3. Jg. | 24.2.2022 | Seiten 1–72 | www.rdz-online.de

Recht der Zahlungsdienste

1. 2022 Betriebs-Berater Geldverkehr

Fachmedien Recht und Wirtschaft | dfv Mediengruppe | Frankfurt am Main

NFC-Zahlungen und mögliche Sicherheitsrisiken*

Abbuchen von Geld im „Vorbeigehen", Auslesen/Kopieren von Karten durch kurzes Auflegen eines Smartphone, Mithören von Transaktionen aus der Ferne; all das sind häufig genannte Angriffsszenarien im Zusammenhang mit Near-Field-Communication-(NFC-)Zahlungen. Doch stellen diese Szenarien ein ernsthaftes Sicherheitsrisiko dar? Gibt es weitere kritische Sicherheitsaspekte? Unterscheiden sich Zahlungen mit der Plastikkarte dahingehend von jenen mit dem Smartphone? Der nachfolgende Beitrag gibt einen Überblick über NFC-Zahlungen und deren potenzielle Sicherheitsrisiken.

Dr. Michael Roland

Einleitung

NFC wurde 2002 von Philips (heute NXP) und Sony als drahtlose Kommunikationstechnologie für den Austausch von Daten über sehr kurze Distanzen, speziell in Unterhaltungs- und Haushaltselektronik, vorgestellt. NFC ist dabei interoperabel zu den marktführenden Radio-Frequency-Identification-(RFID-)Chipkartensystemen NXP MIFARE und Sony FeliCa. Bei diesen Systemen kommuniziert eine aktive Komponente (Lesegerät) mit einer oder mehreren passiven Komponenten (smartcards). Letztere haben bei kontaktlosen Systemen nicht unbedingt die Form einer Karte und werden daher oft als Tags oder Transponder bezeichnet. Egal ob Smartcard, Tag oder Transponder haben diese i. d. R. keine eigene Energiequelle und werden stattdessen während der Interaktion vom Lesegerät über Funk mit Energie versorgt.

NFC hebt die strikte Trennung in Lesegerät und Transponder auf, und integriert beides in einem Gerät. Dadurch können zwei NFC-Geräte, z. B. Smartphones, direkt miteinander kommunizieren. Dieses neue Kommunikationsprotokoll (Near Field Communication – Interface and Protocol [NFCIP-1], Norm ISO/IEC 18092) wird auch als Peer-to-Peer-Modus (P2P) bezeichnet. Neben P2P unterstützen NFC-Geräte noch zwei weitere Betriebsarten: den Reader/Writer-Modus (RW) und den Card-Emulation-Modus (CE). Im RW-Modus agiert ein NFC-Gerät als Lesegerät für kontaktlose Chipkarten. Aktuelle NFC-Geräte implementieren dabei eine Reihe von RFID- und Chipkartenstandards im 13,56-MHz-Band. Im CE-Modus agiert ein NFC-Gerät als Chipkarte und kann so mit anderen Lesegeräten, z. B. Bezahlterminals, kommunizieren. Im Gegensatz zum RW-Modus wird für den CE-Modus gewöhnlich nur ein einzelnes Funkprotokoll (in Europa vorw. ISO/IEC 14443 Typ A) unterstützt.

In den letzten 20 Jahren hat der Begriff NFC eine deutliche Wandlung erfahren. Heute versteht man darunter nicht mehr nur P2P-Kommunikation, Interoperabilitätsmechanismen und das Bezahlen mit dem Handy (CE-Modus), sondern ordnet praktisch alle kontaktlosen Chipkartensysteme im 13,56-MHz-Band NFC zu. Dazu zählen neben kontaktlosen Bank- und Kreditkarten auch viele weitere Bezahl-, Zutritts- sowie Ticketing-Systeme und elektronische Identitätsdokumente wie der Reisepass.

Bezahlen mit NFC

Über die Jahre entstanden verschiedenste NFC-Bezahlsysteme. Die prominentesten Vertreter sind die EMV Contactless Payment Systeme der globalen Kredit- und Debitkartenschemata. Einige Zeit versuchte man auf Smartphones auch NFC-P2P-Bezahlsysteme sowie individuelle (meist) Insellösungen umzusetzen. Der P2P-Modus weist aber in vielen modernen Smartphones Einschränkungen im Hinblick auf User-Experience und Funktionalität auf, welche solche Lösungen erschweren. Auch aufgrund der Komplexität des CE-Modus in Mobiltelefonen setzte man lange Zeit auf Lösungen abseits von EMV Contactless. Der CE-Modus erforderte zunächst stets einen dedizierten Smartcard-Chip, das Secure Element (SE). Dieses kann fest im Telefon verbaut (Embedded SE) oder als Karte ins Telefon einlegbar sein (z. B. als Teil der Subscriber-Identity-Module- [SIM-]Karte, die ja ohnehin eine Smartcard ist). Im Unterschied zu herkömmlichen Smartcards ist beim SE vorgesehen, dass mehrere Dienstanbieter ihre Applikationen nebeneinander in einem SE betreiben. Die Kontrolle über das SE muss dazu in die Hand eines Broker (trusted service manager – TSM) gelegt werden, der anschließend jedem Anbieter einen eigenen, streng abgeschotteten Bereich im Chip zur Verfügung stellt. In der Praxis wurde dieses Modell jedoch bisher nicht wirklich angenommen (*Alattar/Achemlal*, in: 2014 IEEE Intl Conf on High Performance Compu-

* Dieser Text entstand im Rahmen von Digidow, dem Christian Doppler Labor für Private Digitale Authentifizierung in der Physischen Welt, gefördert durch das österreichische Bundesministerium für Digitalisierung und Wirtschaftsstandort und die Nationalstiftung für Forschung, Technologie und Entwicklung.

ting and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), 506–509). Abgesehen von kleineren Roll-Outs, bei denen sich oft Mobilfunkbetreiber zu einem TSM zusammenschlossen, um spezifische Anwendungen auf die SE (typischen SIM-Karten) ihrer Kunden zu bringen (vgl. Bankomatkarte Mobil in Österreich oder OPTIMOS in Deutschland), fand das Konzept keine weite Verbreitung. Lediglich Apple, Google und Samsung gelang der Einstieg in SE-basierte NFC-Wallets auf den eigenen Geräten. Vielen Drittanbietern bleibt der Zugang zum SE jedoch weiterhin versperrt.

Ein Ansatz, um Card Emulation auch ohne SE und damit verbundene Hürden zu ermöglichen, ist Host-Based Card Emulation (HCE). Mit HCE kann jede beliebige App am Smartphone die Rolle der Smartcard-Applikation übernehmen. Ein Smartcard-Chip ist dazu nicht nötig. Schon länger als mögliche Alternative zum SE bekannt, wurde die Emulation von Chipkarten-Applikationen mit reinen Softwarelösungen als unsicher und nicht praxistauglich eingestuft (*Reveilhac/Pasquet*, in: Langer u. a. [Hrsg.], 2009 First International Workshop on Near Field Communication, 75–80). Erst 2011, mit dem Blackberry 7, wurde HCE in einem Smartphone verfügbar. Zwei Jahre später öffnete auch Google den HCE-Modus für Android Apps. Google nutzt seither selbst HCE für das eigene Google Pay NFC-Wallet. Damit kann Google eine EMV-Bezahlapplikation anbieten, ohne diese in ein SE bringen zu müssen. Das ist von Bedeutung, weil das SE bei Android-Geräten oft unter Kontrolle eines anderen Geräteherstellers liegt. Apple setzt hingegen beim Apple Pay NFC-Wallet bis heute ausschließlich auf ein SE. Dieses steht bei allen Apple-Geräten jedoch stets unter der Kontrolle von Apple selbst, wodurch das komplexe SE-Ökosystem umgangen wird.

Ablauf einer Kontaktlos-Zahlung

Zur Beurteilung von Sicherheitsrisiken bei NFC-Zahlungen ist es hilfreich, den groben Ablauf einer Transaktion zu kennen. Die EMV Contactless Payment Systeme fassen Protokolle unterschiedlicher Schemata zusammen. Aktuelle Protokollvarianten laufen weitgehend nach demselben Muster ab; dabei ist es grundsätzlich egal, ob mit Karte oder Smartphone bezahlt wird:

1. Das Bezahlterminal erkennt, aktiviert und durchsucht die Karte nach einer passenden Bezahlapplikation.
2. Anschließend werden statische Kartendaten ausgelesen. Dazu zählen Informationen über das Zahlungsverfahren (z. B. zulässige Transaktionsarten und Nutzerverifikation) und Daten über die Karte und den Aussteller (z. B. Kartennummer/Primary Account Number (PAN), Gültigkeitszeitraum, Zertifikatkette bzw. öffentliche kryptographische Schlüssel zur Kartenverifikation).
3. Das Terminal sendet die von der Karte geforderten Transaktionsdaten (u. a. Betrag samt Währung, Zeitstempel, Informationen über das Terminal) an die Karte. Diese erstellt eine digitale Signatur über statische Kartendaten und die Transaktionsdaten.
4. Durch Verifikation dieser Signatur mithilfe der Zertifikatkette erhält das Terminal die Bestätigung, dass die Transaktion mit einer echten Bezahlkarte durchgeführt wurde.
5. Anhand von Risikomanagementparametern aus Karte und Terminal entscheidet das Terminal, ob die verifizierte Transaktion zur späteren „Einlösung" offline gespeichert wird, oder ob eine unmittelbare Autorisierung online über das Payment-Netzwerk erforderlich ist. Bei NFC-Zahlungen könnte ein solcher Schritt auch die PIN-Verifikation umfassen, welche bei NFC üblicherweise online durch den Kartenherausgeber erfolgt.

Der Ablauf einer Zahlung lässt sich also mit der Unterschrift eines Dokuments mittels digitaler Signatur vergleichen. Der geheime Signaturschlüssel ist dabei nur der Chipkarte bekannt und kann aus dieser nicht ausgelesen werden. Lediglich der öffentliche, zur Verifikation notwendige Teil des Schlüssels ist frei lesbar.

Angriffsflächen

Häufig genannte Angriffsszenarien im Zusammenhang mit NFC-Zahlungen sind das Abbuchen von Geld im Vorbeigehen, das Auslesen/Kopieren von Karten und das Mithören von Transaktionen aus der Ferne. Es stellt sich daher die Frage, ob diese Angriffe auch in der Praxis ein Sicherheitsrisiko darstellen.

Viele der Daten im Chip sind tatsächlich frei lesbar. Es ist keine Authentifizierung notwendig, und die Kommunikation mit dem Chip erfolgt unverschlüsselt. Während die Kommunikation bei NFC nur über kurze Distanzen möglich ist, kann ein Mithören auch mehrere Meter entfernt noch möglich sein. Es ist daher nicht auszuschließen, dass ein Angreifer vor dem Supermarkt steht und bei Transaktionen an der Kasse mitlauscht. Dadurch können Details über die Transaktion (z. B. bezahlter Betrag) und die verwendete Karte (z. B. PAN) ermittelt werden. Nachdem die PAN den Karteninhaber effektiv identifiziert, stellt diese Form des Angriffs ein Privatsphäre-Problem dar. Der Name des Karteninhabers ist aber üblicherweise nicht per NFC auslesbar.

Die belauschte Kommunikation reicht jedoch nicht aus, um weitere Transaktionen durchzuführen. Die digitale Signatur unterschreibt nur genau die belauschte Transaktion und ist nicht für weitere Transaktionen wiederverwendbar. Lediglich die PAN und andere Kartendaten könnten ein Sicherheitsproblem darstellen, wobei die PAN nur als Identifikationsmerkmal gedacht und ohne

zusätzliche Authentifizierung nicht zur Autorisierung von Transaktionen vorgesehen ist. Für Zahlungen in Onlineshops sollte bspw. zusätzlich zumindest der Verifikationscode auf der Kartenrückseite für die Authentifizierung erforderlich sein. Dieser Code ist allerdings nicht im Chip gespeichert und steht dem Angreifer somit auch nicht zur Verfügung. Manche Onlineshops akzeptieren Kreditkarten dennoch alleine durch Eingabe von PAN und Ablaufdatum. In diesen Fällen wäre ein erfolgreicher Angriff mittels der abgehörten Daten möglich. Allerdings könnte man diesem Szenario durch den Einsatz mehrerer PAN pro Karte entgegenwirken. Prinzipiell könnte eine Karte als Aufdruck eine PAN, im Chip aber eine andere PAN haben. Man spricht in diesem Zusammenhang von Tokenization, weil die eigentliche PAN durch mehrere Token-PAN repräsentiert wird. Token-PAN werden dann im Payment-Netzwerk wieder auf die eigentlichen PAN zurückgeführt.

Ähnliche Möglichkeiten ergeben sich für einen Angreifer mit (kurzzeitigem) Zugriff auf die Karte. Dabei sind alle frei lesbaren Daten kopierbar – auch jene, die bei einer Transaktion nicht übertragen werden. Der Chip ist allerdings kein einfacher Datenspeicher. Es handelt sich dabei vielmehr um einen Minicomputer, dessen Software zwar geheime Signaturschlüssel für die Berechnung digitaler Signaturen im Chip verwendet, aber zuverlässig verhindert, dass diese von außen gelesen werden. Damit kann ein Angreifer auch keine voll funktionsfähige Kopie (skimming) eines EMV-Chip erstellen. Dennoch gab es bereits erfolgreiche Angriffe dieser Art. Am EMV-Chip sind z. B. auch Daten analog zum Magnetstreifen gespeichert. Während darin erforderliche Verifikationscodes fehlen und die Daten für die Reproduktion eines Magnetstreifens ungeeignet sein sollten, konnten Forscher aufzeigen, dass diese Codes bei der Freigabe von Transaktionen z. T. nicht geprüft werden (*Galloway*, It Only Takes a Minute to Clone a Credit Card, Thanks to a 50-Year-Old Problem, 2020, 20–21). Auch wir selbst konnten die Abwärtskompatibilität zu älteren Protokollen ausnutzen, um eine funktionsfähige Kartenkopie durch Vorausberechnen von Autorisierungscodes mit einer echten Karte zu erzeugen (*Roland/Langer*, in: 7th USENIX Workshop on Offensive Technologies [WOOT '13], 2013).

Ein weiteres Angriffsszenario ist das Abbuchen von Geld im Vorbeigehen. Dabei ist es wesentlich zu verstehen, dass am Chip selbst kein Guthaben gespeichert ist. Aus einer Karte ist also kein Geldwert extrahierbar bzw. auf eine andere Karte übertragbar. Der Chip ist lediglich der Schlüssel, der Transaktionen zwischen dem Karteninhaber-Konto und einem Händlerkonto autorisiert. Für das Abbuchen von Geld ist daher ein Bezahlterminal samt verknüpftem Händlerkonto erforderlich. Angriffe lassen sich also zu einem bestimmten Händlerkonto und damit auch zu einer Person zurückverfolgen. Der direkte Angriff mit eigenem Bezahlterminal ist daher nicht praktikabel.

Anders könnte es sich bei einer indirekten Variante, dem Relay-Angriff, verhalten. Beim Relay-Angriff hält ein Angreifer mit einem NFC-Lesegerät direkten Kontakt zur Karte, während ein weiterer Angreifer mit einem Kartenemulator an der Kassa zahlt. Kartenemulator und Lesegerät leiten die Kommunikation zwischen Bezahlterminal an der Kassa und der echten Karte über einen anderen Kanal, z. B. das Internet, weiter (*Francis u. a.*, in: Lo/Li [Hrsg.], Radio Frequency Identification System Security, 2011, 21–32). Das Bezahlterminal glaubt somit, direkt mit der echten Karte zu sprechen. Dieses Szenario unterliegt jedoch Einschränkungen. So muss die Zahlung genau dann stattfinden, wenn auch Kontakt zur angegriffenen Karte besteht. Beide Angreifer müssen also zeitgleich agieren. Hinzu kommt die Schwierigkeit des Zugriffs per NFC. Dies ist nur über kurze Distanzen möglich; mit dem Smartphone als Lesegerät nur wenige Millimeter. Münzen, Geldclips und eine Geldbörse mit mehreren NFC-Karten erschweren den Angriff erheblich; moderne Geldbörsen sind zudem oft mit Abschirmungen ausgestattet. Ein Angriff im Vorbeigehen ist also in der Praxis nicht so einfach umsetzbar. Praktikabel könnte dieser Angriff werden, wenn man z. B. alle Sitze in einem öffentlichen Verkehrsmittel mit Lesegeräten ausstattet. Durch die gleichzeitige Verfügbarkeit mehrerer Karten über einen längeren Zeitraum werden die Erfolgschancen für einen Angriff erheblich gesteigert. Während die Detektion von Relay-Angriffen nicht trivial ist, sieht man mittlerweile erste Ansätze in den EMV-Protokollen.

Mehr-Faktor-Authentifizierung schwächt mögliche Angriffsflächen signifikant ab. Ist neben der Karte (Besitz) auch noch eine PIN (Wissen) erforderlich, muss auch dieser Wissensfaktor ermittelt werden, bevor ein erfolgreicher Angriff gelingt. Unter dieser Voraussetzung ist aber der Diebstahl der Karte mitunter einfacher als ein Angriff über NFC. Bei NFC-Zahlungen ist allerdings nicht immer eine PIN erforderlich. In Österreich waren PIN-lose Zahlungen lange auf 25 Euro begrenzt, und nach maximal fünf Transaktionen musste zumindest eine kontaktbehaftete Transaktion durchgeführt werden.

Der Gewinn aus einem Angriff muss stets deutlich über den eingesetzten Kosten liegen, damit ein Angriff rentabel wird. Entscheidend ist dabei, welche Waren mittels Angriffs erworben und wie diese wieder zu Bargeld gemacht werden können. Vorteilhaft sind der Kauf von Wertkarten oder insbesondere Cashback-Transaktionen, bei denen neben dem Erwerb von Waren direkt Bargeld ausgezahlt wird. Die Limits für PIN-lose NFC-Zahlungen werden jedoch kontinuierlich erhöht, wodurch Angriffe rentabler werden. Cashback sollte allerdings, unabhängig vom Betrag, immer eine PIN erfordern.

Von der Karte zum Smartphone

NFC-Zahlungen mit Smartphone unterscheiden sich praktisch nicht von jenen mit Karte. Die Angriffsflächen gelten daher auch beim Smartphone, jedoch ergeben sich sowohl neue Möglichkeiten als auch neue Risiken. Eine Wallet-App kann die emulierte NFC-Karte ein- und ausschalten; oft ist das auch automatisch an das Ein- und Ausschalten des Bildschirms gekoppelt. Damit wird der Zugriff im Vorbeigehen bzw. das unbemerkte Auslesen erheblich erschwert. Eine Wallet-App kann auch jederzeit neue virtuelle Karten in das Smartphone laden und mehrere Karten parallel verwalten. Die Aktivierung virtueller Karten für eine Transaktion ist zudem an gerätespezifische Sicherheitsmechanismen (z. B. die Verifikation eines Fingerabdrucks direkt am Smartphone) koppelbar.

Neben dem verbesserten Schutz öffnen sich aber gleichzeitig neue Angriffsflächen. Im NFC-Smartphone wird das EMV-Protokoll entweder von einem Smartcard-Chip (dem SE) oder von einer App (HCE) abgewickelt. Der NFC-Controller-Chip ist dabei das Modem für die NFC-Schnittstelle zum Bezahlterminal. Das SE ist direkt mit dem NFC-Controller verbunden. Apps können daher nicht in die Datenübertragung zwischen Terminal und SE eingreifen; auch ein Mitlauschen ist nicht möglich. Allerdings müssen Wallet-Apps Applikationen und Daten auf dem SE verwalten. Es gibt daher auch einen Kanal zwischen SE und Apps. Der Zugriff darauf wird vom Smartphone-Betriebssystem geregelt. Aufgrund seiner Komplexität kann dieses Betriebssystem aber nicht annähernd ein ähnliches Schutzniveau wie das SE bieten. Kann die App eines Angreifers das Betriebssystem überlisten, ist ein direkter Zugriff auf das SE möglich. Applikationen und Daten am SE können so freilich nicht manipuliert werden. Dafür ist ein gesicherter, authentifizierter Kanal zwischen SE und Dienstanbieter/TSM notwendig; die Schlüssel dazu sind aber nie am Smartphone. Ein möglicher Angriff wäre, Authentifizierungsmechanismen des SE durch wiederholte, fehlschlagende Anmeldeversuche dauerhaft zu blockieren. Durch Tunneln der Kommunikation zwischen SE und einem Angreifer über das Internet könnte eine entsprechende App einen Relay-Angriff durchführen (*Roland u. a.*, in: Gritzalis u. a. [Hrsg.], Information Security and Privacy Research, 2012, 1–12). Durch den permanenten Zugriff ist diese Variante deutlich praktikabler als bei NFC-Karten. Allerdings lässt sich dieser Angriff, obwohl in der Vergangenheit bspw. bei Google Wallet erfolgreich aufgezeigt, durch einfache Softwaremodifikationen im SE zuverlässig verhindern.

Anders sieht die Situation bei HCE aus. Der NFC-Controller leitet die Kommunikation vom Bezahlterminal direkt an eine App weiter. Das Smartphone-Betriebssystem kümmert sich um die Auswahl der richtigen App, und darum, dass keine andere App die Kommunikation belauscht oder übernimmt. Ist letzteres möglich, kann eine manipulierte App im schlimmsten Fall wie ein Lesegerät auf die HCE-App zugreifen und so einen Relay-Angriff durchführen. Die Sicherheit von HCE baut also wesentlich auf der Sicherheit des Betriebssystems auf. Der fehlende Schutz durch das SE muss anderweitig ausgeglichen werden. Aus diesem Grund ist auch die Speicherung von Kartendaten (insbesondere geheimer Schlüssel) bei HCE komplexer. Erste HCE-Konzepte gingen vom Betrieb eines SE in der Cloud aus. Die HCE-App leitet, ähnlich dem Relay-Szenario, die Kommunikation zwischen Terminal und Cloud-SE weiter. Der Zugriffsschutz auf das Cloud-SE ist damit die einzige Barriere vor einem Angreifer.

Moderne HCE-Konzepte gehen davon aus, dass man Kartendaten und geheimes Schlüsselmaterial zumindest kurzfristig auch auf dem Smartphone speichern darf. Dazu kommt Tokenization zum Einsatz: Statt einer einzelnen Karte wird gleich eine Serie von Token-PAN (mit individuellen Signaturschlüsseln) ins Smartphone geladen. Jedes Token ist eine eigene, virtuelle Karte; jedoch mit zeitlicher und eventuell örtlicher Begrenzung. So kann ein Token nur wenige Minuten lang gültig oder auf eine einzige Transaktion beschränkt sein. Gelingt es einem Angreifer, Token zu extrahieren, kann damit nur eine begrenzte Menge an Zahlungen ausgeführt werden, was die Praktikabilität eines Angriffs deutlich schmälert. Nur wenn kontinuierlich neue Token beziehbar sind, könnten diese gewinnbringend eingesetzt werden.

Zur weiteren Risikominderung werden oft Techniken wie Whitebox-Cryptography und Root-Detection eingesetzt. Whitebox-Cryptography versucht, nach dem Security-by-Obscurity-Prinzip kryptographische Algorithmen und Schlüssel ineinander zu verweben und die Funktionsweise von Programmcode zu verschleiern, um die Extraktion geheimer Schlüssel zu erschweren. Root-Detection versucht zu ermitteln, ob Sicherheitsmechanismen des Betriebssystems durch den Benutzer (un-)wissentlich ausgehebelt wurden. Während Root-Detection ein sinnvolles Kriterium bei der Risikoabwägung zur Zahlungsautorisierung im Backend darstellen kann, wird die Effektivität beider Techniken von der Sicherheitsforschung in Frage gestellt.

Nachdem HCE vergleichsweise neu ist, wird erst die Zukunft zeigen, ob auch in der Praxis erfolgreiche Angriffe auftauchen.

AUTOR

Dr. Michael Roland *ist Post-Doc am Institut für Netzwerke und Sicherheit der Johannes Kepler Universität Linz. Er forscht im Bereich digitaler Identitäten, NFC, Chipkarten und Drahtlostechnologien mit den Schwerpunkten Sicherheit und Privatsphäre.*