IP Based Smart Services: Combining Big Data and Networking to Create Innovative New Applications

Mason R. Allen, mason.r.allen at gmail.com (A paper written under the guidance of Prof. Raj Jain)

Download 

Abstract:

IP Based Smart Services are the result of applying Big Data style analytics on network information inwards in order to improve and strengthen the network itself. These improvements include changes to the ways that data is collected, used, and reused. Changes in data collection enable new smart applications that address changes in network conditions before they even occur (predictive) and even seek to make changes before they ever become an issue or concern (proactive). Other applications seek to increase the integration between conventional networks and existing systems like supply chain or climate control. Lastly, security applications of smart services are discussed including encryption and attack detection.

Keywords: Big Data Networking, Network Management, Smart Technologies, Internet Protocol

Table of Contents:

1. Introduction

As networks become more complicated and traffic volumes continue to increase at high speeds, the ability to understand the needs, uses, and performance of a network has never been greater. However this doesn't tell the

entire story; in the era of Big Data simply monitoring this information is not enough. Networking firms like cisco have realized that this information is not being leveraged to its true potential and that by further understanding this information powerful new tools can be created.

This interest in increased analytics has formed the topic of Smart Services. Smart Services is a general term that applies to any sort of "intelligent" analytics performed on data that provides a more proactive, predictive, or pervasive approach than before. For this paper I am going to be focusing on Smart Services using the IP stack, but they aren't necessarily limited to that in the greater scheme of things. There are new and exciting services leveraging the newly accessible data for all kinds of purposes. [Cisco12]

When understanding IP smart services you need to understand the ways the data is collected (including the specific types of data), the way that data is used and reused, the applications that those things bring us, and some of the security concerns and solutions relating to the process.

2. Capturing Knowledge

The applications of Smart Services involve not only better tools to analyze data but also improving the tools we have to collect it as well. Being able to not only monitor but to visualize the network (whether it be bandwidth, up/down status, power usage, or something else entirely) is the basis of Smart Services. The quality of Data Analytics is limited by the quality of the data itself.

NetFlow is a powerful tool used to capture IP Traffic for analysis that is made by Cisco. It was originally developed in the early 90s as a tool for packet switching. For a number of years it was a proprietary protocol used only on certain Cisco routers and only for certain uses. As time went on they realized that its use as an accounting tool was more important than its ability to direct packets.

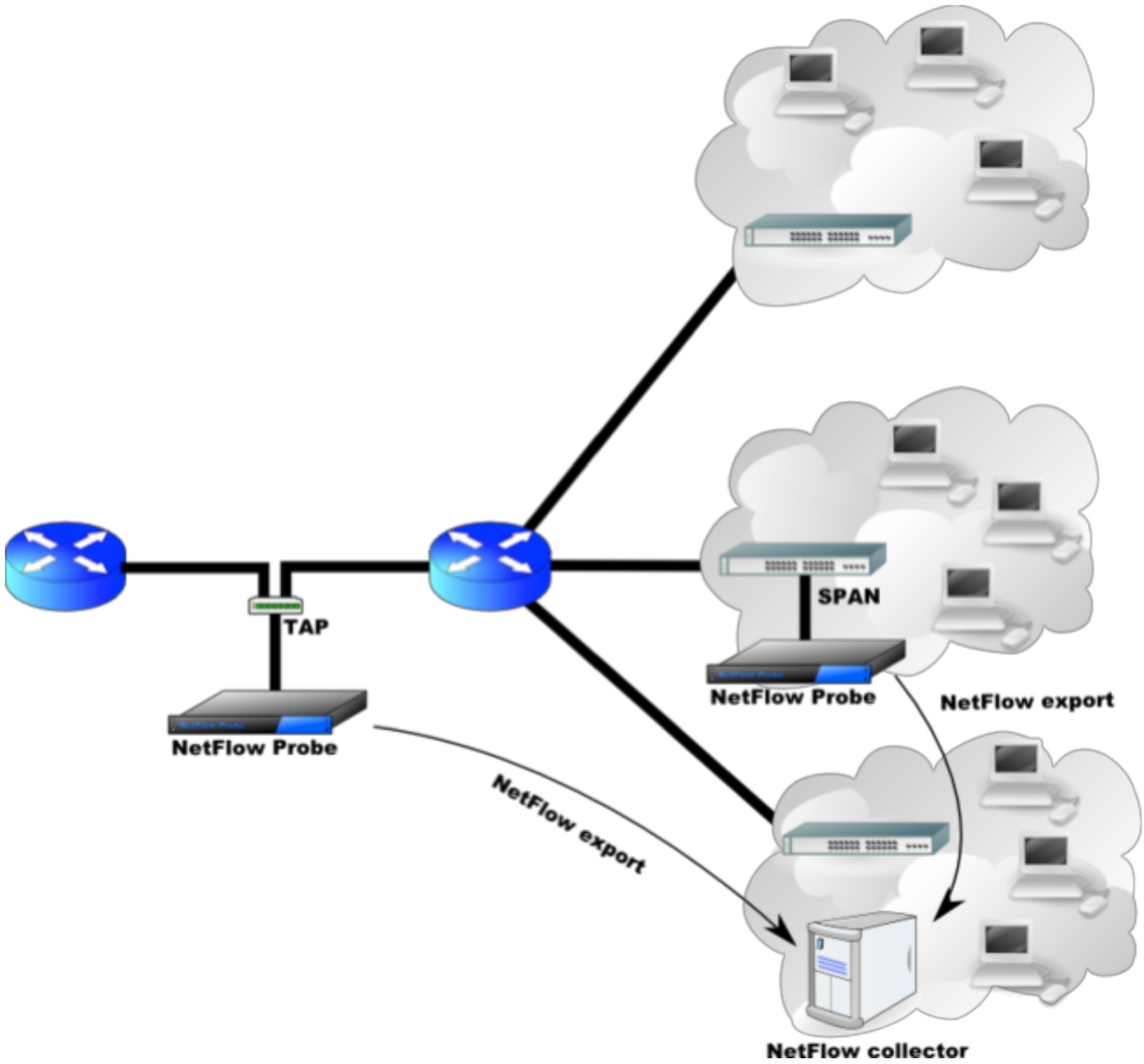

Figure 1: NetFlow Implementation Using Dedicated Probes
Source: Wikipedia Creative Commons Public Domain Images

NetFlow works by identifying what it calls "flows." These flows are one-way streams of packets that share certain key field. These fields include Source IP, Destination IP, IP protocol version, source and destination ports, type of service, and ingress interface. As this data is collected it is sent to a NetFlow controller, which is a server where that data is collated and analyzed. [Cisco06]

Over the two decades since its release, NetFlow would become the standard in monitoring network traffic and recording it for use by administrators. The features it included were increased over time to include IPv6, MPLS, and better aggregation tools. However, it is still limited to NetFlow enabled hardware, which is often expensive. Moreover the actual analysis is somewhat limited but that is changing as NetFlow is expanded and improved.

2.1 Advances in NetFlow

There are two important ways that NetFlow is being advanced. These involve increasing the scope of networks

which can be monitored using the tool and increasing the level of analytics performed by the collector server. These two approaches are almost entirely separate in their purpose - one is trying to massively increase the scope of the data to perform analytics while the other is decreasing the scale to try to get more fine-grained insight into the traffic.

As mentioned above NetFlow is limited to certain expensive hardware tools that probe the network data. This hardware was often confined to core networks or very large edge services. Recently tools have been made to bring a form of NetFlow analytics to smaller domains and allow more widespread use of its monitoring. In 2011 NetFlow lite was introduced. It removes the ASIC hardware requirement and allows NetFlow to monitor a massive new set of data sources. Smaller organizations which cannot implement expensive routers can now receive the benefits of this robust collection tool, but the complete opposite benefits also. This allows NetFlow to penetrate deep into data centers where space constraints made implementing NetFlow specific routers difficult. In some cases it can even be used to monitor L2 switches in addition to L3 traffic. This increase in data available for monitoring increases the accuracy of existing applications but could allow entirely new services which exclusively monitor in the data enters but correlate with outside services. [Deri11] [nTop11]

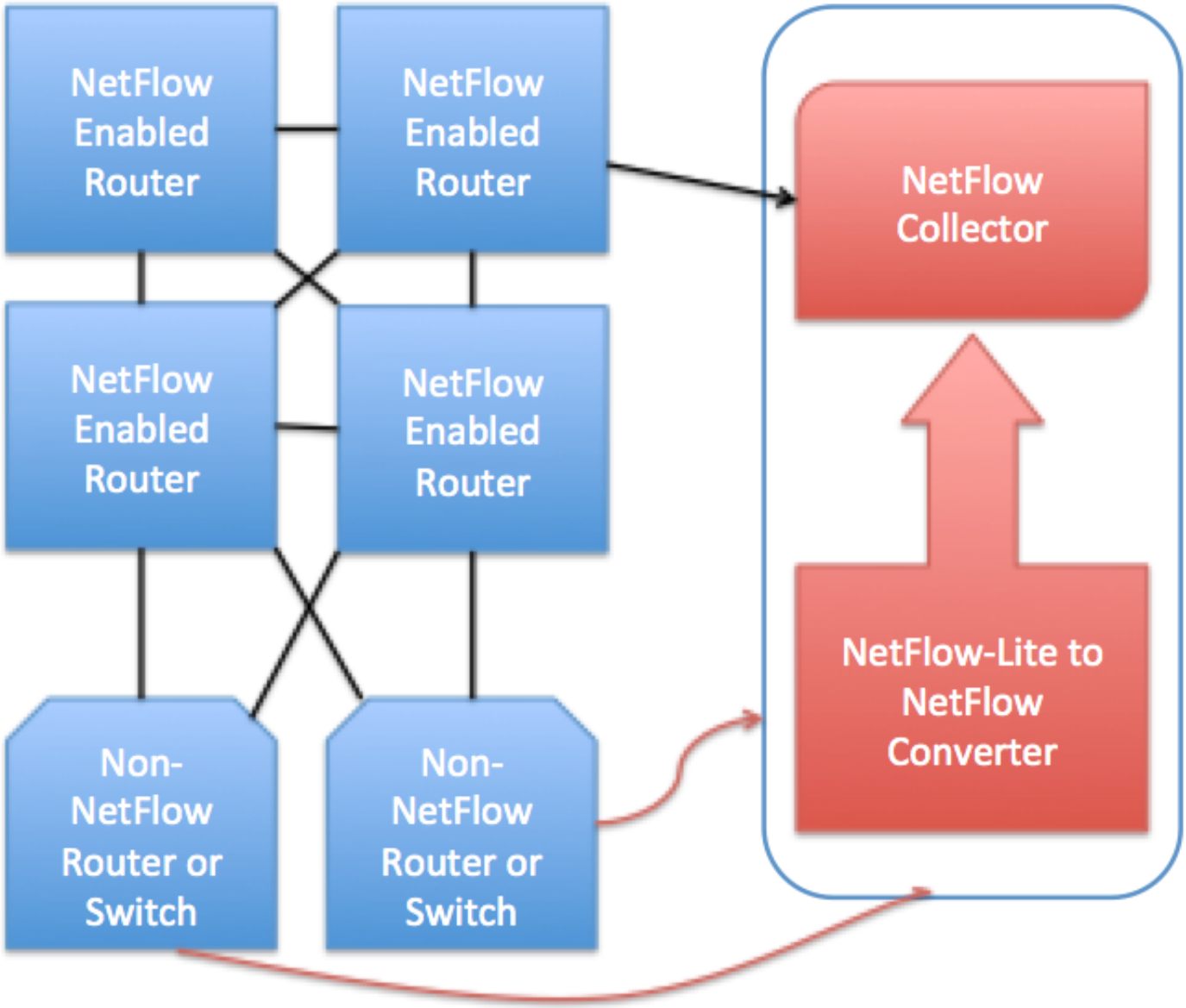

Figure 2: NetFlow-Lite Operation Diagram

The other can be thought of like a microscope. It allows incredibly fine-grained automated analysis of data at the per-packet and even per-byte level. This allows the data to be used in traffic classification in addition to just strict accounting. Due to recent changes in the nature of web applications it has become difficult to use deep packet inspection or port-based classification to make meaningful inferences about the traffic. However by using a machine learning algorithm looking for behavioral patterns in records, researchers in France were able to create a model that could successfully classify data as P2P streaming, P2P file sharing, and more traditional services. An interesting feature of this is that despite being an IP service and relying entirely on an IP collection protocol their model is constructed using byte analysis without any changes to existing infrastructure or policies. This deeper understanding of the actual type of traffic allows network managers to better understand what services and activities are using and in some cases taxing their network. [Rossi10]

2.2 IPFIX - Standard NetFlow

One of the major barriers to NetFlows wider adoption is its proprietary nature. NetFlow is created and owned by Cisco and works using their routers and hardware. The Internet Engineering Task Force (IETF) has created a standard system called IPFIX, or Internet Protocol Flow Information Export. This protocol is similar to NetFlow version 9, which it was based on, but administrators are free to change or add fields and parameters as they see fit. This increased customizability allows for network managers to collect whatever statistics they desire about their network [WorkingGroup13]

This expanded tailorability allows for new applications relating to the Internet of Things. An implementation from the Technical University of Munich allows for collector of data from Wireless Sensor Networks, IP enabled hardware, and other sensors by mining compressed IPFIX flows. This allows devices that are massively limited in memory to perform specialized analysis in cases like home security, health, or traditional network monitoring. Moreover, distributed or localized analysis of data can reduce the load needed to be sent to a central data center or collector. [Schmitt10]

3. Reusing Knowledge

Another interesting aspect of Smart Services comes not from the ways they collect data but the ways they avoid collecting it. With limited resources and space only certain meters and monitors can be installed within the network fabric. As new technologies or applications emerge the network must adapt to meet their needs, however sometimes an actual change isn't really needed.

The massive amount of data that is recorded and monitored but not really used contains a wealth of information that could be used to solve certain problems or provide certain services without having to spend money and time installing new hardware or even changing policies. This is one of the key aspects of Smart Services as defined by Cisco - making better use of the data already available or that already has to be collected. [Cisco12]

In a some regards, the fine-grained analysis of NetFlow is a great example of such a service. By employing a cloud based learning strategy the researchers were able to take data that is procedurally generated by network accounting and use it for more advanced network analytics. Changes such as these can be implemented on existing servers and using existing network tools making them incredibly cheap and efficient ways to bring new functionality to existing networks. [Rossi10]

A great example of one of the ways that network providers can reach outside strict routing services is a new

serviced called Sentinel. Sentinel uses existing source IP information from Wireless routers within a network to physically locate people within the building or office where the network is present. There are many possible applications for such a function including security and safety, but Sentinel uses this data to efficiently run heating and cooling equipment for commercial spaces.

By using the gateway IP for mobile users the service performs a sort of approximate location service. Each wireless router is assigned within one of a buildings existing heating or cooling zones allowing you to get real time information about the number of people in each region and to interface with the cooling element accordingly. [Balaji13]

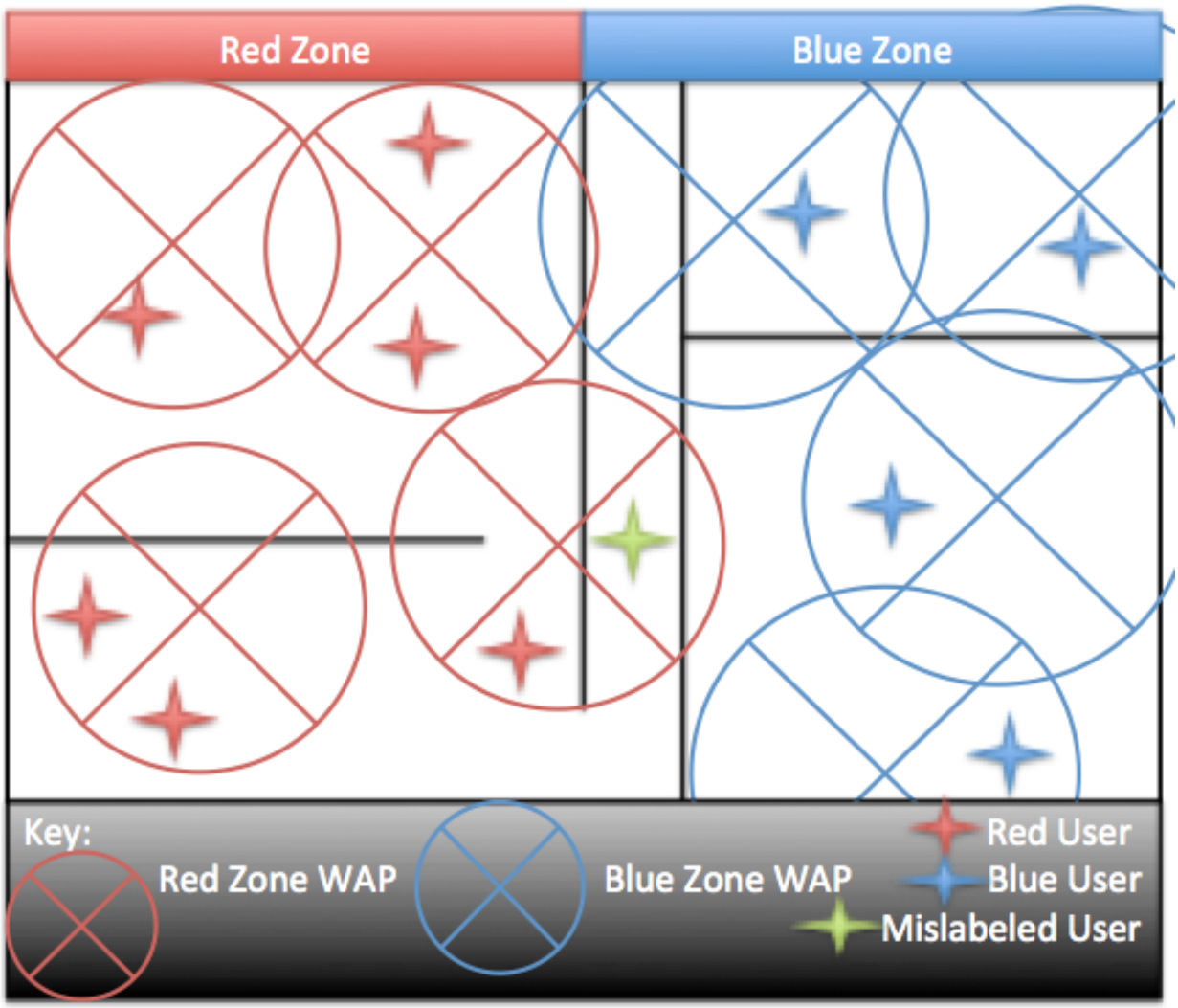

Figure 3: WiFi Based Location Services for Controlling HVAC Systems

This location is approximate as WiFi signals for a single router may reach into other HVAC (Heating, Ventilation, and Air Conditioning), regions and therefore misrepresent the location of persons in edges between WAPs, but this system has proven to be fairly effective in reducing heating and cooling costs. In Figure 3 above the green star represents a person who is in the blue zone but has been mis-classified in the red zone. Certain baselines were kept in the thermal range for the building but overall usage of power for the HVAC decreased by over 10% during the usage period making this a compelling way to reduce costs for companies. [Balaji13]

This solution becomes even more appealing knowing that a massive number of businesses already have WLANs serving their employees and as such the tools needed to implement this system are almost entirely extant already.

All that is needed is an interface to the HVAC system and the Sentinel system to coordinate. This efficiency comes from looking at information which is always collected making this another great reuse of network data for intelligent applications.

4. Predictive Applications

Perhaps the most exciting use for Smart Services is in performing actions or providing alerts in proactive ways to assist the management of the network and related services. These applications leverage not only the existing information about the network but also interface with new devices and sensors in real time to create new stream of intelligence about things that have not yet previously been controlled or in some cases even managed through network services. [Cisco12]

As part of the explosion of the internet of things the ability to communicate with entire new systems has emerged for IP services. Similar to the way in which the Sentinel system was able to bridge the gap between the information web and the physical world, dozens of new applications allow internet control and management of anything from the supply chain to social media.

4.1 Incorporating Supply Chain

The massive reach of IP networks allows for IP based cloud hosted services to extend their reach into the physical. One exciting outlet for this advancement is in the control and management of the supply chain. Being able to observe, direct, and control the systems used in creation and distribution of a product is only the beginning of the picture.

After you have access to these systems you can really start to do some "Smart" things. By bringing the web of data closer to the web of materials that the supply chain relies on you have the most accurate up to date data possible with which to make decisions. The ability to quickly react to changes in supply or demand, sentiment on social networks, and other factors as determined by any number of analytic services combined with the ability to rapidly manage and direct operations in the field allows businesses to operate more efficiently. [Ping11]

Furthermore by using the vast web of information about an entire system to apply context to certain sub-elements managers gain the ability to create powerful application tools. Understanding not only that something is happening but what contributed to it or the things that it is causing allow creation of services to better predict, avoid and compensate. Services related to context are not limited to supply chain thought they can be applied in this area also. In general understanding the context of a device is a task best tackled by data analytics and smart services. [He12]

4.2 Smarter Fault and Error Detection

Other research in predictive systems has looked for ways to use distributed, P2P, or cloud computing systems in order to better assess risk conditions or errors before they even occur. One team looking at Ad Hoc networks created a semi-centralized IP service that simulates the network and its future status under semi-ideal conditions and used that as a system which could effectively predict faults before they occurred. It also allowed the identification of bandwidth jams before other systems could detect. [Bejerano06]

This system is just one rudimentary example of what is called a PMA, or Proactive Management Algorithm.

These are algorithms which look at network status and other data and can actually change the policies (and in the case of SDN, topology) of a network based on its beliefs about the system. As Smart Services advance PMAs are likely to become more sophisticated and take greater actions without any input from administrators or humans. [Iskander08]

5. Security

For any issue in networking there will always be security concerns. Smart Services implore and expand data collection about any number of incredibly sensitive subjects. By integrating into heating, cooling, power, security, and supply chain systems the risk of damage from intrusion or interception becomes enormous. However just as Smart Service cause a problem in security, so too can they be the solution to that issue.

By using new and intelligent services to secure data and to detect attacks, Smart Services can not only provide increased functionality to web services but also increased security over standard practices. Interesting advancements in this field include the ability to data mine traffic to find examples of possible intrusion more accurately than ever before possible and new ways of efficiently encrypting traffic coming from massive numbers of devices and sensors. [Viera13]

5.1 Securing Data

The easiest way to prevent unauthorized use or access of your data is to make sure that it is secured as well as possible. For many applications this involves some form of encryption services, however the new scale of applications causes certain forms of encryption to become less viable. Creating a many-to-many encryption system in order for your service nodes to communicate with each other is an incredibly difficult problem that is being researched. The overhead of existing systems in this area combined with the volume of data being used makes such implementations less than ideal. [Boldyreva08]

One proposal to address this concern is the use of identity-based encryption. The result is a system that can use known identifiers used to route traffic as a sort of public key system, bypassing the need for certificates in some cases. A centralized system serves as a private key generator and controls the decryption for the network [Viera13]

By cutting out the need to store public keys in the system you remove a significant step in the handling of data. No longer will senders need to look up the public key of the destination, instead using its identity as a substitute. Given the huge volume of messages being sent through Smart Services and large networks the benefits of this multiply greatly. [Boldyreva08]

5.2 Detecting Attacks

A very important application of Smart Services is in identifying new patterns and models for detecting attacks. Looking at the recorded traffic data of a network as a massive set of training data for a machine learning algorithm, recent research has begun trying to find models that can detect or predict attacks much earlier than conventional systems.

One such approach is the mining of SNMP information to try to determine when ARP requests are germane versus malicious. ARP, or Address Resolution Protocol, is a vehicle of attack by creating a false association

between a MAC address and an IP address. An ARP is used when routers want to find the appropriate MAC associated with a given IP address. Upon issuing an ARP the server listens for a response which tells it where to send it. Unfortunately ARP is a stateless protocol and there is no way for an agent to know if it has recently requested an ARP of a given address. Users can then send out a false ARP reply message associating their MAC with another person's IP and therefore intercept their traffic. [Hsiao09]

Using certain common data mining tools such as Support Vector Machines (SVM), Decision Trees, and Naive Bayesian classification one can create a model that detects ARP attacks. They found some differences in the accuracy of their models but noted they beat the naive method and therefore are an improvement over current systems. Additionally they have a very low false alarm rate even if they are only around 60% accurate. This is just one example of how traffic abnormalities can be monitored by simply studying the data that is already collected. [Hsiao09]

6. Conclusion

The ability to better utilize data in computers is a never ending task that is only just even starting to mature. The advent of Big Data technologies and machine learning systems have enabled a new wave of services and tools in every field of computing. The area of IP Based Smart Services is a budding field that aims to leverage the huge quantity of transactional data and extant IP infrastructure to create powerful new ways to collect, manage, use, and protect data. By implementing Smart Services networks improve not only their ability to provide better Fault, Configuration, Accounting, Performance, and Security (FCAPS) management tools and systems.

References

[Iskander08] Adel F. Iskander, Proactive Management Algorithm for Sealf-Healing Mobile Ad Hoc Networks, International Journal of Network Management, 2008, P229-250, http://doi.wiley.com/10.1002/nem.654
Details a new approach to managing networks that enables them to better adjust for faults and errors.

[Boldyreva08] Alexandra Boldyreva, Identity-Based Encrpytion with Efficient Revocation, Proceedings of the 15th ACM Conference on Computer and Communications Security, 2008, P417-426, http://dl.acm.org/citation.cfm?doid=1455770.1455823
Discusses a new method of using identity encryption in a machine-to-machine network that is able to work better at large scale than conventional methods.

[Viera13] Barbara Viera, A Security Protocol for Information-Centric Networking in Smart Grids. Proceedings of the first ACM Workshop on Smart Energy Grid Security, 2013, P1-10, http://dl.acm.org/citation.cfm?doid=2516930.2516932

[Balaji13] Bharathan Balaji, Sentinel: Occupancy Based HVAC Actuation using existing WiFi Infrastructure within Commercial Buildings, Proceedings of the 11th ACM Conference on Embedded Netwroked Sensor Systems, 2013, http://dl.acm.org/citation.cfm?doid=2517351.2517370
Describes a system that uses the locator of the Wireless Access Point of mobile devices and laptops to create a regional occupancy count and control the heating/cooling systems in a building.

[Cisco06] Cisco Information Technology. Cisco Information Technology At Work Case Study: Cisco IOS NetFlow Technology. 2006,

http://www.cisco.com/application/pdf/en/us/guest/tech/tk812/c1482/ccmigration_09186a00802de684.pdf

[Cisco12] Cisco Research. IP Based Smart Services. 2012, http://www.cisco.com/web/about/ac50/ac207/crc_new/university/RFP/rfp12074.html
Describes the Cisco research initiative on IP Smart Services.

[Rossi10] Dario Rossi. Ffine-grained traffic calssification with netflow data. Proceedings of the 6th international wireless communications and mobile computing conference 2010, http://dl.acm.org/citation.cfm?id=1815507
Explains how you can mine NetFlow collected information to be able to group the traffic into P2P, streaming, and other classes to gain insight into network activity.

[Hsiao09] Han-wei Hsiao. Constructing an arp attack detection system with snmp data. Proceedings of the 11th international conference on electronic commerce 2009, p341-345, http://dl.acm.org/citation.cfm?id=1593309
Uses data mining of transactional information to attempt to quickly identify a particular type of man in the middle attacks.

[He12] Jing He. A Smart Web Service Based on the Context of Things. ACM Transactions on Internet Technology. 2012, http://dl.acm.org/citation.cfm?id=2078321

[Ping11] Lou Ping. Agile Supply Chain Management over the Internet of Things. International Conference on Management and Service Science. 2011, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5998314"&"abstractAccess=no"&"userType=inst
Describes an interface that can monitor and alter supply chain operations in rapid time based on various metrics.

[Deri11] Luca Deri. Increasing Data Center Network Visibility with Cisco NetFlow-Lite. International Conference on Network and Service Management. 2011. P1-6, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6104026
NetFlow-Lite is an alteration on traditional NetFlow that requires less specific and expensive hardware.

[nTop11] nTop and Cisco. Say Hello to NetFlow-Lite. 2011,http://www.ntop.org/nprobe/say-hello-to-netflow-lite-nflite/
Further describes the creation and use cases of NetFlow-Lite.

[IPFIX] IP Flow Information Export Working Group. Working Group Charter. Updated 2012, http://datatracker.ietf.org/wg/ipfix/charter/.
Details the goals and purpose of the IETF working group in charge of IPFIX. Gives a nice top level feature list and milestone guide.

[Schmitt10] Corinna Schmitt, Lothan Braun, Georg Carie. Collecting Sensor Data Using Compressed IPFIX. 2010, http://dl.acm.org/citation.cfm?doid=1791212.1791269

[Bejerano06] Yigal Bejerano. Robust Monitoring of Link Delays and Faults in IP Networks. IEEE/ACM Transactions on Networking. 2006, http://dl.acm.org/citation.cfm?doid=1217722

Appendix A - List of Acronyms

ARP - Address Resolution Protocol
ASIC - Application Specific Integrated-Circuit
FCAPS- Fault, Configuration, Accounting, Performance, Security
HVAC - Heating, Ventilation, and Air-Conditioning
IP - Internet Protocol
IPFIX- IP Flow Information Export
LAN - Local Area Network
MAC - Media Access Control
NIST - National Institute of Standards and Technology
NFLite- NetFlow-Lite
P2P - Peer-to-Peer
PMA - Proactive Management Algorithm
QoS - Quality of Service
SNMP - Simple Network Management Algorithm
TLS - Transport Layer Security
WAP - Wireless Access Point
WLAN - Wireless LAN

Last Modified: December 10, 2013
This and other papers on latest advances in computer networking are available on line at
http://www.cse.wustl.edu/~jain/cse570-13/index.html
Back to Raj Jain's Home Page