitels werden wichtige Mechanismen und Maßnahmen diskutiert; zusammengetragen aus verschiedenen Quellen [1], [3], [8], [11].

Im Edge-Computing Paradigma gibt es viele Aktoren und Komponenten, welche ständig miteinander im Austausch stehen. Auch kann das Endgerät seinen Aufenthaltsort wechseln. Dadurch verbindet es sich zur nächst näheren Basisstation und nützt somit einen anderen EC-Knoten. Um nur bekannten und befugten Identitäten Zugriff auf eine jeweilige Trust-Domäne zu geben, ist es wichtig ein sicheres Identitäts- und Access-Management System (IAM) zu verwenden. Da es verschiedene Provider für EC-Knoten geben kann, ist eine Föderations- oder Inter-Realm-Unterstützung nötig. Auch ist eine Peer-to-Peer Authentifizierungsmethode vorzuziehen, da verzögerungslose Wechsel von Trust-Domänen nötig und die Abhängigkeit von der Verfügbarkeit eines zentralen Servers unerwünscht ist. Des Weiteren müssen Zugriffsberechtigungen fein und per Trust-Domäne definierbar sein. Ein zusätzlicher Faktor im EC-Paradigma ist auch das Kontext-Bewusstsein (z.B. den aktuellen Standort des Endgeräts), welche in die Entscheidung über die Zugriffsberechtigungen einfließt. Nützlich könnte auch ein zusätzliches System für Vertrauensmanagement sein: Einem Endgerät stehen verschiedenste Serviceanbieter zur Verfügung. Um die Nutzung von böswilligen oder egoistischen Services zu entgehen, ist ein System für einfacherer und sicherer Entscheidungsfindung nötig.

Ein anderer Aspekt im EC-Paradigma ist die Sicherheit von Netzwerken und Protokollen. Da verschiedenste Technologien für lokale (bzw. drahtlose) Netzwerke, Kernnetzwerke und der Internetinfrastruktur verwendet werden, dazu noch verschiedenste Betreiber und Wartungstechniker die EC-Infrastruktur verwalten, ist es wichtig einheitliche Standards, Richtlinien und Konfigurationen zu verwenden, damit auch in einem heterogenen Szenario wie dem EC-Paradigma sicher Verbindungen möglich sind. Darüber hinaus sollte es auch eine absolute Netzwerkisolation der einzelnen Services bestehen.

Aber für eine sichere EC-Infrastruktur sind nicht nur Präventivmaßnahmen, sondern auch Einbruchs- und Mustererkennung vonnöten. Eine gut geschützte und regelmäßig aktualisierte Datenbank von Angriffen kann helfen, solche frühzeitig zu erkennen und Schaden zu verhindern. Hier geht es auch nicht nur um Einbrüche ins System, sondern auch um Missbrauch der die EC-Infrastruktur zu verhindern. So sollte z.B. ein Erkennungssystem Verhaltensmuster boshafter Services, welche auf der Virtualisierung-Infrastruktur laufen, erkennen und den entsprechend Service anhalten. Oben drein ist es wichtig, dass jeder einzelne EC-Knoten sich und seine Umgebung überwacht. Das heißt, dass EC-Knoten auch zusammenarbeiten müssen, um eine globale Überwachung der Systemsicherheit über mehrere Ebenen der Infrastruktur zu ermöglichen.

Ein anderer Hauptbestandteil des EC-Paradigmas ist Virtualisierung. Darum ist es wichtig diesen mit verschiedenen Maßnahmen zu schützen. Solche Maßnahmen sind bereits aus der Cloud bekannt, z.B. eine Isolations-Strategie, Hypervisor Hardening, Netzwerk Abstraktion u.s.w. [12]. Da aber die virtuellen Maschinen oft von einem Knoten zum anderen wandern, ist die Umsetzung einiger erschwert.

Des Weiteren ist davon auszugehen, dass nicht nur böswillige Angreifer Daten von Nutzer einsehen möchten, sondern auch autorisierte, jedoch neugierige Aktoren. Darum ist es nicht nur erforderlichen Daten auf Übertragungswegen vor fremden Zugriff zu schützen, sondern die Daten auch auf den vermeintlich sicheren EC-Knoten zu verschleiern. Nur so kann die Privacy der Nutzer garantiert werden. Die Herangehensweisen bleiben die gleichen, wie auch schon für die Cloud bekannt (z.B.

Multiparty Computation, Proxy Re-Encryption, Homomorphic Encryption, Identity- und Attribute-Based Encryption, u.s.w.)[13].

4. Fazit

In diesem Bericht wurde eine systematische Analyse zur Identifizierung von sicherheitsrelevanten Gefahren durchgeführt. Dafür wurden gefährdete Güter und Komponenten, aber auch Gefahrenquellen identifiziert. Die homogene Struktur des Edge-Paradigmas ergibt für den Endnutzer zwar viele Vorteile, sicherheitstechnisch aber auch viel Angriffsfläche für potenzielle Angreifer. Darum sind die Sicherheitsanforderungen viele und nicht alle sind noch zur Genüge gedeckt bzw. können nur durch einen größeren Performance-Trade-Off gedeckt werden. Das betrifft speziell Bereiche wie Privacy und automatisierte Muster- und Einbruchserkennung. Auch müssen noch Erfahrungen in der breiten öffentlichen Nutzung von Edge-Computing gesammelt werden. Darum setzen viele Firmen (die es sich leisten können) aus verständlichen Gründen noch meist auf ihre private Edge-Infrastruktur[i].

5. Bibliographie

[1] R. Roman, J. Lopez, and M. Mambo, “Mobile edge computing, Fog et al.: A survey and analysis of security threats and challenges,” *Future Generation Computer Systems*, vol. 78, pp. 680–698, Jan. 2018, doi: 10.1016/j.future.2016.11.009.

[2] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, and A. Neal, “Mobile-edge computing introductory technical white paper,” *White paper, mobile-edge computing (MEC) industry initiative*, vol. 29, pp. 854–864, 2014.

[3] A. Reiter, “Hybrid Edge Computing Enable Processing of Sensitive Data in Mobile Edge Computing Scenarios,” 2017.

[4] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the internet of things,” in *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, 2012, pp. 13–16.

[5] N. Bessis and C. Dobre, *Big data and internet of things: a roadmap for smart environments*, vol. 546. Springer, 2014.

[6] R. W. Shirey, “Internet Security Glossary, Version 2,” *RFC*, vol. 4949, pp. 1–365, 2007, doi: 10.17487/RFC4949.

[7] N. V. Juliadotter and K.-K. R. Choo, “Cloud Attack and Risk Assessment Taxonomy,” *IEEE Cloud Comput.*, vol. 2, no. 1, pp. 14–20, 2015, doi: 10.1109/MCC.2015.2.

[8] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, “Edge Computing Security: State of the Art and Challenges,” *Proceedings of the IEEE*, 2019, doi: 10.1109/JPROC.2019.2918437.

[9] Bundesamt für Sicherheit in der Informationstechnik (BSI), “Die Lage der IT-Sicherheit in Deutschland 2016,” Bonn, Oct. 2016.

[10] Y. Kim *et al.*, “Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors,” in *ACM/IEEE 41st International Symposium on Computer Architecture, ISCA 2014, Minneapolis, MN, USA, June 14-18, 2014*, 2014, pp. 361–372. doi: 10.1109/ISCA.2014.6853210.

[11] M. Yahuza *et al.*, “Systematic Review on Security and Privacy Requirements in Edge Computing: State of the Art and Future Research Opportunities,” *IEEE Access*, vol. 8, pp. 76541–76567, 2020, doi: 10.1109/ACCESS.2020.2989456.

[12] G. Pék, L. Buttyán, and B. Bencsáth, “A survey of security issues in hardware virtualization,” *ACM Comput. Surv.*, vol. 45, no. 3, pp. 40:1–40:34, 2013, doi: 10.1145/2480741.2480757.

[13] A. Abbas and S. U. Khan, “A Review on the State-of-the-Art Privacy-Preserving Approaches in the e-Health Clouds,” *IEEE J. Biomed. Health Informatics*, vol. 18, no. 4, pp. 1431–1441, 2014, doi: 10.1109/JBHI.2014.2300846.

[i] https://www.rcrwireless.com/20211216/telco-cloud/how-is-mobile-edge-computing-infrastructure-being-deployed, aufgerufen am 13.01.2021