2010-10-19(Tue) マルウェア対策研究人材育成ワークショップ(MWS2010)



マルウェア対策のための研究用データセット ~MWS 2010 Datasets~

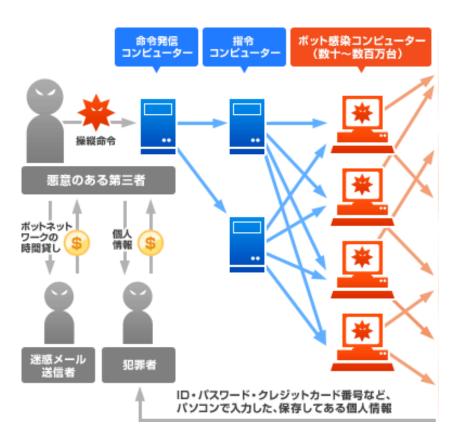
畑田充弘(NTTコミュニケーションズ)、中津留勇(JPCERT/CC)、 秋山満昭(NTT PF研)、三輪信介(NICT)



目次

- > 課題と目的
- MWS 2010 Datasets
 - CCC DATAset 2010
 - MARS for MWS 2010
 - ▶ D3M 2010
- ▶ 実績(速報)
- おわりに

課題



(https://www.ccc.go.jp/bot/index.html)

- ・ 共通の教材がないこと
- 研究用データを入手する こと自体が難しくなってい ること
- DARPAの最新は2001年、 CDX Datasetsは2009年だ がマルウェアによる攻撃 ではない

目的





- 研究用データセットの提供
 - CCC DATAset 2008
 - CCC DATAset 2009
- 研究成果を共有する場、 切磋琢磨する環境作り
 - MWS 2008
 - MWS 2009

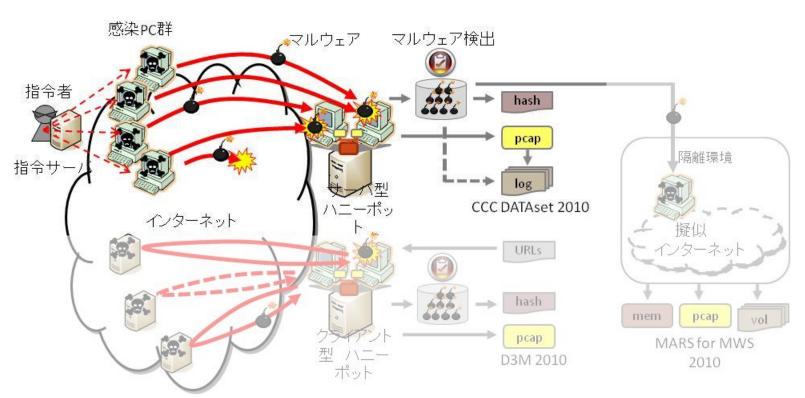


MWS 2010 Datasets

- ► CCC DATAset 2010 (NTTコミュニケーションズ/CCC)
 - ▶ CCCが提供するサーバ型ハニーポットで収集したデータ群
- ▶ MARS for MWS 2010 (CSEC研究会/NICT)
 - CCC DATAset 2010のマルウェア検体を擬似環境で動作させて 得た一次解析結果のデータ群
- ▶ D3M 2010 (NTTコミュニケーションズ/NTT PF研)
 - ▶ クライアント型ハニーポットで収集したWeb感染型マルウェアに関するデータ群

CCC DATAset 2010

- マルウェア検体
- ▶ 攻撃通信データ
- 攻撃元データ



CCC DATAset 2010 – マルウェア検体

8x2017 18016x 88743 Sex 8616 below 881 ma

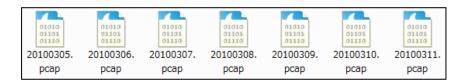
it カッシュ保険 photomorphism の間で回避(肝形種情)」 ■技術的に自動機してもらいたは機能、利益的な技術 Ro-Alignative Schiller (Vecca) Schiller (Veta) et al. DEFENDANCE FROM THE PROPERTY AND THE PROPERTY AND ALTO MAKE THE PROPERTY OF THE [6] U. Lamando P. B. Control (1998) 44 (49) 47 (1998) 74 (49) 54 (49) 化基础 医阿克尔氏氏腺炎 经收益股份 计代码设计 医阿克姆氏管 医阿耳伯氏 with temptransaction the pharm to contact the migra-SANDERS SCHOOL PARTICIPATION OF HARMAN PROBE. 6.7 m C (1 m C) 1 m C (1 m 91207936-1186-6931283118-148824 Gallet 7 9 5 00 485 875 87 1 7 6 89 4 5 4 5 99 5 6 4 6 7 6 DIMENSIONAL PROPERTY AND PROPERTY OF STREET, S Principle School Section 1994 February 1994 and advisor to Starting Start elle SP Schille CS Physics Subs DP Mild on Start SB S adh/2541 de l a 1724/7158 has Sheath di 26a In endate bricks and between the title of the between the case of 21883-860-6018-64-019-12-14-15-418-20 Make 1 Fig. 1 and 10 for property of the case 1 and 20 for the case 2 and ||東京19月1日||東京1月までのお砂様作工 Andrew Chief School Co. Sale McCo. Set 198 etc. 14 Falls F. S. \$16.641 \$10.052 \$48.548 \$250 \$100 \$100 \$100 \$100. ASSESSA 427 0 000 044 674 8 1 2 10 4 5 00 5 10 4 4 7 5 7 6 2 0 0 1 6 4-400 57:000 700 40-001 F02:140 1443 5:1000 \$187383945607164040869410118756111154841071 2114217406-11-0002011933339-0-0-0-193 Compared to the Contract of th \$3 mil ("Sanda FOA ED FOR ED AD AND SECTION SECTION SEC 2002/04/14 10 14600642 01 17 1440007 60 00 460 60 00 \$15600440104060874718840454545646 40% cm 650 04 1 M0000 550 7 x 654 5 10 da 25 c 40 5 x 5 5 a s. 808 104 c 7 6 7 8 4 8 7 c 8 1 8 n 1 2 3 c 8 0 8 n 8 c 0 0 0 0 But 719 1996 Serio ceto e toero del e 1416 Sero e 1993 sel s 9-100 26 26-26-120 15 197 903 97 97 16-109 2.644 do 044 go 6.74 58 573 and 6 to go 6 46 50 2 and 7 to 67 5 表示: 4.4.2.6.4.6.17以下的17.3.4.17。17.3.14.4.18.1.18.1.18.1.1 0.04/54/04056-010-0158414710-04-00-04-12072-0-0 47401904c05c44000a00a0223d4002048 RECORDED TO THE CONTRACTOR AND DRIVE THE PROPERTY OF SHEET 4 (42) 27 (40) (40) 24 (5) 170 (6) (4) 187 (4) 199 (6) A shall be the title to the first that it will be the first to a first to \$2.0 13/Se 1 13/10/15/94 10:5440 56 4000 doi:010/56/12/9 1 1961 100 S Feet a Strike a PM 1 Polesteet Decides a 2 F 3 60 C 50 F 7 F F F I was 7 T for 1 60 to 60 60 de 1 d 2 60 S de 1. 2 add 65 add 1 a 2 88 a 2 1 82 882 8 88 a 2 1 a 4 5 a 004 (\$75700000x7567) (\$007) (\$344000000 mint Fort Sinc Prioritical to 00 filter by 54 Fine by 6 or 5 me of 8 ROBOTH STATES THE PARTIES AND ADDRESS OF THE PARTIES. A 4 8 1 1 4 1 1 4 4 7 1 4 1 7 1 4 1 5 1 7 4 1 5 1 7 4 1 5 1 7 4 1 6 1 7 4 1 7 and the made \$10.000 from \$10.00 for \$100 Medical States from A STREET BOOK OF STREET AND AND ADDRESS OF STREET AND ADDRESS OF A STREET AND n Sattle Bright (1964) and the Sattle Section (1964) in 4 FO Local CO 1 NO REALEST INCIDENT COMPANY REALEST 4883371-3-48-488318023548-6877756-788831-3 \$6.00\$\$6.00\$146.0\$156.0\$4.0\$2.0\$444.1067.20156.0\$6.001. 440 (3x/5N/5005700x44/6754) (xx/615247) IS REPORTED TO THE PARTY OF THE PROPERTY OF THE PROPERTY OF THE PARTY SHEED AND RECEIVED FOR THE ADMINISTRATION AND 10 art 2015 bit 14 Liv 15 als 10 1 all 50 art 2 bit 500 interfoliace earth intermediate and but select in each see. \$5.00 FE BURGERS 1975 TO BE \$46.00 A THE USE SE 3.686/0387.668886-com/FT by 186-642-0687318 0 by 0.1 with a field in the field of th Associate for that the different State and different states. 246/266 (a) 714/666 (166/8987) (b) 10/251/8 FIRE DESIGNATION INCOMES TRANSPORTED IN THE [1] (T-4)(1) (Fe-1)(1) (T-5)(1) (T-5)(1) (Fe-1)(1) (Fe-1)(1) (Fe-1)(1) 3 of 11 No Station 1 to Cod 3 to F4 of P4 57 of P4 to 2 of 2 ft 1 ft of 5 8.4a 1 69.3 6a 5a 1 83 644 67 1 96 9a 8a 87 8 83 \$65,700,000,007,544,685,660,007,562,246,790,740,40,40,46,565,646,6 #5500#End2x050040740#9ac7#54c0#c 42-44-1901 1-3172-41876-6333-6-2191-6-7112-6-4-4-31 strandonomic belle ast the period in the factors is four BY NACTORACY TO A DIS FOR HIS BEST CONTRACTOR IN Physics SEC \$100 and all \$100 by \$100 at \$100 for \$100. 等文字的 \$P\$ (\$P\$) \$P\$ (h 1927/2011/27/2011/1988 he/h/908/2804 h/24/40 N SWISSER AND BRISING SER ENT PRODUCT INS SRI \$70.044 and SACK. 44 00 040 74 784 FB 846 10 044 25 400 675 an I SECURE PROBABILIZATION AND PROPERTY AND DESIGNATION OF THE CONTRACTOR OF THE CONTR \$5.00 PROBLEMS TAXABLE NO. 177 LATER AND In 1 a 2 6 6 7 9 6 6 6 7 7 6 6 6 6 6 7 1 2 6 6 6 6 6 6 6 6 7 7 6 6 6 6 7 7 6 6 6 6 7 7 6 6 6 6 7 7 6 6 6 6 7 7 6 6 6 6 7 7 6 6 6 6 7 7 6 6 6 7 7 6 7 24/24 sans 118 Sept n/24/83 male 63 n/18/n/24/86/14/8 is 47% by 77% 4115 15 1144 582 9145 410 646 51182 92 9 化有限分子 经收入的 医多种多种 医克雷氏病 医电影中枢 经租赁的 医二甲状腺 177819600e-45544504-54-64544616-4-179-286-04-15-1 \$15 GEOGRAPHICA - CONTROL 1-177 From the first # FIREMANDEPER AND AND ADDRESS OF SECOND SEC. 4000041100004014004141444461781 positive entire the property of the property o 50検体のハッシュ値

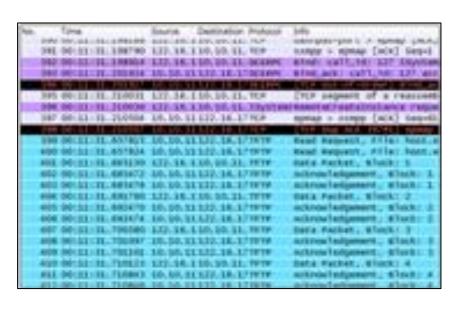
解析結果を照合できる検 体:10検体

▶ 未知検体: 40検体

REPORT OF THE PROPERTY OF THE 21244255252524466Te018e1e8174660644262e

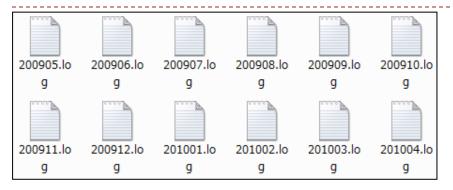
CCC DATAset 2010 – 攻撃通信データ





- サーバ型ハニーポット2台 のパケットキャプチャデー タ
- Windows XP SP1+α
- ▶ 2010年3月5日~11日
- ▶ 2200万パケット、3.5GB

CCC DATAset 2010 – 攻撃元データ





- ハニーポット92台のマルウェア取得時のログ
- ▶ 116万件(TCP:105万件)
- ▶ 3万種(ハッシュ)、1000種 (ウイルス名称)の検体

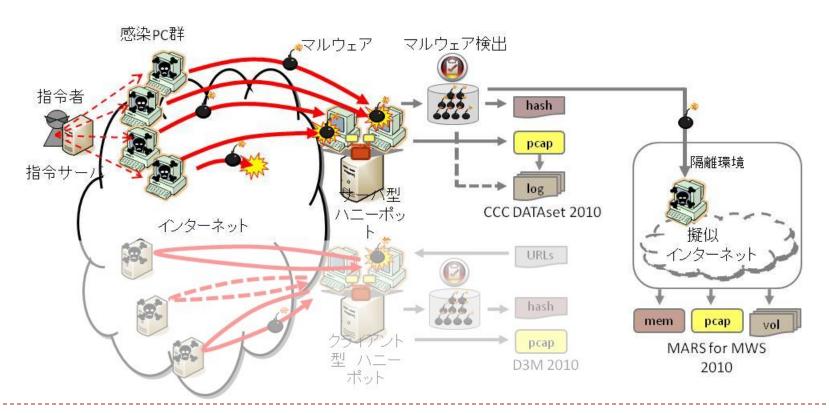
ログ項目	例(一部を*でマスク)	
マルウェア検体の取得時刻	2010-03-05 03:02:41	
送信元IPアドレス	honey001	
送信元ポート番号	1028	
宛先IPアドレス	**. 243.167	
宛先ポート番号	5824	
TCPまたはUDP	TCP	
マルウェア検体のハッシュ値	****bc3c850cf68a39c9	
(SHA1)	e8013f2169d408a9d90	
ウイルス名称	WORM_DOWNAD.AD	
ファイル名	C:¥WINDOWS¥system32	
	¥dhnlr.dll	

CCC DATAset 2010 - 2008/2009比較

	項目	2008	2009	2010				
マ	マルウェア検体							
	検体数	1	10	50				
	選定条件	多機能,解読困難	解析結果あり、関連性の	解析結果あり、特徴的				
			ある複数検体、特徴的な	な機能, 2010年1月~3				
			機能	月に収集した未知検体				
攻	攻撃通信データ							
	ハニーポット	honey001, honey002	honey003, honey004	honey001, honey002				
	収集日	2008/4/28, 2008/4/29	2009/3/13, 2009/3/14	2010/3/5~2010/3/11				
	総パケット数	15,901,943	3,511,850	22,486,674				
攻	攻撃元データ							
	ハニーポット	112台	94台	92台				
	数							
	ハニーポット なし(ダウンロードホスト ID と通信方向のみ)		あり	あり				
	収集期間	2007/11/1~2008/4/30	2008/5/1~2009/4/30	2009/5/1~2010/4/30				
	全レコード数	2,942,221	2,470,766	1,162,093				

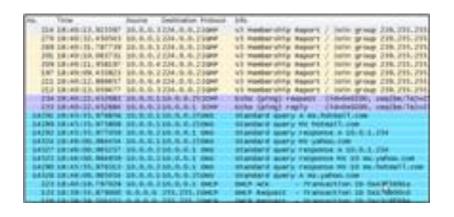
MARS for MWS 2010

- CCC DATAset 2010のマルウェア検体を擬似環境で動作させて得た一次解析結果のデータ群
- ▶ 解析結果の応用によるマルウェア解析技術の研究



MARS for MWS 2010



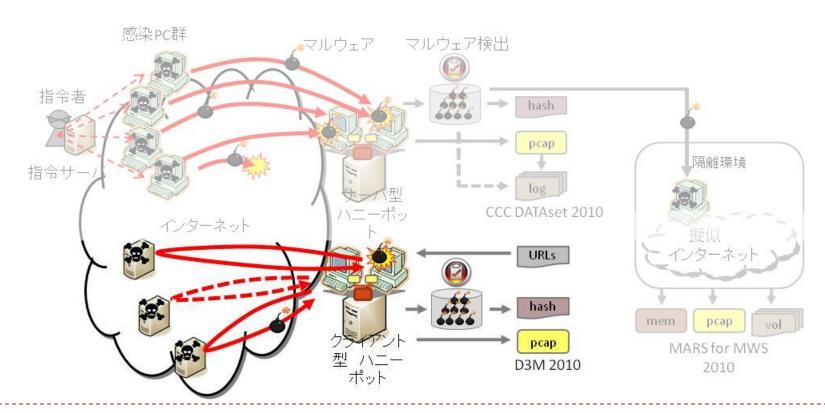


分類情報

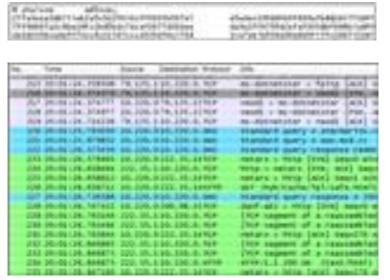
- 検体ファイル名、ハッシュ 値
- 動的解析結果
 - 検体動作時のメモリダンプ、 パケットダンプ、模倣DNS へのアクセス記録
- ▶ 静的解析結果
 - ▶ strings結果、メモリダンプに 対するVolatility Framework による解析結果

D3M 2010

- クライアント型ハニーポットで収集したWeb感染型マルウェアに関するデータ群
- ▶ マルウェア解析技術、感染手法の検知技術の研究



D3M 2010



- ▶ マルウェア検体
 - ▶ 3検体のハッシュ値
 - ▶ Gumblar:8080系
- ▶ 攻撃通信データ
 - Windows XP SP2
 - ▶ 2010年3月8、9、11日
 - ト 入り口URL malwaredomainlist.com
 - ▶ 各日で205、180、172個
 - ▶ 133MB



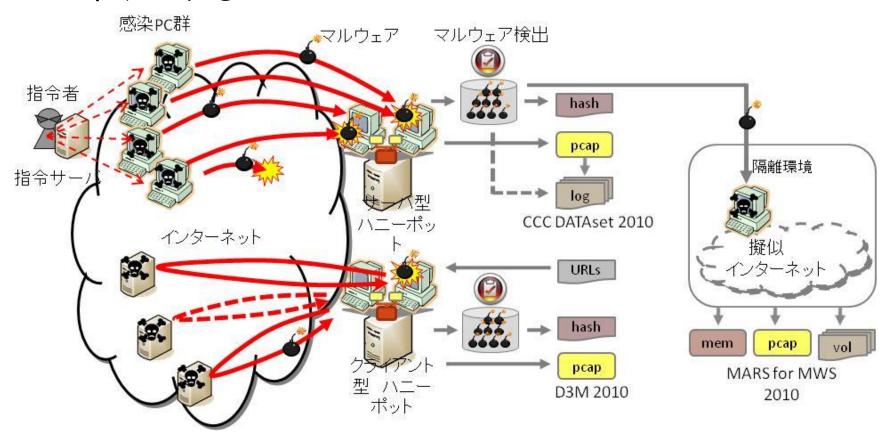
研究用データセットの利用実績

		2008	2009	2010
CCC DATAset	マルウェア検 体	5	7	6
	攻撃通信デー タ	9	14	5
	攻撃元データ	8	6	5
MARS				1
D3M				4
総括			1	1
合計		22(8)	28(15)	22(10)



今後の課題

- ▶ データ種類、収集環境の網羅性、期間、運用/構成情報
- ▶ 正常データも



MWS201x これからが本番

MWS2010 in Okayama

参考文献

- MIT Lincoln Laboratory, DARPA Intrusion Detection Evaluation Data Sets, http://www.ll.mit.ed u/mission/communications/ist/corpora/ideval/index.html
- B. Sangster, et al.: Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets, 18th USENIX Security Symposium CSET'09 (2009.08)
- ▶ サイバークリーンセンター, https://www.ccc.go.jp/
- ▶ 畑田充弘, 他: マルウェア対策のための研究用データセットとワークショップを通じた研究成果の共有, 情報処理学会シンポジウムシリーズ, Vol.2009, No.11, CSS2009(MWS2009), pp.1-8 (2009.10)
- 畑田充弘: 研究用データセットを用いたマルウェア対策研究人材育成ワークショップ,情報処理, Vol.51, No.3, pp.284-287 (2010.03)
- ▶ 竹森敬祐, 他: MWS Cup 2009, 情報処理, Vol.51, No.3, pp.296-299 (2010.03)
- ▶ マルウェア対策研究人材育成ワークショップ, http://www.iwsec.org/mws/2010/
- 三輪信介, 他: 小規模攻撃再現テストベッドによる動作記録データセットの生成, 情報処理学会シンポジウムシリーズ, Vol.2009, No.11, CSS2009(MWS2009), pp.931-936 (2009.10)
- N. Petroni, et al: FATKit: A Framework for the Extractionand Analysis of Digital Forensic Data from Volatile System Memory, Digital Investigation Journal 3(4) (2006.12)
- Mitsuaki Akiyama, et al: Design and Implementation of High Interaction Client Honeypot for D rive-by-download Attacks, IEICE Transactions on Communication, Vol.E93-B No.5 pp.1131-113 9 (2010.05)