Bundesamt
für Sicherheit in der
Informationstechnik

Lernen neuer Malware-URL-Muster in einem Hostblocking-System

Thomas Hungenberg, Dr. Timo Steffens

Referat 121 CERT-Bund, BSI

DFN-Workshop
09.02.2010

Agenda

Malware-Trends

Hostblocking

Lernen neuer Muster

Experimentelle Evaluation

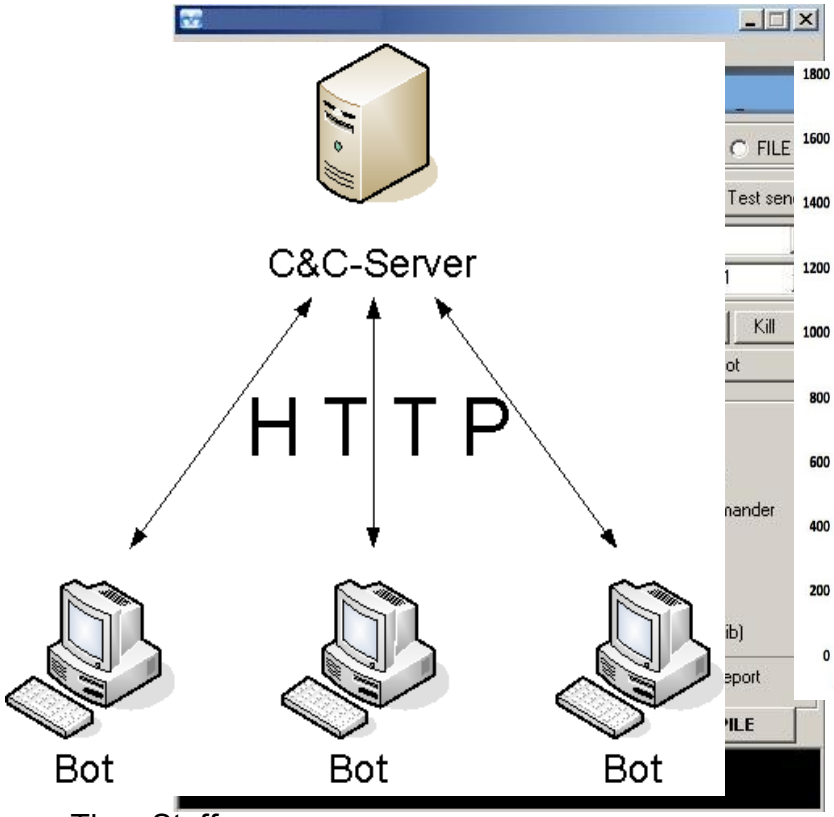
Zusammenfassung

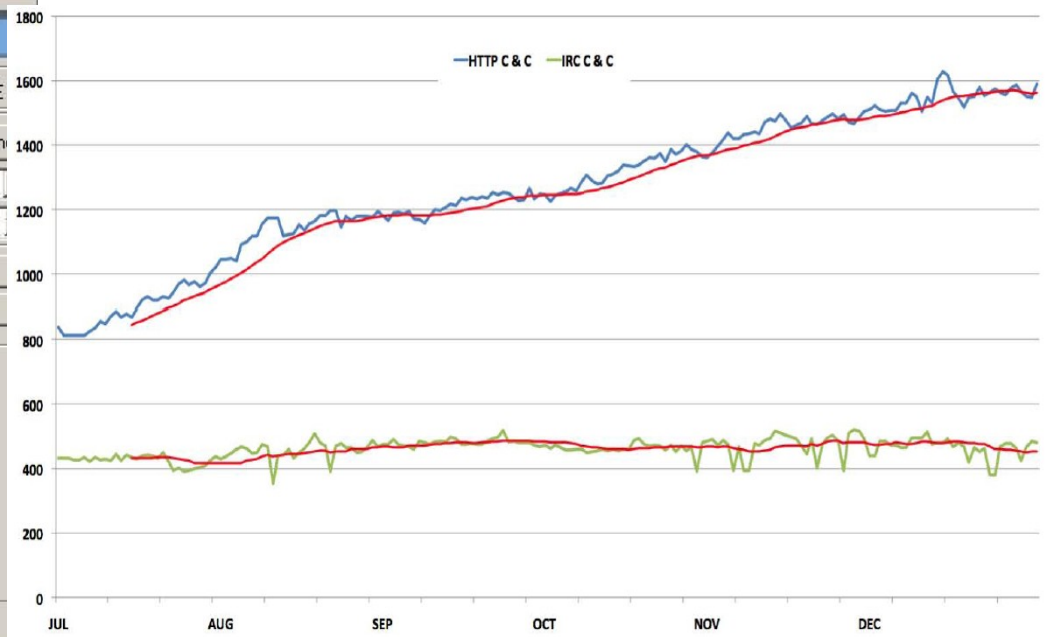
Malware-Trends

Malware wird zunehmend über Drive-By-Exploits verbreitet

Professionalisierung durch „Malware-Kits“

C&C-Server kommunizieren zunehmend über HTTP





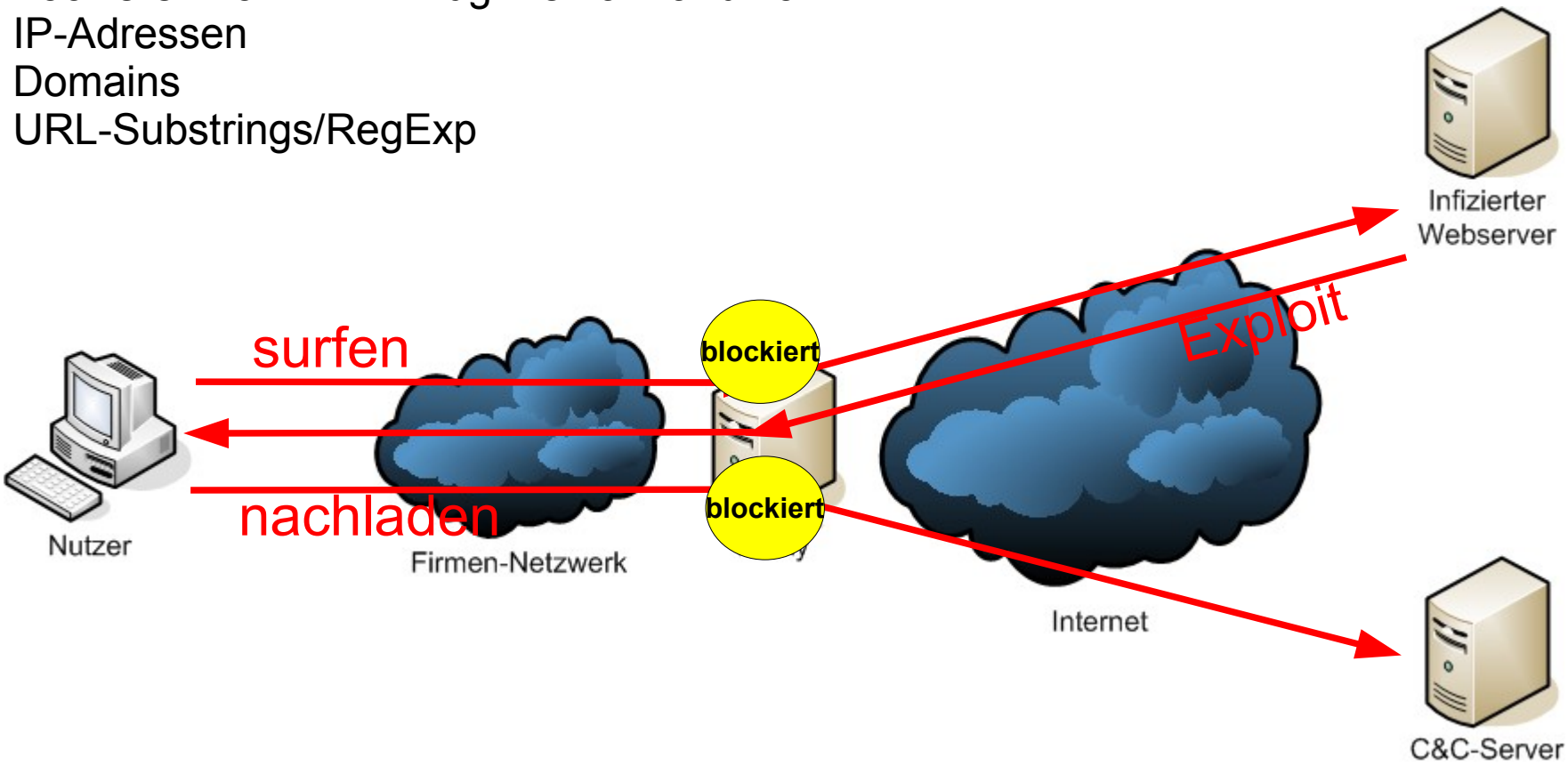
Quelle: Team Cymru

Malware-Trends

- Drive-By-Exploits häufig massenhaft verbreitet
- Binaries werden täglich neu erstellt
- Kommunikationswege (URLs) sind stabiler
  - Gumblar:
    http://xxx.com:8080/**landig.php?id=**8
- Kommunikation von Bots mit dem Command-and-Control-Server zunehmend per HTTP
- Häufig konstante Substrings in der URL
  - Koobface:
    http://xxx/**ld/gen.php**
  - Bredolab:
    http://xxx/yyy/**controller.php?action=bot&entity_list=**&**uid=**1&**first=**1&**guid=**336535724&**rnd=**981633

Hostblocking

Blockieren von HTTP-Zugriffen anhand von
- IP-Adressen
- Domains
- URL-Substrings/RegExp



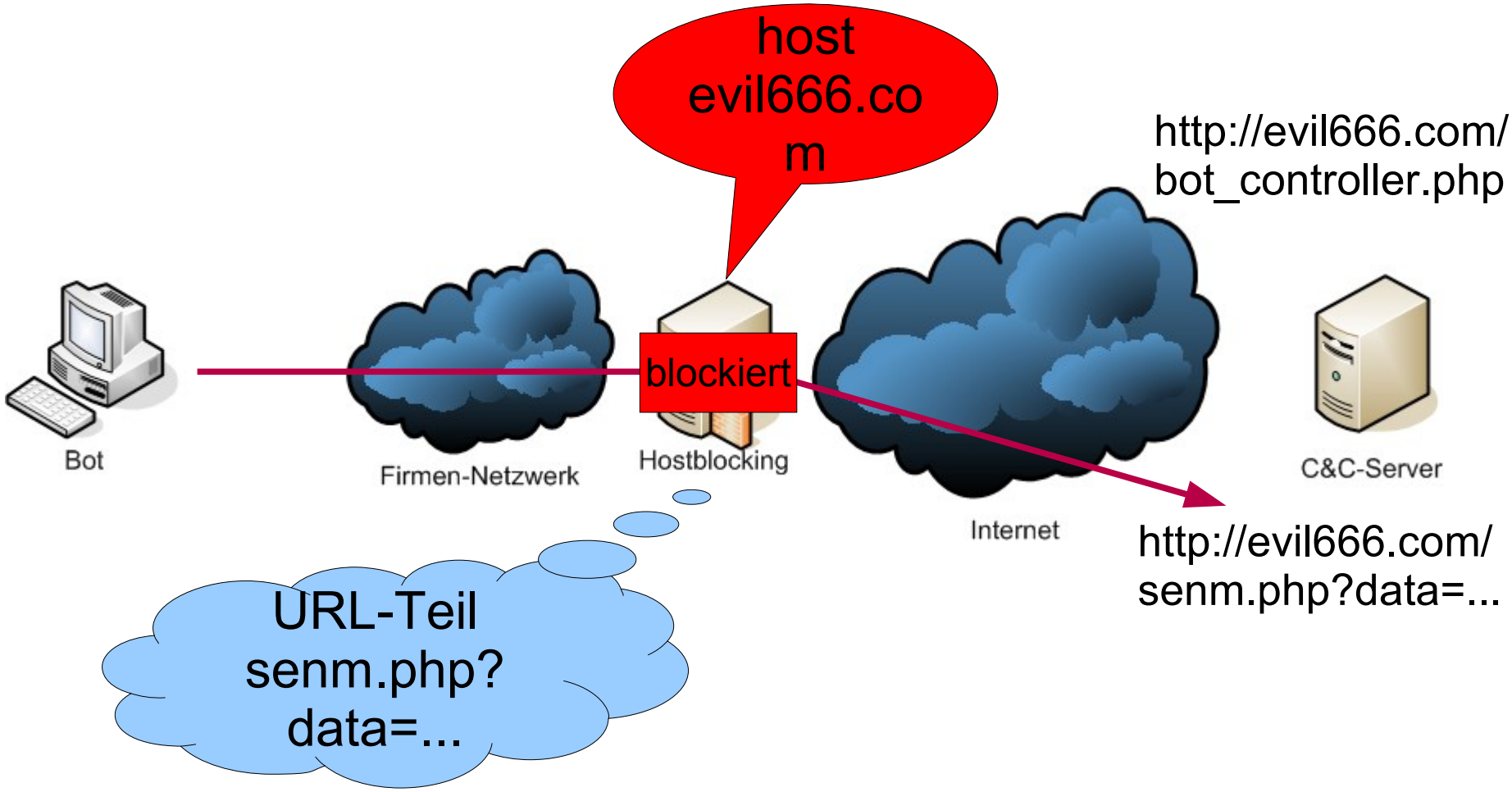
Auswerte-Modul

- Schädliche Seiten blockieren ist bereits ein Vorteil
- Protokollierte Zugriffe erlauben zudem Identifikation infizierter Rechner

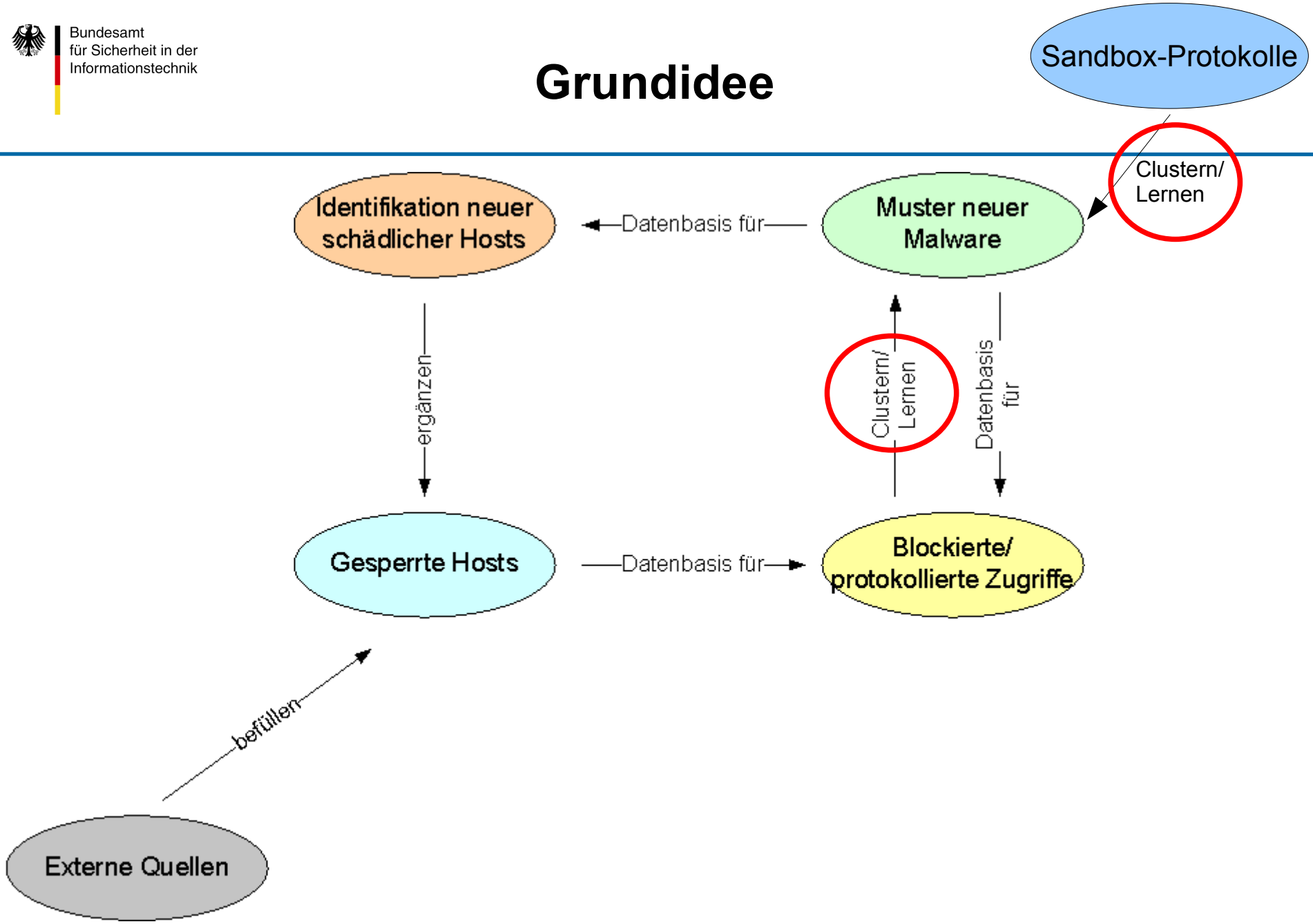
Zugriffe auf BLOCKIERTE Hosts MIT EVTL. MELDEBEDARF					
URL	**Aufrufende IP**	**Zeit**	**Methode**	**Response**	**Transfer**
213.175.193.142	Meldebedarf: individuell entscheiden			Kritikalität: hoch	**top**
http://213.175.193.142/click.php?	[redacted]	07:28:43	GET	TCP_DENIED/403	1464
http://213.175.193.142/click.php?	[redacted]	07:45:11	GET	TCP_DENIED/403	1464
91.207.8.242	Meldebedarf: dringend melden			Kritikalität: sehr hoch	**top**
http://91.207.8.242/spm/page.php?	[redacted]	07:01:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:06:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:11:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:16:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:21:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:26:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:31:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:36:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:41:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:46:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:51:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	07:56:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	08:01:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	08:06:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	08:11:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	08:16:43	GET	TCP_DENIED/403	1464
http://91.207.8.242/spm/page.php?	[redacted]	08:21:43	GET	TCP_DENIED/403	1464

Domain-Mehrfachnutzung

Malware-Server werden häufig wiederverwendet



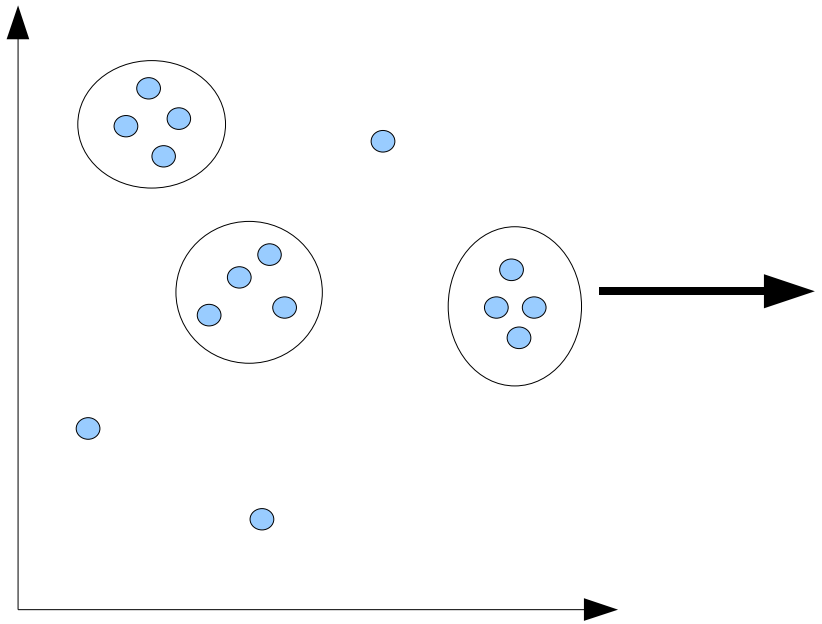
Grundidee


Sandbox-Protokolle
Clustern/
Lernen
Identifikation neuer
schädlicher Hosts
Datenbasis für
Muster neuer
Malware
ergänzen
Clustern/
Lernen
Datenbasis
für
Gesperrte Hosts
Datenbasis für
Blockierte/
protokollierte Zugriffe
befüllen
Externe Quellen

Ansatz

Substring-/RegExp-Erkennung auf allen URLs ist zu teuer

Pipeline aus Clustering und Substring-Erkennung



hXXp://evil666.com/**bot_controller.php?mac=**08-00-20-AE-FD-7E**&id=**123456

hXXp://www.7766.org/scripts/**bot_controller.php?mac=**23-A4-AE-10-00**&id=**27494

hXXp://kasberski.com/update/**bot_controller.php?mac=**44-A4-FE-08-23**&id=**39921

hXXp://a23.tonji.cn/**bot_controller.php?mac=**08-72-E5-66-53**&id=**40941

/bot_controller.php?mac=...&id=...

Ähnlichkeitsmaß für Clustering

Basiert auf Levenshtein / Edit-Distance

$s(A$

Nachteil:
Ähnlichkeit hängt nicht nur
von gemeinsamen Teilen ab,
sondern von gesamter URL.

levenshtein(„aab“, „ab“) = 1, s(„aab“, „ab“)=0.66

levenshtein(„aaab“, „ab“) = 2, s(„aaab“, „ab“)=0.5

Parameterwert-Maskierung

```
A = „http://host.com/bot_controller.php?mac=08-00-20-AE-FD-7E&id=123456“
B = „http://a.com/scripts/bot_controller.php?mac=12-34-56-78-9A-BC&id=899988“
```

s(A,B) = 0.56

```
A' = „http://host.com/bot_controller.php?mac=<MAC>&id=<ID>“
B' = „http://a.com/scripts/bot_controller.php?mac=<MAC>&id=<ID>“
```

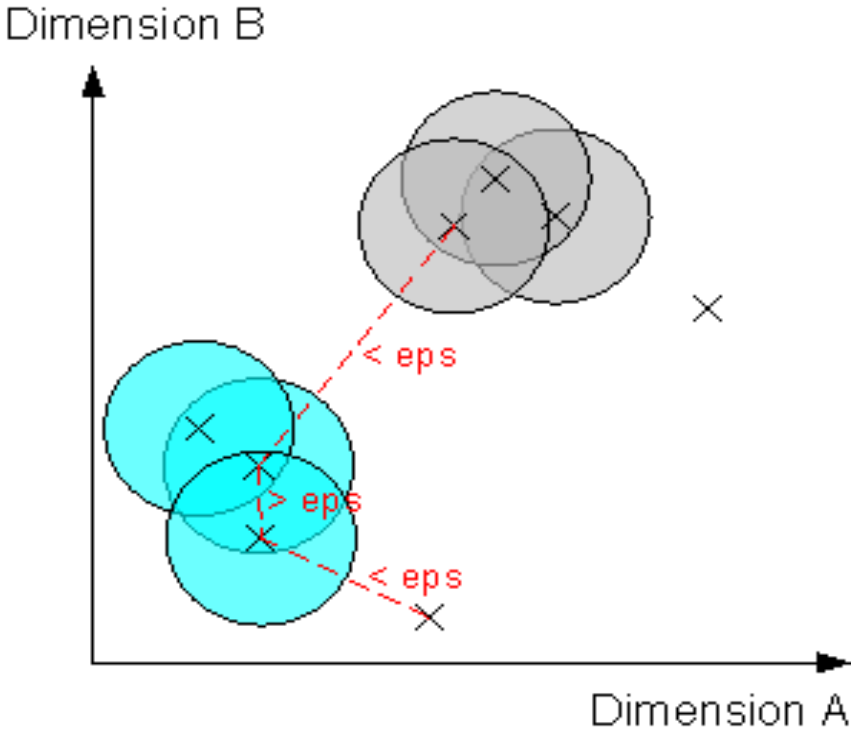
s(A,B) = 0.82

Clustering

DBSCAN

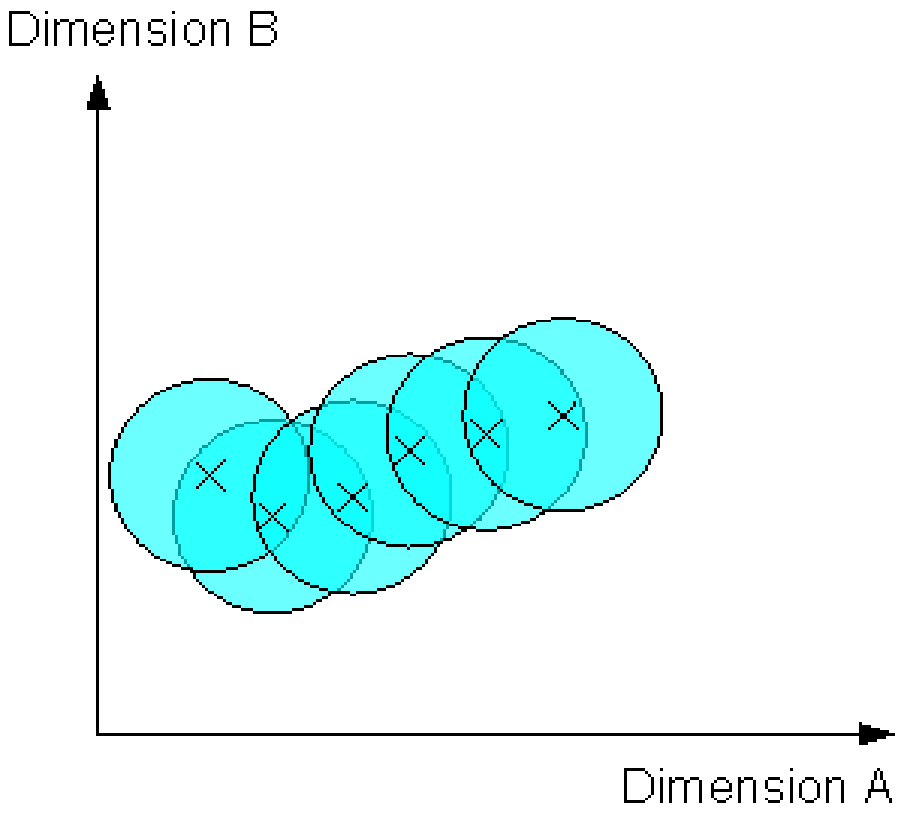
Iteriert einmal durch die Menge der Elemente

- Prüft, ob die Ähnlichkeit zu anderen Elementen > eps
- Fügt Element ggf. einem Cluster hinzu



Schwächen des Verfahrens

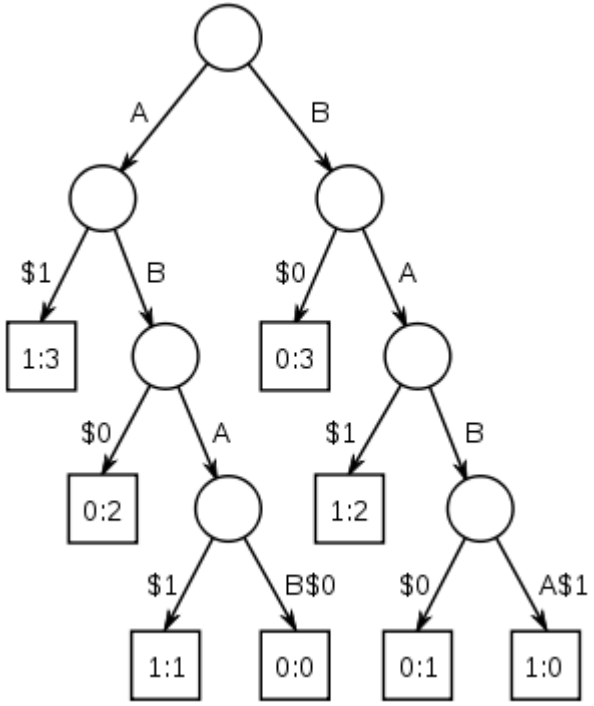
Ähnlichkeit ist nicht transitiv → Entstehung transitiver Ketten
(dem wird bei der Substring-Generierung gegengesteuert)



Substring-Erkennung

Für jedes Cluster wird ein Generalized Suffix Tree erstellt
Im Baum ist jeder einzelne Substring enthalten
Ablesbar, wie viele URLs den Substring enthalten

Um unreinen Clusters entgegenzusteuern, wird ein Substring als „gemeinsam“ für ein Cluster bewertet, wenn er in 66% der Elemente enthalten ist.



Evaluation

Das Verfahren ist im Einsatz und liefert regelmäßig neue Muster für das Hostblocking

Zusätzlich kontrollierte Evaluation:

Datengrundlage: 2437 protokollierte URLs von 7 aufeinanderfolgenden Tagen

Zusätzlich 50 bekannte Muster (349 URLs)

Ergebnisse

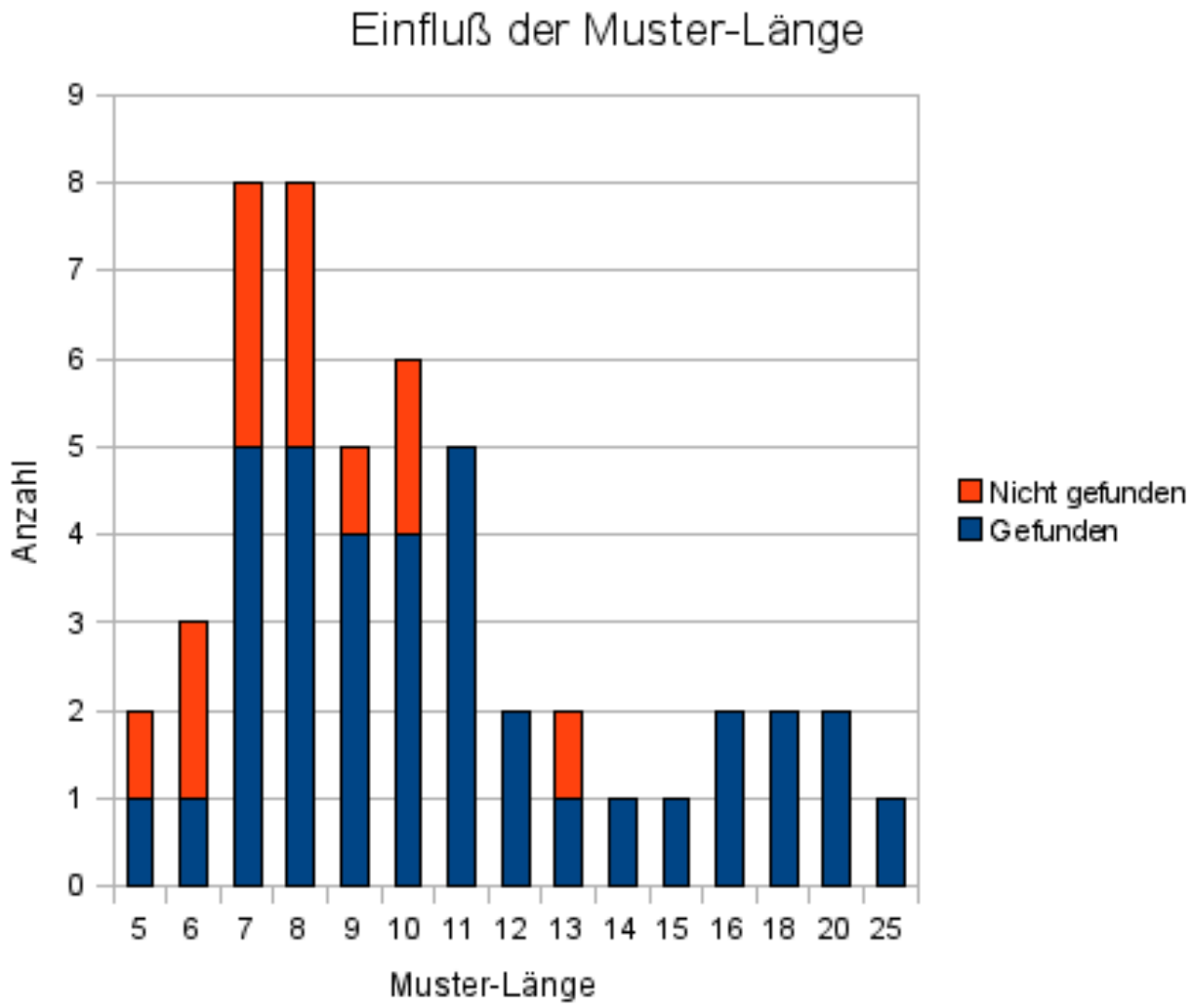
74 % der Ziel-Muster gefunden

Zusätzlich relativ viele Muster, die nicht sperrbar sind

- /main.php?id=
- /ptserver/ok.php?id=

Manuelle Kontrolle notwendig

Einfluss der Muster-Länge



Zusammenfassung

Web-basierte Botnetze und Drive-By-Exploits sind die verbreitetsten Malware-Phänomene

Automatisches Generieren von URL-Mustern kann die Zeit bis zur Sperrung neuer Malware-Seiten reduzieren

Das Verfahren ist nicht perfekt, hat aber eine hohe Erkennungsrate, die sich in der täglichen Arbeit bereits als nützlich erwies

Kontakt

Bundesamt für Sicherheit in der Informationstechnik

Dr. Timo Steffens
Godesberger Allee 185-189
53175 Bonn

Tel. : +49 (0)22899-9582-5822
Fax : +49 (0)22899-10-9582-5822

timo.steffens@bsi.bund.de
www.bsi.bund.de
www.bsi-fuer-buerger.de