

Business Intelligence på Internet

Mikael Simovits
ICL Svenska AB
mikael.simovits@li.icl.se

Tomas Forsberg
ABB Infosystems AB
tomas@datw235.seinf.abb.se

***Sammanfattning:** Business Intelligence (BI) på Internet är idag en realitet, och på många sätt ett växande problem. Datasäkerhet är inte längre en fråga om att bara förhindra intrång. En kartläggning och analys av tillgängligt material på Internet kan vålla ett företag betydande skador. Artikeln beskriver övergripande problematiken samt tekniker och metoder som idag används inom Business Intelligence på Internet.*

1 Inledning

Datorsäkerheten har idag blivit ett kluvet ämne. En del anser att databrottslighet i kommersiellt syfte knappast existerar, medan andra målar upp en hotbild som i vissa fall är överdriven. I USA har DoD (Department of Defence) undersökt säkerheten i sina system, genom att själva attackera 38000 av sina UNIX maskiner. Attackerna lyckades i två tredjedelar av fallen. Endast i 988 fall upptäcktes intrången/försöken och endast 267 av dessa fall rapporterades[1]. På många sätt fokuseras intrång som det enda säkerhetshotet från Internet, vilket leder till att andra hot mot informationen glöms bort. Förutom ”stöld” av information, så har Internet också gett nya möjligheter till att manipulera, marknadsföra och inhämta information utan att göra intrång eller skadegörelse. Det är den senare aspekten som kommer att belysas i denna artikel. För att problemet skall kunna ses i sitt rätta sammanhang, beskrivs först innebörden av termen Business Intelligence, för att sedan anpassa problemställningen till dagens situation på Internet.

2 Business Intelligence (BI)

I t ex Sverige, tänker inte företagen på industrispionage i konkreta termer, emellertid kan det i andra länder vara ett legitimt sätt att skaffa nödvändig information. Industrispionage går oftast under den kollektiva termen "Business Intelligence". Business Intelligence definieras som en aktivitet för att övervaka ett företags externa omgivning med syfte att hitta information som kan införas i företagets beslutsprocess[4]. Vissa hävdar att Business Intelligence inte består av illegala metoder och att det är ett passivt sätt att inhämta och använda information. Men i sitt sammanhang är det svårt att dra en gräns mellan vad som är ”Business Intelligence” och vad som är illegalt industrispionage[2]. Det är lagen och moraliska värderingar som bestämmer var gränsen går för inhämtningen och användningen av information. Nedan följer en gradering av Dr Worth Wade, på metoder som man kan tänkas använda för att skaffa information om konkurrenter [3]:

1. Publicerat material offentliga dokument, typ rättegångsprotokoll.
2. Information som konkurrenter själva publicerar.
3. Marknadsundersökningar och konsultrapporter.
4. Finansiella rapporter och mäklares prognoser.
5. Mässor, utställningar och broschyrer.
6. Analys av konkurrenters produkter.
7. Rapporter från försäljningsavdelning och inköpsavdelningen.
8. Legitima anställningsintervjuer med personer som arbetat hos konkurrenter.
9. Kamouflerade "utpumpningar" av konkurrenter vid tekniska seminarier.
10. Direkta observationer under hemliga förhållanden.
11. Falska anställningsintervjuer med personer som arbetat hos konkurrenter. (m.a.o. utan någon verklig intention att anställa personen).
12. Falska förhandlingar med konkurrenter om t ex samarbete eller licensavtal.
13. Anställa professionella utredare att hämta specifik typ av information.
14. Rekrytera medarbetare från konkurrent, för att få specifik "know how".
15. Gå in i konkurrentens byggnader, fabriker och egendom utan tillstånd.
16. Muta konkurrenters leverantörer eller anställda.
17. "Plantera" en agent på konkurrentens lönelista.
18. Avlyssning av konkurrenter.
19. Stöld av ritningar, prov, dokument och liknande egendom.
20. Utpressning och hotelser.

De första sju metoderna anses som moraliskt riktiga och lagliga. De resterande punkterna är sorterade i nedstigande ordning med hänsyn till moral och lag. Det intressanta är att gränsen för lag och moral placeras på olika ställen i graderingen, beroende på land och företagskultur. I vissa länder i Östeuropa är det inte ovanligt att man anser att hela listan är legitim[5]. Förändringar av sociala värden och trygghetsförhållanden kan snabbt leda till att nya etiska värderingar sprids som i sin tur kommer att medföra en mer tydlig informationskrigsföring[6].

3 Business Intelligence på Internet

"Business Intelligence" har fått en ny innebörd på grund av Internet och andra multinationella nätverk. Informationsinsamlingen förenklades när de nationella gränserna raderades bort och tack vare företags respektlöshet mot den information de presenterar på Internet. Det saknas dessutom en fungerande lagstiftning och bevisbarhet för "tveksamma" aktiviteter på Internet. Nedan följer en fri anpassning av Wade's gradering till den moderna informationsteknologin och Internet:

1. Periodisk hämtning och spegling av konkurrenters WWW-sidor och filarkiv.
2. Hämtning av öppen information på Internet i form av tidningsarkiv.
3. Kartlägga ett företag med hjälp av sökverktyg samt söka de sidor som har "länkar" till presentationerna för att finna intressenter, typ kunder och leverantörer.
4. Avlyssna och filtrera informationsflödet i nyhetsgrupper och mailinglistor.
5. Öppen analys av konkurrentens datanät och användare, undersöka nätet utifrån med hjälp av bland annat kataloginformation.
6. Kartlägga användarnas intressen på Internet, dvs vilken information söker de mest.
7. Delta i konkurrents mailinglistor, interna WWW-sidor mm under falskt namn eller identitet som disponeras via konkurrentens kunder.

8. Kartlägga konkurrentens interna datanät i samband med samarbetsprojekt eller via konsultarbeten.
9. Kamp om domännamn och uttag av domännamn i förebyggande syfte.
10. Dataintrång och skadegörelse.

Det är idag inget ovanligt att företag tillämpar de metoder som beskrivs ovan. Särskilt i konkurrensen om marknaden för internetprodukter har många av ovan nämnda metoder iakttagits. Punkterna 1-3 kan betraktas som normala BI-aktiviteter. Punkterna 4-6 utgör en aggressivare hållning som i vissa situationer kan ifrågasättas ur ett moraliskt och juridiskt perspektiv. Punkterna 7-9 är direkt tveksamma och punkt 10 består enbart av illegala verksamheter.

4 Genomgång av punkterna

1. Periodisk hämtning och spegling av konkurrenters WWW-sidor och filarkiv.

Allt eftersom att en marknad blir mättad, riktas företagens uppmärksamhet mot att bevaka sina egna marknadsandelar eller rent av ta marknadsandelar från konkurrenter. Det är därför viktigt att trender, kunder och konkurrerande företags aktiviteter bevakas. Det kan även vara intressant att bevaka organisationer verksamma inom konsument, miljö och andra intresseområden.

Internet är ett nytt medium för företag att sprida information om sina verksamheter, tjänster och produkter. Företagen utnyttjar Internet också för att presentera samarbeten, forskningsrapporter, rekrytering osv. Man bryter normalt inte mot den praxis man har för klassning av information, men koncentrationen av informationen och dess tillgänglighet gör den värdefull.

Tekniskt sett finns det flera ambitionsnivåer. Den enklaste lösningen är att bara använda cacheservers och ange lämpliga parametrar för det företag man vill bevaka. Med cache avses här möjligheten att lokalt behålla inhämtat data en kortare tid, ifall detta data efterfrågas på nytt. Denna lösning kan göras transparent för den egna organisationen. Man kan också undvika konkurrentens inställningar - vissa företag anger att deras dokument inte skall sparas genom sin "expire" tidpunkt. Genom att inte reagera på denna uppgift kan den egna organisationens tillgång till det önskade materialet förbättras och minska antalet synliga förfrågningar.

I nästa nivå speglas ett företags maskiner. En normal sökrobot hämtar en hemsida och följer sedan länkarna i dessa. I en mer aggressiv spegling försöker man hämta samtliga filer. Antag att hemsidan heter <http://www.bolag.se> och en länk i denna leder till dokumentet <http://www.bolag.se/produkt1/beskrivning.html>. I detta läge försöker roboten även hämta dokumentet <http://www.bolag.se/produkt1> som t ex kan vara en indexfil, samt hämta <http://www.bolag.se/produkt1/> som beroende på inställningarna i WWW-servern kan ge en fullständig filförteckning av denna katalog.

En tredje nivå kan även inkludera sökning efter olika former av indexfiler eller filer som beskriver WWW-servern. Många av dessa filer är avsedda för sökrobotar. Det enklaste sättet att bekanta sig med namnen och funktionerna på dessa filer är att gå igenom förteckningarna över robotar som finns spridda på Internet.

Man kan slutligen använda robotar avsedda för att spegla gopher och ftp servrar för att se om man kan få mer av ett filsystem jämfört med vad företagets WWW-server lämnar ut.

2. Hämtning av öppen information på Internet i form av tidningsarkiv.

Ett sätt att enkelt bevaka ett företags aktiviteter, är att söka genom arkiv och tidningars hemsidor, efter nyckelord som berör det företag som skall undersökas. Tekniken för detta är samma som i punkt ett med den skillnaden att endast informationen som passar nyckelorden laddas ned på den egna servern. Med tidningsarkiv avses här även fackpress, forskningsartiklar, företagstidningar och lokala tidningar. Speciellt företagstidningar ger ofta mer detaljrika beskrivningar av vad man sålt, köpt eller inlett för samarbete med andra företag.

Ifall man inte vill utföra denna sökning själv kan man köpa information från en befintlig leverantör, antingen en traditionell leverantör av mediabevakning eller en Internetleverantör med en sökmotor. I det andra fallet köper man antingen vissa sökord eller hela databasen för att själv filtrera denna.

3. Kartlägga ett företag med hjälp av sökverktyg samt söka de sidor som har "länkar" till presentationerna för att finna intressenter, typ kunder och leverantörer.

Förutom att ha kännedom om ett företags utseende kan det också vara intressant att kunna kartlägga företagets kunder, partners och samarbetsprojekt. Ett företags presentationer innehåller ofta länkar som pekar på kunder och partners. Det är möjligt att göra en enkel omvänd sökning där länkarna till företaget söks, dvs att man ger en sökmotor uppdraget att söka efter presentationer som innehåller strängen "<http://www.bolag.se>". Normalt finns denna söksträng i dokument som ligger utanför mål-företagets egna WWW-server och kontroll.

Det är viktigt att förstå betydelsen av sökverktygen rätt, kapaciteten har ökat dramatiskt under några få år. I takt med förbättrad funktionalitet kommer deras betydelse öka ytterligare. De funktioner som saknas idag är främst möjligheterna att undertrycka flera träffar från samma katalogstruktur, ange geografisk och språklig avgränsning samt möjligheten att spara en sökning för vidare bearbetning i form av nya sökningar eller hämtning av hela materialet för en egen analys.

Det bör också understrykas att man söker på all text, alltså även kommentarer och anrop av olika script. Det är till och med möjligt att söka WWW-servrar med `cgi-bin` svagheter för att sedan göra datainrång i dessa datorer.

4. Avlyssna och filtrera informationsflödet i nyhetsgrupper och mailinglistor.

Nyhetsgrupper, också kallat för Usenet News, har fått ny aktualitet och kan inte liknas med äldre metoder att förmedla information. För varje intresseområde finns en nyhetsgrupp där intresserade kan diskutera olika problem. Många anställda använder detta medium för att diskutera eller för att lösa problem. Med tanke på att ett företag i snitt har tillgång till 2000-3000 nyhetsgrupper och varje grupp i snitt har 50-100 inlägg varje dag, handlar det om stora informationsmängder som är svåra att överblicka.

Om ett företag t ex vill bevaka sin konkurrent som har domännamnet `bolag.se` kan företaget i sin egen news-server filtrera alla inlägg som innehåller `bolag.se` i headern och lägga upp dessa inlägg i egna databaser. Man kan på så sätt dra slutsatser om konkurrentens problem, kompetens och pågående projekt. Många anställda utnyttjar dessutom sin tillgång till nyhetsgrupper i privata syften, även denna informationen kan utnyttjas.

5. Öppen analys av konkurrentens datanät och användare, undersöka nätet utifrån med hjälp av bland annat kataloginformation.

När en säljare kommer till ett nytt företag, är oftast det första målet att erhålla den interna telefonkatalogen. Syftet är att få en klarare bild över kundens organisation. En intern telefonkatalog innehåller uppgifter om personers organisatoriska ställning och ansvarsområden. En säljare kan vid förfrågningar från en person inom detta bolag bedöma sina egna "presale" insatser, allvaret i en förfrågan eller om han behöver hjälp av andra från det egna bolaget vid säljaktiviteter. Ännu bättre är om säljaren vid en upphandling har tillgång till konkurrenternas interna telefonkataloger. Han kan utifrån dessa bedöma vilket hot de konkurrerande företagen utgör och kan knyta nödvändiga allianser med underleverantörer för att stärka sin egen position i upphandlingen.

Den elektroniska motsvarigheten till en intern telefonkatalog är kataloginformation över e-mail adresser, x-400 kataloger och liknande information. Dessa brukar vara öppna för att förenkla kommunikationen mellan företaget och dess kunder och leverantörer. Ett företags datanät kan också undersökas på domännivå. Det finns allmänt tillgänglig programvara som förenklar kartläggningen[7].

- `nslookup` - ger enkel interaktiv access mot en DNS-server (Domain Name Service.)[8]
- `dig` - "Domain Internet Grouper", är ett program som liknar `nslookup`.
- `doc` - "Domain Obscenity Control", är ett script som utnyttjar `dig`, för att felsöka och ställa diagnos på felaktiga DNS-uppsättningar.
- `host` - är inte lika avancerad i sin funktion som `nslookup` och `dig`, men räcker gott och väl för att hämta uppgifter från en DNS-server.

Om ett företags inre nät är öppet mot Internet på ICMP-nivå (Internet Control Message Protocol), så kan en utomstående med `traceroute`[7] och liknande funktioner bestämma antalet adresser och hur dessa fördelas över olika subnät. Indelningen i subnät kan typiskt ge indikationer om vilka datasystem som finns i olika byggnader. Denna information kan kompletteras med uppgifter från DNS-servrar och mailservrar, och därmed ge en klarare bild om nätet och företagets organisation.

6. Kartlägga användarnas intressen på Internet, dvs vilken information söker de mest.

I takt med att Internet kommersialiseras ökar behovet av att samla statistik över användningen av resurser. Detta gäller särskilt WWW-hotell, dvs att ett företag står för utrustningen där andra företag kan hyra in sig för att visa sitt material. Det blir samtidigt allt mer vanligt att dessa leverantörer säljer reklamutrymme för populära sidor och baser. På en sida som besöks ofta är det möjligt att sälja reklamplatser. WWW-serverar börjar idag uppvisa funktionaliteter där man kan följa en användares gång i servern och dessutom kunna se vart användaren går sedan. Denna information kan filtreras på personnivå, av debiteringsskäl, eller för att se vilken sida som är mest besökt. Idag betraktas denna funktionalitet som "state of the art", men tendensen är klart markerad.

Vad som kan betraktas som skrämmande ur ett BI-perspektiv är att uppgifter i loggar på ett effektivt sätt kan användas för att samla information om ett företag och dess anställda. Om t ex denna funktionalitet skulle införas i de populäraste sökdatasaserna, skulle informationen kunna säljas till tredje part eller utnyttjas i eget syfte. Användarna lämnar ut sin adress och vad de är intresserade av och detta skulle kunna användas för att kartlägga ett företag.

7. Delta i konkurrents mailinglistor, interna WWW-sidor mm under falskt namn eller identitet som disponeras via konkurrentens kunder.

Genom att ta del av mailinglistor samt interna WWW-sidor eller nyhetsgrupper kan mer information om ett företag erhållas. Möjligheten ges också att agera aktivt, dvs en möjlighet att ställa frågor inför konkurrentens övriga kunder, t.ex. "*när kommer funktionen x i era produkter den finns ju redan i produkterna y och z*". Normalt väljer man sannolikt att bara lyssna. Medlen kan vara falskt namn eller identitet, som disponeras via konkurrentens kunder.

Ibland tillåter ett företag att endast vissa domäner, t ex skolor eller det egna landet, att läsa vissa WWW-sidor, detta kan typiskt gälla examensarbeten eller känslig information som t ex berör kryptoteknik. I dessa fall kan en lämplig publik proxyserver användas eller en identitet som disponeras inom detta land för att förbija restriktionen.

8. Kartlägga konkurrentens interna datanät i samband med samarbetsprojekt eller via konsultarbeten.

Kartläggning av datanäten i samband med samarbetsprojekt eller via konsultarbeten kan syfta till att inventera nätverken, kopplingar till externa leverantörer och för att se vad som finns tillgängligt för att i nästa steg hämta önskad information. Normalt frågar en konsult efter vilken information han bör ta del av (produktinformation, processarbetsformer, administration) så denna utforskning behöver inte vara så besvärlig eller riskabel att göra.

Bland den viktigaste informationen finns marknadsanalyser och företagets strategier både på kort och lång sikt. Denna information finns normalt i databaser, fileservers för grupp-programvara och i rapporteringssystem. Oftast lagras informationen på flera ställen.

9. Kamp om domännamn och uttag av domännamn i förebyggande syfte.

Internet har hittills präglats av en kaotisk tillväxt, där det varit möjligt att ta ut domännamn relativt fritt. Detta har speciellt gällt så kallade .com adresser. Den vanligaste konkurrensen har gällt uttag av produkt eller märkesnamn, t.ex. `www.sportcars.com` som skulle kunna vara ett bra namn för en försäljare eller tillverkare av exklusiva sportbilar. Namnet har fördelen att vara lätt att komma ihåg och samtidigt utestänger man delvis sina konkurrenter.

Ett företag med en stor eller dominerande marknadsandel kan frestas att registrera det allmänna produktnamnet som ett eget domännamn. Det är också vanligt att sådana domännamn registreras i förebyggande syfte. Det anses kanske inte korrekt att använda domännamnet, men man vill inte riskera att en konkurrent eller organisation gör det.

En mer aggressiv variant är kamp om domännamn eller uttag av domännamn man tror sig kunna sälja i framtiden eller domännamn som skulle vålla skada om de kom till användning. Exempel på det senare är namn som anknyter till avregistrerade företags eller varunamn som fortfarande är allmänt kända.

En registrering av ett domännamn kan även utnyttjas för att spåra vilka som försöker hämta information från denna domän. Detta kan utnyttjas i BI-sammanhang för att hitta kundämnen.

10. Dataintrång och skadegörelse.

Målet kan vara att i det tysta skaffa tillgång till strategisk information, order eller offertuppgifter samt tekniskt kunnande. Skadegörelse av system kan användas för att vid ett lämpligt tillfälle skada en konkurrent, kanske i syfte att slå ut denne från en viktig marknad. En beskrivning av den metodik som kan användas finns i "Hur arbetar crackers?" [9].

Dock är inte stöld av information genom intrång i ett företags interna miljöer det enda hotet. Det finns t ex möjligheter att påverka informationens trafikvägar på Internet genom att manipulera routrar och DNS-servrar utanför företagets kontroll. Det innebär att det går att avlyssna trafik på Internet tillhörande ett visst företag, utan att göra några "fysiska" ingrepp.

5 Slutsats

Internet är ett relativt nytt fenomen och det tar tid innan alla möjligheter och konsekvenser kan inses. Business Intelligence är ett område som tillämpas i allt större omfattning på Internet och leder begreppet datorsäkerhet till en mycket mer komplicerad problemställning. Säkerheten på Internet kan inte lösas enbart med tekniska medel, utan kräver ett större ansvar och inblick från företagets ledningar så att både presentation och inhämtning av information sker på ett strukturerat sätt.

Det är framförallt viktigt att avgöra vem som hanterar dessa frågeställningar och vilka krav som ska ställas på vederbörande. Är det en datasäkerhetsfråga, en uppgift för IT-avdelningen och webmaster eller bara en ny uppgift för företagets traditionella informationsavdelning?

Referenser

- [1] *"INFORMATION SECURITY Computer Attacks at Department of Defense Pose Increasing Risks"*, United States General Accounting Office, May 1996.
- [2] *Föreläsningsskpendium i "Business Intelligence and Security"*, Stevan Dedijer, Department of Business Administration, School of Economics and Management Lund University.
- [3] *"Chemical Engineering: Article by Dr Worth Wade"*, 23rd May 1966, McGraw-Hill Inc.
- [4] *"The Business Intelligence System: A New Tool for Competitive Advantage"*, Benjamin Gilad, Tamar Gilad, AMACOM, American Management Association, 1988.
- [5] *"Brottslighet i det post-sovjetiska samhället - råd till svenska affärsmän och företag"*, Joakim von Braun, Näringslivets Beredskapsbyrå, Svenska Arbetsgivareföreningen, December 1994.
- [6] *"Changing Social Values and Their Implications for the Ethics of Information Warfare"*, Christine A.R. MacNulty, 31 July 1996. (www.infowar.com)
- [7] *"TCP/IP Illustrated Volume 1, The Protocols"*, Richard Stevens W, Addison-Wesley Publishing Company, 1994.
- [8] *"DNS and BIND"*, Albitz P, Liu C, O'Reilly and Associates, Sebastopol, California.
- [9] *"Hur arbetar crackers?"*, Mikael Simovits, Institute for International Research, Maj 1996. (Exemplar av detta dokument kan erhållas från författarna.)