به نام خدا

ساخت یک سیستم تشخیص نفوذ برای محیط اینترنت اشیا با استفاده از تکنیک های یادگیری ماشین

نویسندگان:
K. V. V. N. L Sai Kiran[a] , R. N. Kamakshi Devisetty[a], , N. Pavan Kalyan[a], K. Mukundini[a], R. Karthi[a],*

مترجم: ماهان شکرانی

نام استاد: سمیه جعفرعلی جاسبی

خلاصه مقاله

شبکه اینترنت اشیا در برابر حملات آسیب پذیر است و شناسایی این رفتارهای مخرب در مراحل اولیه می تواند داده ها را از حملات نجات دهد. هدف اصلی این کار ساخت مدل های یادگیری ماشین برای شناسایی حملات در شبکه اینترنت اشیا است. برای ساخت یک مدل ، داده های عادی و حمله از محیط اینترنت اشیا تولید می شود. یک بستر آزمایش برای شبیه سازی محیط اینترنت اشیا با استفاده از Node MCU ESP8266 ، سنسور DHT11 و روتر بی سیم ساخته شده است. یک سیستم حمله کننده با استفاده از یک سیستم لپ تاپ ساخته می شود که اقدامات حملات sniffing و Poisoning را انجام می دهد. داده های گرفته شده از سنسورها دما ، رطوبت و نقطه هشدار بودند که با استفاده از درگاه بی سیم به سیستم عامل Think Speak منتقل می شوند. در فاز نرمال ، مقادیر سنسور توسط Node MCU گرفته شده و به سرور Think Speak انتقال داده می شود که به عنوان داده های عادی ذخیره و برچسب گذاری می شوند. در مرحله حمله ، از سیستم حمله کننده ، مهاجم مخفیانه داده ها را رهگیری می کند ، هنگام انتقال داده ها بین سرور Node MCU و Think Speak ، داده ها را اصلاح می کند. در مرحله حمله ، حمله Man in the Middle با استفاده از ARP Poisoning در شبکه انجام می شود و داده های گرفته شده به عنوان داده های حمله برچسب گذاری می شوند. طبقه بندی یادگیری ماشین مانند Naïve Bayes، SVM ، درخت تصمیم ، Adaboost برای دسته بندی داده ها ساخته شده اند.

معرفی مراحل بررسی مقاله

در محله ی اول این آزمایش دارای پیکربندی سخت افزاری و نرم افزاریست:

برای ایجاد سخت افزار از Node MCU ESP8266 که با سنسور DHT 11 تجهیز شده، از طریق WIFI به سرور Think Speak متصل شده. داده ها از طریق روتر بی سیم که به عنوان دروازه ای در شبکه عمل می کند به بستر سرور ThinkSpeak منتقل می شوند.

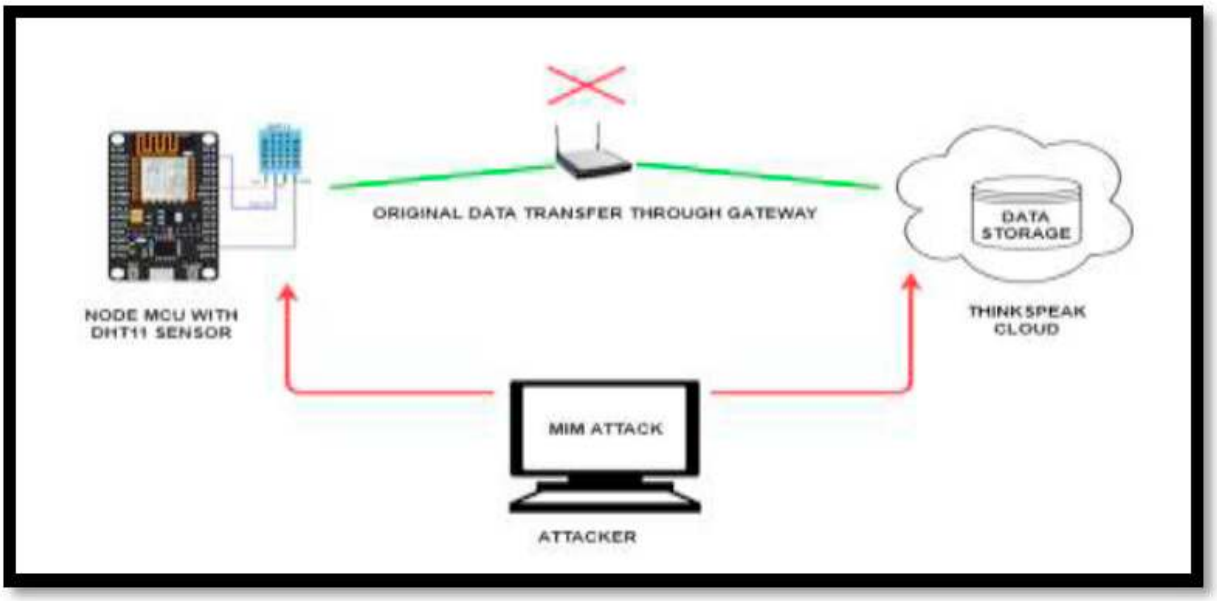

Fig. 1. Architecture of the proposed IoT system

ThinkSpeak کانالهایی را فراهم می کند که تمام داده های ارسال شده (دما،رطوبت و رطوبت نقطه ای (نقطه ی شبنم)) توسط برنامه را ذخیره می کند. هر کانال با استفاده از شناسه کانال خود شناسایی می شود و از کلیدهای API برای نوشتن و خواندن داده ها از کانال ها استفاده می شود. با ایجاد یک کانال جدید ، کلید API به طور خودکار ایجاد و اضافه می شود.

دیتا های ارسال شده با استفاده از کتابخانه های ESP8266 به صورت خودکار کد گذاری می شوند.

برای پیکربندی نرم افزار، نیاز به یک لپتاپی که حداقل دارای رم 4 گیگ و پردازنده ی intel core i7 می باشد. بر روی این لپتاپ نرم افزار kali linux رو نصب می کنیم و برای شناسایی آدرس های IP های ارسالی به wire shark تجهیز می کنیم.

به صورت کلی در شکل ۲ نحوه ی حمله کردن به سیستم و خواندن اطلاعات و عوض کردن دیتا ها به صورت شماتیک نشان داده شده است.

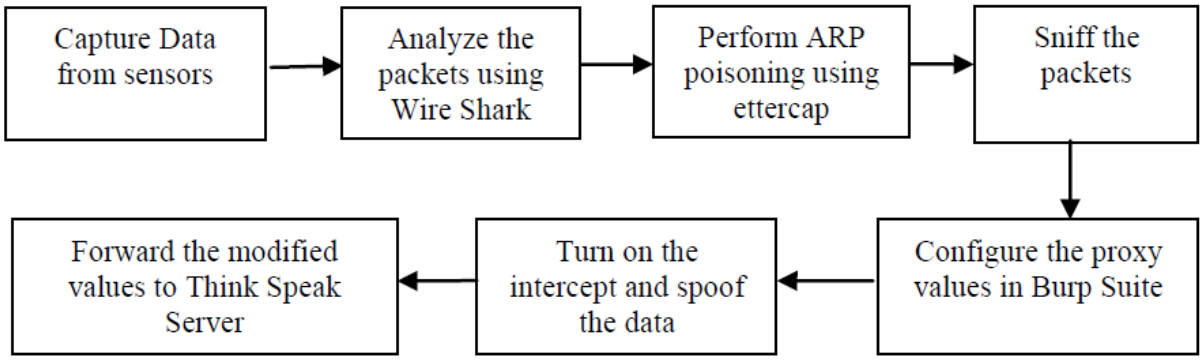

Fig. 2 Design of attacks in Test bed

زمانی که با استفاده از wire shark آدرس IP و MAC و مشخصات و داده ها استخراج شد از Ettercap (برای انجام حملات Man in the Middle کاربرد دارد)در kali Linux استفاده می کنیم که به صورت زیر عمل می کند.

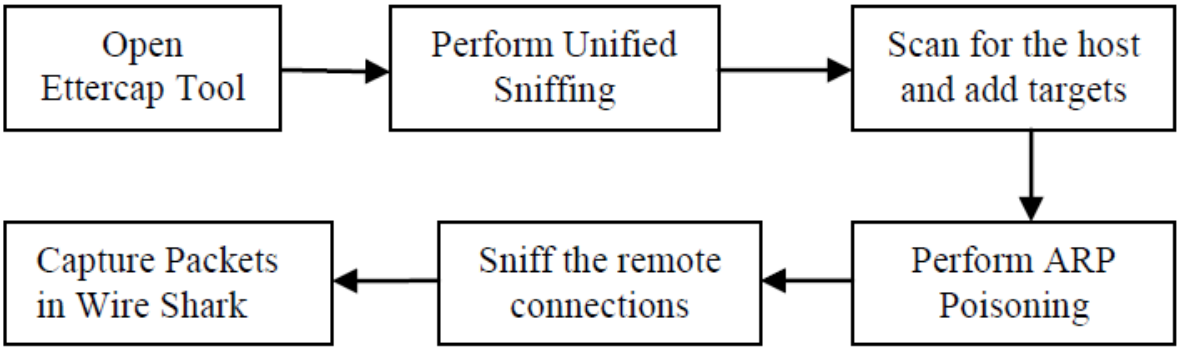

Fig. 3 Attack generation using Ettercap

برای گرفتن اطلاعات در سمت سرور از Burp Suite (ابزار امنیتی منبع باز است که برای انجام و آزمایش ویژگی های امنیتی کاربرد دارد) استفاده می شود که به شکل زیر اطلاعات را دریافت و بعد از تغییرات، تحت پوشش Node MCU به Think Speak می فرستد.

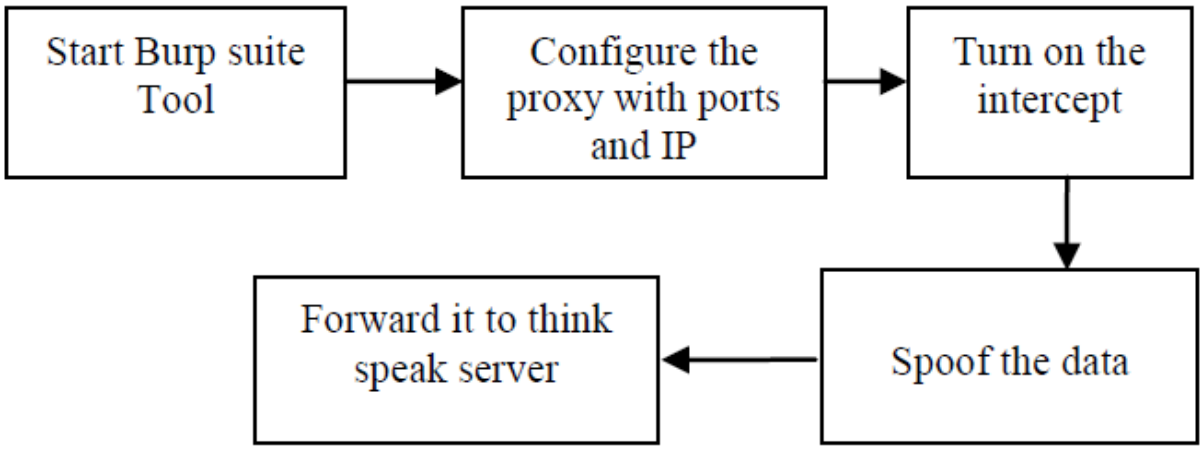

Fig. 4 Attack generation using Burp suite

دیتا های دریافتی بدون تغییرات را در این شکل می بینید:

دیتا هایی که مورد حمله قرار گرفتن و تغییر کردن به صورت زیر است:

برای طبقه بندی داده های حمله شده و عادی در بحث یادگیری ماشین به چهارالگوریتم SVM، Naïve Bayes، درخت تصمیم وadaboost استفاده شده.

- الگوریتم SVM داده های خطی را انتخاب می کند که بیشترین برخورد را با حاشیه اطمینان داشته باشد، به این صورت طبقه بندی می کند. (داده ها به صورت کامل و درست طبقه بندی شدند به جز ۱ مورد)

Table 2. Confusion Matrix for SVM algorithm.

		Actual Class	
		Normal	Attack
Predicted Class	Normal	54	0
	Attack	1	40

- الگوریتم درخت تصمیم، مدل درختی از تمام حالات ایجاد می کند و تک تک آن ها را بررسی می کند و طبقه بندی می کند. (داده ها به صورت کامل و درست طبقه بندی شدند)

Table 3. Confusion Matrix for Decision Tree algorithm

		Actual Class	
		Normal	Attack
Predicted Class	Normal	54	0
	Attack	0	41

- الگوریتم Naïve Bayes طبقه بندی با استفاده از احتمالات قبلی، تلاش ها و خطاهاست. (داده ها به جز ۲ نقطه، به صورت کامل و درست طبقه بندی شدند)

Table 4. Confusion Matrix for Naïve Bayes algorithm

		Actual Class	
		Normal	Attack
Predicted Class	Normal	54	0
	Attack	2	39

- الگوریتم Adaboost یک الگوریتم تقویت کننده است که با استفاده از درخت پایه ضعیف برای ساخت یک طبقه بندی قوی ترکیب می کند. (داده ها به صورت کامل و درست طبقه بندی شدند به جز ۱ مورد)

Table 5. Confusion Matrix for Adaboost algorithm

		Actual Class	
		Normal	Attack
Predicted Class	Normal	54	0
		1	40

در این مطالعه ، مدل طبقه بندی بر روی ۸۰٪ از داده ها ساخته شده و ۲۰٪ باقیمانده برای آزمایش عملکرد طبقه بندی استفاده شده است.

در مقایسه ی کلی این ۴ الگوریتم، با توجه به جدول ۶، درخت تصمیم بالاترین دقت را داشته و الگوریتم Native Bayes و SVM کمترین خطاهارا داشتند.

Table 6. Performance Measures of classifiers

	Accuracy	Precision	Sensitivity	Specificity	F1	Detection rate	False Alarm Rate
SVM	0.9895	1.0000	0.9818	1.0000	0.9908	1.0000	0.02439
Naïve Bayes	0.9789	1.0000	0.9643	1.0000	0.9818	1.0000	0
Decision Tree	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.04878
Adaboost	0.9895	1.0000	0.9818	1.0000	0.9908	1.0000	0.02439

نتیجه گیری مقاله

با توجه به اینکه شبکه های اینترنت اشیاء در برابر حملات بسیار آسیب پذیر هستند، نیاز به یک IDS (Intrusion Detection System) دارند، که می توان این را با الگوریتم های یادگیری ماشین و یک مجموعه ی Data Set معتبر و واقع بینانه ایجاد کرد.

نکات مقاله

- این مقاله صرفا از پروتکل ارتباطی WIFI استفاده و بررسی کرده که یکی از ضعیفترین شبکه های ارتباطی نسبت به حمله شدن به این نوع دیوایس هاست.
- در این مقاله از پروتکل های شبکه دیگه ای برای تست امنیت مثل LAN و یا GSM می توانست بررسی کند.
- از طرفی هم می توانست در مقاله به کارهایی که در زمینه ی امنیت در این حوزه انجام شده اشاره کند.
- به سادگی روش پیاده سازی این مقاله با استفاده از منابعی که ذکر شده می توان اشاره کرد.
- در مقاله به این موضوع اشاره شد که دیتا ها با استفاده از کتابخانه ESP8266 رمزنگاری می شوند، ولی به دیکد کردن این رمز ها در قسمت حمله یا اینکه به چه صورت این الگوریتم رو تشخیص می دهد اشاره ای نکرد.