VoIP-Sicherheit versus Sprachqualität

Sprach-barriere

Peter Backs, Norbert Pohlmann, Claas Rettinghausen

Kritik an Voice over IP bezieht sich meist auf zwei Aspekte: die mangelnde Sprachverständlichkeit und das Fehlen eines Abhörschutzes. *iX* geht der Frage nach, ob und wie sich beide Forderungen erfüllen lassen.

Oft berücksichtigen Entscheider bei der Wahl von VoIP-Lösungen nur die potenzielle Kostenersparnis sowie die flexible und einfache Handhabung, vernachlässigen aber einen anderen wesentlichen Aspekt: die Sicherheit. Die Autoren haben im Rahmen einer Forschungsreihe am Institut für Internet-Sicherheit der FH Gelsenkirchen untersucht, inwieweit unterschiedliche Ansätze helfen, Sicherheitslücken effektiv und einfach zu schließen, und wie stark sie die Benutzung von VoIP beeinflussen.

Voice over IP kapselt Sprachdaten in IP-Pakete und versendet sie über das Internet oder LAN. Üblicherweise kommt dabei das User Datagram Protocol (UDP) zum Einsatz, das im Gegensatz zu TCP keine Liefergarantie gibt. Es eignet sich jedoch besser für Echtzeitkommunikation: Das Neuanfordern verlorener Pakete, wie es TCP durchführt, führt zu Verzögerungen, die die Sprachqualität verschlechtern können. Allerdings genügt UDP allein nicht für eine isochrone Audio- oder Video-Übertragung. Dafür zeichnet das auf UDP aufsetzende Realtime Transport Protocol (RTP) verantwortlich. Das dazugehörige RTP Control Protocol (RTCP) misst während des Telefonats periodisch die Übertragungsqualität des Transportnetzes. Beide sind in RFC 3550 spezifiziert (siehe Kasten „Onlinequellen“).

Zur Anrufsteuerung verwendet man meist das Session Initiation Protocol (SIP, RFC 3261). Es signalisiert dem anderen Teilnehmer, dass ein Gespräch stattfinden soll, handelt die Kommunikationsparameter aus – etwa UDP-Ports und die zu verwendende Sprachkodierung – und beendet die Verbindung wieder, sobald ein Teilnehmer den Hörer auflegt.

Wartezeiten unerwünscht

Einfluss auf die Sprachqualität einer VoIP-Verbindung nimmt unter anderem der verwendete Codec. Er verwandelt das analoge Sprachsignal in ein digitales und umgekehrt. Steht nur eine geringe Übertragungskapazität zur Verfügung – etwa im WAN –, kann er die Sprachdaten zusätzlich komprimieren. Wichtiger ist jedoch, dass sich das Signal mit minimaler Verzögerung kodieren und dekodieren lässt. Tabelle 1 führt die bei der Untersuchung berücksichtigten Codecs auf.

Auch die Eigenschaften des Netzes – gemeinhin unter dem Schlagwort „Quality of Service“ (QoS) zusammengefasst – beeinflussen die Qualität der Sprachübertragung. Für gute Verständlichkeit sind drei Kriterien zu erfüllen:

– Die benötigte Übertragungskapazität muss während des gesamten Gesprächs zur Verfügung stehen.

– Paketverluste (Packet Loss) müssen verhindert oder durch den Codec ausgeglichen werden.

– Pakete müssen rechtzeitig und regelmäßig beim Empfänger eintreffen (Delay, Jitter).

Je nach Entfernung der Endgeräte und Anzahl der Router auf dem Übertragungsweg kann die Verzögerung (Delay) stark schwanken. Die Signallaufzeit (Propagation Delay) – die Zeit, die ein Signal dafür benötigt, eine Leitung zu durchqueren – ist proportional zur Entfernung, genauer gesagt zur Leitungslänge. Hinzu kommt die Vermittlungsverzögerung (Switching Delay). Netzkoppelelemente wie Router oder Switches müssen

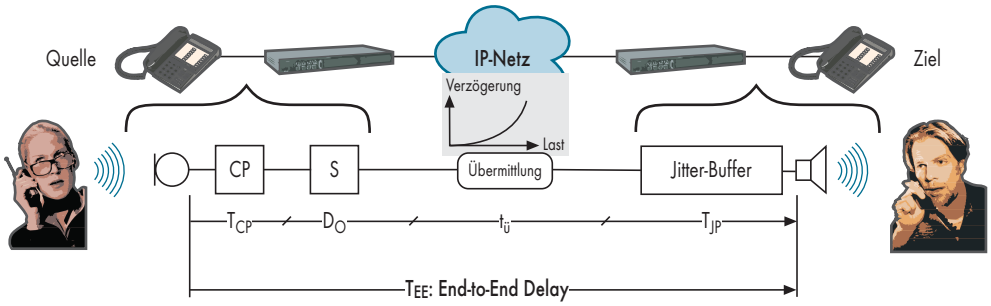
Sprach-Codecs

Codec	Delay (ms)	Bitrate (kBit/s)	MOS
G.711 µ-Law	10	64	4,1-4,5
G.711 A-Law	10	64	4,1-4,5
G.726	1	32	3,85-4,2
G.723.1	30	6,3	3,9
G.729a	10	8	3,9
GSM	20	13,2	3,75

in der Regel warten, bis alle Bits eines Pakets eingetroffen sind. Anschließend wählen sie eine Route zum Empfänger und serialisieren die Bits wieder (Store-and-Forward). Die Warteschlangen eines Routers verzögern die Verarbeitung zusätzlich (Queueing Delay). Muss ein Rechner aufgrund des Zugriffsverfahrens des Netzes – etwa CSMA/CD bei Ethernet – warten, bis er Daten senden darf, spricht man von Zugriffsverzögerung (Access Delay). Sie ist jedoch in der Regel vernachlässigbar.

Der maximale Durchsatz ergibt sich aus der verwendeten Hardware, wobei das schwächste Glied die Kapazität einer Übertragungsstrecke bestimmt. Theoretisch sind Verzögerung und Durchsatz voneinander unabhängig. Messungen zeigen jedoch, dass mit steigendem Verkehrsaufkommen auch die Verzögerung steigt.

Jitter steht bei VoIP für Schwankungen der Verzögerungszeit (Delay Jitter), die ihrerseits die Sprachqualität herabsetzen. Üblicherweise enthalten VoIP-Endgeräte deshalb einen Jitter-Puffer, der die Unregelmäßigkeiten ausgleicht. Er erhöht jedoch wiederum die Gesamtverzögerung (siehe Abbildung 1).



Lange Reise: Netze, VPN-Router und Endgeräte tragen zur Verzögerung des Sprachsignals bei (Abb. 1).

Das Ohr als Messgerät

Leider liefern die Zahlen keine direkte Aussage, ob ein VoIP-Gespräch verständlich ist oder nicht. Darum wurden Verfahren entwickelt, die die Verbindungsqualität durch eine einzige Zahl bewerten:

– Mean Opinion Score (MOS) ist der am häufigsten zu findende Wert zur Qualitätsbeschreibung. Man ermittelt ihn, indem man eine repräsentative Auswahl von Testpersonen Sprachproben auf einer Skala von 1 bis 5 bewerten lässt (siehe Tabelle 2). Dabei steht der Wert 5 für optimale Sprachqualität; ISDN-Verbindungen erreichen einen MOS von etwa 4,5, analoge Festnetzverbindungen ungefähr 3,5. Andere Verfahren, die ohne die subjektive Bewertung durch Testpersonen auskommen, können den MOS rechnerisch herleiten.

– Sprachmusterorientierte Verfahren senden ein definiertes Sprachmuster über ein Netz und vergleichen das an der Gegenstelle eintreffende Signal mit dem ursprünglichen. Das für VoIP wichtigste Verfahren Perceptual Evaluation of Speech Quality (PESQ) hat die ITU-T in der Empfehlung P.862 ratifiziert. Der Vorgänger PSQM (Perceptual Speech Quality Measurement) berücksichtigte nur den Einfluss des Codecs auf die Sprachqualität, PESQ lässt darüber hinaus die QoS-Parameter in die Bewertung einfließen. Es funktioniert allerdings bei „schlechten" Netzen mit hohem Delay oder Packet Loss nicht zuverlässig. Als Ergebnis liefert PESQ unter anderem den hergeleiteten MOS.

– Netzbasierte Verfahren bewerten die Verbindungsqualität passiv, ohne das Einspielen spezieller Sprachmuster. Das häufig verwendete E-Modell produziert anhand der gemessenen und erwarteten QoS-Parameter sowie des verwendeten Sprachcodecs den sogenannten R-Faktor, der die Qualität der Sprachübertragung angibt. Er liegt zwischen 0 und 100, wobei 100 optimale Sprachqualität bedeutet; 94 entspricht ISDN-Qualität. Der R-Faktor lässt sich ebenfalls in den entsprechenden MOS umrechnen.

Nicht für aller Ohren

Telefonie ist vor E-Mail und Post nach wie vor das wichtigste Kommunikationsmedium der heutigen Gesellschaft. Geradezu selbstverständlich verlassen sich Nutzer von analogen oder ISDN-Telefonen auf die Sicherheit der Telefonnetze, deren Schutz ausschließlich von den Beschränkungen des physischen Zugangs zur Telefonleitung abhängt.

Bei VoIP muss der Angreifer ebenfalls Zugang zu den Systemen oder Übertragungsmedien haben, die die VoIP-Telefonate übermitteln. Durch einfache Spoofing-Angriffe und die für den Benutzer nicht vorhersehbaren Routingwege der IP-Pakete erhöht sich die Zahl der potenziellen Angreifer im Vergleich zur herkömmlichen Telefonie jedoch erheblich. Zudem ist ungesichertes VoIP mit einem normalen PC und frei erhältlicher Software kompromittierbar, während man für das Abhören von ISDN-Gesprächen zumindest spezielle – obgleich nicht komplizierte – Hardware benötigt. Wer die Sicherheit eines VoIP-Gesprächs gewährleisten will, muss daher zusätzliche Maßnahmen ergreifen.

Sicherungsmechanismen können sowohl VoIP-spezifisch als auch allgemeiner Natur sein. Zu Letzteren zählen Virtual Private Networks

iX-TRACT

- VoIP-Verbindungen lassen sich mit protokollspezifischen oder allgemeinen Verfahren sichern.
- Beide Ansätze erhöhen die zu übertragende Datenmenge, wirken sich jedoch nur unwesentlich auf QoS-Parameter wie die Verzögerung aus.
- In langsamen Netzen empfiehlt es sich, VoIP-spezifische Sicherheitsmaßnahmen wie Secure RTP (SRTP) einzusetzen, da sie weniger Overhead erzeugen.

Die MOS-Skala (Mean Opinion Score)

MOS	Rating	Bedeutung
5	excellent	keinerlei Anstrengung zum Verständnis der Sprache notwendig; totale Entspannung möglich
4	good	keine Anstrengung, aber Aufmerksamkeit notwendig
3	fair	leichte, moderate Anstrengung nötig
2	poor	merkbare, deutliche Anstrengung nötig
1	bad	keine Verständigung möglich

(VPN). Sie können den gesamten Verkehr zwischen zwei oder mehreren Netzen absichern.

Da der Verschlüsselungskanal (Security Association, kurz SA) bei einem VPN meist permanent besteht, muss man die verwendeten Schlüssel regelmäßig ändern. Das sogenannte Re-Keying kann während eines Gesprächs Einfluss auf die Qualität der Verbindung und damit auf die VoIP-Sprachqualität haben, wenn der Aufbau einer neuen SA erst mit dem Ablaufen der alten stattfindet. Daher sollte die VPN-Lösung in der Lage sein, neue SAs rechtzeitig vor dem Ablaufen der momentan verwendeten zu erzeugen.

Keine Infrastruktur für Public Keys

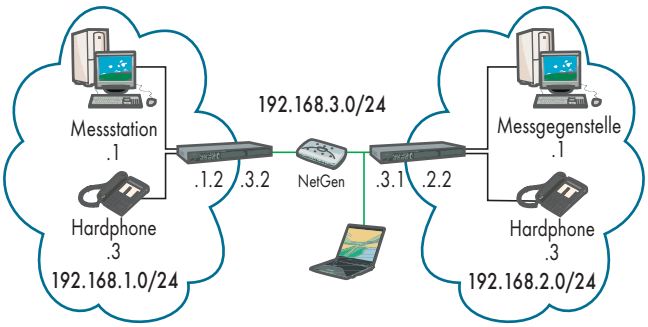
Darüber hinaus hat eine generische Lösung wie ein VPN den Nachteil, dass die Sicherung erst an der Grenze des lokalen Netzes beginnt beziehungsweise endet. Das LAN transportiert alle Sprachdaten unverschlüsselt, ein Angreifer kann sie daher leicht lesen. Außerdem sind organisationsübergreifende VPNs – etwa eine Kopplung unterschiedlicher Firmennetze – nur mit erheblichem Aufwand realisierbar, weshalb eine flächendeckende VoIP-Absicherung oft scheitert. In erster Linie eignet sich der VPN-Ansatz daher für den internen VoIP-Verkehr einer Organisation.

VoIP-spezifische Sicherheitsmaßnahmen erweitern die existierenden VoIP-Protokolle. Üblicherweise überträgt man SIP-Daten per SSL/TLS gesichert (SIPS) von einem Hop zum nächsten und verwendet Secure RTP (SRTP, RFC 3711) zur Verschlüsselung des Medienstroms zwischen den Teilnehmern (End-to-end).

Von entscheidender Bedeutung ist dabei die Vertrauenskette zwischen den an der Signalisierung beteiligten Intermediären, etwa den SIP-Gateways der jeweiligen Provider. Die Teilnehmer tauschen die für SRTP verwendeten Schlüssel im Klartext innerhalb der SIP-Signalisierungsnachrichten aus. Dadurch kommen die Intermediäre in den Besitz der Schlüssel, und es besteht die Gefahr eines Man-in-the-middle-Angriffs (MITM Attack).

Des Weiteren ist eine organisationsübergreifende Public-Key-Infrastruktur (PKI) oder eine andere Infrastruktur für kryptografische Schlüssel Voraussetzung, um Kommunikationspartner zu authentifizieren. Angesichts der Tatsache, dass sich eine weitflächige PKI etwa für den E-Mail-Dienst bislang nicht durchsetzten konnte, bleibt jedoch zweifelhaft, ob sich verbreitetes sicheres Telefonieren mit den etablierten VoIP-Standards durchsetzen wird.

Um den organisatorischen Aufwand zu umgehen, hat PGP-Erfinder Philip R. Zimmermann ein Protokoll entwickelt, das ohne PKI auskommt: ZRTP führt eine ungesicherte Diffie-Hellman-Schlüsselaushandlung durch und baut anhand der daraus abgeleiteten Schlüssel einen gesicherten SRTP-Kanal auf. Um die konzeptionell bedingte Anfälligkeit des Verfahrens gegen Man-in-the-middle-Angriffe zu kompensieren, stellt ZRTP einen Hash-Wert der öffentlichen Diffie-Hellman-Keys bereit, den die Gesprächspartner mündlich („In-band“) vergleichen und sich dadurch authentifizieren können. Nachteilig bei dem Verfahren ist, dass die Authentifizierung durch den Menschen erfolgt, was naturgemäß nie völlig sicher ist. Außerdem lassen sich fremde Stimmen – im Geschäftsumfeld die Regel – prinzipiell nicht authentifizieren.



Messaufbau zur Untersuchung des Einflusses von VPN-Tunneln auf die Quality of Service (QoS) (Abb. 2).

Sicher und trotzdem verständlich

Den Einfluss der Sicherungsmaßnahmen auf die Sprachqualität haben die Autoren im Labor untersucht. Alle Messungen fanden in einem isolierten LAN statt, um Störquellen im Netzwerk auszuschließen und unverfälschte Messergebnisse zu garantieren. Der Testaufbau bestand aus drei unterschiedlichen Subnetzen: den Sender- und Empfänger-Netzen *192.168.1.0* und *192.168.2.0* sowie dem Transportnetz *192.168.3.0*. Die Subnetze mit den VoIP-Komponenten sind durch Security Gateways gekoppelt (siehe Abbildung 3).

Als VPN-Gateways kamen die frei verfügbare IPsec-Implementierung OpenSWAN und das ebenfalls freie OpenVPN (SSL/TLS) zum Einsatz. Ein Netz mit gewöhnlichen Routern anstelle der VPN-Gateways diente als Referenz. Die Messungen wurden mit einer Beta-Version des QoS-Messsystems NetGage, entwickelt vom Projekt QoSSIP der FH Köln, und der VoIP Test Suite von ITD Informationstechnologie durchgeführt. Die Bridge-Software NetGen – ebenfalls eine Entwicklung des QoSSIP-Projekts – simulierte zwischen den Gateways ein reales Netz mit einstellbarem Delay, Jitter und Packet Loss. Mit dem Netzbenchmark Iperf erhöhten die Tester schrittweise die Last im Netz.

Sowohl mit NetGage als auch mit der VoIP Test Suite ließ sich der Einfluss der VPN-Tunnel auf Jitter und Packet Loss messen und ergaben eine zusätzliche Ver-

Onlinequellen

Institut für Internet-Sicherheit der FH Gelsenkirchen	www.internet-sicherheit.de
RFC 3550 – RTP: A Transport Protocol for Real-Time Applications	tools.ietf.org/html/rfc3550
RFC 3261 – SIP: Session Initiation Protocol	tools.ietf.org/html/rfc3261
RFC 3711 – The Secure Real-time Transport Protocol (SRTP)	tools.ietf.org/html/rfc3711
ZRTP: Media Path Key Agreement for Secure RTP	zfoneproject.com/docs/ietf/draft-zimmermann-avt-zrtp-06x.html
OpenSWAN	www.openswan.org
OpenVPN	openvpn.net
QoSSIP-Projekt	www.qossip.de/index.php?catid=25
ITD VoIP Test Suite	www.trafficlyser.de/hp/itd/front_content.php?idcat=306
Iperf	dast.nlanr.net/projects/Iperf/

zögerung von ein bis zwei Millisekunden, die für die Sprachverständlichkeit nahezu unerheblich ist.

Beide VPN-Gateway-Implementierungen können die übertragenen Daten zusätzlich komprimieren. Verwendet man den PCM-Codec G.711, gleicht die Kompression den durch die Gateways entstehenden Overhead aus. Kommt ein komprimierender Codec zum Einsatz, bleibt die zusätzliche Kompression im Gateway ohne Wirkung. In beiden Fällen erhöht sich die Verzögerung um weniger als eine Millisekunde. Da die Gateways bei aktivierter Kompression mehr Arbeit leisten müssen, ist allerdings zu erwarten, dass der Gesamtdurchsatz sinkt.

Beim VoIP-spezifischen Ansatz ließen sich mit der eingesetzten Software keine detaillierten Messungen durchführen – sie beherrscht die verwendeten Sicherungsprotokolle noch nicht. Allein den Datendurchsatz beziehungsweise die Bitrate kann man zuverlässig messen. Daraus ergibt sich ein Overhead von rund rund 10 %, bei Verwendung von SRTP sowie den Codecs G.723.1 und G.729a. Die VPN-Gateways verursachen in derselben Situation mit oder ohne Kompression einen Overhead zwischen 75 und 90 % relativ zu einer ungesicherten Verbindung. Allerdings können mehrere Parameter die Messungen beeinflussen:

– Paketierung der Sprachsegmente: Überträgt man mehrere Sprachsegmente pro IP-Paket, verringert sich der Protokoll-Overhead. Die Verzögerung nimmt jedoch zu.

– Codec-Bitrate: Die Codecs unterscheiden sich hinsichtlich der benötigten Übertragungskapazität und der Sprachqualität voneinander.

– Codec-Segmentgröße: Abhängig von der Segmentgröße übertragen Codecs mehr oder weniger Pakete auf einmal, was den Protokoll-Overhead beeinflusst.

– Blockgröße der verwendeten Verschlüsselung: Die verwendeten Algorithmen verschlüsseln Daten blockweise. Stehen nicht genug Daten zur Verfügung, müssen sie Null-Bytes als „Füller“ einfügen.

Fazit

Die untersuchten Sicherheitsmaßnahmen wirken sich vorrangig auf die benötigte Übertragungskapazität aus, nicht jedoch auf Delay, Jitter oder Packet Loss. Abhängig von den Einstellungen der VoIP-Software und der Security-Gateways kann sich die Größe der VoIP-Pakete mehr als verdoppeln. Anwendern mit langsamer Internetanbindung sowie Unternehmen mit vielen VoIP-Nutzern kann dies erhebliche Schwierigkeiten bereiten. Durch geeignete Einstellungen der VoIP- und Verschlüsselungssoftware lässt sich jedoch die Bitrate beeinflussen. Ist die zur Verfügung stehende Übertragungskapazität gering, sollte man jedoch das VoIP-spezifische SRTP einem VPN-Tunnel vorziehen. (mr)

PETER BACKS

ist IT-Consultant/Developer bei der Sirrix AG in Saarbrücken.

PROF. DR. NORBERT POHLMANN

ist Informatikprofessor für Verteilte Systeme und Informationssicherheit sowie Leiter des Instituts für Internet-Sicherheit – if(is) an der Fachhochschule Gelsenkirchen.

CLAAS RETTINGHAUSEN

ist VoIP System Architekt bei der Carpo Deutschland GmbH in Ratingen.

iX-Link ix0806110