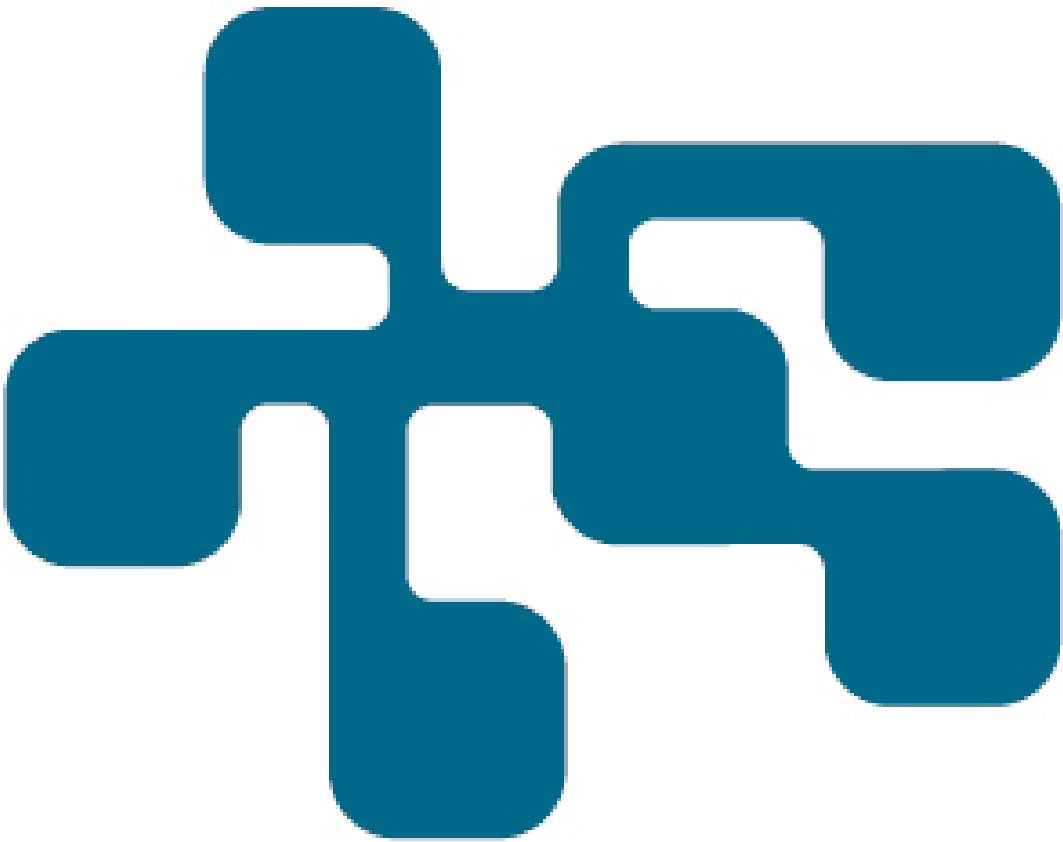
NCS3 Internetanslutna styrsystem i Sverige

En studie av Censys och Shodan

HANNES HOLM

FOI
MSB



FOI-R--4415--SE
MSB 2016-5134
ISSN 1650-1942

Mars 2017

Hannes Holm

NCS3: Internetanslutna styrsystem i Sverige

En studie av Censys och Shodan

Titel	NCS3: Internetanslutna styrsystem i Sverige
Title	NCS3: Industrial control systems in Sweden with Internet connections
Rapportnr/Report no	FOI-R--4415--SE
Månad/Month	Mars
Utgivningsår/Year	2017
Antal sidor/Pages	43
ISSN	1650-1942
Kund/Customer	MSB
Forskningsområde	4. Informationssäkerhet och kommunikation
FoT-område	
Projektnr/Project no	E72121
Godkänd av/Approved by	Christian Jönsson
Ansvarig avdelning	Ledningssystem

Sammanfattning

Samhällskritiska tjänster såsom elkraft, transport och sjukvård är beroende av informationsteknologier för att fungera effektivt. Om dessa teknologier är internetuppkopplade kan även aktörer utan större teknisk kunskap interagera med dem och därmed potentiellt kompromettera deras funktion. Denna rapport beskriver en studie av de teknologier som realiserar samhällskritiska tjänster som är internetanslutna i Sverige. Studien nyttjade databaserna Shodan och Censys för identifiering av internetanslutna system och geodata för kategorisering av dessa i sektorer som tillhandahåller samhällskritiska tjänster. Resultatet visar att det finns internetanslutna komponenter inom flera sektorer, och att elkraft är den klart mest exponerade sektorn. Totalt identifierades 47 komponenter som rörde industriella informations- och styrsystem, 9118 potentiella kontorssystem (varav cirka bedömdes vara 3 % interna) och 1967 potentiella kommunikationssystem.

Nyckelord: Samhällskritiska tjänster, kritiska infrastrukturer, industriella informations- och styrsystem, kartläggning, empiri

Summary

Critical infrastructures such as energy, transportation and healthcare are dependent on the support of various information technologies. If these technologies are connected to the Internet, even actors with minor technical knowledge are able to interact with them and thereby potentially compromise their functionality. This report describes a study of the extent to which technologies that support critical infrastructures in Sweden are connected to the Internet. The databases Shodan and Censys were used to identify systems with internet connections and geodata was employed to relate these systems to critical infrastructure sectors. The results show that systems in various sectors, especially the energy sector, are connected to the Internet. A total of 47 industrial information- and control system components, 9118 potential office system components (of which 3 % were judged as internal), as well as 1967 potential communication components, were identified.

Keywords: Critical infrastructure, industrial information and control systems, survey, empirical

Innehållsförteckning

Förkortningar

API	Application Programming Interface
BGP	Border Gateway Protocol
COTS	Commercial Off The Shelf
DNP3	Distributed Network Protocol
DNS	Domain Name System
DUC	Dataundercentral
FTP	File Transfer Protocol
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IANA	Internet Assigned Numbers Authority
ICMP	Internet Control Message Protocol
IED	Intelligent Electronic Device
IPSec	Internet Protocol Security
IRC	Internet Relay Chat
ISP	Interner Service Provider
NAT	Network Address Translation
NetBIOS	Network Basic Input/Output System
OSI	Open Systems Interconnection
PLC	Programmable Logical Controller
RTU	Remote Terminal Unit
REST	REpresentational State Transfer
SCADA	Supervisory Control and Data Aquisition
SIP	Session Initiation Protocol
SMB	Server Message Block
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
SOAP	Simple Object Access Protocol
SSH	Secure Shell
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol
UPS	Uninterruptable Power Supply
VPN	Virtual Private Network
XML	Extensible Markup Language

1 Inledning

Vårt samhälle är beroende av olika kritiska infrastrukturer, såsom elkraft, vattendistribution och sjukvård, för att fungera. De industriella informations- och styrsystem som nyttjas inom dessa tjänster var ursprungligen specialiserade komponenter som var fysiskt och logiskt avskilda från andra system. I samband med allt större krav på sänkta kostnader och ökad funktionalitet, exempelvis gällande övervakning och styrning, så har dock specialisering och isolering ersatts med generalisering och sammankoppling.

Från ett användarperspektiv medför nya funktioner och ökad tillgänglighet såklart stora fördelar. Problem kan dock uppstå då dessa egenskaper också förenklar möjligheten för hotaktörer att framgångsrikt tillämpa IT-angrepp. Detta är särskilt problematiskt om komponenten i fråga är nåbar från Internet och exempelvis styr över en brytare som kan slå av strömmen för hundratusentals individer.

Det finns idag globala kartläggningar av Internet som är publikt tillgängliga via webbsidor. Dessa gör att även aktörer utan större teknisk kunskap enkelt kan identifiera och interagera med internetanslutna kritiska system, vilket medför att även nyfikna aktörer utan elakartade intressen kan orsaka stora samhällsproblem.

Det är okänt i vilken omfattning industriella informations- och styrsystem i Sverige är anslutna till Internet. Detta arbete ämnar studera just denna fråga, nämligen:

> ***Fråga 1****: Vilka industriella informations- och styrsystem i Sverige är internetanslutna*?

Detta arbete ämnar dock inte enbart identifiera styrsystemsspecifika komponenter. Det försöker även identifiera vilka typer av komponenter som samhällskritiska verksamheter har exponerat mot Internet (se kapitel 2):

> ***Fråga 2****: Vilka komponenter inom svenska samhällskritiska verksamheter är internetanslutna?*

Denna frågeställning medför en stor skillnad mellan detta arbete och tidigare studier av samma ämne. Tidigare studier beskrivs i kapitel 3 och har huvudsakligen ämnat identifiera styrsystemskomponenter, men inte de internetanslutna system som på ett eller annat sätt är kopplade till dessa. Metoden bakom detta arbete beskrivs i kapitel 4. Kapitel 5 presenterar resultatet, och kapitel 6 konkluderar rapporten och diskuterar framtida arbete.

Studien hade två huvudsakliga begränsningar. Det ställdes inga direkta frågor till komponenter. Kartläggningsarbetet utfördes istället genom att fråga publika databaser som indexerar historiska kartläggningar av Internet. Det genomfördes heller inga försök att identifiera tekniska sårbarheter i kartlagda komponenter. Utöver dessa begränsningar var det inte inom ramen för studien att kontakta ägare och ansvariga för identifierade komponenter.

2 Industriella informations- och styrsystem

Industriella informations- och styrsystem kopplar samman den digitala och analoga världen, och kallas därför ofta för cyber-fysiska system. Exempel på tillämpningsområden är inom elkraft för att mäta ström och slå på eller av brytare, och inom transportsektorn för att kommunicera med tåg och säkerställa att de inte spårar ur eller krockar med något.

Detta kapitel sammanfattar industriella informations- och styrsystem i kontexten för denna studie, nämligen ”*Vilka industriella informations- och styrsystem i Sverige är internetanslutna?”*. Mer omfattande beskrivningar ges av exempelvis NIST 800-82 [1] och [2]–[6].

En enkel kategorisering av industriella informations- och styrsystem efter teknologier och accesskontroll ger fyra olika områden: *process*, *kontrollcenter*, *kommunikation* och *kontor*.

- **Processen** innefattar de teknologier som övervakar och styr fysiska maskiner. Exempel på sådana teknologier är Remote Terminal Units (RTU), Programmable Logical Controllers (PLC), Intelligent Electronic Devices (IED) och dataundercentraler (DUC). Dessa datorer har i regel specialskrivna uppgifter, är robusta mot väder och är väl testade gällande vanliga programvarufel. De är dock oftast inte härdade mot IT-attacker.
- **Kontrollcentret** involverar centraliserad loggning, övervakning och styrning av processen. Oftast innebär detta en kombination av datalagring, dataprocessande och mänsklig övervakning. Typiskt nyttjas vanliga commercial-off-the-shelf (COTS) produkter såsom Microsoft Windows och Linux Red Hat för detta ändamål. Supervisory Control and Data Aquisition (SCADA) är en vanlig benämning för system som används inom kontrollcentret.
- **Kommunikationen** innefattar de teknologier som nyttjas för att slussa data i ett informations- och styrsystem. Exempel på sådana teknologier är fältbussar[1], routrar, switchar och modem. I kontexten för denna rapport ses även teknologier som realiserar säker kommunikation, såsom IP Security (IPSec)[2] och Virtual Private Network (VPN)[3], som kommunikationsteknologier.

[1] Ett seriellt protokoll som kopplar sensorer till PLC:er.
[2] IPSec tillför autentisering och kryptering av IP-paket.
[3] VPN möjliggör för datorer som befinner sig på publika nätverk att koppla upp sig till varandra som om de satt på ett eget privat nätverk.

- **Kontoret** handlar om ”vanliga” teknologier som nyttjas i kontorsmiljöer. Normalt finns det någon eller några kopplingar mellan system i kontoret och SCADA-systemet i kontrollcentret för överföring av data eller fjärrkommandon. Kontoret kan delas upp i *externa* och *interna* system. Externa system förmedlar en tjänst till Internet, t.ex. en publik filserver, mailserver eller webbserver. Ett internt system är exempelvis en kontorsdator eller en skrivare.

Rörande internetanslutningar så är det rimligt att anta att kontoret har någon typ av koppling mot Internet, t.ex. för att kunna surfa eller skicka e-post. Endast vissa tjänster på externa system bör dock vara direkt nåbara från Internet. Interna system, såsom kontorsdatorer, bör i regel enbart kunna initiera sessioner mot Internet och inte tvärtom, då detta skulle utgöra en stor säkerhetsrisk. Med andra ord, en typisk kontorsdator bör inte kunna identifieras via en kartläggning av Internet. På samma sätt bör komponenter inom process och kontrollcenter i regel vara helt isolerade från Internet då de inte har behov av en sådan anslutning för att fungera. Vissa tjänster inom kommunikation måste vara internetanslutna för att realisera den; andra inte. Dessa hypoteser sammanfattas av Tabell 1 och denna studie ämnar utvärdera hur ofta komponenter inom dessa områden kan nås från Internet.

Tabell 1. Åtkomst från Internet av olika typer av komponenter.

Område	Bör kunna nås från Internet
Process	Nej
Kontrollcenter	Nej
Kommunikation	Ja (vissa tjänster)
Kontor (externa system)	Ja (vissa tjänster)
Kontor (interna system)	Nej

3 Kartläggning av internetanslutna styrsystem

Det finns två huvudsakliga metoder för att kartlägga internetanslutna datorer. Det första alternativet är att aktivt sända nätverkspaket mot olika IP-adresser och portar (se avsnitt 3.1). Fördelarna med denna metod är att det ger ”färsk” data, samt att frågorna enkelt kan anpassas för olika ändamål. En stor nackdel är dock att det inte på förhand är känt vilka datorer som är intressanta, vilket medför att oerhört många datorer som är ointressanta för den aktuella forskningsfrågan ändå måste kartläggas. Att ställa dessa frågor är problematiskt eftersom de skapar en belastning för de utsatta nätverken och datorerna; vissa känsliga komponenter såsom PLC:er kan till och med skadas av dem [7]. Dessutom kan skanning uppfattas som kränkande, vilket gör användning av det etiskt diskutabelt.

Det andra alternativet är att ställa frågor till en databas som innefattar resultatet från utförda aktiva skanningar av andra aktörer (se avsnitt 3.2). Fördelen med denna metod är att forskaren personligen inte behöver ställa frågor direkt mot komponenter. Nackdelen är att forskaren måste leva med de begränsningar som finns i den nyttjade databasen. En annan nackdel gentemot aktiv kartläggning är att potentiellt känsliga söktermer röjs [8].

Som komplettering bör det nämnas att det även är teoretiskt möjligt att passivt lyssna på nätverkstrafik och analysera den för att identifiera olika typer av datorer och applikationer [9]. Det vill säga, att inspektera nätverkstrafik utan att de inblandade systemen märker det. Eftersom inga frågor ställs så finns det heller ingen risk att känsliga system störs ut. Det stora problemet med denna metod, utöver att den kan ses som mycket etiskt diskutabel, är att det krävs access till antingen någon av de kommunicerande datorerna eller någon komponent som vidarebefordrar deras trafik (t.ex. en switch, router eller brandvägg).

Empiriska resultat som erhållits via aktiva kartläggningar eller databassökningar beskrivs i avsnitt 3.3.

3.1 Publika verktyg för aktiva kartläggningar

Det finns ett stort antal olika publika verktyg som kan nyttjas för att aktivt kartlägga internetanslutna datorsystem. Detta avsnitt fokuserar på de idag mest aktivt nyttjade verktygen för global kartläggning av internetanslutna system.

Det kanske allra vanligaste verktyget, och det de flesta nya kartläggningsverktyg nyttjar som referensdesign, är *Nmap*. Basfunktionen för Nmap är att skicka frågor till portar på IP-adresser. Nmap innefattar ett stort antal olika färdiga moduler som möjliggör protokollanpassade frågor till olika portar och tjänster, samt tolkning av

erhållna svar. Forskare har över lag inte försökt förbättra korrektheten i Nmap, utan istället försökt erhålla samma datakvalitet fast till en kortare kartläggningstid.

Durumeric m.fl. [10] föreslår ett verktyg för kartläggning av IPv4 kallat *ZMap*. Den största skillnaden gentemot tidigare verktyg såsom Nmap är att ZMap är mycket snabbare, utan att för den sakens skull påverka resultatens reliabilitet eller validitet. Exempelvis kartlade Durumeric m.fl. hela IPv4-adressrymden från en enda dator inom loppet av 45 minuter, vilket gör ZMap 1300 gånger snabbare än Nmap. Huvudanledningarna till att ZMap kan vara så mycket snabbare utan att kompromissa med resultaten är följande:

- Det specialkonstruerar paket.[4]
- Det sparar inte statusen för initierade uppkopplingar, utan skickar ut paket till en förutbestämd mängd adresser så fort det går utan att vänta på svar. ZMap hanterar förlorade paket genom att skicka upprepade frågor mot samma mål.

Masscan[5] [11] är ett annat publikt verktyg för globala kartläggningar av internetanslutna system. Masscan producerar resultat lika de från Nmap och ZMap och påstås av utvecklarna vara världens snabbaste verktyg för kartläggningar av Internet. En jämförelse av verktygen genomfördes av Myers m.fl. [12]. De huvudsakliga skillnaderna mellan Masscan och ZMap som identifierades av författarna var:

- ZMap erbjuder möjlighet att tillämpa både svartlistor och vitlistor; Masscan endast svartlistor.
- ZMap är till skillnad från Masscan modulärt[6].
- Med Masscan kan hastigheten för kartläggningen enbart begränsas via en parameter (paket per sekund). Med ZMap kan hastigheten även begränsas baserat på tid eller som ett utfall av observerade resultat.

3.2 Publika databaser med resultat från utförda kartläggningar

Den idag mest kända databasen för information om internetanslutna datorer är Shodan[7]. Shodan fungerar lite som Google – både Shodan och Google samlar aktivt in information från internetanslutna datorer. Den huvudsakliga skillnaden mellan Shodan och Google är att Google primärt indexerar webbsidor och deras

[4] Det nyttjar inte TCP/IP-stacken utan genererar paketramar direkt.
[5] https://github.com/robertdavidgraham/masscan
[6] ZMap är byggt för att det skall vara enkelt att ta bort/lägga till egna moduler.
[7] https://www.shodan.io/

innehåll, medan Shodan indexerar data om allehanda serverapplikationer. Forskning har visat att Shodan kartlägger en komponent inom fyra dagar efter att den blivit internetansluten [13].

Censys[8] [14] är en nyare sökmotor som likt Shodan genomför aktiva sökningar av IPv4 och sedan indexerar identifierade banners i en publik databas. Censys fungerar från en användares perspektiv likartat Shodan. Shodan och Censys nyttjades av denna studie, och beskrivs mer ingående i avsnitt 4.1.

Utöver Shodan och Censys finns även den kinesiska tjänsten ZoomEye[9], som innefattar liknande datamaterial. ZoomEye identifierades tyvärr för sent under projektets gång för att kunna inkluderas, men vore spännande att nyttja för framtida studier.

Utöver de tre ovan nämnda tjänsterna publicerades resultatet från en global studie av internetanslutna system av en eller flera anonyma individer under 2013 [15]. Dessa individer inspirerades av studierna utförda av Heidemann m.fl. [16] och genomförde flera kartläggningar av hela IPv4 under sex veckors kalendertid. Kartläggningarna utfördes med hjälp av Nmap konfigurerat med alla funktioner aktiverade (frågor till 632 TCP-portar och 110 UDP-portar), där varje kartläggning utfördes inom loppet av 24 timmar. De snabba genomsökning realiserades genom användning av ett botnät[10] kallat *Carna* som skapades av författarna genom utnyttjande av default-lösenord i telnet[11] på internetanslutna datorer. Studiens resultat validerades senare av Krenc m.fl. [17]. Denna validering visade att studien var äkta, men att det fanns en del metodproblem som påverkade dess resultat. Exempelvis stämde inte siffrorna i rapporten med det publicerade rådatat, och kartläggning av vissa IP-adresser favoriserades före andra (i verkligheten skannades vissa adresser en enda gång, och andra uppåt 600 gånger, snarare än de 13 gånger/adress som nämns i rapporten). De metodmässiga problemen i kombination med studiens ålder medförde att resultatet från [15] inte nyttjades av denna studie.

3.3 Empiriska studier av internetanslutna system

Detta avsnitt behandlar de empiriska studier som hittills genomförts av internetanslutna datorer, och fokuserar särskilt på studier av internetanslutna styrsystem. Dessa studier kan kategoriseras i två huvudsakliga klasser – studier som utfört aktiva kartläggningar (se avsnitt 3.3.1) och studier som studerat publika

[8] https://censys.io/
[9] https://www.zoomeye.org/
[10] En samling kompromitterade datorer som kan övervakas och styras av hotaktören.
[11] En applikation som erbjuder fjärraccess till en dator.

databaser (se avsnitt 3.3.2). Ingen av dessa studier har försökt identifiera datorer som har indirekta kopplingar till styrsystem, såsom vanliga kontorsdatorer med fjärraccess till SCADA. Denna frågeställning, ”*Vilka komponenter inom svenska samhällskritiska verksamheter är internetanslutna?”*, är ett fokusområde för denna studie.

3.3.1 Studier som genomfört aktiva kartläggningar

Heidemann m.fl. [16] utförde under åren 2003 till 2007 kartläggningar av hela IPv4-adressrymden och utförde upprepade sökningar mot en delmängd av alla identifierade adresser under 2006 och 2007. Frågor ställdes med hjälp av echo-requests, vanligen benämnt ”ping”, över ICMP eller TCP. Författarna kartlade därmed enbart om olika datorer svarade eller inte; utnyttjandet av olika portar eller metadata från de svar som erhölls studerades ej. Kartläggningarna tog mellan 24 och 191 dagar, medan de upprepade sökningarna gjordes på mellan 6 och 12 dagar. Författarna observerade att endast 3,6 % av alla allokerade adresser faktiskt var ockuperade av synliga datorer, och att fördelningen över adressutrymmet var högst ojämnt. Författarna observerade också att cirka 16 % av alla aktiva adresser hade mycket få driftstopp.

Durumeric m.fl. [10] exemplifierade ZMap:s funktionalitet genom att undersöka olika IT-säkerhetsproblem som tidigare inte studerats i stor skala. Dessa undersökningar inkluderade: (1) studier av hur ofta Transport Layer Security (TLS) certifikat som används för säker kommunikation på Internet är felaktigt utgivna, (2) studier rörande vilken omfattning HTTPS används, (3) identifiering av sårbara maskiner, (4) identifiering av icke-annonserade tjänster (t.ex. Tor-bryggor), samt (5) studier av IT-avbrott.

Nancy m.fl. [18] föreslår ett verktyg för kartläggning av SCADA-system vilken de kallar *WiScan*. WiScan bygger på ZMap, men nyttjar en annan algoritm för att bestämma i vilken ordning olika IP-adresser skall skannas[12]. De sökte igenom hela IPv4-adressrymden under 33 dagar med hjälp av åtta virtuella WiScan-maskiner. Totalt identifierades cirka 19,7 miljoner unika IPv4-adresser. SCADA-komponenter antogs vara komponenter som exponerade någon av portarna 102 eller 502 (dock ej i kombination)[13]. Cirka 3,5 miljoner sådana träffar erhölls, varav de flesta fanns i USA (1,5 miljoner), Kina (0,6 miljoner) och Belgien (0,6 miljoner). Sverige var inte ens med på listan över de 25 länderna med flest exponerade SCADA-komponenter. Författarna noterade även att de erhållna resultaten väl överlappade med de Shodan och Censys identifierade för samma sökrymd.

[12] Ordningen spelar roll eftersom att en intensiv frågefrekvens mot ”närliggande” adresser kan göra att skannern blir svartlistad.

[13] Dessa antogs röra antingen SCADA-komponenter från Siemens eller Schneider Electrics.

Tiilikainen [8] byggde ett kartläggningsverktyg kallat KATSE och använde detta för att kartlägga styrsystem i Finland. KATSE nyttjade Nmap, ett beslut som dels baserades på att extrem snabbhet inte bedömdes som viktigt och dels på att Nmap täcker in fler protokoll och funktioner. Författaren nyttjade sedan geodatabasen MaxMind[14] för att relatera identifierade IPv4 adresser till geografiska platser. Identifiering av styrsystemskomponenter genomfördes med hjälp av en databas med styrsystemsinformation kompilerad från källor såsom Metasploit och Shodan. Totalt identifierades 91 komponenter som bedömdes röra industriella informations- och styrsystem.

Williams [19] nyttjade Shodan för att identifiera internetanslutna PLC:er av typen Allen-Bradley RSLogix 5000. Varje identifierad PLC skickades anpassade frågor för att identifiera deras användningsområden. Identifiering gjordes huvudsakligen genom manuella analyser av två experter. Av 154 identifierade PLC:er lyckades experterna kategorisera 73 % till specifika sektorer. Den absolut största sektorn bedömdes vara avloppsvatten (32 % av alla PLC:er).

3.3.2 Studier av databaser

Leverett [20] analyserade internetanslutna styrsystem mellan 2009 och 2011 (cirka två kalenderår) med hjälp av Shodan. Författaren identifierade mer än 7500 styrsystemskomponenter, såsom SCADA-servrar, byggnadsautomationssystem och PLCer. Av dessa komponenter bedömdes 442 ha funnits i Sverige. Utöver att identifiera styrsystem korrelerade Leverett dessutom resultatet mot ExploitDB, en databas med färdiga attackkoder.

Projekt SHINE (SHodan INtelligence Extraction) [21] analyserade, likt studien av Leverett, internetanslutna styrsystem genom Shodan. SHINE pågick mellan april 2012 och oktober 2014 och innefattade en relativt omfattande lista med söktermer (886 söktermer jämfört med de 29 som Leverett nyttjade). Bland annat försökte SHINE inte enbart identifiera direkta styrsystemskomponenter såsom SCADA-servrar, utan även supportsystem för SCADA såsom enheter för avbrottsfri kraft och vissa typer av switchar. Totalt identifierades cirka 2,1 miljoner komponenter.

Populärvetenskapliga studier av Shodan lika de av Leverett och SHINE har genomförts av diverse journalister och bolag, såsom Shodan självt [22], Kaspersky [23], Trend Micro [24] och TripWire [25]. Resultaten från dessa pekar i liknande riktning som resultaten från Leverett och SHINE.

Testaaja [26] studerade resultaten från den anonyma kartläggningen av Internet som genomfördes under 2013 [15] med ett fokus på Finlands adressrymd. Författaren identifierade bland annat över 2900 datorer som bedömdes vara del av

[14] https://www.maxmind.com/

SCADA-system. Likt de andra studierna av internetanslutna styrsystem inkluderar dessa dock inte datorer som har indirekta kopplingar till SCADA.

4 Datainsamlingsmetod

Detta kapitel beskriver den datainsamlingsmetod som nyttjades för att svara på forskningsfrågorna ”*Vilka industriella informations- och styrsystem i Sverige är internetanslutna?*” samt ”*Vilka komponenter inom svenska samhällskritiska verksamheter är internetanslutna?*”. Avsnitt 4.1 beskriver vilka datakällor som nyttjades och avsnitt 4.2 hur dessa användes. Avsnitt 4.3 presenterar studiens metodmässiga begränsningar.

4.1 Datakällor

Detta projekt utnyttjade tre huvudsakliga datakällor för att analysera internetanslutna styrsystem: Shodan, Censys och IP-API.

4.1.1 Shodan

Shodan lanserades 2009 för att möjliggöra enkla analyser av internetanslutna system och har idag vuxit till en tjänst som mer än häften av alla Fortune500-företag betalar för att använda. Shodan indexerar mer än en miljard sökningar varje månad och under november 2016 kartlade Shodan enligt dess application programming interface (API) 257 olika portar. Denna studies resultat visade dock att långt fler portar kartläggs i verkligheten, då totalt 366 unika portar enbart i Sverige identifierades med hjälp av Shodan (se kapitel 0). Flera av dessa portar rör specifika SCADA-protokoll såsom Modbus (blandade system), Siemens Step7 (blandade system), DNP3 (främst elkraft och vattendistribution), BACnet (främst byggnadsautomation) och IEC 60870-5-104 (främst elkraft). Dessa kartläggs med hjälp av specifika protokollfunktioner (totalt 151 varianter). Modbus kartläggs exempelvis via Modbus-protokollets funktion 17 och 43 [27]. Det är dock inte känt för författaren vilka faktiska teknologier (såsom Nmap) som Shodan nyttjar för kartläggningar.

Shodan är åtkomligt antingen via en webbsida, ett representational state transfer (REST) API och ett streaming API. Skillnaden som finns mellan dessa metoder är att det via API:erna även går att söka igenom historisk data, exempelvis hur en port svarat vid olika kartläggningar. Skillnaden mellan REST och ”streaming” är att det senare erhåller realtidsdata (”rådata”) direkt från aktiva kartläggningar, medan det tidigare rör behandlad data i Shodans databas. Det är också möjligt att mot en kostnad nyttja Shodan för aktiva kartläggningar.

Det går att nyttja en delmängd av Shodans funktionalitet helt gratis. För att få access till alla sökfilter, samt för att genomföra mer omfattande sökningar, krävs det dock ett abonnemang. Studien nyttjade ett abonnemang som medgav komplett täckning av sökfilter.

4.1.2 Censys

Censys lanserades under 2015 av utvecklarna av ZMap och samlar med hjälp av ZMap in en delmängd av informationen som Shodan innefattar. Totalt 19 portar/protokoll kartläggs, varav fem rör styrsystemsutrustning (Modbus, BACnet, Niagara Fox, DNP3 och Siemens S7). En annan skillnad mellan Censys och Shodan är att det finns en koppling till sårbarheter i Shodan, vilket det inte gör i Censys. Precisionen gällande identifierade sårbarheter är dock tveksamt i Shodan, då även kommersiella sårbarhetsskanners har relativt låg precision [28].

Till skillnad från Shodan är Censys helt gratis. Det krävs dock att ett användarkonto registreras för att kunna genomföra fler än fem sökningar per dag, något som utvecklarna lagt till för att minska missbruk av tjänsten. Censys är, likt Shodan, åtkomligt antingen via en webbsida eller ett REST-API[15]. Skillnaden mellan dessa ligger likt Shodan i sökningar av historisk data.

4.1.3 Geodata via IP-API

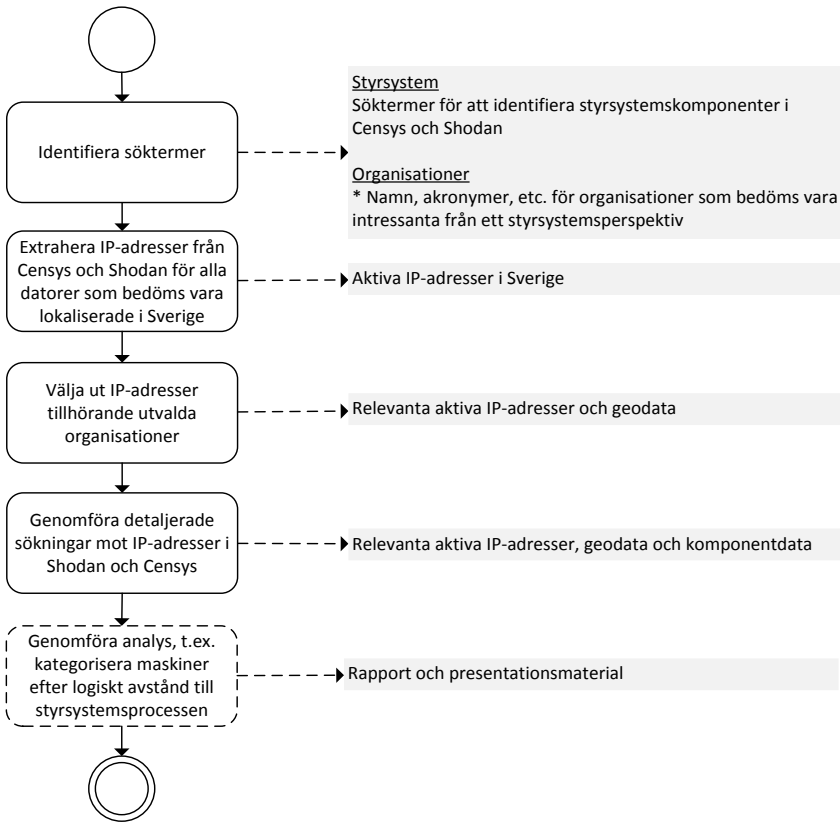
Utöver att identifiera internetanslutna datorer i Sverige och deras tekniska egenskaper var en viktig del för projektet att identifiera ägare/ansvariga för dessa system, samt vart de kan tänkas finnas rent geografiskt. Dessa informationskällor kan tillsammans potentiellt sett nyttjas för att identifiera vilket syfte en dator faktiskt har inom en särskild verksamhet, samt hur kritisk den är för infrastrukturen ifråga. Den argumenterbart mest korrekta publika informationen av detta slag tillhandahålls av så kallade geodatabaser. Geodatabaser baserar sina resultat på information från de regionala internetregister som allokerar och distribuerar IP-adresser till organisationer, t.ex. RIPE Network Coordination Centre[16]. De korrelerar sedan denna information med diverse olika andra datakällor, och försöker även korrigera felaktigheter genom en blandning av algoritmer och manuellt arbete.

Censys använder geodatabasen *MaxMind GeoLite2*, vilken tyvärr inte var tillräcklig för denna studie eftersom den har mycket begränsad information om organisationer. För att identifiera organisationer bakom adresser kartlagda av Censys användes geodatabasen IP-API[17]. IP-API erbjuder ett REST API som enkelt kan nyttjas för att identifiera organisationsinformation för IP-adresser. Shodan innefattar geodata som bedömdes fullgod för studiens syfte, dock av okänd härkomst. Inga sökningar mot IP-API gjordes därför för IP-adresser identifierade via Shodan.

[15] https://censys.io/api
[16] https://www.ripe.net/
[17] http://ip-api.com

4.2 Metod

En översikt av metoden beskrivs i Figur 1. Först identifierades söktermer rörande organisationer och teknologier (se avsnitt 4.2.1). Sedan extraherades alla IPv4-adresser som kartlagts av Censys och Shodan i Sverige (se avsnitt 4.2.2). Adresser från Censys förädlades med geodata från IP-API (se avsnitt 4.2.3). De adresser som rörde organisationer som matchade mot söktermerna kartlades genom specifika sökningar i Censys och Shodan (se avsnitt 4.2.4).



Figur 1. Översikt av metod.

Resultatet analyserades sedan med hjälp av de tekniska söktermerna. Processen är inbyggd i en applikation vilken möjliggör exekvering av hela händelsekedjan genom en knapptryckning. Applikationen ifråga är byggd i Python och omfattar MongoDB och ett REST API vilket används för att styra den. Genom att köra applikationen många gånger över tid är det möjligt att analysera internetanslutna datorer inom kritiska infrastrukturer i Sverige över tiden. Ett sådant arbete är dock utanför denna studies omfattning, vilket enbart omfattar de datorer i Sverige som var internetanslutna under början av december 2016.

4.2.1 Urval av söktermer

Det första steget inför datainsamlingen var att ta fram lämpliga söktermer för identifiering av teknologier samt organisationer som ansvarar för kritiska infrastrukturer involverande industriella informations- och styrsystem.

För identifiering av teknologier nyttjades en kombination av de kända söktermer som tidigare forskningsstudier nyttjat och publicerat (se kapitel 3), då särskilt arbetena av SCADAStrangeLove [29] (39 söktermer) Leverett [20] (30 söktermer), SHINE[18] (6 söktermer), Tiilikainen [8] (6 söktermer) och Williams [19] (1 sökterm). Dessa kombinerades med Shodans officiella söktermer för SCADA och ICS (14 söktermer), samt de inofficiella söktermer för SCADA och ICS som delats av användare av Shodan (63 termer). Ytterligare 45 söktermer tillfördes av författaren efter manuella studier av det resulterande datamaterialet. Den manuella studien resulterade också i ett flertal revideringar av de söktermer som andra tagit fram. Totalt nyttjades 204 unika söktermer för teknologier. En översikt av dessa söktermer presenteras i Tabell 2.

Tabell 2. Översikt av tekniska söktermer.

Område	Söktermer
Process	132
Kommunikation	15
Kontrollcenter	25
Kontor (externa system)	24
Kontor (interna system)	8
Summa	*204*

[18] Härledda baserat på projektets slutrapport.

Gällande organisationer utgick söktermsframtagningen från de av MSB beskrivna sektorerna för samhällskritisk verksamhet, med fokus på elkraft, transport, sjukvård samt vatten och avlopp.

- **Elkraft** inkluderar de aktörer som producerar eller distribuerar elektricitet. Inom elkraft används industriella informations och styrsystem exempelvis för att mäta ström och stänga av/på industriella strömbrytare.
- **Transport** inkluderar de aktörer som äger och underhåller infrastrukturer som realiserar transportsystem, såsom väg och järnväg. Inom transport används industriella- informations och styrsystem exempelvis för växling av järnvägsspår, kommunikation med tåg och styrning av trafikinformationsskyltar.
- **Sjukvård** inkluderar de sjukhus som finns i Sverige. Aktörer som enbart sysslar med läkemedelsproduktion är därmed inte inkluderade. Inom sjukvård används industriella- informations och styrsystem exempelvis för operationsrobotar, lungventilatorer samt hjärtstimulatorer.
- **Vatten och avlopp** inkluderar de aktörer som arbetar med vattenrening och distribution. Inom vatten och avlopp används industriella-informations och styrsystem exempelvis för att styra vattenpumpar samt kontrollera pH-värdet i vattenbassänger.

Då det inte fanns någon komplett lista med relevanta organisationer inom dessa sektorer, eller lämpliga böjningar av deras namn, genomfördes ett manuellt arbete för att identifiera sådana. Detta arbete nyttjade katalogtjänster såsom hitta[19], eniro[20] och allabolag[21], samt IP-API. Totalt identifierades 92 söktermer rörande organisationer. Av dessa var 22 generella söktermer kopplade till de olika sektorerna. Exempelvis innefattade sektorn elkraft generella söktermer såsom ”elkraft”, ”energi” och ”power”.

4.2.2 Aktiva IP-adresser i Sverige

Tidigare forskning har påpekat att Censys och Shodan till stor del överlappar gällande kartlagda IP-adresser [18]. Som grund i detta genomfördes först en analys av IP-adresser insamlade med Censys som bas. Analysen av detta resultat visade dock att Censys och Shodan inte överlappade på ett tillfredsställande sätt, vilket medförde att resultatet kompletterades med IP-adresser insamlade med Shodan.

Under denna studie fanns det 1 156 915 IP-adresser i Censys senaste kartläggning av Sverige. Motsvarande antal adresser i Shodan för samma tidpunkt var 1 161

[19] www.hitta.se
[20] www.eniro.se
[21] www.allabolag.se

274. Dataextraktionen från Censys tog cirka två timmar; från Shodan tog samma insamling cirka sju dygn. Tidsskillnaden berodde på att det gick att ställa fler frågor per sekund till Censys, och att varje svar från Censys returnerade tio gånger fler IP-adresser. Frågor mot Censys kan dessutom skräddarsys för att returnera mer precis data (t.ex. enbart IP-adresser).

4.2.3 Urval av IP-adresser

De av Censys identifierade 1.2 miljoner IP-adresserna i Sverige förädlades med geodata från IP-API. Detta steg tog cirka fem och ett halvt dygn att genomföra. Shodan innefattar fullgod organisationsinformation, vilket innebar att IP-API inte behövdes för dessa IP-adresser. Av de identifierade IP-adresserna föll 18 421 ut som relevanta enligt organisationssöktermerna. Dessa adresser rörde 81 olika organisationer.

De identifierade organisationerna granskades manuellt för att ta bort irrelevanta poster som råkat följa med på grund av de mycket inkluderande söktermerna. Av de 81 organisationerna och 18 421 IP-adresserna bedömdes 40 organisationer som relevanta. Dessa 40 organisationer omfattade 13 449 IP-adresser. En översikt ges i Tabell 3. Som kan ses rörde de allra flesta IP-adresserna organisationer verksamma som distribuerade eller producerade elkraft. Detta betyder dock inte att aktörer inom elkraftssektorn är särskilt sårbara, eller att aktörer inom vatten och avlopp är särskilt duktiga på att konfigurera sina system. Det är sannolikt så att aktörer som inte arbetar med elkraft oftare har en extern internetleverantör (Internet Service Provider [ISP]), såsom Telia eller Bredbandsbolaget, som står bakom sina IP-adresser. Då sökningar mot IP-API för dessa adresser därmed inte visar på kritiska infrastrukturer är de avgränsade från denna studies analys. Aktörer inom elkraftssektorn agerar ibland även själva ISP, vilket medför att de hanterar adresser för externa abonnenter. Detta diskuteras mer i avsnitt 4.3.

Tabell 3. Översikt av studerade sektorer och organisationer.

Sektor	Organisationer	IP-adresser
Elkraft	33	12695
Sjukvård	3	591
Transport	3	160
Vatten och avlopp	1	3
Totalt	*40*	*13449*

4.2.4 Insamling av konfigurationsinformation

De 13 449 IP-adresserna som rörde organisationer som bedömts som relevanta genomgick detaljerad datainsamling angående konfigurationsinformation i

Shodan och Censys. Totalt sett erhölls det konfigurationsinformation om 11 701 IP-adresser. Ett kort utdrag från en IP-adress som kartlagts enligt denna studies metod beskrivs av Figur 2. Denna datainsamling tog cirka 2 dygn, mycket tack vare begränsningar i Shodan.


```
{
    "ip": "127.0.0.1",
    "scans": [
        {
            "Censys": {
                "443": {
                    "https": {
                        […]
                    }
                },
                […]
            },
            "IP-API": {
                "org": "AB Elkraftsproducent",
                […]
            },
            "Shodan": {
                "ports": [
                    1723
                ],
                […]
            },
            "date": "2016-12-02 12:10:11"
        }
    ]
}
```


Figur 2. Exempel på kartlagd IP-adress.

Flest adresser och träffar erhölls av Censys, men som visas i Tabell 4 var överlappet mellan databaserna förvånande lågt. Censys och Shodan var enbart överens om 35 % av alla IP-adresser och delade enbart konfigurationsinformation om 38 % av dessa adresser. Särskilt förvånande var att det inte fanns någon konfigurationsinformation i någon av databaserna för 13 % av de IP-adresser som påstods finnas i Censys eller Shodan. Det ställdes upprepade frågor mot dessa IP-adresser över loppet av en kalendervecka för att bekräfta dessa problem. Det skickades också e-post till skaparna av både Censys och Shodan men inga svar erhölls.

Tabell 4. Överlapp mellan Censys och Shodan.

	Källa till IP-adress			
Träff	**Censys och Shodan**	**Censys**	**Shodan**	***Summa***
Censys och Shodan	4156	834	79	*5069*
Censys	273	3719	35	*4027*
Shodan	215	64	2326	*2605*
Ingen träff	14	954	780	*1748*
Summa	*4658*	*5571*	*3220*	*13449*

4.3 Begränsningar

Det finns flera faktorer som begränsar resultatet beskrivet i denna rapport. Detta avsnitt belyser de största begränsningarna.

Många organisationer nyttjar teknologin Network Address Translation (NAT) för att exponera många datorer mot Internet via få publika IP-adresser. NAT medför att en fråga mot samma IP-adress och port kan få olika svar beroende på när frågan ställs, eftersom olika tjänster/datorer kan sitta bakom samma publicerade port vid olika tillfällen.

Vissa organisationer har IP-adresser som rör helt olika tjänster bakom samma organisationsdetaljer. Exempelvis kan en organisation både distribuera ström och bredband och koppla IP-adresserna bakom dessa tjänster till samma namn. I detta fall är det inte möjligt att sålla ut om en identifierad Windows-dator ägs av organisationen eller om den ägs av en individ som är abonnent av organisationens bredbandstjänst. Den manuella analysen av datamaterialet identifierade ett antal organisationer där det var särskilt svårt att särskilja på organisationen och dess kunder. Dessa exkluderades från analysen (se avsnitt 4.2.3).

Shodan och Censys täcker enbart in en delmängd av alla protokoll som finns (och är relevanta). Vid denna studies utförande täckte Censys in 19 portar/protokoll, och Shodan 257 portar[22] och 151 protokoll (samma protokoll testades på flera portar). Av protokollen i Shodan rörde 41 styrsystemsutrustning, givet en generös tolkning av styrsystem. Exempelvis inkluderade denna siffra protokollet *Wemo Link*, vilket nyttjas för att styra belysning i ”smarta hem”. Detta är långt färre än de 200 styrsystemsprotokoll som nämns av American Gas Association [30]. Täckningen av protokoll verkar dessutom hårt vinklad mot de som används i USA. Många av de styrsystemsprotokoll som är vanliga i Sverige stöds inte alls, såsom

[22] Enligt Shodan:s API. I verkligheten kartläggs fler portar. Exempelvis identifierade denna studie 376 portar enbart i Sverige. Det är oklart varför inte API:t stämmer överens med de faktiska kartlagda portarna.

de protokoll som utvecklats av *Cactus*[23]. Detta medför att det sannolikt finns styrsystemsutrustning i Sverige som helt enkelt inte kan identifieras på grund av begränsningar i databaserna.

Att basera kopplingen mellan en organisation och en IP-adress på en geodatabas såsom IP-API är inte helt precist. Exempelvis kan en internetansluten dator stå i ett kraftverk medan dess externa IP-adress är skriven på en extern organisation (såsom Telia), vilken sedermera är beskriven för adressen i geodatabaser. Det är okänt hur ofta detta är fallet för geodatabaser generellt och IP-API i synnerhet. Det är dock möjligt att jämföra hur mycket styrsystemsspecifik utrustning som baserat på geodata fanns i Sverige som helhet jämfört med vad som kunde härledas till specifika organisationer och därmed analyserades av denna studie. En jämförelse av träffar för särskilt vanliga sådana protokoll som kartläggs av både Censys och Shodan presenteras i Tabell 5. Som kan ses var det endast möjligt att härleda en liten delmängd (0,7 %) av alla träffar till specifika organisationer som arbetar med kritiska infrastrukturer.

Tabell 5. Träffar i Sverige som kunde härledas till specifika organisationer.

		Totalt i Sverige		Kritisk infrastruktur	
Protokoll	**Port**	**Censys**	**Shodan**	**Censys**	**Shodan**
Modbus	502	980	426	8	2
BACnet	47808	159	69	1	0
Niagara Fox	1911	101	99	0	0
Siemens S7	102	71	27	2	1
Summa	-	*1311*	*621*	*11*	*3*

Slutligen är studiens resultat enbart relevant för en tidpunkt, samt beroende av att rätt söktermer valts ut. Att välja ut bra söktermer är en icke-trivial process då det inte finns något facit, och olika forskare har kommit fram till olika listor. Många studier, som projekt SHINE, har dessutom hemlighållit söktermer. Det är rimligt att anta att endast en liten delmängd av alla relevanta söktermer identifierades av studien beskriven i denna rapport.

[23] http://cactus.se/

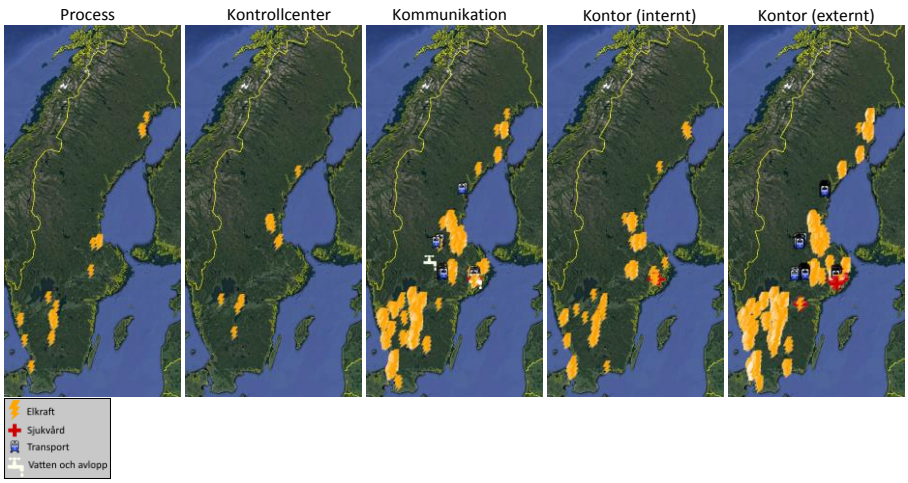
5 Resultat och analys

En översikt av resultatet beskrivs i Tabell 6. Totalt kunde 11 132 av 11 701 IP-adresser (95 %) kategoriseras enligt studiens tekniska söktermer. De allra flesta träffar rörde elkraftssektorn (94 %) samt externa kontorssystem (80 %). Det var också endast inom elkraftssektorn som det detekterades komponenter inom processen och kontrollcenter. Som beskrivet i avsnitt 4.2.3 betyder detta dock inte att aktörer inom elkraftssektorn är särskilt sårbara, utan snarare att de oftare själva står bakom sina IP-adresser istället för en extern internetleverantör.

Tabell 6. Översikt av resultat.

	Sektor				
Verksamhetsområde	**Elkraft**	**Sjukvård**	**Vatten och avlopp**	**Transport**	***Summa***
Process	30	0	0	0	*30*
Kontrollcenter	17	0	0	0	*17*
Kommunikation	1935	9	2	21	*1967*
Kontor (interna system)	229	3	0	0	*232*
Kontor (externa system)	8253	508	1	124	*8886*
Summa	*10464*	*520*	*3*	*145*	*11132*

En geografisk översikt av de identifierade komponenterna, genererade med hjälp av geodata från IP-API och Google Earth, ges av Figur 3. Som kan ses är de spridda över större delen av landet oavsett vilket verksamhetsområde eller vilken sektor som det handlar om, från Norrbotten till Skåne. Dessa figurer bör dock tas med en nypa salt då de geografiska platserna är härledda från IP-adresser och därmed sannolikt skiljer sig en hel del från de faktiska fysiska platserna där komponenterna verkligen finns.



Figur 3. Geografisk lokalisering av internetanslutna system.

Korrektheten i klassificeringssystemet kontrollerades genom att manuellt granska alla processkomponenter och kontrollsystemskomponenter (eftersom de var få samt bedömdes viktigast att kategorisera rätt), 20 kommunikationskomponenter, 20 interna kontorssystem och 20 externa kontorssystem. Resultatet presenteras av Tabell 7. Alla komponenter som kunde bestämmas inom process, kontrollsystem och kommunikation var rätt kategoriserade. De sex komponenter inom processen som inte kunde bestämmas på grund av begränsningar i dataunderlaget hade portar för Modbus (fyra fall), BACnet samt Ethernet/IP öppna. Av de interna kontorssystemen kunde åtta inte bestämmas. Tio av de tolv komponenter som kunde bestämmas bedömdes vara korrekt klassificerade. Dessa bestod av nätverksdiskar samt en TV. De två som bedömdes felaktigt klassificerade var routrar som hade dynamiska portar (port 49152 och uppåt) samt Network Basic Input/Output System[24] (NetBIOS, port 137) öppna. Gällande externa kontorssystem kunde elva komponenter bestämmas och nio inte bestämmas. Av de som kunde bestämmas bedömdes fem vara korrekt klassificerade. De övriga sex bestod av två routrar, en brandvägg och tre nätverksdiskar. De nio komponenter som inte kunde bestämmas hade alla vanliga externa portar aktiverade, såsom webb (Hypertext Transfer Protocol [HTTP], port 80) och filöverföring (File Transfer Protocol [FTP], port 21).

Korrektheten för klassificeringen kan ses som godkänd för alla områden utom externa kontorssystem. Att öka korrektheten för externa kontorssystem är dock

[24] NetBIOS är ett generellt protokoll för sessionslagret i Open Systems Interconnection (OSI)-modellen.

mycket svårt. Då detta område inte bedömdes som kritiskt för att svara på studiens forskningsfråga lämnades förbättringar av precisionen till framtida arbete.

Tabell 7. Korrekthet för klassificeringar.

Område	Rätt	Fel	Okänd	Total
Kontrollcenter	17	0	0	17
Process	24	0	6	30
Kommunikation	20	0	0	20
Interna kontorssystem	10	2	8	20
Externa kontorssystem	5	6	9	20

5.1 Process

Totalt identifierades 30 internetanslutna processnära komponenter, alla för organisationer inom elkraftssektorn. En översikt av portarna och protokollen som identifierades för dessa komponenter ges av Tabell 8.

Webbservrar (HTTP och HTTPS), filöverföring (FTP) och fjärrstyrning (Telnet och Secure Shell [SSH]) är generella gränssnitt för konfigurationsändringar av komponenter. Modbus, S7Comm, EtherNet/IP och BACnet är mer specifika gränssnitt för styrning och övervakning. Simple Network Management Protocol (SNMP) är ett generellt protokoll som fyller ungefär samma syfte som de sistnämnda.

Lantronix discovery är till för att hitta Lantronix-produkter på ett nätverk. Moxa Broadcast möjliggör kommunikation med multipla seriellt anslutna komponenter.

NetBIOS på port 137 används för att koppla samman namn och IP-adresser, t.ex. namnet ”HEMDATOR” med IP-adressen 192.168.0.10. NetBIOS har innefattat flera kritiska sårbarheter, särskilt för Windows-baserade operativsystem, och läcker dessutom potentiellt känslig konfigurationsinformation.

Ett av protokollen kunde inte helt identifieras. Detta protokoll nyttjade port 2323 på en komponent av typen i.LON SmartServer 2.0.

I styrsystemssammanhang nyttjas e-post (Simple Mail Transfer Protocol [SMTP]) ibland för att sända information om inkomna larm. Den manuella granskningen indikerar dock att mailservern som identifierades för en processkomponent sannolikt var ett fel i Shodan och Censys; förmodligen har den korresponderande IP-adressen huserat en Microsoft IIS-server tidigare. Detta gällde även VPN och NetBIOS (se ”*” i Tabell 8).

Tabell 8. Portal och protokoll inom process.

Protokoll	Port	Antal
HTTP	80	16
Modbus	502	9
Lantronix discovery	30718	7
Telnet	23	6
FTP	21	4
HTTPS	443	4
Moxa Broadcast	4800	4
Telnet	9999	4
HTTP	81	2
S7Comm	102	2
SSH	22	1
SMTP*	25	1
NetBIOS*	137	1
VPN*	500	1
VPN*	1723	1
Okänt (i.LON SmartServer)	2323	1
HTTP	8081	1
EtherNet/IP	44818	1
BACnet	47808	1

* Sannolikt fel i Censys och Shodan (samma IP-adress kan ha använts till olika komponenter).

Det genomfördes ett manuellt arbete för att identifiera modeller på komponenter; detta lyckades för 24 av de 30 komponenterna. En översikt av dessa ges av Tabell 9. Lantronix XPort och Moxa NPort (5100, 5210 och 5210A) används för att koppla upp seriekopplade komponenter till ethernet-nätverk. En uppkopplad komponent kan sedan övervakas och styras via olika gränssnitt i modulen såsom Telnet, webb eller SNMP. Climatix Advanced Web Module är en insticksmodul till en fläktstyrenhet som möjliggör fjärrstyrning via en webbserver. i.LON SmartServer 2.0 är en generell styrenhet för t.ex. elkraft och trafikljus. Corrigo E är en styrenhet för fastighetsautomationssystem. Dranetz DataNode är en styrenhet för elkraftssystem. Siemens Simatic S7-300 är en PLC som används för diverse ändamål. Vipa Speed7 315SB är en något modifierad variant av en Siemens Simatic S7-300 PLC och fyller ett liknande syfte.

Tabell 9. Identifierade modeller inom processen.

Modell	Antal
Lantronix XPort	7
Moxa NPort 5110	3
Climatix Advanced Web Module	3
i.LON SmartServer 2.0	2
Corrigo E	2
Dranetz DataNode	2
Vipa Speed7 315SB	1
Siemens SIMATIC S7-300 (315-2 DP)	1
Moxa NPort (okänd modell)	1
Moxa NPort 5210	1
Moxa NPort 5210A	1
Summa	*24*

5.2 Kontrollcenter

Sjutton kontrollcenterkomponenter identifierades. Dessa hade vanliga tjänster för manuell uppdatering, styrning och övervakning påslagna (se Tabell 10).

Tabell 10. Portar och protokoll inom kontrollcenter.

Protokoll	Port	Antal
HTTP	80	16
HTTPS	443	16
SSH	22	4
FTP	21	1
SNMP	161	1

Alla 17 komponenter kunde bestämmas av den manuella granskningen. Som visas i Tabell 11 rörde dessa tre typer av SCADA-system för fastighetsautomation. Av de tre modellerna används ofta TAC Xenta 511 för lite mindre system och Schneider Electric StruxureWare Building Operation Webstation 1.8 samt Honeywell i30-SPC för lite större system.

Tabell 11. Identifierade modeller inom kontrollcentret.

Modell	Antal
Schneider Electric StruxureWare Building Operation Webstation 1.8	14
TAC Xenta 511	2
Honeywell i30-SPC	1

5.3 Kommunikation

Vilka portar och protokoll som var vanliga inom kommunikation skilde sig väldigt lite mellan de olika studerade sektorerna. Av denna anledning analyserades resultatet för alla sektorerna i kombination. De tio vanligaste portarna presenteras i Tabell 12.

Flera av dessa portar bör inte vara nåbara från Internet. Exempelvis har webbservrar (HTTP och HTTPS) i kommunikationsutrustning i regel syftet att möjliggöra fjärradministration av komponenten, något som enbart bör vara möjligt från komponentens lokala nätverk. Detsamma gäller för Simple Object Access Protocol (SOAP) Gateway, vilket är ett Extensible Markup Language (XML) baserat protokoll som kan användas för att konfigurera utrustning på ett liknande sätt som via HTTP/HTTPS. SNMP bör inte vara nåbart från nätet, då det i bästa fall avslöjar viktig konfigurationsinformation, och i värsta fall kan nyttjas för att fjärrstyra komponenten. Apropå fjärrstyrning är SSH (port 22) aktiverat för 77 komponenter och Telnet (port 23) för 26 komponenter. Dessa portar bör heller inte vara nåbara från Internet.

Border Gateway Protocol (BGP) är vanligt routingprotokoll som behöver vara nåbart från Internet för att kunna routa nätverkstrafik (given antagandet att komponenten faktiskt är tänkt för ett sådant ändamål). En proxyserver används för att förmedla trafik från en sändare till en mottagare. Sändaren förblir ”osynlig” för mottagaren samt för alla komponenter som förmedlar trafik mellan mottagaren och proxyservern.

Tabell 12. Portar och protokoll inom kommunikation.

Protokoll	Port	Antal
HTTPS	443	624
VPN	500	568
HTTP	80	526
VPN	4500	443
VPN	1723	427
SOAP Gateway	7547	267
HTTP	8080	144
SNMP	161	137
BGP	179	123
Proxyserver	3128	123

5.4 Kontor (externa system)

Likt kommunikationssystemen skiljer det sig mycket lite mellan sektorerna gällande externa kontorssystem. De vanligaste portarna och protokollen för externa kontorssystem ges i Tabell 13.

Tabell 13. Portar och protokoll för externa kontorssystem.

Protokoll	Port	Antal
HTTP	80	5854
HTTPS	443	3579
SSH	22	682
FTP	21	615
SIP	5060	591
DNS	53	528
HTTP	8080	508
SMTP	25	284
Okänt	12345	249
Telnet	23	169

De flesta portarna är rimliga givet antagandet att externa användare nyttjar deras tjänster. Det vill säga, webb (HTTP, HTTPS), domännamnstjänster (DNS), filöverföring (FTP) och mail (SMTP). SSH och Telnet bör dock i regel inte vara

åtkomligt direkt från Internet. Session Initiation Protocol (SIP) används av voice over IP-servrar för IP-telefoni.

Port 12345 (över TCP) är intressant eftersom få kända godartade applikationer lyssnar på den och kartläggningarna innefattade väldigt lite information om den. Censys kartlade den inte alls och Shodan identifierade den med en Telnet-modul, och gav inga detaljer kring vilken applikation som egentligen lyssnade på den. Av de 249 externa kontorssystemen som hade port 12345 exponerad kategoriserade Shodan 16 som Linux. De övriga 233 var okända för Shodan. Internet Assigned Numbers Authority (IANA) listar enbart en applikation som nyttjar port 12345 över TCP (och ingen över UDP). Denna applikation är Italk[25], en chatserver som inte uppdaterats sedan 2006. Mer okända applikationer (t.ex. en Internet Relay Chat [IRC] bouncer[26]) och specifika lösningar (t.ex. port forwarding av SSH-trafik) nyttjar ibland port 12345. För skadliga koder är port 12345 däremot ett vanligt val, med NetBus 1.7 som det mest kända exemplet. NetBus släpptes på 90-talet och är kanske den första skadliga koden för Windows-datorer som fick ordentlig spridning. Dess huvudsakliga funktion är att erbjuda dold fjärrstyrning av en dator. Annan skadlig kod som är (ännu) svårare att med säkerhet identifiera men skulle kunna tänkas finnas i datamaterialet är: *W32.Axatak* (port 8888-8889, 11 träffar) och *Backdoor.Graybird* (port 8001, 23 träffar).

5.5 Kontor (interna system)

Likt kommunikation och externa kontorssystem skiljer det sig inte mycket mellan sektorerna gällande interna kontorssystem och analysen görs därför på sektorerna i kombination. De vanligaste portarna och protokollen beskrivs av Tabell 14.

Port 49152 och 49153 är de första vanligen använda dynamiska/privata portarna (49152-65535). Dessa brukar nyttjas av applikationer som använder olika portar vid varje start, i synnerhet Torrent-applikationer.

Server Message Block (SMB), ofta kallat Samba, används för delade mappar, skrivare och seriell kommunikation. SMB har likt NetBIOS innefattat många kritiska sårbarheter under åren och bör inte vara exponerat mot Internet.

SSH är vanligt även för interna kontorssystem, och bör inte vara åtkomligt från Internet. Windows Powershell server fungerar ungefär som SSH och används för fjärrstyrning av Windows-maskiner.

Det kan tyckas förvånande att webbservrar (HTTP och HTTPS) var vanliga även för interna kontorssystem. Anledningen till detta är att nätverksdiskar och skrivare

[25] http://italk.sourceforge.net/
[26] http://www.psybnc.dk/www.psybnc.at/about.html

ofta har inbyggda webbservrar för fjärrstyrning. Filservrar (FTP) är ofta installerade i nätverksdiskar.

Tabell 14. Portar och protokoll för interna kontorssystem.

Protokoll	Port	Antal
Dynamisk port	49152	83
HTTP	80	82
NetBIOS	137	61
HTTPS	443	44
SMB	445	35
Dynamisk port	49153	22
FTP	21	18
Kerberos server	88	16
SSH	22	14
Windows Powershell	5985	14

6 Slutsatser och framtida arbete

Denna studie analyserade Internetanslutna datorer i Sverige med hjälp av databaserna Censys och Shodan för att svara på frågorna ”*Vilka industriella informations- och styrsystem i Sverige är internetanslutna?*” samt ”*Vilka komponenter inom svenska samhällskritiska verksamheter är internetanslutna?*”.

Studiens resultat kategoriserades i fyra olika kritiska samhällssektorer (elkraft, transport, sjukvård samt vatten och avlopp), och fem teknikområden (process, kommunikation, kontrollcenter, interna kontorssystem samt externa kontorssystem).

Resultatet visade att klart flest internetanslutna komponenter kunde härledas till elkraftssektorn (94 % av totalen). Flest komponenter identifierades för externa kontorssystem (80 %), följt av kommunikation (18 %), interna kontorssystem (2 %), process (0,27 %) samt kontrollcenter (0,15 %). Bortsett från processen och kontrollcentret, där det enbart identifierades komponenter inom elkraft, var fördelningen av teknologier relativt lika för de olika sektorerna.

Gällande styrsystemsspecifik utrustning inom elkraft så är det lätt att tro att den primärt rör elkraftsspecifika funktioner såsom mätning av spänning och ström. Resultatet visar dock på att övervakning och automation av fastigheter (t.ex. gällande belysning och temperatur, se [5]) är vanligt inom elkraftssektorn. Av de 41 styrsystemsspecifika komponenter vars modeller kunde identifieras rörde *minst*[27] 19 fastighetsautomation.

Det är uppenbart att många av komponenterna som identifierades av denna studie inte bör vara internetanslutna. Det är dock svårt att veta exakt vilket syfte en identifierad komponent har i en verksamhet. Exempelvis kan en PLC av typen Siemens SIMATIC S7-300 användas för allt mellan honeypots[28] och labbtester till styrning av belysning i en sporthall eller en strömbrytare i ett elkraftverk. Det valida sättet att ta reda på en komponents verkliga syfte är att involvera systemägaren. Kvalitativa analyser av studiens resultat visade också att många av de identifierade komponenterna troligen inte realiserade någon kritisk infrastruktur, utan istället hanterades av externa parter som hade en organisation som arbetade med kritisk infrastruktur som internetleverantör i kombination med att denna organisation inte särskilde på IP-blocken för olika systemfunktioner (t.ex. egna system jämfört med bredbandskunder). Det enklaste sättet att hantera detta problem vore att erhålla sådan information direkt från de studerade organisationerna. Att involvera organisationerna är för övrigt det sannolikt

[27] Mer generella modeller såsom Siemens SIMATIC S7-300 kan också tänkas nyttjas för byggnadsautomation.

[28] En honeypot simulerar en komponent som en hotaktör kan tänkas vilja kompromettera för att därmed lura hotaktören att genomföra angrepp mot den. Dessa angrepp analyseras och kan nyttjas för att skapa mer effektiva skydd.

viktigaste framtida arbetet för att minska antalet internetanslutna kritiska system. Eftersom organisationerna i denna studie är kända skulle detta arbete enkelt kunna automatiseras. Lämpligen skulle datainsamlingsverktyget varje vecka eller månad genomföra en ny sökning och sedan automatiskt meddela alla berörda parter angående deras internetanslutna system. Att samla in data över tiden är också en viktig aktivitet för att kunna analysera trender. Trender är dock i allmänhet mycket svårt att helt analysera eftersom det är okänt hur många komponenter som är i drift i relation till de som är internetanslutna.

Extraheringen av data under studien tog mycket lång kalendertid (cirka 15 dygn från början till slut). Anledningen till detta var begränsningar i API:erna för Shodan, Censys och IP-API i kombination med att studien genomfördes med hjälp av en dator på en IP-adress och via ett konto på Shodan och ett annat på Censys. Datainsamlingstiden skalar linjärt med antalet datorer, IP-adresser och konton i webbtjänsterna. Om snabba analyser är av vikt bör det därför investeras i att bygga ut infrastrukturen för datainsamlingsapplikationen som beskrivs i avsnitt 4.2.

Det finns mycket arbete att göra med att skapa bättre sökfilter, både gällande organisationer och teknologier. Gällande organisationer var det enbart ett mycket begränsat antal IP-adresser (cirka 0,7 %) som kunde härledas till organisationer via geodatabaser. Gällande teknologier finns det idag få publikt tillgängliga sökfilter, och att skapa nya är ett krävande arbete.

Studien nyttjade även enbart ”passiv” datainsamling via existerande databaser, och lever därmed med dessa databasers begränsningar. Primärt rör begränsningarna de frågor som Shodan och Censys ställer till publicerade portar: de har enbart stöd för frågor rörande ett fåtal funktioner hos ett urval av olika protokoll. Ett sätt att lösa detta problem vore att genomföra ett aktivt kartläggningsarbete, där forskaren själv styr över de frågor som ställs till olika komponenter. Detta arbete skulle kunna baseras på ett existerande verktyg såsom Nmap eller Zmap, utökat med stöd för fler protokoll och protokollfunktioner som är vanliga i Sverige.

Gällande databaser innefattade studien ej tjänsten ZoomEye, vilken bör inkluderas av framtida studier av samma område.

Det var utanför studiens omfattning att identifiera tekniska sårbarheter i de studerade komponenterna. För systemägare vore det dock värdefullt att veta om komponenter är sårbara för olika IT-angrepp. En inventering av sårbarheter skulle till viss del kunna automatiseras.

Slutligen har tidigare forskning [18] påstått att kartläggningar av Censys och Shodan väl överlappar. Denna studie visar på motsatsen, och det vore intressant att göra mer omfattande studier för att bedöma datakvaliteten för Censys, Shodan och ZoomEye givet olika kriterier.

7 Referenser

[1] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams, and A. Hahn, “NIST 800-82 (r2). Guide to Industrial Control Systems (ICS) Security,” 2015.

[2] A. Daneels and W. Salter, “What Is Scada ?,” in *International Conference on Accelerator and Large Experimental Physics Control Systems, Trieste, Italy*, 1999, pp. 339–343.

[3] ISA, “Security for Industrial Automation and Control Systems Part 1: Terminology, Concepts, and Models,” 2007.

[4] CPNI, “Good Practice Guide - Process Control and SCADA Security.” Center for the Protection of National Infrastructure (CPNI), 2008.

[5] K. M. Sonnek, H. Holm, J. Lindgren, F. Lindgren, and E. Westring, “NCS3 - informations- och styrsystem inom spårbunden trafik. En kartläggning (FOI-R--4029--SE),” Linköping, Sweden, 2015.

[6] GAO, “Cybersecurity for Critical Infrastructure Protection.” United States General Accounting Office (GAO), 2004.

[7] S. Lüders, “Cern tests reveal security flaws with industrial network devices,” 2006.

[8] S. Tiilikainen, “Improving the National Cybersecurity by Finding Vulnerable Industrial Control Systems from the Internet,” Aalto University, Esbo, 2015.

[9] S. Jeon, J.-H. Yun, S. Choi, and W.-N. Kim, “Passive Fingerprinting of SCADA in Critical Infrastructure Network without Deep Packet Inspection,” *ArXiv e-prints*, pp. 1–8, 2016.

[10] Z. Durumeric, E. Wustrow, and J. A. Halderman, “ZMap: Fast Internet-wide scanning and its security applications,” in *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*, 2013, pp. 605–620.

[11] R. Graham, “Masscan: designing my own crypto,” *Errata Security*, 2013. [Online]. Available: http://blog.erratasec.com/2013/12/masscan-designing-my-own-crypto.html. [Accessed: 01-Dec-2016].

[12] D. Myers, E. Foo, and K. Radke, “Internet-wide scanning taxonomy and framework,” in *Proceedings of Australasian Information Security Conference (ACSW-AISC), 27-30 January 2015*, 2015.

[13] R. Bodenheim, J. Butts, S. Dunlap, and B. Mullins, “Evaluation of the ability of the Shodan search engine to identify Internet-facing industrial control devices,” *Int. J. Crit. Infrastruct. Prot.*, vol. 7, no. 2, pp. 114–123, 2014.

[14] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, “A search engine backed by Internet-wide scanning,” in *Proceedings of the 22nd*

ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 542–553.

[15] "Internet Census 2012 Port scanning /0 using insecure embedded devices," 2013.

[16] J. Heidemann, Y. Pradkin, R. Govindan, C. Papadopoulos, G. Bartlett, and J. Bannister, "Census and survey of the visible internet," in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, 2008, pp. 169–182.

[17] T. Krenc, O. Hohlfeld, and A. Feldmann, "An internet census taken by an illegal botnet: a qualitative assessment of published measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 103–111, 2014.

[18] J. François, A. Lahmadi, V. Giannini, D. Cupif, F. Beck, and B. Wallrich, "Optimizing internet scanning for assessing industrial systems exposure," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2016 International*, 2016, pp. 516–522.

[19] P. M. Williams, "Distinguishing internet-facing ICS devices using PLC programming information," 2014.

[20] E. P. Leverett, "Quantitatively assessing and visualising industrial system attack surfaces," 2011.

[21] B. Radvanovsky and J. Brodsky, "Project SHINE (SHodan INtelligence Extraction) Findings Report," 2014.

[22] J. Matherly, "Shodan ICS Radar," *Shodan*, 2017. [Online]. Available: https://ics-radar.shodan.io/. [Accessed: 22-Feb-2017].

[23] O. Andreeva, S. Gordeychik, G. Gritsai, O. Kochetova, E. Potseluevskaya, S. Sidorov, and A. Timorin, "Industrial cybersecurity threat landscape," 2016.

[24] N. Huw, S. Hilt, and N. Hellberg, "US Cities Exposed in Shodan," 2017.

[25] S. Merdinger, "Locating ICS and SCADA Systems on .EDU Networks with SHODAN," 2014. [Online]. Available: https://www.tripwire.com/state-of-security/government/locating-scada-and-ics-systems-on-edu-networks-with-shodan/. [Accessed: 01-Mar-2017].

[26] T. Testaaja, "Analysis of the Internet Census data The Finnish Cyber Landscape," 2013.

[27] Modbus, "Modbus application protocol specification v1.1b," 2006.

[28] H. Holm, T. Sommestad, J. Almroth, and M. Persson, "A quantitative evaluation of vulnerability scanning," *Inf. Manag. Comput. Secur.*, vol. 19, no. 4, p. 2.

[29] G. Gritsai, A. Timorin, Y. Goltsev, and R. Ilin, "ICS/SCADA/PLC

Google/Shodanhq Cheat Sheet," *SCADAStrangeLove*, 2012. [Online]. Available: http://www.slideshare.net/qqlan/icsscadaplc-googleshodanhq-cheat-sheet . [Accessed: 20-Nov-2016].

[30] AGA, "Cryptographic protection of scada communications - retrofitting serial communications. Technical Report 12," 2006.

Security in Industrial Control Systems

Nationellt Centrum för säkerhet i styrsystem för samhällsviktig verksamhet (NCS3) är ett kompetenscentrum med uppdraget att bygga upp och sprida medvetenhet, kunskap och erfarenhet om cybersäkerhetsasspekter inom industriella informations- och styrsystem. Centrumets är ett samarbete mellan de svenska myndigheterna FOI och MSB och dess verksamhet fokuserar på aktörer som äger och/eller driver samhällsviktig verksamhet där industriella informations- och styrsystem ingår.

The National Centre for increased security in industrial control systems is a centre of excellence focused at building and disseminating awareness, knowledge and experience about cyber security aspects in ICS. The Centre is a cooperation between the Swedish Defence Research Agency and the Swedish Civil Contingencies Agency and the activities is focused on actors that owns and/or operates critical infrastructure where ICS are a part.

FOI
Swedish Defence Research Agency
SE-164 90 Stockholm

Phone +46 8 555 030 00
Fax +46 8 555 031 00

www.foi.se

Swedish Civil Contingencies Agency
SE-651 81 Karlstad

Phone: +46 (0) 771-240 240
Fax: +46 (0) 10-240 56 00

www.msb.se