اولین کنفرانس بین المللی و ششمین کنفرانس ملی
کامپیوتر، فناوری اطلاعات و کاربردهای هوش مصنوعی
۱۳ اسفندماه ۱۴۰۱

بهبود تشخیص حملات بات‌نت در اینترنت اشیا با استفاده از الگوریتم پنگوئن امپراطور و شبکه‌های عصبی

داریوش چینی ساز[1]،کارشناس ارشد شبکه ،جهاد دانشگاهی ،اصفهان ،ایران
مهدی اکبری [2]،دانشکده مهندسی کامپیوتر، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد،ایران

چکیده

موازی با گسترش استفاده از اینترنت اشیاء، تعداد دستگاه‌های اینترنت اشیاء آسیب‌پذیر یا محافظت نشده، همراه با میزان فعالیت‌های مشکوک، مانند بات‌نت اینترنت اشیا و حملات سایبری، به شدت افزایش یافته است. رویکرد پیشنهادی به روش انتخاب ویژگی مبتنی بر الگوریتم فراابتکاری پنگوئن امپراطور به منظور تعیین مرتبط‌ترین ویژگی‌ها متکی است. الگوریتم پنگوئن امپراطور یک روش انتخاب ویژگی بسته‌بندی است که برای تعیین ویژگی‌های مهم و کنار گذاشتن ویژگی‌های نامربوط از خطای طبقه‌بندی و صحت درخت تصمیم استفاده می‌کند. اثربخشی چنین روش‌هایی بسیار وابسته به مجموعه ویژگی‌های ارائه شده به آن‌ها است. در این مدل برای شناسایی حملات از شبکه عصبی یادگیری افراطی استفاده شده است. نتایج تجربی در پایگاه داده N-BaIOT نشان می‌دهد که مدل پیشنهادی به ترتیب در حالت دو کلاسی و چند کلاسی به دقت ۹۹,۹۹٪ و ۹۵,۳۳٪ دست یافته است.

کلمات کلیدی: حملات بات‌نت، اینترنت اشیا، الگوریتم پنگوین امپراطور، شبکه عصبی یادگیری افراطی.

[1] **daroush.chinisaz.sh@gmail.com**
[2] **mehdi_akbari@hotmail.com**

۱- مقدمه

شبکه مجموعه‌ای از دستگاه‌هایی است که از طریق اینترنت به هم متصل می‌شوند و هر روز بر تعداد کل این دستگاه‌ها افزوده می‌شود. بی‌تردید شبکه‌های مؤسسات مالی و تجاری به‌طور مستمر در معرض خطر امنیتی هستند که علاوه بر خسارت مالی کلان ، شهرت آنها را نیز خدشه دار می‌کند. افزایش تعداد کاربران تحت تأثیر نرم افزارهای مخرب ، در حال تبدیل شدن به یک مشکل حیاتی است. بات‌نت‌ها به نگرانی اصلی تبدیل شده‌اند، زیرا یکی از بزرگترین تهدیدها برای سیستم‌های امنیتی محسوب می‌شوند[۱]. بات‌نت یک شبکه پوششی است که توسط میزبان‌های زیادی که توسط ربات‌ها آلوده شده‌اند و توسط یک مهاجم (botmaster) با هدف فعالیت‌های مخرب کنترل می‌شوند، تشکیل شده است [۲, ۳]. ربات اصلی یا botmaster می‌تواند سرور را کنترل کند تا انواع حملات سایبری را آغاز کند، مانند انکار سرویس توزیع شده (DDoS)، هرزنامه، فیشینگ، کلاهبرداری کلیک و سرقت اطلاعات که یکی از جدی ترین تهدیدات امنیتی اینترنت است [۴].

امروزه فناوری اینترنت اشیا (IoT) بدلیل کاربردهای مختلفی که دارد ،توجه بسیاری از جوامع تحقیقاتی و صنعتی را به خود جلب کرده است.اینترنت اشیا را می‌توان به عنوان شبکه‌ای از چیزهای زیادی در نظر گرفت که هر یک دارای یک سیستم محاسباتی (یعنی CPU، حافظه، منبع انرژی) و یک رابط ارتباطی مانند رادیو یا اترنت هستند. این دستگاه‌ها به اینترنت متصل هستند و با آدرس منحصر به فردشان، قابل شناسایی هستند. اینترنت اشیا مجموعه وسیعی از کاربردها دارد و ثابت کرده است که برای جامعه مفید است. با این حال، نگرانی‌های امنیتی مرتبط با اینترنت اشیا وجود دارد. اقدامات امنیتی متعددی وجود دارد که می‌توان برای محافظت از سیستم‌های شبکه‌ها ودستگاه‌های اینترنت اشیا استفاده کرد[۵, ۶].

با این حال، کارشناسان به این نتیجه رسیده‌اند که نمی‌توان به طور کامل از انواع حملات جلوگیری کرد [۷].

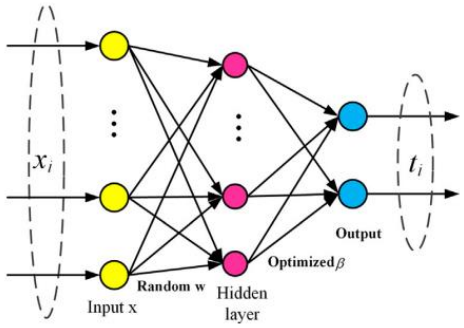
چندین روش برای شناسایی حملات بات‌نت تاکنون ارائه شده است. به منظور پرداختن به این مشکل امنیتی، محققان روش‌های یادگیری ماشینی و عمیق را برای شناسایی حملاتی که دستگاه‌های IoT در معرض خطر را هدف قرار می‌دهند، به کار گرفته‌اند. علی‌رغم این تلاش‌ها، توسعه یک رویکرد تشخیص حمله کارآمد و مؤثر ،برای این منظور، همچنان یک وظیفه چالش برانگیز برای جامعه تحقیقاتی امنیتی است [۸]. مطالعه انجام شده توسط الحربی و همکاران در سال ۲۰۲۱ از جمله مطالعاتی است که در راستای تشخیص حملات بات‌نت و کنترل مشکلات امنیتی دستگاه‌های اینترنت اشیاء انجام شده است [۹]. در روش آن‌ها از الگوریتم بهینه‌سازی خفاش [۱۰] برای تنظیم پارامترهای شبکه عصبی پیشخور و انتخاب ویژگی‌های بهینه استفاده شده است که در بهترین حالت در تشخیص حملات به صحت ۹۰٪ دست یافته است. شبکه‌های عصبی پیشخور فرآیند یادگیری زمان‌بری دارند که با بزرگ شدن سایز داده‌ها این مشکل جدی تر می‌شود. از طرف دیگر فرآیند انتخاب ویژگی توسط الگوریتم‌های فرااکتشافی در شرایطی که الگوریتم بهینه‌سازی دچار بهینه محلی شود و به بهینه سراسری دست نیابد، باعث می‌شود فرآیند انتخاب ویژگی به درستی انجام نشود و تشخیص حملات به خوبی انجام نپذیرد و همچون روش الحربی و همکارانش مدل صحت مناسبی از خود نشان ندهد. از اینرو در این تحقیق با الهام از روش الحربی و همکارانش و در راستای بهبود عملکرد آن‌ها در تشخیص حملات بات‌نت، هدف استفاده از الگوریتم جدید بهینه‌سازی پنگوئن امپراطور[۱۱] جهت انتخاب ویژگی بهینه و نوع خاصی از شبکه‌های عصبی پیشخور، مثل ماشین یادگیری افراطی [۱۲] برای تشخیص حملات بات‌نت است.

۲- مروری بر دو روش مورد استفاده در تحقیق

در این پژوهش روشی برای تشخیص حملات باتنت معرفی شده است که قادر به کاهش ابعاد بهینه داده‌های شبکه است. در ادامه ، ملزومات معرفی روش پیشنهادی شامل مفاهیم بنیادی و شبکه‌های عصبی و انتخاب ویژگی بیان می‌گردد.

2-1-شبکه عصبی ماشین یادگیری افراطی

برخلاف شبکه‌های عصبی پیشخور که فرآیند یادگیری بسیار زمانبری دارند، در سال ۲۰۰۴ یک شبکه عصبی جدید به نام ماشین یادگیری افراطی (ELM) پیشنهاد شد که به‌طور تصادفی وزن‌های ورودی را انتخاب می‌کند و وزن‌های خروجی را به صورت تحلیلی ،تعیین می‌کند. معماری شبکه افراطی در شکل(۱) ارائه شده است.



شکل (۱)ساختار ELM

اساساً آموزش ELM شامل دو مرحله است: (الف) نگاشت فیلتر تصادفی و (ب) حل پارامتر خطی.

در مرحله اول وزن‌های $w = [w_1, w_2, \dots, w_L]$ و بایاس‌های $b = [b_1, b_2, \dots, b_L]$به‌طور تصادفی تولید می‌شوند .

می‌شوند سپس گره‌های لایه پنهان را می‌توان از رابطه (۱) محاسبه کرد:

$$H = \begin{bmatrix} h(x_1) \\ . \\ . \\ . \\ h(x_m) \end{bmatrix} = \begin{bmatrix} h_1(w_1^T x_1 + b_1), & \cdots, & h_L(w_L^T x_1 + b_L) \\ \vdots & \ddots & \vdots \\ h_1(w_1^T x_m + b_1), & \cdots, & h_L(w_L^T x_m + b_L) \end{bmatrix} \quad (۱)$$

در مرحله دوم وزن خروجی β با به حداقل رساندن خطای ماتریس خطای مربع در داده‌های آموزشی محاسبه می‌شود. هدف شبکه ELM رسیدن به حداقل خطای مرحله یادگیری و حداقل وزن‌های خروجی نرمال با استفاده از رابطه (۲) است [۱۳].

$$\min \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\|H\beta - Y\|^2 \quad (۲)$$

در رابطه بالا بخش اول نرم وزن‌ها و بخش دوم خطای یادگیری است. C ضریب جریمه ای است که موازنه بین دو بخش فوق را برقرار می‌کند. در رابطه بالا $Y = [y_1, y_2, \dots, y_m]$ کلاس حقیقی داده‌های آموزشی است. هنگامی که پارامتر تابع خطا در معادله فوق با داده‌های آموزشی محاسبه می‌شود، مقدار خروجی پیش بینی شده نمونه های آزمایشی جدید را می‌توان به صورت رابطه (۳) محاسبه کرد[۱۳]:

$$T = h(x)\beta = \begin{cases} h(x) \times \left(\left(H^T H + \frac{I}{C}\right)^{-1} H^T Y\right) & if\ m \geq L \\ h(x) \times \left(\left(H^T H + \frac{I}{C}\right)^{-1} Y\right) & if\ m < L \end{cases} \quad (۳)$$

در رابطه بالا m تعداد نمونه‌های داده‌های آزمایشی و L تعداد نورون‌های لایه پنهان است. I ماتریس تطبیق با ابعاد L است. $H^T H$ ماتریس معکوس H است، و واضح است که وقتی گرادیان معادله (۲) برابر با صفر است، خطای آموزشی ELM می‌تواند به حداقل برسد [۱۳].

2-2-الگوریتم فرا اکتشافی پنگوین امپراطور

الگوریتم فراابتکاری الهام گرفته از رفتار پنگوئن‌های امپراتور که کلونی پنگوئن‌های امپراتور (EPC) نامیده می‌شود، معرفی می‌شود. این الگوریتم توسط گرمای بدن پنگوئن‌ها و حرکت مارپیچ مانند آنها در کلونی کنترل می شود [۱۱]. پنگوئن امپراطور (Aptenodytes forsteri) بزرگترین گونه پنگوئن است [۱۴]. پنگوئن‌های امپراتور لانه‌ای برای پرورش جوجه‌های خود ندارند. آن ها در طول پرورش جوجه و در اواسط زمستان گروه‌هایی از هزاران پنگوئن به نام هادل [3] می‌سازند. هادل یا توده در برابر سْرما و باد از آن‌ها محافظت می‌کند. در مرکز هادل، گرمای بسیار بیشتری وجود دارد. برای استفاده از گرما توسط همه، پنگوئن‌ها حرکتی مارپیچ مانند به سمت مرکز انجام می دهند شکل (۲) یک حرکت مارپیچ مانند هماهنگ را در هادل نشان می‌دهد.

شکل(۲)حرکت مارپیچی گرمای وسط هادل

برای این الگوریتم قوانینی به شرح زیر وجود دارد [۱۱]:

۱) همه پنگوئن‌ها در جمعیت اولیه دارای تابش گرما هستند و به دلیل ضریب جذب به یکدیگر جذب می‌شوند.

۲) سطح بدن همه پنگوئن ها برابر با یکدیگر در نظر گرفته می شود.

۳) پنگوئن، تابش گرمای کامل را جذب می کند و تأثیر سطح زمین و جو در نظر گرفته نمی‌شود.

۴) تابش گرمایی پنگوئن ها خطی در نظر گرفته می‌شود.

۵) جذب پنگوئن با توجه به میزان حرارت در فاصله بین دو پنگوئن انجام می‌شود. در مسافت طولانی تر، گرمای کمتری دریافت می‌شود و در فاصله کوتاه تر، گرمای بیشتری دریافت می‌شود.

۶) حرکت مارپیچ پنگوئن در طول فرآیند جذب یکنواخت نیست و دارای انحراف با توزیع یکنواخت است.

۳- معرفی روش پیشنهادی

روش‌های طبقه‌بندی برای تشخیص کلاس داده‌ها از دو فاز اصلی یادگیری و ارزیابی تشکیل شده‌اند که هر یک از این دو فاز با توجه به عملیاتی که انجام می‌دهند داده‌های مورد نیاز خود را دریافت کرده و خروجی خاص خود را تولید می‌کنند.

1-3-گام یادگیری

۱) اولین ورودی در گام یادگیری، داده‌های حملات بات نت همراه با کلاس حملات است. روش‌های طبقه‌بندی اعم از شبکه‌های عصبی و روش‌های یادگیری ماشین برای تکمیل مرحله یادگیری داده‌های ورودی را همراه با کلاس هدف (در این مسئله نوع حملات باتنت) دریافت می‌کنند. در مدل‌های طبقه‌بندی داده‌ها به دو بخش آموزش و آزمایش تقسیم می‌شوند. داده‌های بخش آموزش همراه با برچسب کلاس داده‌ها به مدل در مرحله یادگیری تحویل داده می‌شود و داده‌های بخش آزمایش بدون برچسب داده‌ها به فاز ارزیابی ارائه می‌شود.

۲) دومین گام در مدل ،مرحله انتخاب ویژگی است. با توجه به اینکه در این تحقیق انتخاب ویژگی مبتنی بر الگوریتم پنگوئن امپراطور است در نتیجه تعداد ویژگی، تعداد جمعیت اولیه و تعداد گردش الگوریتم فرااکتشافی از جمله پارامترهای مدل در این مرحله است.

۳) در مرحله یادگیری ،پارامترهای شبکه عصبی نیز دریافت می‌شود که شامل تعداد نرون‌های لایه پنهان در شبکه افراطی و نوع تابع هسته شبکه است.

۴) خروجی مرحله یادگیری، مدل آموزش دیده شبکه عصبی ماشین یادگیری افراطی است. این مدل برای فاز ارزیابی استفاده می‌شود.

[3] Emperor Penguins Colony (EPC)

[4] huddles

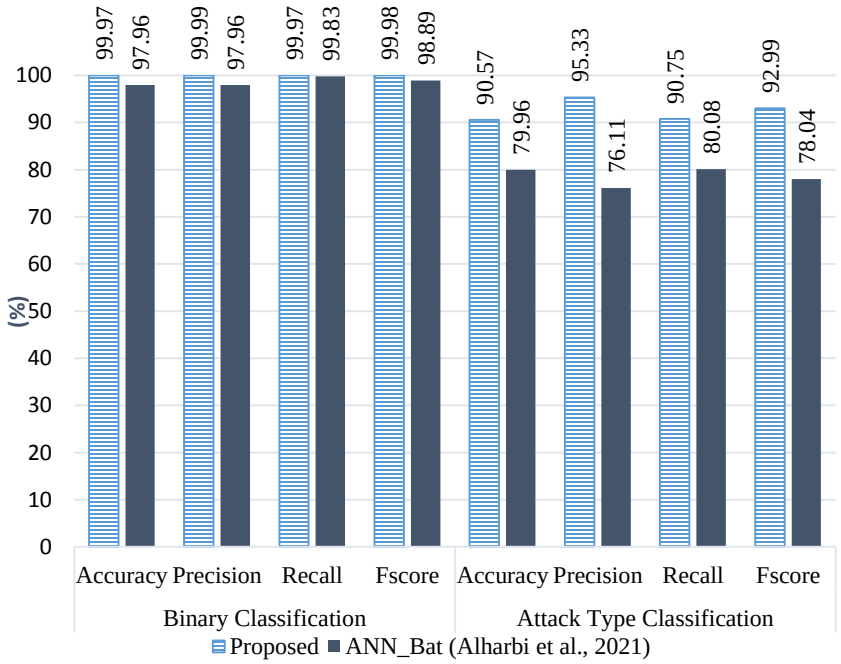
2-3-گام ارزیابی مدل

۱) اولین ورودی مدل آموزش دیده در گام قبل است.
۲) دومین ورودی در این گام داده‌های بخش آزمایش است. تا نوع حملات را برای هر داده جدید پیش‌بینی کند.
۳) خروجی فاز ارزیابی دو ماتریس درهم ریختگی (برای طبقه‌بندی دو کلاسی و چند کلاسی) است که برای محاسبه معیارهایی مثل دقت، فراخوانی، صحت، تشخیص که از جمله متغیرهای وابسته این تحقیق هستند، استفاده می‌شود.

3-3-طبقه‌بندی و ارزیابی با ماشین افراطی

پس از انتخاب ویژگی و رسیدن به زیر مجموعه مناسب از ویژگی‌ها که حصول دقت و صحت بهتر مدل را ممکن می‌کنند. در این بخش تشخیص نوع حملات با استفاده از شبکه عصبی یادگیری افراطی انجام می‌شود. داده‌های بخش آموزش همراه با برچسب حملات به شبکه عصبی یادگیری افراطی ارائه می‌شود و شبکه پس از تکمیل فرآیند یادگیری، داده‌های آزمایش را بدون برچسب کلاس آن‌ها دریافت می‌کند تا کلاس حملات داده‌های آزمایش را پیش‌بینی کند. در مدل این تحقیق تعداد نرون برای لایه پنهان در شبکه یادگیری افراطی برای تشخیص حملات به صورت دو کلاسی برابر با ۶۰۰ و برای تشخیص حملات به صورت چند کلاسی برابر با ۱۸۰۰ تعریف شد. تابع فعال‌سازی برای نرون‌ها نیز تابع 'ReLU' تعریف گردید. پس ازپیش‌بینی کلاس حملات توسط شبکه عصبی،کلاس حملات پیش‌بینی شده باکلاس حملات درداده های واقعی مقایسه می‌شود. از طریق این مقایسه ماتریس درهم ریختگی تشکیل می‌شود و با کمک آن صحت، دقت و فراخوانی مدل حاصل می‌شود.

۴- مقایسه نتایج روش پیشنهادی و روش پایه



نمودار (۱) مقایسه نتایج روش‌ها

نتایج نمودار (۱) نشان می‌دهد که هر دو روش در طبقه‌بندی باینری یا دو کلاسی به نتایج تقریبا نزدیکی رسیده‌اند. اما همچنان مدل در طبقه‌بندی دو کلاسی نتایج بهتری ارائه داده است. برای تشخیص حملات (طبقه‌بندی باینری) مدل به صحت ٪۹۹,۹۷، دقت ٪۹۹,۹۹ و فراخوانی ٪۹۹,۹۷ و معیار F ٪۹۹,۹۸ رسیده است. که در مقایسه با مدل الحربی و همکاران موفق‌تر است. اما بیشترین موفقیت مدل در طبقه‌بندی چند کلاسی (تشخیص نوع حملات) است. در این حالت مدل توانسته است به دقت ٪۹۵,۳۳ دست یابد و این معیار را ٪۱۹,۲۲ افزایش دهد. همچنین مدل ما توانسته به صحت ٪۹۰,۵۷ در تشخیص نوع حملات دست یابد که این متغیر را در مقایسه با روش الحربی ٪۱۰,۶۱ افزایش داده است. علت حصول این نتایج تفاوت در مرحله انتخاب ویژگی و طبقه‌بندی است.

۵- نتیجه گیری و پیشنهادات

اگر از ترکیب پنگوئن امپراطور و ماشین یادگیری افراطی استفاده شود، آنگاه مدل در تشخیص دو کلاسی و چند کلاسی موفق عمل می‌نماید. نتایج مقایسه روش‌ها ، نشان داد که ترکیب دو روش فوق باعث شده نتایج مدل در طبقه‌بندی دو کلاسی و چند کلاسی موفق‌تر باشد؛ اما مدل در طبقه‌بندی چند کلاسی به مراتب به موفقیت بیشتری رسیده است. در طبقه‌بندی دو کلاسی دقت و صحت در حدود ٪۲ نسبت به روش الحربی و همکارانش افزایش یافته است اما در طبقه‌بندی چند کلاسی به ترتیب در حدود ۱۹ و ٪۱۰ افزایش داشته است.

در ادامه این تحقیق از ترکیب دو روش فرااکتشافی برای انتخاب ویژگی می‌توان استفاده نمود. انتظار می‌رود استفاده از دو روش فرااکتشافی برای رسیدن به مجموعه ویژگی‌های مناسب بتواند در طبقه‌بندی چند کلاسی (تشخیص نوع حملات) صحت و دقت مدل را بهبود ببخشد. همچنین استفاده از مدل‌های دیگر شبکه‌های عصبی عمیق مثل شبکه عصبی عمیق چند لایه پرسپترون از دیگر کارهایی است که می‌تواند باعث بهبود نتایج شود.

۶- مراجع

[1] K. Shinan, K. Alsubhi, A. Alzahrani, and M. U. Ashraf, "Machine learning-based botnet detection in software-defined network: a systematic review," *Symmetry,* vol. 13, p. 866, 2021.

[2] B. Fang, X. Cui, and W. Wang, "Survey of botnets," *Journal of Computer Research and Development,* vol. 48, pp. 1315-1331, 2011.

[3] G. Vormayr, T. Zseby, and J. Fabini, "Botnet communication patterns," *IEEE Communications Surveys & Tutorials,* vol. 19, pp. 2768-2796, 2017.

[4] A. Karim, R. B. Salleh, M. Shiraz, S. A. A. Shah, I. Awan, and N. B. Anuar, "Botnet detection techniques: review, future trends, and issues," *Journal of Zhejiang University SCIENCE C,* vol. 15, pp. 943-983, 2014.

[5] M. A. Razzaq, S. H. Gill, M. A. Qureshi, and S. Ullah, "Security issues in the Internet of Things (IoT): A comprehensive study," *International Journal of Advanced Computer Science and Applications,* vol. 8, p. 383, 2017.

[6] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet of Things Journal,* vol. 4, pp. 1250-1258, 2017.

[7] H. Sedjelmaci, M. Hadji, and N. Ansari, "Cyber security game for intelligent transportation systems," *IEEE Network,* vol. 33, pp. 216-222, 2019.

[8] M. Alqahtani, H. Mathkour, and M. M. Ben Ismail, "IoT botnet attack detection based on optimized extreme gradient boosting and feature selection," *Sensors,* vol. 20, p. 6336, 2020.

[9] A. Alharbi, W. Alosaimi, H. Alyami, H. T. Rauf, and R. Damaševičius, "Botnet attack detection using local global best bat algorithm for industrial internet of things," *Electronics,* vol. 10, pp. 1-24, 2021.

[10] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature inspired*

[11] S. Harifi, M. Khalilian, J. Mohammadzadeh, and S. Ebrahimnejad, "Emperor Penguins Colony: a new metaheuristic algorithm for optimization," *Evolutionary Intelligence,* vol. 12, pp. 211-226, 2019.

[12] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing,* vol. 70, pp. 489-501, 2006.

[13] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mechanical systems and signal processing,* vol. 133, p. 106272, 2019.

[14] G. L. Kooyman, C. Drabek, R. Elsner, and W. Campbell, "Diving behavior of the emperor penguin, Aptenodytes forsteri," *The Auk,* pp. 775-795, 1971.