Projekt CAMNEP – systém detekce průniku ve vysokorychlostních počítačových sítích

Pavel Čeleda, Karel Bartoš, VojtěchKrmíček, Pavel Minařík, ÚVT MU

1 Úvod

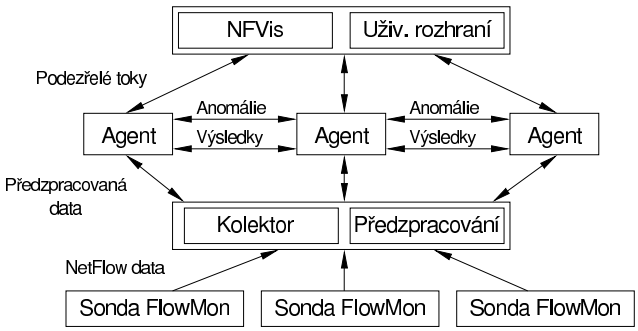
Neutuchající nárůst uživatelů a služeb na Internetu vede k postupnému rozšiřování vysokorychlostních počítačových sítí na rychlostech do 10 Gb/s. Kromě jednoznačně plynoucích výhod jako je navýšení přenosových kapacit, dochází i k logickému zvětšování objemu přenášených dat. Větší množství uživatelů vzájemně propojených vysokorychlostní sítí představuje optimální podmínky pro šíření škodlivého kódu (nevyžádané e-maily – SPAM, útoky odmítnutí služby – DoS, získávání citlivých údajů uživatelů atd.).

Komunikace na páteřní desetigigabitové lince může představovat přenos až několika milionů paketů za vteřinu. V tomto prostředí je prakticky nemožné manuálně provádět kontrolu a odhalování bezpečnostních incidentů na počítačové síti. V důsledku této situace vzniklo několik metod snažících se o automatické rozpoznávání anomálního provozu a nepovoleného chování uživatelů. Obecnou vlastností, kterou se tyto metody vyznačují, je (*i*) nízká datová propustnost (problém bezztrátového získávání statistik ze sítě) a (*ii*) vysoká míra špatně oklasifikovaného provozu, která znesnadňuje jejich praktické nasazení.

V oblasti bezpečnosti počítačových sítí se angažuje řada komerčních subjektů, státních organizací a výzkumných pracovišť. Nelze se proto divit, že jednou ze zájmových oblastí, kterou v rámci podpory výzkumu financuje Armáda Spojených států, je problematika síťové bezpečnosti. Armáda Spojených států navrhla v roce 2005 spolupráci Ústavu výpočetní techniky MU a Gerstnerově laboratoři na ČVUT. Výsledkem spolupráce bylo sestavení řešitelského týmu projektu CAMNEP [1], který měl za cíl vytvořit systém pro detekci průniku ve vysokorychlostních počítačových sítích. Skupina z Ústavu výpočetní techniky MU zastřešovala problematiku měření síťového provozu a inteligentní vizualizaci výstupů systému. Skupina agentních systémů z Gerstnerovy laboratoře se věnovala problematice rozpoznávání anomálního provozu a snížení míry špatně klasifikovaného provozu.

2 Architektura systému

Systém CAMNEP (*Cooperative Adaptive Mechanism for Network Protection*) byl od prvopočátku rozdělen do tří vrstev zobrazených na obrázku 1. Nejnižší vrstva systému je optimalizována na co největší výkon, aby umožňovala zpracovávat obrovské objemy dat přenášených v dnešních vysokorychlostních sítích (1-10 Gb/s) a předzpracovávat je pro vyšší vrstvy systému. Prostřední vrstva je zaměřena na využití předzpracovaných dat pro vyhledávání anomálního síťového provozu a určování míry důvěryhodnosti jednotlivých toků v síti. Nejvyšší vrstvu systému tvoří rozhraní směrem k bezpečnostnímu správci sítě, kterému poskytuje informace o detekovaném anomálním provozu.



Obrázek 1: Architektura systému CAMNEP.

2.1 Vrstva sběru dat a jejich předzpracování

Nejnižší vrstva se skládá z hardwarově akcelerovaných sond FlowMon [4], které vytvářejí NetFlow statistiky o síťovém provozu a odesílají je na kolektor. Kolektorová část provádí sběr přijatých dat a jejich předzpracování pro vyšší vrstvy. Tento přístup poskytuje v reálném čase přehled o všech síťových spojeních na monitorované lince. Předzpracování dat je zaměřeno především na výpočetně náročné operace (agregace dat, výpočty entropií provozu) a odlehčuje vyšším vrstvám od časově náročných výpočtů nad neagregovanými NetFlow daty.

2.2 Agentní vrstva pro detekci anomálií

Prostřední vrstva se skládá z několika specializovaných agentů. Jednotliví agenti reprezentují různé metody detekce anomálií a vyhledají anomálie v NetFlow datech za pomoci rozšířených důvěryhodnostních modelů [5]. K celkovému rozhodování o důvěryhodnosti jednotlivých toků se využívá kolektivní rozhodování s reputačním mechanismem. Agenti jsou spouštěni uvnitř agentní platformy AGLOBE [2] a využívají jejích pokročilých vlastností jako je migrace agentů a klonování z důvodu adaptace celého systému na aktuální provoz a vyskytující se hrozby.

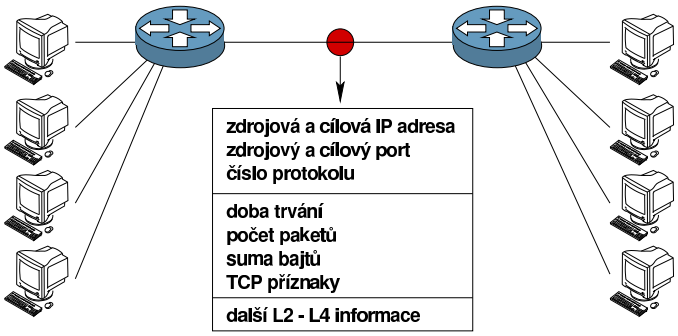
2.3 Uživatelské rozhraní síťového operátora

Hlavní role nejvyšší vrstvy je interakce s operátorem. Základní komponentou je vizualizační nástroj NFVis (*NetFlow Visualizer*), který pomáhá operátorovi analyzovat výstup z detekční vrstvy pomocí přehledné prezentace podezřelých toků doplněných o další relevantní informace. Jakmile je v detekční vrstvě odhalen podezřelý provoz, je předán vizualizační vrstvě, která jej společně s relevantními informacemi z datových zdrojů na síti (DNS jména, informace o portech, WHOIS databáze atd.) zobrazí operátorovi. Operátor má možnost daný provoz zkoumat z různých aspektů, případně se zaměřit na konkrétní podmnožinu či charakteristiky ve zkoumaných datech.

3 Sběr dat ve vysokorychlostních sítích

Sběr dat z počítačové sítě je založen na získávání informací o IP tocích. Tok je definován jako sekvence paketů se shodnou pětící údajů: cílová/zdrojová IP adresa, cílový/zdrojový port a číslo protokolu. Vytvářené statistiky poskytují informace o rozložení provozu na síti a chování uživatelů.

Současné technické a programové vybavení určené pro sledování síťového provozu pomocí toků využívá směrovačů anebo specializovaných sond exportujících data ve formátu NetFlow verze 5 a 9. V projektu CAMNEP jsou k získávání NetFlow statistik použity hardwarově akcelerované sondy FlowMon. Oproti řešením založeným



Obrázek 2: Princip vytváření NetFlow statistik.

na směrovačích umožňují sondy v libovolném místě sítě bezztrátové měření i na nejvyšších rychlostech. Jsou nezávislé na stávající síťové infrastruktuře a neovlivňují chování sítě. Sondy je možno podle potřeby rozšiřovat o další funkcionalitu, což v případě směrovačů možné není, a navíc je limitující při použití pro výzkumné účely.

Sonda FlowMon vychází z rodiny akceleračních karet COMBO [3] vyvíjených v rámci výzkumného záměru sdružení CESNET (řešitelé z Masarykovy univerzity a VUT v Brně) a mezinárodních projektů EU. Úloha vytváření síťových statistik na bázi toků je rozdělena do dvou částí. První část je založena na programovatelných strukturách s obvody FPGA (*Field Programmable Gate Array*), které slouží k akceleraci časově kritických operací. Jedná se zejména o bezztrátový příjem paketů a jejich agregaci do záznamů o tocích. Druhou část tvoří programové vybavení zodpovědné za obsluhu akcelerační karty, vyčtení záznamů z paměti karty a jejich odeslání protokolem NetFlow na kolektor.

Sběr a vyhodnocování získaných NetFlow dat s využitím FlowMon sond jsou prováděny pomocí specializovaných aplikací, tzv. kolektorů. Funkcí kolektoru je přijímat NetFlow data odesílaná z exportérů na sondě, ukládat je na disk a provádět jejich další zpracování.

Pro potřeby projektu CAMNEP bylo použito open-source kolektorové řešení, založené na sadě nástrojů *NFDUMP* [7] s grafickým rozhraním *NfSen* [8]. Hlavními výhodami tohoto řešení je možnost modifikace zdrojového kódu a dobrá podpora pro zpracování různých verzí NetFlow

protokolu. Dále je možné provádět detailní zpracování získaných NetFlow dat pro různé časové periody, vytvářet profily sběru dat a využívat řady funkcí pro filtraci a agregování.

Zdrojový kód kolektoru byl doplněn o výpočet entropií síťového provozu a vybranou množinu agregačních funkcí, které jsou využívány v metodách pro detekci anomálií. Výpočet se provádí v pětiminutových intervalech. V závislosti na typu sítě se může jednat o statisíce až milióny toků. Komunikace s agentní vrstvou je prováděna po síti protokolem TASI (*Traffic Acquisition Server Interface*), který byl navržen speciálně pro potřeby projektu CAMNEP.

Pro testování a ověření navrženého systému bylo nezbytné získání dat z reálné vysokorychlostní počítačové sítě, kde se vyskytuje jak legitimní, tak nežádoucí síťový provoz. Sonda FlowMon byla zapojena do sítě Masarykovy univerzity, která vzhledem k charakteru provozu poskytuje optimální podmínky pro výzkum v oblasti počítačové bezpečnosti. V síti je zapojeno několik tisíc počítačů využívajících rozličných síťových služeb. Akademický charakter sítě s sebou nese riziko spojené s širokým okruhem zájmů řady uživatelů na Internetu. Ne vždy je však počínání těchto uživatelů vedeno s ušlechtilými cíli. Často je jejich zájem směřován na ovládnutí počítačů v univerzitní síti a jejich využití k dalším nelegálním aktivitám.

4 Detekce škodlivého síťového provozu

Detekční vrstva má za úkol vybrat z množiny všech zaznamenaných spojení pouze ta, která představují neobvyklou či nežádoucí aktivitu v síti. Tyto tzv. *incidenty* jsou poté reportovány bezpečnostním administrátorům, kteří mohou následně provést odpovídající reakce v rámci bezpečnostních opatření na síti. Detekční vrstva v systému CAMNEP je založena na multiagentním systému AGLOBE [2], ve kterém pracuje několik agentů (samostatných entit), vykonávajících vlastní detekci. Výsledky jednotlivých agentů jsou v průběhu detekčního procesu agregovány k zajištění komplexního pohledu na daný síťový provoz.

Každý detekční agent je založen na jedné z námi předem vybraných metod detekce anomálií. Výběr těchto metod se řídil podmínkou vzájemné odlišnosti v pohledu na charakteristiky síťového provozu, pomocí kterých lze dosáhnout rozpoznávání incidentů na síti. V současné době využíváme pět základních přístupů, které jsme více či méně modifikovali pro vzájemnou kompatibilitu všech komponent systému.

Metoda MINDS [6] modeluje počty příchozích a odchozích spojení jednotlivých hostů (v kombinaci s číslem portu) v průběhu času a detekuje vzájemné odchylky v těchto časových řadách – velikost odchylky určuje míru anomálie. Metoda autorů Xu et al. [9], za použití statických klasifikačních pravidel, umisťuje jednotlivá spojení do vícedimenzionálního prostoru entropií IP adres a portů. Toto umístění poté rozhoduje o anomálnosti daného spojení. Další metoda [10] využívá k modelování objemu provozu jednotlivých hostů (počty spojení, paketů a bytů) statistickou metodu PCA (*Principal Component Analysis*), pomocí které a na základě sady předchozích a současného pozorování rozděluje provoz na normální a reziduální část. Velikost reziduální části se poté využívá ke stanovení míry anomálie každého hosta. Metoda [11] je analogická s předchozí metodou pouze s tím rozdílem, že PCA používá k separaci reziduální části distribuce provozu, čili entropií IP adres a portů. Poslední metoda [12] je speciálně vyvinuta k detekci různých typů skenování a využívá k tomu velikost poměru počtu odchozích IP adres a počtu různých odchozích portů v kombinaci s metodou sekvenčního testování hypotéz.

V první fázi každý agent na základě své metody detekce anomálií přiřazuje jednotlivým spojením jejich míru anomálie. Pro každé spojení takto získáme od každého agenta jednu hodnotu. Tyto hodnoty se v závěru první fáze agregují v celkovou míru anomálie pro dané spojení.

Metody detekce anomálií mají obecně velkou výhodu v tom, že jsou schopné detekovat i dosud neznámé síťové události. Tato schopnost je však vykoupena obecně horší klasifikací, zejména *false positivů* (legitimního provozu klasifikovaného jako nežádoucího). Proto si každý agent společně s metodou detekce anomálií udržuje

rovněž svůj tzv. důvěryhodnostní model [5], který používá ke stanovení důvěryhodnosti daného spojení na základě dlouhodobého sledování provozu.

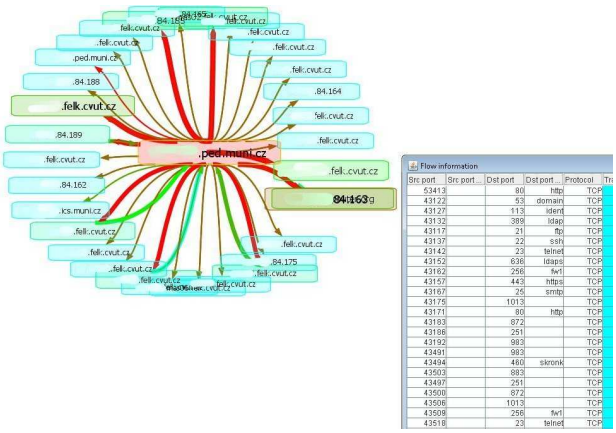
Důvěryhodnostní model využívá ke stanovení důvěryhodnosti vícedimenzionální klastrovací prostor, do kterého umisťuje dané spojení na základě hodnot jeho pozorovaných vlastností a charakteristik, jež byly v první fázi detekce využívány daným agentem k určení míry anomálie. Důvěryhodnost daného spojení je poté určena jeho velikostí anomálie s přihlédnutím na důvěryhodnost okolních spojení s obdobnou charakteristikou síťového provozu.

V závěrečné fázi detekce dochází pro každé spojení k finální agregaci důvěryhodností ze všech důvěryhodnostních modelů. Agregace se provádí pomocí tzv. OWA (*Ordered Weighted Averaging*) operátorů. Jelikož každý agent používá k modelování důvěryhodnosti různé charakteristiky, nese výsledná hodnota informaci o široké škále vlastností síťového provozu. Výsledná důvěryhodnost značí míru legitimnosti každého spojení a dále se zobrazuje ve vizualizační vrstvě pro případnou další analýzu ze strany bezpečnostních administrátorů.

5 Vizualizace síťového provozu

Tradiční způsob zobrazování dění na počítačové síti bývá řešen ve formě statistických grafů (např. koláčový graf). Jednotlivé incidenty jsou zobrazovány jako řádky v tabulce. V projektu CAMNEP jsme zvolili zcela jiný přístup. Naším cílem bylo zobrazovat provoz vyhodnocený jako nedůvěryhodný způsobem, který umožní operátorovi efektivně rozhodnout, zda se jedná o bezpečnostní incident či nikoliv. Vytvořená aplikace NFVis zobrazuje počítačovou síť jako orientovaný graf, ve kterém uzly představují jednotlivá zařízení v síti a hrany reprezentují komunikaci mezi danými zařízeními. Výsledný graf je doplněn o tabulky se statistickými informacemi (celkové objemy přenesených dat, druhy provozu, spojení s daným zařízením v síti atd.) a detaily o přenosech na konkrétní hraně.

Vizualizace ve formě orientovaných grafů podporuje intuici operátora. Je možné využívat barvy a velikosti uzlů i hran k vyjádření různých ukazatelů. Např. barvy je možné využít pro míru důvěryhodnosti daného provozu a velikost k vyjádření objemu dat. Tyto atributy lze navíc efektivně měnit za běhu na požádání uživatelem. Výstup aplikace je zobrazen na obrázku 3. Komponenta sloužící k vizualizaci dat o provozu na počítačové síti je založena na knihovně *Prefuse* [13].

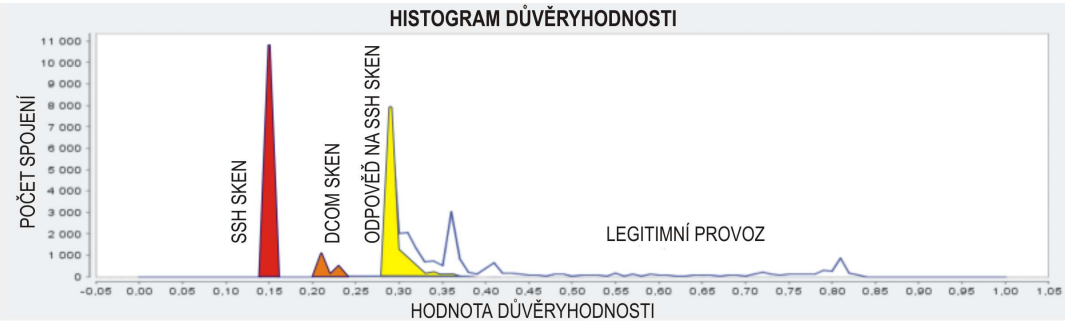


Obrázek 3: Příklad vizualizace provozu na síti ve formě orientovaného grafu doplněného o tabulku detailů provozu.

6 Závěr

Vytvořený systém jsme nasadili na síť Masarykovy univerzity. Provedené testování pomocí několika různých scénářů (např. uměle vytvořenými útoky) sloužilo ke stanovení efektivity detekce a určení celkové výkonnosti systému při zatížení velikými objemy dat. Výsledky systému byly následně manuálně prověřeny síťovým administrátorem.

Na obrázku 4 je uveden jeden z grafických výstupů systému. Jedná se o histogram důvěryhodnosti jednotlivých spojení (z pětiminutového intervalu) získaných při sledování testovací sítě. Graf udává, kolik spojení má danou hodnotu důvěryhodnosti. V levé části histogramu se nachází dva podezřelé incidenty, které systém bezpečně detekoval a prezentoval jako nežádoucí. V pravé části se pak nachází legitimní provoz s vyšší důvěryhodností.



Obrázek 4: Histogramu důvěryhodnosti síťového provozu.

Z experimentů vyplývá, že přidání důvěryhodnostního modelu jako dlouhodobého modelovacího mechanismu k metodě detekce anomálií má za následek redukci false positivů o 50 % až 75 %. Další redukci false positivů při udržení (nesnížení) počtu odhalených incidentů lze dosáhnout pomocí vzájemné OWA agregaci výsledků z jednotlivých anomálních a důvěryhodnostních modelů (až o 90 % v porovnání s individuálními modely). Použití hardwarové akcelerace v podobě FlowMon sond umožňuje nasazení systému na gigabitových linkách s vysokými počty spojení.

Poděkování

Projekt CAMNEP byl vytvořen za podpory evropského úřadu pro výzkum Armády Spojených států – číslo kontraktu N62558-07-C-0001.

Literatura

[1] Martin Rehak, Michal Pechoucek, Karel Bartos, Martin Grill, Pavel Celeda, Vojtech Krmicek: *CAMNEP: An intrusion detection system for high-speed networks*. In Progress in Informatics, number 5, pages 65-74, March 2008.

[2] David Šislák, Martin Rehák, Michal Pěchouček, Milan Rollo, Dušan Pavlíček: *A-globe: Agent Development Platform with Inaccessibility, Mobility Support*. In Software Agent-Based Applications, Platforms and Development Kits, pages 21–46, Berlin, 2005. Birkhauser Verlag.

[3] Projekt Liberouter: *Rodina karet COMBO*. `www.liberouter.org/hardware.php`

[4] Projekt Liberouter: *FlowMon probe*. `www.liberouter.org/flowmon`

[5] Martin Rehak, Michal Pechoucek: *Trust Modeling with Context Representation and Generalized Identities*. In Cooperative Information Agents XI, number 4676, in LNAI/LNCS. Springer-Verlag, 2007.

[6] Levent Ertoz, Eric Eilertson, Aleksandar Lazarevic, Pang-Ning Tan, Vipin Kumar, Jaideep Srivastava, Paul Dokas: *MINDS – Minnesota Intrusion Detection System*. In Next Generation Data Mining, MIT Press, 2004.

[7] Peter Haag: *Kolekce nástrojů NFDUMP*. `nfdump.sf.net`

[8] Peter Haag: *Kolektor NfSen – Netflow Sensor*. `nfdump.sf.net`

[9] Kuai Xu, Zhi-Li Zhang, Supratik Bhattacharrya: *Reducing Unwanted Traffic in a Backbone Network*. In USENIX Workshop on Steps to Reduce Unwanted Traffic in the Internet (SRUTI), Boston, MA, July 2005.

[10] Anukool Lakhina, Mark Crovella, Christophe Diot: *Diagnosis Network-Wide Traffic Anomalies*. In ACM SIGCOMM '04, pages 219–230, New York, NY, USA, 2005. ACM Press.

[11] Anukool Lakhina, Mark Crovella, Christophe Diot: *Mining Anomalies using Traffic Feature Distributions*. In ACM SIGCOMM, Philadelphia, PA, August 2005, pages 217–228, New York, NY, USA, 2005. ACM Press.

[12] Avinash Sridharan, Tao Ye, Supratik Bhattacharyya: *Connectionless Port Scan Detection on the Backbone*. In Malware workshop, held in conjunction with IPCCC, 2006.

[13] Berkeley Institute of Design: *The Prefuse Visualization Toolkit*. `www.prefuse.org` □