روشی برای تشخیص بات‌نت‌ها در مرحله فرمان و کنترل با استفاده از خوشه‌بندی برخط

موسی یحیی‌زاده و مهدی آبادی
آزمایشگاه تشخیص/پیش‌گیری از نفوذ
دانشگاه تربیت مدرس
{m.yahyazaeh,abadi}@modares.ac.ir

رئوس مطالب

- مقدمه
- چرخه‌حیات بات‌نت‌ها
- فرمان وکنترل
- بیان مسئله و فرضیات
- روش OBD و مراحل آن
- ارزیابی
- مقایسات
- نتیجه‌گیری

مقدمه

- گسترش حملات اینترنتی با انگیزه‌های مختلفی از قبیل سرگرمی، خراب‌کاری، کسب شهرت و درآمد
- بدافزارهای اینترنتی مهم‌ترین عوامل حملات در فضای اینترنت
  - تکامل به سمت سازماندهی بهتر و سود-محوری بیشتر
- در مرکز این حملات، یک گروه از میزبان‌هایی قرار دارند که به تصرف مهاجم درآمده و توسط وی از راه دور هدایت می‌شوند.
  - بات‌نت‌ها

مقدمه

- بات
  - برگرفته شده از کلمه روبات
  - بدافزاری پیشرفته که برای انجام کارهایی که فرامین آن از راه‌دور صادر می‌شوند طراحی شده
  - به سیستم آلوده به این بدافزار نیز بات گفته می‌شود
- بات‌نت
  - شبکه‌ای از بات‌ها طبق توپولوژی موردنظر مدیر بات
  - یک گروه از میزبان‌های آسیب‌پذیر کنترل شده از راه‌دور
  - استفاده از آن‌ها برای انجام حملات

چرخه‌حیات بات‌نت‌ها

- بات‌نت‌ها نسبت به سایر بدافزارها چرخه حیات شفاف‌تری دارند
- چرخه حیات بات‌نت‌ها می‌تواند به سه مرحله تقسیم‌بندی شود

- در هر مرحله نوع فعالیت و ترافیک بات‌نت‌ها مختلف است.

فرمان و کنترل

- مهمترین ویژگی بات‌نت‌ها نسبت به سایر بدافزارها است.
- مدیر بات را قادر می‌سازد تا بات‌نت را با فرامین خود هدایت نماید.
- مدیر بات سعی می‌کند تا ساختار مناسبی را برای آن انتخاب کند.
- **تقسیم‌بندی بات‌نت بر اساس ساختار فرمان و کنترل (C&C)**
  - متمرکز (بات‌نت‌های مبتنی بر IRC و HTTP)
  - غیرمتمرکز (بات‌نت‌های P2P)
  - ترکیبی (بات‌نت HTTP2P)

بیان مسئله و فرضیات

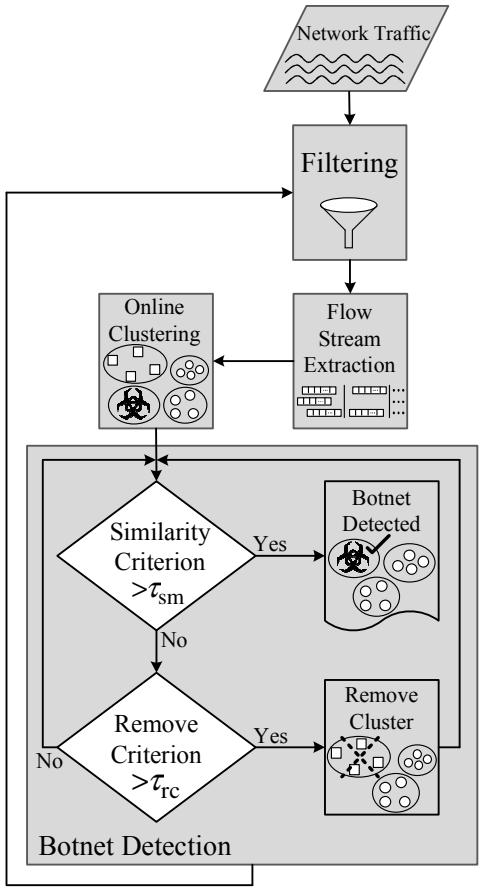
تعریف ۱ – باتنت. هر باتنت یک گروه هماهنگ از باتهایی است که از طریق کانال‌های فرمان و کنترل هدایت شده و فعالیت‌های بدخواهانه‌ای را انجام می‌دهند

- گروه هماهنگ از باتها (حداقل دو بات)
- کنترل از طریق کانال‌ها فرمان و کنترل به صورت یکپارچه

- **روش OBD**
  - ✓ تشخیص باتهای عضو یک باتنت در شبکه تحت نظارت
  - ✓ در مرحله فرمان و کنترل
  - ✓ به صورت برخط

روش OBD

- مراحل روش OBD
  - فیلترسازی ترافیک
  - استخراج جریان‌های دنباله‌ای
  - خوشه‌بندی برخط
  - تشخیص باتنت



مرحله فیلترسازی ترافیک

- تکنیک فهرست سفیدسازی
  - ساخت یک فهرست از میزبان های مورد اعتماد
    - مثل Yahoo و Google
  - حذف بسته‌ها به/از سمت میزبان‌های مورد اعتماد و ثبت سایر بسته‌ها
- مزیت
  - کاهش میزان محاسبات و فضای مورد نیاز برای ذخیره‌سازی ترافیک
  - کاهش نرخ هشدار نادرست

مرحله استخراج جریان‌های دنباله‌ای

- روش OBD بر پایه تحلیل ترافیک شبکه به صورت جریان
  - جریان، به مجموعه‌ای از بسته‌ها با آدرس IP مبدا، درگاه مبدا، آدرس IP مقصد، درگاه مقصد و پروتکل یکسان گفته می‌شود
- روش‌های تشخیص غیربرخط جریان‌ها را پس از اتمام آن‌ها تجمیع و سپس تحلیل می‌کنند.
- در نظر گرفتن جریان ترافیک به صورت دنباله‌ای در روش OBD

مرحله استخراج جریان‌های دنباله‌ای

تعریف ۲ - جریان دنباله‌ای. جریان دنباله‌ای $\vec{F}_i$ یک توالی از بردارهای ویژگی استخراج شده برای هر جریان در دوره‌های زمانی متفاوت با طول یکسان می‌باشد

$$\vec{F}_i = \langle f_i(t_1), f_i(t_2), f_i(t_3),..., f_i(t_j),... \rangle$$

که $f_i(t_j)$ بردار ویژگی استخراج شده برای جریان i در پایان دوره زمانی t_j است و بردار جریان i در این دوره زمانی نامیده می‌شود

$$f_i(t_j) = (x_1, x_2, x_3, ..., x_p)$$

$$x_k = f_i^k(t_j)$$

مرحله استخراج جریان‌های دنباله‌ای

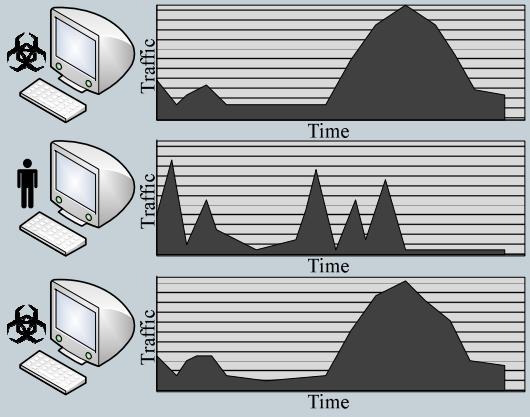
تعریف ۳ – مجموعه جریان. مجموعه همه جریان‌های دنباله‌ای موجود در ترافیک شبکه مجموعه جریان نامتناهی نامیده شده و با S نمایش داده می‌شود

$$S = \{..., \vec{F}_{i-1}, \vec{F}_i, \vec{F}_{i+1}, ...\}$$

مجموعه همه بردارهای جریان در دوره زمانی t با $S(t) \subseteq S$ نمایش داده می‌شود

$$S(t) = \{f_i(t), f_{i+1}(t), f_{i+2}(t), ..., f_{i+n}(t)\}$$

مرحله خوشه‌بندی برخط



- الگوریتم خوشه‌بندی با شعاع ثابت برخط
  - مشابه خوشه‌بندی با شعاع ثابت
  - خوشه‌هایی با شعاع ثابت σ می‌سازد
  - تفاوت
    - داده‌های متغیر به‌طور پیوسته در طول زمان می‌رسند

مرحله خوشه‌بندی برخط

تعریف ۴ – فاصله جریان. فاصله دو بردار جریان $f_i(t)$ و $f_j(t)$ در دوره زمانی t با محاسبه اختلاف بین ویژگی‌های متناظر آن‌ها حاصل می‌شود

$$d(f_i(t), f_j(t)) = \frac{\sum_{k=1}^{p} \delta_{i,j}^{k} \, \Delta(f_i^k(t), f_j^k(t))}{\sum_{k=1}^{p} \delta_{i,j}^{k}}$$

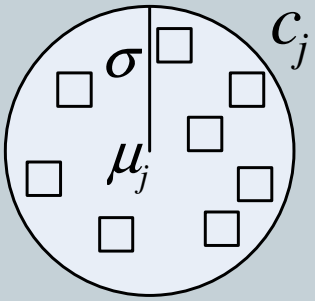
$$\delta_{i,j}^{k} = \begin{cases} 0 & \text{if } f_i^k(t) \text{ or } f_j^k(t) \text{ is missed} \\ 1 & \text{otherwise} \end{cases}$$

$$\Delta(f_i^k(t), f_j^k(t)) = \begin{cases} 0 & \text{if } f_i^k(t) = f_j^k(t) \\ 1 & \text{otherwise} \end{cases}$$

$$\Delta(f_i^k(t), f_j^k(t)) = \frac{| f_i^k(t) - f_j^k(t) |}{f_{\max}^k(t-1) - f_{\min}^k(t-1)}$$

مرحله خوشه‌بندی برخط

تعریف ۵ – خوشه. هر خوشه شامل مجموعه‌ای از بردارهای جریان است که فاصله آن‌ها تا مرکز خوشه از یک شعاع ثابت σ کمتر است. هر خوشه c_j با دوتایی (μ_j, β_j) نمایش داده می‌شود که μ_j مرکز و β_j دوره زمانی تولد خوشه است.



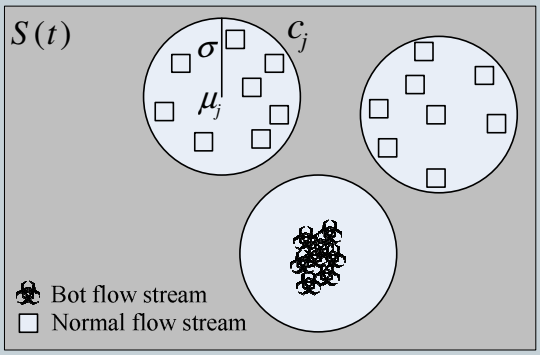
مرحله خوشه‌بندی برخط

input:
$S(t)$: Set of flow vectors
σ : Cluster radius
$C(t-1)$: Set of clusters
output:
$C(t)$: Set of clusters
begin
for $k = 1$ to p **do**
Calculate $f^k_{\max}(t-1)$ and $f^k_{\min}(t-1)$
end for
for each flow vector $f_i(t) \in S(t)$ **do**
if $f_i(t-1) \in c_j$ for some $c_j \in C(t-1)$ **then**
if $d(f_i(t), \mu_j) < \sigma$ **then**
$c_j := (c_j - \{f_i(t-1)\}) \cup f_i(t)$
Update centroid μ_j
else
$c_j := c_j - \{f_i(t-1)\}$
Update centroid μ_j
end if
end if

مرحله خوشه‌بندی برخط

if $f_i(t) \notin c_j$ for all $c_j \in C(t-1)$ **then**
  Find the nearest cluster $c_l \in C(t-1)$ to $f_i(t)$
  if $d(f_i(t), \mu_l) < \sigma$ **then**
    $c_l := c_l \cup \{f_i(t)\}$
  else
    Make a new cluster c_k with centroid μ_k
    $C(t-1) := C(t-1) \cup \{c_k\}$
    $\mu_k := \{f_i(t)\}$
    $\beta_k := t$
  end if
 end if
end for
$C(t) := C(t-1)$
end procedure

مرحله تشخیص بات‌نت



تعریف ۶ – معیار شباهت درون خوشه‌ای

$$sm(c_j) = e^{-\frac{\bar{d}_j}{1+o_j}}$$

$$\bar{d}_j = \frac{1}{m}\sum_{i=1}^{m} d(f_i(t), \mu_j) \qquad o_j = t - \beta_j$$

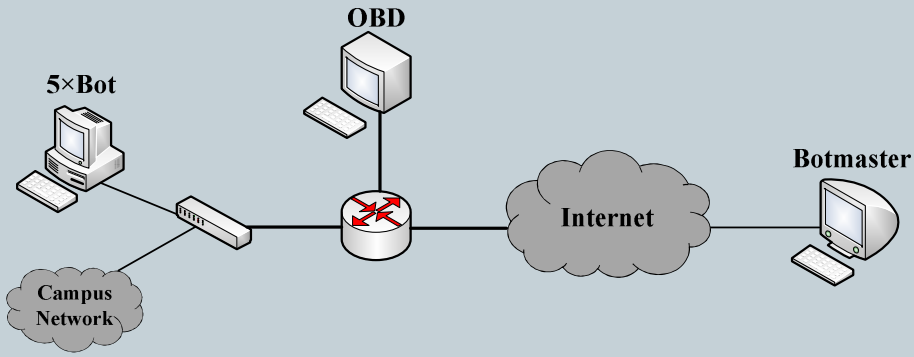
تعریف ۷ – معیار حذف خوشه‌ها

$$rc(c_j) = \bar{d}_j \cdot o_j$$

مرحله تشخیص بات‌نت

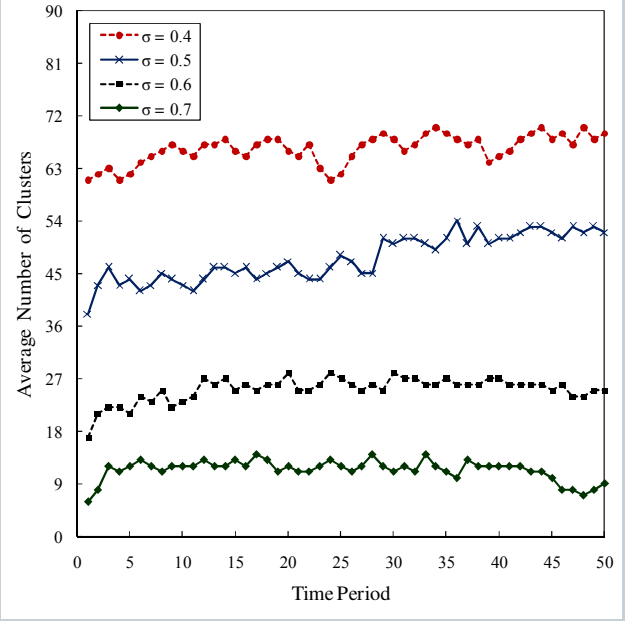
input:
σ : cluster radius
τ_{sm} : similarity threshold
τ_{rc} : remove threshold
begin
for each time period t **do**
Extract a set $S(t)$ of flow vectors from S
$C(t)=ofwc(S(t),C(t-1),\sigma)$
for each cluster $c_i \in C(t)$ **do**
Calculate similarity criterion $sm(c_i)$ using equation (9)
if ($|c_i|>2$ **and** $sm(c_i)>\tau_{sm}$) **then**
Mark flow vectors in cluster c_i as suspicious
Alert "*Bot Detected!*"
end if
Calculate remove criterion $rc(c_i)$ using equation (12)
if $rc(c_i)>\tau_{rc}$ **then**
$C(t):=C(t)-\{c_i\}$
end if
end for
end for
end procedure

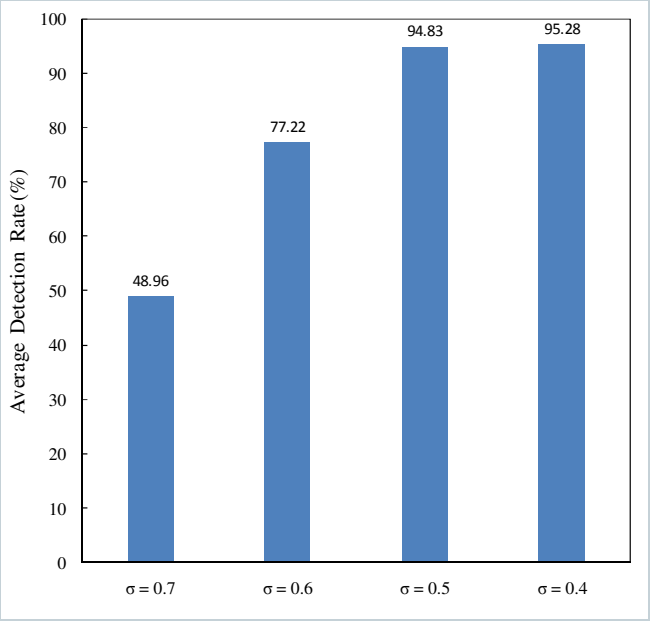
ارزیابی



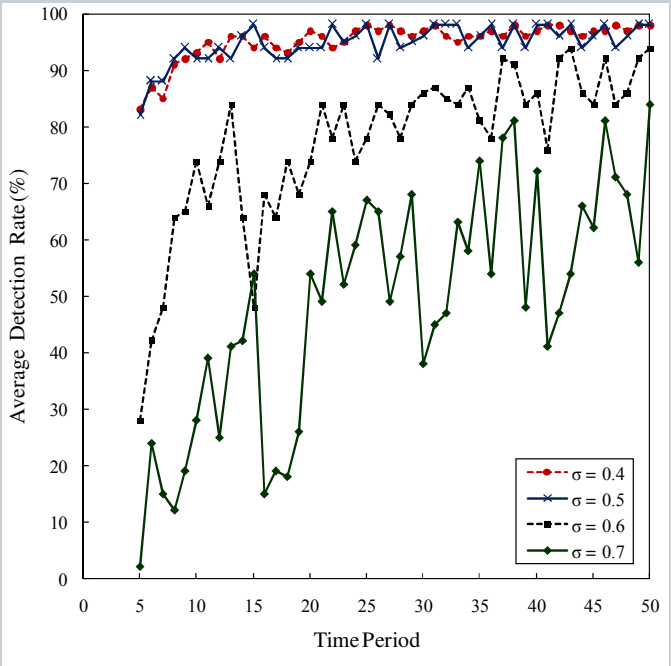
Time period (t)	5(s)
Fixed width cluster radius (σ)	0.5
Similarity threshold (τ_{sm})	0.95
Remove threshold (τ_{rc})	10

ارزیابی

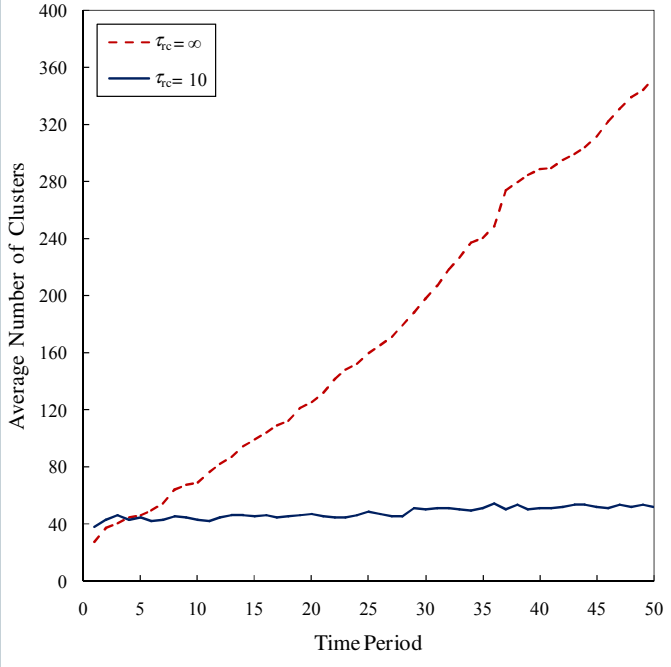




ارزیابی



ارزیابی



مقایسات

روش تشخیص باتنت‌ها	تشخیص باتنت‌های ناشناخته	تشخیص باتنت‌ها با C&C رمزشده	تشخیص برخط باتنت‌ها	نرخ هشدار نادرست پایین
BotGAD [4]	✓	✓	✓	✓
BotMiner [5]	–	✓	–	✓
DataAdaptive [6]	✓	–	✓	✓
Rishi [10]	–	–	✓	–
BotProbe [11]	–	–	✓	✓
BotSniffer [12]	✓	✓	–	✓
OBD	✓	✓	✓	✓

نتیجه‌گیری

- روشی OBD برای تشخیص برخط باتنت‌ها در مرحله فرمان و کنترل پیشنهاد شد
  - با در نظر گرفتن رفتار هماهنگ و گروهی بات‌ها
  - تشخیص به صورت برخط
  - قادر به تشخیص بات‌های ناشناخته
  - قادر به تشخیص بات‌ها با کانال C&C رمز شده