Jurnal Elektro
p-ISSN: 1979-9780
e-ISSN: 2746-4288

Jurnal Elektro Vol.16 No.1 April 2023
journal homepage: https://ejournal.atmajaya.ac.id/index.php/JTE

Analisis *Audio Capture* untuk Pengumpulan Data pada Smart Speaker

Andre Zenik[1], Lukas[2*], Catherine Olivia Sereati[3], Kumala Indriati[4], Linda Wijayanti[5]

[1,2,3,4,5]Prodi Teknik Elektro, Fakultas Teknik
Universitas Katolik Indonesia Atma Jaya

Article Info

Article history:

Received
3-12-2023

Accepted
9-12-2023

Keywords:
Audio capture, security, smart speaker

Abstract

Internet of Things or IoT is a technology that is currently trending, because it can help human in their daily routine. One of the IoT product that used in homes is smart speaker. However, every latest technological innovation does not escape from vulnerabilities and one of them is the microphone on the smart speaker. Several studies and research have found that security vulnerabilities in smart speakers, such as dolphin attacks and audio capture. To find out how this audio capture technique works, an experiment was made to understand how it works and the impact of this technique. This experiment uses a smart speaker with voice assistant Alexa, using the Alexa Skills Kit services and Flask-Ask framework to create an audio capture program. The results of this program testing are expected to be used as a benchmark to prevent smart speakers from becoming the target of this technique anymore.

Info Artikel

Histori Artikel:

Diterima:
3-12-2023

Disetujui:
9-12-2023

Kata Kunci:
Audio capture, security, smart speaker.

Abstrak

Internet of Things atau IoT adalah sebuah teknologi yang sedang trending saat ini, dari awal kemunculannya yang diminati karena dapat membantu pekerjaan manusia. Produk-produk IoT kini sudah dapat ditemukan dimana saja, termasuk di dalam rumah , salah satunya adalah smart speaker. Akan tetapi, setiap inovasi teknologi terbaru tidak luput dari adanya celah keamanan dan salah satunya adalah mikrofon yang terdapat pada smart speaker. Beberapa penelitian dan riset telah menemukan bahwa celah keamanan pada smart speaker, seperti dolphin attack dan audio capture. Untuk membuktikan cara kerja dari teknik audio capture ini, maka dibuat sebuah penelitian untuk memahami tentang cara kerja dan dampak dari teknik tersebut. Penelitian ini menggunakan smart speaker dengan voice assistant Alexa, menggunakan layanan Alexa Skills Kit dan framework Flask-Ask untuk membuat program audio capture. Hasil dari pengujian program ini diharapkan dapat digunakan sebagai patokan untuk mencegah agar tidak ada lagi smart speaker menjadi sasaran teknik ini.

1. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi pada era modern ini sudah sangat maju. Banyak teknologi baru yang dibuat untuk memenuhi keperluan manusia sehari-hari,salah satunya adalah *Internet of Things* (IoT). *Smart Speaker* merupakan salah satu jenis IoT yang sering digunakan di dalam rumah, mempunyai fungsi sama seperti *speaker* pada umumnya yaitu sebagai pemutar lagu. *Smart Speaker* dilengkapi dengan mikrofon, sehingga bisa berinteraksi dengan pengguna seperti berbicara, menerima perintah dari pengguna seperti menyalakan lampu, menyalakan televisi dan masih banyak lagi yang bisa dilakukan oleh *smart speaker* [6].

Audio Capture adalah salah satu teknik serangan yang memanfaatkan mikrofon untuk merekam pembicaraan secara diam-diam. Teknik ini menggunakan program yang dimasukkan dan dijalankan pada *smart speaker* untuk menyalakan mikrofon, kemudian saat ada pembicaraanyang terjadi di dekat *smart speaker,*pembicaraan tersebut direkam dan didengarkan oleh pelaku (*hacker*).

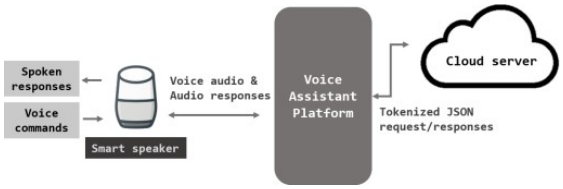
*Corresponding author: Lukas
Email address:lukas@atmajaya.ac.id

Pembahasan kali ini mengenai pemanfaatan celah keamanan pada *smart speaker* untuk melakukan *audio capture. Smart speaker* yang mempunyai program ini, dapat merekam pembicaraan yang ada di sekitar jangkauan *smart speaker* secara diam-diam. Hasil yang sudah didapatkan dari perekaman ini, dikirimkan dan diterima oleh *Command and Control* (CnC) melalui perantara layanan sebuah *voice assistant.*

2. LANDASAN TEORI

2.1 Smart Speaker

Smart speaker adalah sebuah alat yang terdiri dari komponen mikrofondan *speaker* yang ditanam sebuah *voice assistants* berbasis *cloud. Voice assistants* yang tersedia antara lainAmazon Alexa, dan Google Assistant yang sudah berbasis *cloud.* Cara kerja dari *smart speaker* dimulai dari masukan yang diberikandari pengguna berupa instruksi atau respon yang kemudian diterima oleh mikrofon, lalu dikirim menuju *voice assistant platform* dan diteruskan menuju server. Selanjutnya servermengolah sinyal yang diberikan dan memberikan respon atas sinyal tersebut menuju *smart speaker,* dan instruksi tersebut akan dijalankan [6].



Gambar 1. Cara kerja smart speaker

2.2 Audio Capture

Audio capture merupakan salahsatu teknik yang digunakan untuk mendapatkan informasi dari korban dengan memanfaatkan mikrofon.Teknik ini membuat mikrofon selalu menyala untuk mendapatkan informasiyang diterima dari percakapan di sekitar secara-diam-diam [1] [2] [7].

2.3 Automatic Speech Recognition

Speech Recognition atau yangbiasa dikenal dengan automatic speech recognition (ASR) merupakan suatu pengembangan teknik dan sistem yang memungkinkan komputer untukmenerima masukan berupa kata yangdiucapkan.

Teknologi ini memungkinkanuntuk identifikasi kata yang diucapkan dapat dibaca oleh *smart speaker* sebagai sebuah perintah untukmelakukan suatu pekerjaan, misalnya penekanan tombol pada telepon genggam yang dilakukan secaraotomatis dengan komando suara Alexa merupakan sebuah ASR yang dikembangkan oleh Amazon untuk sebuah *smart speaker.* Untuk berinteraksi dengan Alexa, diperlukan sebuah *intent* untuk mengenali kata- kata dari sebuah percakapan yang tertuju ke *smart speaker* [3].

2.4 Skills

Skills adalah sebuah sebutanuntuk menjelaskan sebuah program layanan yang tersedia pada sebuah *voice assistant* yang ada pada *smart speaker* [2] [3] .Adanya *skills* ini membuat sebuah *smart speaker* memiliki berbagai macam kemampuan,yang dapat membantu manusia.

2.5 Flask-Ask

Flask-ask adalah web framework yang dikhususkan untuk perangkat yang menggunakan *voice assistant* Alexa, dalam membuat sebuah *skill* tanpa menggunakan layanan server dari Alexa [5]. Flask-Ask dikembangkan oleh John Wheeler pada tahun 2017 untuk mengembangkan sebuah *skills* pada*voice assistant* Alexa dengan bahasa Pemograman Python yang telah disederhanakan.

2.6 Tunneling Protocol

Tunneling adalah sebuah protocol komunikasi yang menghubungkan antar perangkat dengan jalur pengiriman data yang aman. Pengiriman data yang aman diperoleh dari proses fragmentasi data dan

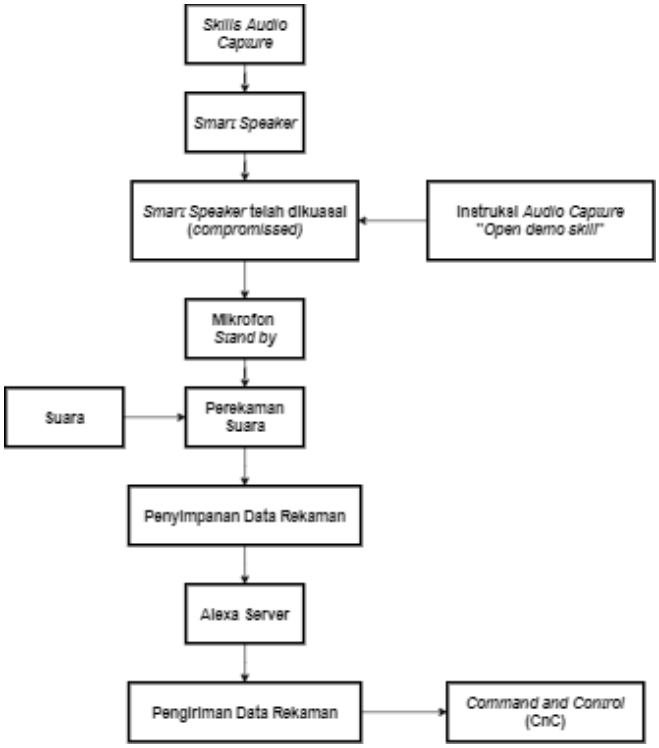
enkripsi yang dilakukan sebelum data dikirimkan [1]. Ngrok adalah sebuah programyang dibuat oleh Alan Shreve yang dapat menghubungkan jaringan publik ke port komputer lokal menggunakan tunneling. Ngrok akan membuat / membuka jaringan private melalui NAT atau firewall untukmenghubungkan *local host* ke internet dengan jalur yang aman menggunakan HTTPS [4].

3. METODOLOGI PENELITIAN

3.1 Diagram Blok Sistem

Skills yang dibuat adalah sebuah program untuk merekam pembicaraan pengguna, dan hasil rekaman tersebut diterima oleh divais Computer numerical Control (CnC) router. *Skills* dimasukkan ke dalam *smart speaker* dan menunggu untuk diaktifkan hingga instruksi telah diberikan. Setelah *skills* aktif, *smart speaker* sudah bisa merekam pembicaraan pengguna yang berada di sekitar *smart speaker.*

Selama *skills* masih berjalan didalam *smart speaker*, suara percakapanyang terdengar diterima dan disimpan sementara di dalam server. Hasil perekaman tersebut lalu dikirimkan menuju CnC menggunakan *tunneling* dari ngrok.



Gambar 2. Blok diagram *audio capture*

3.2 Perancangan Sistem

Persiapan untuk penelitian ini adalah *smart speaker* telah terhubung ke internet dan sudah memiliki *voice assistant* di dalamnya. *Smart speaker* ini telah memasang *skills* untuk melakukan *audio capture* yang sudah bisa dijalankan saat *smart speaker* menyala.

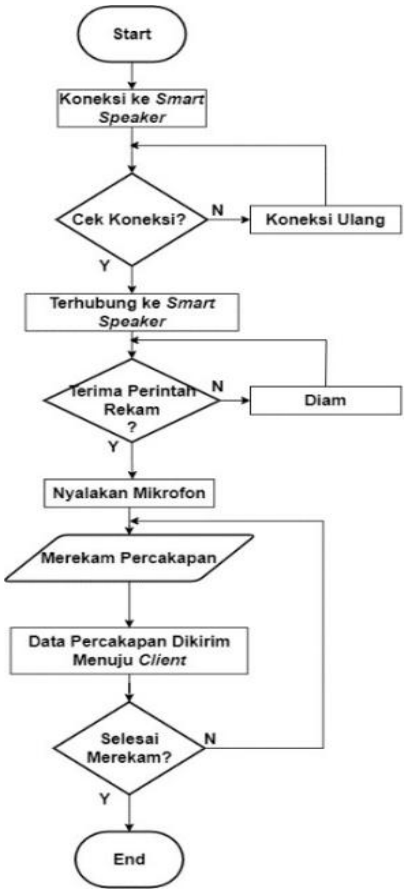
Lalu ada sebuah *server* yang bertugas untuk menjalankan layanan / service dari *skill audio capture* ini. *Server* yang digunakan merupakan komputer lain yang sudah di *tunneling* pada Alexa Skills Kit menggunakan ngrok. *Skills* yang dijalankan menggunakan *framework* dari Python yang bernama Flask-Ask. Flask-Ask digunakan karena memiliki keuntungan dari sisi *syntax* yang digunakan lebih sederhana dibanding *syntax* bawaan dari Alexa, dan dapat dijalankan di luar layanan *cloud* Alexa.

Selanjutnya terdapat komputer yang berperan sebagai CnC untuk menerima hasil dari *audio capture* dari *smart speaker.* CnC menerima data rekaman dari server Alexa yang telah disimpan setiap adanya jeda setelah merekam pembicaraan.

3.3 Perancangan Perangkat Lunak

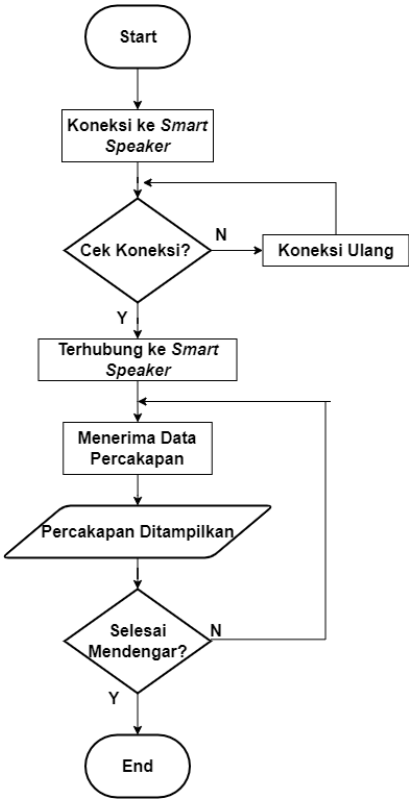
Program *audio capture* menggunakan sistem *skills* dan *client* dan diprogram dengan menggunakan bahasa pemrograman Python dan memanfaatkan layanan Alexa Skills Kit yang tersedia dari Amazon. Program *skills* bertugas untuk menerima suara yang didengar oleh *smart speaker.* Saat *skills* dijalankan, *skills* memulai untuk mengaktifkan mikrofon dan siap merekam pembicaraan. Mikrofon menunggu adanya pembicaraan yang sedang terjadi selama beberapa detik sebelum *timeout*, apabila mikrofon tidak menerima suara percakapan sama sekali maka *skills* berhenti. Saat mikrofon sudah mendengarkan semua pembicaraan yang terjadi di sekitar *smart speaker*, data percakapan tersebut disimpan, lalu dikirimkan secara berkala ke CnC.

Setiap ada jeda pembicaraan, data rekaman sebelumnya langsung ditampung dalam sebuah variabel sementara dan dikirimkan ke CnC. Setiap selesai mengirim data pembicaraan sebelumnya, *skills* melanjutkan tugasnya untuk merekam percakapan selanjutnya dan dikirimkan lagi ke CnC. *Skills* akan terus bekerja sampai tidak ada pembicaraan yang bisa didengar oleh *smart speaker.*



Gambar 3. Diagram alir program *audio capture skills*

Kemudian terdapat program *client* yang berfungsi untuk menampilkan hasil *audio capture* dari *smart speaker*. Program ini berjalan di dalam CnC untuk menunggu adanya rekaman yang terjadi di sekitar *smart speaker.*Setiap terjadinya rekaman pada *smart speaker,* data yang didapatkan kemudian dikirimkan melalui *tunneling*menggunakan ngrok menuju CnC. Hasil pengiriman ini lalu ditampilkan pada layar CnC. Setiap data yang masuk ditampilkan berdasarkan waktu diterimanya data rekaman dari *server.*



Gambar 4.Diagram Alir Program Audio Capture Client

4. PEMBAHASAN DAN HASIL

Pengujian *audio capture* ini bertujuan untuk mengetahui apakahdata hasil *audio capture* sama dengan yang telah direkam *smart speaker.*Pengujian ini dilakukan dengan cara merekam berbagai percakapan yang terjadi pada *smart speaker*. Perangkat yang digunakan antara lain sebuah komputer, dan sebuah *smart speaker.* Komputer digunakan sebagai pusat kontrol dari*skills* yang dijalankan dan sebagai penerima hasil dari *audio capture.* Kemudian *smart speaker* yangdigunakan saat ini menggunakan *voice assistant* Alexa. Alexa memiliki sebuah fitur yang digunakan untuk mengenali percakapan yaitu *intent. Intent* adalah sebuah tempat untuk menampungpercakapan yang memiliki objek yang sejenis. Di dalam *intent,* terdapat sebuah *slot* yang merupakan variabel yang mengelompokkan tipe objeksesuai kategori tertentu. Hal ini bertujuan untuk mempermudah Alexa dalam mengetahui objek yangdiucapkan dari pengguna.

Selain itu, diperlukan beberapa *tools* untuk menjalankan program ini antara lain Python sebagai bahasa pemograman, Flask-Ask sebagai *framework* untuk *skills* dari Alexa,ngrok sebagai *tunneling server* menuju Alexa Skills Kit dan Alexa Skills Kit sebagai layanan pengelola *skills* pada Alexa.
Penelitian dimulai dengan menjalankan program *client* pada CnC dan *skills* pada server terlebih dahulu, aktivasi *skill* pada *server* dan menghubungkan *skills* ke Alexa Skills Kit .Setelah *skills* sudah berjalan, dilanjutkan dengan memberikan instruksi untuk menjalankan *audio capture* pada *smart speaker*. Baik instruksi yang diberikan dan data yang diucapkan menggunakan Bahasa Inggris, karena saat ini Alexa belum menyediakan layanan dalam Bahasa Indonesia. Data yang digunakan dalam pengujian ini antara lain, nama orang, posisi kota, negara asal, kode pin dan nomor telepon. Data-data digunakan untuk penelitian ini karena merupakan informasi yang cukup penting dan rahasia, untuk mendapatkan identitas seseorang, mengetahui lokasi dan negara tempat tinggal, mengetahui kode pin kartu ATM dan nomor telepon.

4.1 Perekaman Data Berupa Nama Orang

Uji coba ini bertujuan untuk mendapatkan nama seseorang yang mungkin terucap pada saat berbicara. Nama yang telah direkam oleh *smart speaker.* Uji coba ini dilakukan masing-masing sebanyak 1 (satu) kali pengucapan oleh penutur Bahasa Indonesia, bukan penutur asli Bahasa Inggris. Hasil penelitian dapat dilihat pada Tabel 1.

Tabel 1. Uji coba *Audio capture* untuk mendapatkan informasi berupa nama seseorang

Data yang Direkam	**Hasil *Audio Capture***	**Status**
Anne	Anna	X
John Cena	John Cena	OK
Steve Jobs	Steve Jokes	X
Alvin	Alvin	OK
Barack Obama	Barack Obama	OK
Albert Einstein	Albert Einstein	OK
Michael	Error	X
Kevin	Error	X
Jack	Error	X
Octavius	Error	X

Dari data hasil uji coba pada Tabel 1, dapat dilihat bahwa tidak semua nama bisa dikenali oleh *smart speaker*. Hal ini terjadi karena adanya masalah dalam pengucapan, dalam hal ini disebabkan oleh penutur bukan penutur asli Bahasa Inggris, sehingga menyebabkan kesalahan dalam menerima nama yang berujung pada error yang dihasilkan.

4.2 Perekaman Data Berupa Posisi Kota

Uji coba selanjutnya dilanjutkan dengan mendapatkan informasi lokasi kota dengan menggunakan data kota- kota besar di Indonesia.

Tabel 2. Uji coba Audio capture untuk mendapatkan informasi nama-nama kota

Data yang Direkam	**Hasil *Audio Capture***	**Status**
Jakarta	Jakarta	OK
Medan	Mentor	X
Palembang	Hollenberg	X
Manado	Manado	OK
Pontianak	Pompeiana	X
Merauke	Marathi	X
Ternate	Ternate	OK
Ambon	Ambon	OK

Dari data hasil pengujian padaTabel 2, hasil yang didapatkan sudah cukup baik, namun masih terdapat beberapa nama kota yang salah. Hal inidisebabkan karena nama-nama kota di atas dibuat dengan

penuturan BahasaInggris, sehingga Alexa masih belummengenali penuturan dalam BahasaIndonesia dan menyebabkan kesalahan penerimaan data.

4.1 Perekaman Data Berupa Lokasi Negara

Uji coba selanjutnya adalahmendapatkan informasi lokasi negaradari hasil *audio capture* pada *smart speaker*. Hasil penelitian dapat dilihat pada Tabel 3.

Tabel 3. Uji coba Audio capture untuk mendapatkan informasi berupa nama-nama negara

Data yang Direkam	**Hasil *Audio Capture***	**Status**
Indonesia	Indonesia	OK
Kamboja	Cambodia	OK
Laos	Laos	OK
Rusia	Russia	OK
Jepang	Japan	OK
Spanyol	Spain	OK
Belgia	Belgaum	X
Jerman	Germany	OK

Hasil uji coba pada Tabel 3 menunjukkan hampir semua nama negara dapat diterima dari *audio capture* dengan baik. Kesalahan yang terjadi juga disebabkan oleh kesalahan pengucapan, sehingga menyebabkan penerimaan suara dari *smart speaker* berbeda dari seharusnya.

4.2 Perekaman Data Berupa Kode Pin

Penelitian selanjutnya adalah mendapatkan kode pin 4 (empat) digit dan 6 (enam) digit dari *audio capture* pada *smart speaker*.

Tabel 4. Uji coba *Audio capture* untuk mendapatkan informasi kode pin

Data yang Direkam	Hasil Audio Capture	Status
3129	3129	OK
7719	7719	OK
3571	3571	OK
4576	4576	OK
7919	7919	OK
629156	629156	OK
912479	912479	OK
817614	817614	OK
961409	961409	OK

Dari hasil uji coba pada Tabel 4, data pin yang diucapkan bisa diterima dengan baik oleh *smart speaker.* Percobaan ini berhasil karena jumlah digit pin yang sedikit, yaitu 4 (empat) dan 6 (enam) digit, sehingga *smart speaker* dapat mendengar dengan baik.

4.3 Perekaman Data Berupa Nomor Telepon

Uji coba selanjutnya adalah mendapatkan nomor telepon yang didapatkan dari *audio capture* pada *smart speaker*.

Tabel 5. Uji coba *Audio capture* untuk mendapatkan informasi berupa nomor telepon

Data yang Direkam	Hasil *Audio Capture*	Status
081212345437	081212345437	OK
081242112917	081242112917	OK
081282188214	081282188214	OK
089612479127	089612479127	OK
089691244911	089691244911	OK

Dari hasil uji coba pada Tabel5, *smart speaker* dapat mendengar pengucapan nomor telepon dan hasilnya sama seperti yang diucapkan. Akan tetapi untuk pengujian nomor telepon memiliki kesulitan tersendiri, bila pengucapan nomor telepon tiba- tiba terhenti, maka nomor telepon yang diterima tidak lengkap.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Teknik *audio capture* bias mendapatkan informasi penting, seperti identitas seseorang, nomor telepon dan kode pin dengan memanfaatkan *smart speaker* dalam mengumpulkan informasi. Teknik ini memanfaatkan celah keamanan pada *smart speaker,* khususnya pada mikrofon untuk mengumpulkan informasi. Hasil dari perekaman data suara ini yang diambil oleh CnC ini digunakan untuk merugikan orang lain dan termasuk pengguna yang telah menjadi korban.

Keterbatasan dari percobaan ini adalah tidak tersedianya Bahasa Indonesia dalam konfigurasi bahasa pada Alexa, sehingga percobaan ini menggunakan Bahasa Inggris agar data yang diucapkan bisa diterima oleh *smart speaker.* Keterbatasan lainnya adalah Alexa telah membuat sebuah pembatasan pada data yang dapat diterima oleh *smart speaker,* sehingga mempersulit dalam mengumpulkan informasi dari keseluruhan kalimat yang diucapkan.

5.2 Saran

Saran yang bisa diberikan terkait dengan pembahasan seputar *audio capture* pada *smart speaker* antara lain:

Pemilik / pengguna *smart speaker*:

1. Selalu mematikan mikrofon pada saat *smart speaker* tidakdigunakan.
2. Selalu mewaspadai *skills* yang ingin ditambahkan pada *smartspeaker.*
3. Memastikan agar jaringan internet yang digunakan selalu aman dan ter-*update*.

Peneliti yang ingin mendalami topik ini:

1. Pengembangan lebih lanjut untuk program *audio capture*, terutama pada cara untuk mendapatkan informasi utuh dari pembicaraan.
2. Implementasi pada *smart speaker* dengan *voice assistant* berbeda, seperti Google.
3. Jangan menyalah gunakan program ini untuk tindak kriminal.

Memberikan edukasi kepada orang awam tentang bahaya dari *audio capture,* dan cara mengamankan *smart speaker*dari teknik *audio capture.* Section headings should be concise and numberred sequentially, using a decimal system for subsections. Emphasized *words* should be italicized, but such emphasis should be sparingly used.

DAFTAR PUSTAKA

[1] A. Ohri, "Tunneling Protocol a Complete Overview," Jigsaw academy, 2021. [Online]. Available: https://www.jigsaw academy.com/blogs/cyber-security/tunneling-protocol/. [Accessed 1- 7- 2022]

[2] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, . A. Bates and M. Bailey, "Emerging Threats in Internet of Things Voice Services," *IEEE,* vol. 17, no. 4, pp. 18-24, 2019.

[3] H. Chung, M. Iorga, J. Voas and S. Lee, "Alexa, Can I Trust You?," *IEEE,* vol. 50, no. 9, pp. 100-104, 2017.

[4] K. Casey, "Deploying ngrok in Production," ngrok, 3 5 2022. [Online]. Available: https://ngrok.com/blog-post/deploying-ngrok-in-production. [Accessed 2 7 2022].

[5] K. Reitz, "Flask-Ask Rapid Alexa Skills Kit Developments for Amazon Echo Devices," Flask-Ask documentation, 2016. [Online]. Available: https://flask-ask.readthedocs.io/en/latest/. [Accessed 30 6 2022].

[6] N. Malkin, J. Deatrick, A. Tong and P. Wijesekera, "Privacy Attitudes of Smart Speaker Users," *Proceedings on Privacy Enhancing Technologies,* vol. 4, pp. 250-271, 2019.

[7] Y. Park, H. Choi, S. Cho and Y. G. Kim, "Security Analysis of Smart Speaker: Security Attacks and Mitigation," *Computers, Materials & Continua,* vol. 61, no. 3, pp. 1075-1090, 2019